Inter-American Development Bank

Best Practices for Transparent, Reproducible, and Ethical Research

Office of Strategic Planning and Development Effectiveness

Fernando Hoces de la Guardia
Jennifer Sturdy

TECHNICAL NOTE Nº
IDB-TN-01635

February 2019

Best Practices for Transparent, Reproducible, and Ethical Research

Fernando Hoces de la Guardia
Jennifer Sturdy

February 2019

IDB
Inter-American Development Bank

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Hoces de la Guardia, Fernando.
Best practices for transparent, reproducible, and ethical research / Fernando Hoces de la Guardia, Jennifer Sturdy.
p. cm. — (IDB Technical Note ; 1635)
Includes bibliographic references.
1. Social sciences-Research-Moral and ethical aspects. 2. Reproducible research. I. Sturdy, Jennifer. II. Inter-American Development Bank. Office of Strategic Planning and Development Effectiveness. III. Title. IV. Series.
IDB-TN-1635

1300 New York Ave NW, Washington DC 20577

Best Practices for Transparent, Reproducible, and Ethical Research[1]

Fernando Hoces de la Guardia and Jennifer Sturdy[2]

Abstract

The social science landscape is changing rapidly; alongside existing standards for ethics in research there are rising standards for what is considered credible and rigorous research, including the transparency and reproducibility of the research. This technical note provides a summary of the main innovations in best practices behind transparent, reproducible, and ethical (TRE) research and recommends applications to the knowledge products generated by the IDB.

Keywords: Reproducibility; Knowledge products; Reproducible research; Ethical research; Transparent research.

JEL codes: C81, C83, C89, C90, Y10

[1] The preparation of this Technical Note was commissioned by the Office of Strategic Planning and Development Effectiveness, under the direction of Carola Alvarez and Sebastian Martinez, with support from the Berkeley Initiative for Transparency in the Social Sciences (BITSS). Hoces de la Guardia led Chapters 2 & 3; Sturdy led Chapter 4; Executive Summary and Chapter 1 was co-written by both authors. This Note benefited from valuable discussions and feedback from IDB staff: Paloma Acevedo, Oliver Azuara, Roseangela Bando, Samuel Berlinski, Diether Beuermann, Ana Cuesta, Suzanne Duryea, Sebastian Gallegos, Oscar Mitnik, Cecilia Vidal, Solis Winters, Ana Yarygina, and Giulia Zane; as well as external peer reviewers Aleksandar Bogandoski (BITSS), Fiona Burlig (University of Chicago), Scott Desposato (UC San Diego), Katherine Hoeberling (BITSS), and Sara Stoudt (UC Berkeley).
[2] Berkeley Initiative for Transparency in the Social Sciences (BITSS) within the Center for Effective Global Action (CEGA)

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The social science landscape is changing rapidly; alongside existing standards for ethics in research there are rising standards for what is considered credible and rigorous research, including the transparency and reproducibility of the research. This technical note provides a summary of the main innovations in best practices behind transparent, reproducible, and ethical (TRE) research and recommends applications to the knowledge products generated by the IDB. In addition to helping maintain its position as a leader in knowledge generation, these best practices will be relevant for safeguarding the credibility of all IDB knowledge products. Examples of knowledge products that could benefit from these best practices include impact evaluations, economic analyses, surveys, research projects, and other empirical research activities[3]. The practices described in this document are broadly applicable to both quantitative and qualitative research methods, though some distinctions are discussed, as well as for research that has operational or academic purposes.

Transparent research refers to the set of practices and tools used to disclose all methods and data behind an analysis. Particular emphasis is placed on tracking the entire body of research using *registration*, disclosing key decisions like the formulation of hypotheses and specific research design details in *pre-analysis plans*, and facilitating the accumulation of knowledge using standardized *reporting guidelines*. ***Reproducible research*** refers to the ability of the research community to access and recreate the final results of the analysis from raw data with minimal effort. ***Ethical research*** refers to practices following the ethical principles of beneficence, respect for persons, and justice, with particular emphasis on *informed consent*, *independent review* (institutional or by other board), and proper *data de-identification and management*.

The menu of TRE research practices described in this document are complementary and should be considered in conjunction, although emphasis on any single topic will depend on the specific nature and circumstances of each individual research activity. They are intended to expand the project team's toolbox for designing, implementing, and disseminating high quality, credible research. That may mean applying just one, or two, or all of these best practices, depending on the requirements of the study, context, and other external factors.

The benefits of incorporating TRE practices are many. Transparent research enables correct (null) hypothesis testing, increases visibility and discoverability of research, and shifts attention away from statistically significant results to the quality and relevancy of the research itself. Reproducible research practices facilitate collaboration with other researchers and provide a strong foundation for future researchers to build on (increasing the likelihood of citations). Complying with ethical research standards demonstrates a commitment to respect for the rights and welfare of those on which research projects rely (research subjects), as well as of those on whom results will impact (the public).

The costs of not adopting TRE practices are also many. Opaquely conducted research may contribute to a loss of trust in research findings, as well as bias in meta-analyses and systematic reviews relying on

[3] For projects in preparation, this Technical Note may inform the "transparency and credibility" criteria of the Development Effectiveness Matrix (DEM) and the proposed plan should be described in the project's Monitoring and Evaluation Plan. A sample text is provided in Appendix B.

past research. Irreproducible research violates basics scientific principles and makes it impossible to detect coding errors. Unethically conducted research increases the likelihood of harming research subjects (through insufficient protocols), can harm the reputation of the IDB, and threatens the reputation, employment, and funding of project teams. Additionally, failure to comply with certain elements of transparency (registration), reproducibility (code and data sharing), and ethics (IRB approval) may prevent researchers from publishing in academic journals.

Specific best practices for TRE research presented in this technical note include the following:

Transparent Research:

- **Study Registration**, a brief, documented description of a study before data is available for analysis.
- **Pre-Analysis Plans (PAP)**, the extensive description of a study before data is available for analysis.
- **Registered Reports**, a method of submitting an article to a journal for review whereby editors and peer reviewers assess a manuscript prior to data analysis.
- **Reporting Guidelines** to define which content is to be presented in a manuscript in order to facilitate systematic reviews and meta-analysis.

Computationally Reproducible Research:

- **File Management** to ensure clear structure and format of the files in a research workflow. Additionally, a proper data back-up plan should specify how the data will be stored.
- **Version Control strategy** to ensure tracking of the complete history of all code used in a project. This is critical to provide reproducibility of current and previous iterations of the analysis.
- **Code Readability and Dynamic Documents** to standardize coding style and legibility within a team. These practices are meant to facilitate the reproduction of the analysis across multiple researchers (including the original researcher on a later occasion).
- **Code Sharing and Data Sharing**, practices and tools recommended for the sustainable archival and dissemination of data and code.

Ethical Research:

- **Community Stakeholder Engagement (CSE)** to inform project teams of the needs and priorities of the communities with whom they will engage to inform decisions for ethical research.
- **Understanding of legal requirements**, including international and national data privacy and protection laws, as well as any legal regulations on research and protection of human subjects, to guide research practices.
- **Independent review** that is culturally and contextually sensitive to assure, both in advance and by periodic review, that appropriate steps are taken to protect the rights and welfare of those impacted by the research. This is often done through, but not limited to, Institutional Review Boards (IRBs).
- **Research Protocol** to document the ethical research design and implementation plan. A well-done research protocol can also facilitate future reproducibility efforts.
- **Data Management Plan**, often a section of a research protocol, to document who will have access to what data, when and how.
- **Informed consent** to ensure research subjects are aware of the objectives, duration, and description of the research; its expected benefits and risks; and that their participation is voluntary. Highly relevant

for transparency and data sharing, this is also where promises of confidentiality are made, or not, as well as clarifications of who has access to what data and when.

- **Proper data collection, storage, transfer, and disposal** to ensure privacy protection and risk mitigation.
- **De-identification** of documents and data to ensure privacy protection and risk mitigation**.**
- **Data sharing and dissemination** activities that adhere to research protocol and data management plan.

As a companion to this document, the one-pager summarizes all these practices, organized according to the broad phases of a research project: (1) Project or Research Design and Preparation, (2) Implementation, and (3) Dissemination.

CHAPTER 1: INTRODUCTION

CREDIBILITY CRISIS IN SOCIAL SCIENCE RESEARCH

The scientific community has documented a series of methodological challenges across fields leading to calls for more transparent, reproducible and ethical (TRE) research practices to bring scientific practice in closer alignment with scientific norms[4]. This section presents an overview of those scientific norms, the magnitude of challenges facing research - including cases of scientific misconduct, p-hacking and publication bias - and discusses their consequences. Understanding each of these challenges has also led to a much more critical examination of the ability to replicate landmark studies. This section also discusses the reproducibility crisis that emerged when large fractions of studies were found to be non-replicable.

SCIENTIFIC NORMS AND SCIENTIFIC MISCONDUCT

Merton (1942) postulated that researchers understanding of good science complies with the following four principles, known as *Mertonian norms*:

- **Disinterestedness**: a key motivation in a researcher's agenda is the search for the truth. Financial interest cannot dictate the research agenda;
- **Organized skepticism**: a system of peer review and replication should be in place to verify the consistency and veracity of the claims advanced by the scientific community;
- **Communality**: researchers should share their knowledge. All the scientific output (papers, data, code and materials) should be available to the research community; and
- **Universalism**: scientific validity should be determined by the veracity or quality of a scientific finding and not by the hierarchies of those making the claims.

Current incentive structures for academic tenure and publishing tend to reward research results that are flashy and statistically significant, not necessarily those that abide by these norms. Additionally, program managers and funders also have strong incentives to push for a research agenda that finds strong positive effects of their interventions. This has driven researchers to employ practices that no longer align with these norms, including selective reporting, failing to share underlying analysis plans, data, and code. These practices range from a small number of researchers committing outright fraud to a much larger number of researchers overlooking questionable research practices and scientific misconduct.

The most well-known case of scientific fraud in recent years is of former professor of psychology Diederik Stapel. An academic star in his field, he was found guilty of fabricating much of his data, collected over decades, and was expelled from Tilburg University in 2011. Another emblematic case is that of Michael LaCour, a PhD student in political science who produced a high-profile study on the effect of in-person canvassing on voter attitudes regarding gay rights (LaCour and Green, 2014), for which it was found that data was likely fabricated (Broockman, Kalla, and Aronow, 2015) resulting in Science retracting the publication[5]. While there has not yet been a high-profile case of fraud in economics, a recent survey

[4] In Economics, the term Credibility Crisis is mainly identified with Angrist and Pischke (2010) where they discuss improvements around research design and quality of data. The term credibility crisis is here used to refer to another set of problems and solution (with some overlap).

[5] It is worth noting that LaCour was caught in part because Science requires data publication.

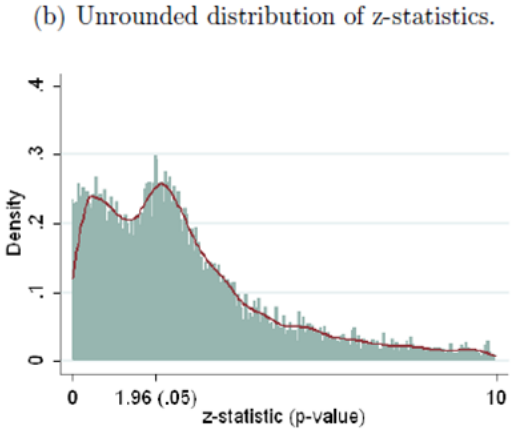
reported that 25% of editors of economics journals have seen cases of plagiarism (Enders and Hoover 2004).

It is difficult to estimate the -most likely very low- incidence of fraud in empirical work, but these examples outline how current practices and norms of opaqueness and publish-or-perish incentives can produce profoundly problematic cases such as these. A much more prevalent phenomenon is that of Questionable Research Practices (QRP) a term used to denote the much more prevalent grey areas of research that are used to increase the likelihood of obtaining a statistically significant result (John et al., 2012). Examples of QRPs include: not reporting all the dependent variables analyzed, rounding off p-values (reporting 0.054 as less than 0.05), and excluding data after looking at the original results. In recent years the term *p-hacking* has emerge as an encompassing terminology for behind this type of practices.

P-HACKING

P-hacking occurs when researchers, intentionally or not, select a subset of the possible analyses in a study based on whether those analyses generate statistically significant results. Typically, statistical significance is defined by a p-value less than 0.05, hence the term “p-hacking”. This problem was first theoretically outlined in economics by Leamer (1983), but until recently, there has been little quantitative evidence to document the problem. Brodeur et al. (2016) collected z-statistics for 641 articles in three top journals in economics (the American Economic Review, Quarterly Journal of Economics, and Journal of Political Economy) from 2005 to 2011, and examined the distribution of z-statistics. They found strong evidence of p-hacking as the empirical distribution had an abnormal hump around the common threshold of statistical significance of 1.96, as depicted in Figure 1.1.

Figure 1.1: Distribution of z-statistics in top journals in economics. Reproduced from Brodeur et al. (2016)



Gerber and Malhotra (2008) have documented a similar problem in other social science fields. They use a caliper test that, similar to Brodeur et al. (2016), to examine the empirical distributions of z-statistics in top journals for sociology and political science. They too find an abnormal number of p-values right before the 0.05 cutoff as compared to just after the cut off (two times as many for sociology and three times as many for political science).

Gelman and Loken (2013) provide a theoretical framework for a general type of p-hacking. In what they call the “garden of forking paths,” researchers will p-hack, intentionally or not, by choosing specifications,

data sets, hypotheses, and others modeling decisions. It is important to emphasize that p-hacking is not addressed by adding robustness checks, as these tend to verify stability of the results over highly reduced set of all the possible analytical choices.

The main consequence of p-hacking is that it increases drastically the chances of false positives and produces biased results within a single study and across a body of literature. The problem can be understood as a version of multiple hypothesis testing where the researcher does not know, or does not report, the true number of underlying hypothesis. Ioannidis (2007) calibrates a model with different levels of p-hacking-type of manipulations by the researchers (among other components) to argue provocatively that most published research is probably false.

One key way to address p-hacking is through the use of transparently reported Pre-Analysis Plans (PAPs) as discussed in Chapter 2 Transparent Research.

PUBLICATION BIAS

Empirical research suffers from publication bias when results in published studies are systematically unrepresentative of conducted studies. The most common manifestation of such bias occurs when studies with positive results (i.e., those that reject the null hypothesis) have a much higher likelihood of being published than studies with null results[6].

Publication bias is the result of three forces operating in the same direction. First, researchers might perform or report a subset of analyses that produces positive results, and not report null or unintuitive results. This is the problem of p-hacking discussed above. Second, editors or referees might favor papers with positive results over null results and therefore not accept papers with null results., Third, researchers might not write up the final paper for a study that finds null results – a problem labeled as the "File Drawer Problem" (Rosenthal, 1979).

The mechanism of the File Drawer is particularly relevant for the IDB as a producer and contractor of hundreds of impact evaluations. Each of these evaluations represents a potential data point in a body of evidence, yet a large fraction of these results could end up in the (electronic) file drawer of many researchers and analysts inside and outside the bank. Given its position, the IDB has in its reach the ability to both measure the severity of the problem, and to address it directly with the tools describe in this document.

The evidence for publication bias is growing. One of the first studies that provided suggestive evidence of publication bias in economics was a meta-analysis on the minimum wage literature (Card and Krueger, 1995) that showed that t-test and sample size were negatively correlated (for a given effect, this relation should be positive). More recent and conclusive evidence comes from research that tracks the publication status of originally planned studies in a given field by the type of results – strong, weak, or null. In medicine, Turner et al. (2008) review all of the studies that were submitted to the Federal Drug

[6] In can be argued that null studies are less prevalent in the published literature due to a conservative stance from editors and referees to identify true null results, as oppose to underpowered or poorly implemented interventions. However, this is unlikely as discussions of power have been, until recently, absent from empirical studies. For a discussion on the lack of power treatments in economics see McCloskey (1996)

Administration for the approval of a set of antidepressants. They found that, among the studies submitted, 97% of those with positive results were published, while only 33% of those with negative results were published[7]. Franco et al. (2014) conducted a similar analysis among studies in economics, political science, sociology, and psychology that were awarded highly competitive resources by the National Science Foundation. They found that 22% of studies with null results were published, while 61% of those with strong results were published. More recently, Andrews and Kasy (2017) develop non-parametric estimations of publication bias, finding that null results have a 3-4% chance of being published relative to results that reject the null.

Publication bias matters because the IDB and its partners make decisions on policies and investments based on evidence that comes from a body of literature. If the literature is systematically biased, as Ioannidis et al. (2017) recently suggest for the economics literature, then those decisions may not be the right ones.

In Chapter 2 Transparent Research registrations and reporting guidelines are discussed as a solution to prevent publication bias. For a more extensive discussion of publication bias, see chapters 5 and 6 in Christensen, Freese, & Miguel (Forthcoming).

LOW REPRODUCIBILITY

Perhaps the most infamous component of the credibility crisis in science has to do with failed replications[8]. The inability to replicate the work of previous scholars is in direct contradiction with scientific norms outlined above. Failures to replicate have manifested across fields. Unfortunately, the terminology across fields has created some confusion regarding what is meant by “replication”.

Clemens (2015) identifies dozens of different definitions across fields and suggests his own taxonomy. For simplicity, this document will refer two key concepts: *replicability* and *computational reproducibility*. We refer to replicability as the practice of repeating a methodology using new data sets with similar characteristics (related to the idea of external validity). We refer to computational reproducibility as the practice of running the same code over the same data and obtaining the same results as those presented in the original reported analysis (starting from raw data to the final output). Whenever the term reproducibility is used throughout this document, it refers to computational reproducibility.

Before reviewing the evidence on replications, it is important to mention that a finding might not replicate for several reasons. A common, and plausible, explanation is that the original study was a false positive. But other explanations are also valid: the intervention might not have been delivered properly, measurement could have not been comparable, or the true underlying parameter could differ across populations. Current research on replications is actively exploring these issues, pointing towards a large fraction of false positives (Klein et al., 2018), but much research is still needed to provide any general conclusion regarding the relative importance of this reasons.

[7] In this context the authors define negative results as those that fail to reject the null.
[8] The credibility crisis is often referred to as the “reproducibility crisis” (Baker, 2016)

Regarding the evidence behind replications: a large-scale effort in psychology attempted to replicate the results of 100 studies. While 40 of these replicated, 30 studies failed to do so and evidence was deemed inconclusive for the remaining 30 (Collaboration et al., 2015). In a similar exercise in behavioral economics, 11 of 18 studies were replicated (Camerer et al., 2016). More recently, a replication of all social science experiments published in *Science* and *Nature* found that only 13 out of 21 experiments had similar results (Nosek et al., 2018).

Regarding the evidence behind reproducibility: previous evidence had suggested that the ability to reproduced were alarmingly low in a specific journal (Dewald et at., 1986). More recently studies have begun to systematically quantify the reproduction rates across fields and journals. Chang and Li (2015) assesses the computational reproducibility of 67 papers in macroeconomics and were able to obtain qualitatively similar results for 29 of them[9]. Gertler et al. (2018) similarly attempted to re-run the analysis code from a sample of 203 empirical papers from leading journals in economics and was able to obtain the same results for only 14% of the papers.

In Chapter 3, this document focuses on Computational Reproducibility as a required first step in improving researcher workflow.

ETHICS

Before there was a credibility crisis that called for more transparency and reproducibility in social science research, there were a series of ethical crises that called for stronger governance in research. Notably, the creation of the US federal government's Belmont Report (1979) was directly informed by the Nuremberg Trials (and Nuremberg Code[10]), the 1932-1972 Tuskegee syphilis study[11] (Brandt, 1978), as well as several other high level ethical concerns in research (Beecher, 1966).

The Belmont Report defined three ethical principles - beneficence, respect for persons, and justice – laying a foundation for protection for human subjects in research given the vulnerabilities of research subjects and the type of data required for the research - particularly personally identifiable and sensitive data. These principles, as well as the subsequent US regulations for human subject research established in 1991 (the Federal Policy for Protection of Human Subjects, also known as the Common Rule), aimed to mitigate risks to research subjects from harm and/or exploitation.

Although the establishment of these principles and regulations have likely mitigated risks to research subjects, they have not eliminated unethical research. For example, the case of using biospecimens from the Havasupi Tribe in Arizona for research purposes beyond the original scope without consent (Sterling, 2011) highlights that research may be cleared by an Institutional Review Board (IRB) and in line with regulations, but not ethical. This is a useful case for highlighting two challenges to conducting ethical research: first it highlights that IRB review alone may not be sufficient for contextually and culturally sensitive consideration of the proposed research and its methods; and second it highlights that the individual research subjects are not the only ones for which the research may pose a risk – the general

[9] 6 papers could not provide proprietary data.
[10] https://history.nih.gov/research/downloads/nuremberg.pdf
[11] Tuskegee study involved United States Public Health researchers who actively withheld treatment from 400 African American men with syphilis to study the natural course of untreated syphilis.

subject population (i.e. all members of the Havasupi tribe not just those in the study population) or other bystanders may also be negatively (or positively) impacted by the research. Although a more relevant consideration for project teams is identifiable survey data - rather than biospecimen data - sharing, these two challenges remain relevant when considering the ethics of IDB research.

A third additional challenge for conducting ethical research is that risks may evolve and change over time. At the point of IRB review and research implementation, there may be little to no risk to research subjects (or other bystanders), however depending on the purpose of the study and the type of data collected, risk may increase at a future date if certain contextual and/or cultural changes occur (Knott, forthcoming). How project teams consider and monitor risk over time is also an important element of ethical research. Given that IDB research is conducted in dozens of countries with varying laws and regulations, this document aims present best practices for ethical research to consider these challenges alongside any necessary local requirements.

CONDUCTING TRANSPARENT, REPRODUCIBLE, AND ETHICAL RESEARCH

Reproducible and ethical research practices contribute to the quality of the research design, implementation, and dissemination, while transparency allows the scientific community to assess the research quality and credibility. Consider the following:

- Research that is transparent, reproducible, and ethical. This occurs if project teams follow international standards – and when applicable national standards - for ethics, transparency and reproducibility through the public, limited, and/or restricted-access sharing of study design materials, data, and code.
- Research that is reproducible and ethical, but not transparent. This occurs if project teams follow international standards for ethics and reproducibility, through proper documentation of study design and analysis and version control, but do not, or are unable to, fully disclose and share study design elements, data and code required for full transparency. The analysis can be reproduced, but only if researchers can obtain access to the documentation, data, code.
- Research that is transparent and ethical, but not reproducible. This occurs if project teams follow international standards for ethics, are fully transparent with study design elements, and share de-identified data and code. However, if the analysis relies on identifiable data that cannot be shared due to human subjects' privacy protection[12], the provision of data and code does not facilitate computational reproducibility. Similarly, if project teams have not created a reproducible workflow, computational reproducibility may not be feasible even with access to the data.
- Research that is ethical, but not transparent and not reproducible. This occurs if project teams determine that study design elements, data and code cannot be shared due to human subjects' privacy protection, or the team is otherwise not transparent with its materials. It may also occur when best practices in reproducibility, through proper documentation of study design and analysis and version control, are not followed.
- Research that is transparent and reproducible, but is not ethical. This occurs if project teams follow best practices in terms of transparency and reproducibility, but fail to follow international standards in ethical research, such as sharing data and code that does not adhere to promises of confidentiality.

[12] It should be noted that for some data, this may vary by time. It may be that the risks associated with disclosure decrease over time so that the identifiable data may be released or otherwise available in the future.

Given that best practices for transparency and reproducibility may create tension for best practices for ethical research, this document includes all three so project teams may consider for their context which transparent and reproducible practices they can implement while also maintaining ethical practices. With this goal in mind:

Chapter 2 presents best practices and tools for transparent research, including:

- **Study Registration**, a brief, documented description of a study before data is available for analysis.
- **Pre-Analysis Plans (PAP)**, the extensive description of a study before data is available for analysis.
- **Registered Reports**, a method of submitting an article to a journal for review whereby editors and peer reviewers assess a manuscript prior to data analysis.
- **Reporting Guidelines** to define which content is to be presented in a manuscript in order to facilitate systematic reviews and meta-analysis.

Chapter 3, on Computationally Reproducible Research, includes:

- **File Management** to ensure clear structure and format of the files in a research workflow. Additionally, a proper data back-up plan should specify how the data will be stored.
- **Version Control strategy** to ensure tracking of the complete history of all code used in a project. This is critical to provide reproducibility of current and previous iterations of the analysis.
- **Code Readability and Dynamic Documents** to standardize coding style and legibility within a team. These practices are meant to facilitate the reproduction of the analysis across multiple researchers (including the original researcher on a later occasion).
- **Code Sharing and Data Sharing**, practices and tools recommended for the sustainable archival and dissemination of data and code.

Finally, Chapter 4 presents best practices and tools for ethical research, including:

- **Community Stakeholder Engagement (CSE)** to inform project teams of the needs and priorities of the communities they will engage with during the research.
- **Understanding of legal requirements**, including international and national data privacy and protection laws, as well as any legal regulations on research and protection of human subjects, to guide research practices.
- **Independent review** that is culturally and contextually sensitive to assure, both in advance and by periodic review, that appropriate steps are taken to protect the rights and welfare of those impacted by the research. This is often done through, but not limited to, Institutional Review Boards (IRBs).
- **Research Protocol** to document the ethical research design and implementation plan. A well-done research protocol can also facilitate future reproducibility efforts.
- **Data Management Plan**, often a section of a Research Protocol, as a useful tool for documenting who will have access to what data, when, and how.
- **Informed consent** to ensure research subjects are informed on the objectives, duration, and description of the research, its expected benefits and risks, and that their participation is voluntary. Highly relevant for transparency and data sharing, this is also where promises of confidentiality are made, or not, and who has access to what data when is described for the research subject.

- **Proper data collection, storage, transfer and disposal** to ensure privacy protection and risk mitigation.
- **De-identification** of documents and data to ensure privacy protection and risk mitigation**.**
- **Data sharing and dissemination** activities that adhere to research protocol and data management plan.

The transparent, reproducible, and ethical research practices described in this document are complementary and should be considered in conjunction, although emphasis on any single topic will depend on the specific nature and circumstances of each individual research activity. The best practices presented in this document are intended to expand the project team's toolbox for how to design, implement, and disseminate high quality, credible research. That may mean applying just one or two or all of these best practices, depending on the requirements of the study, context, and other external factors. For example, as a first step the IDB could suggest that all of its impact evaluations contain the following elements of TRE research:

Table 1: Suggested Prioritization of TRE Research Practices for IDB

TRE Practices/Deliverables	Strongly Recommended	Recommended	Optional
(T) Study Registration	X		
(T) Pre-Analysis Plan (PAP)	X		
(T) Registered Report			X
(T) Reporting Guidelines	X		
(T) Code and Data Sharing		X	
(R) File Management	X		
(R) Version Control Strategy	X		
(R) Code Readability and Dynamic Documents		X	
(R) Code Sharing and Data Sharing		X	
(E) Community Stakeholder Engagement (CSE)		X	
(E) Understanding of legal requirements	X		
(E) Independent review (IRB and/or other)	X		
(E) Research Protocol	X		
(E) Data Management Plan		X	
(E) Informed consent	X		
(E) Proper data collection, storage, transfer and disposal	X		
(E) Data De-identification of Documents and Data		X	

CHAPTER 2: TRANSPARENT RESEARCH

Chapter 1 described the main problems driving the credibility crisis in science, including publication bias, p-hacking, and failures to replicate or reproduce findings. Chapters 2 and 3 present project teams with best practices for mitigating these specific problems.

This chapter first discusses the concept of registration as brief record of the study before it is conducted. Second, pre-analysis plans are introduced as a more in depth complement to registration with the goal of specifying in detail the analyses ahead of time. Third, registered reports are described as a mechanism to incentivize the generation of pre-analysis plans. Finally, reporting guidelines are introduced as a set of standardized procedures to write down the study results. Standardized reporting facilitates tracking entire bodies of knowledge within a topic[13], reducing publication bias.

All four sections of this chapter are meant to be used in prospective studies, where access to the final data set has not been obtain yet, and are not uniquely restricted to randomized controlled trials (RCTs).

REGISTRATION

DESCRIPTION

The main purpose of study registration is to track the complete body of knowledge in a topic of research, regardless of the direction and magnitude of the results. In a world without publication bias, this could be achieved by surveying the published literature.

Unfortunately, the file drawer problem, or the tendency of negative or null results to go unpublished, generates significant publication bias. One solution is the use of registries to catalogue all research conducted on a given topic. The best way to ensure that this record is created is to require registration, before the analysis is carried out. These registries are meant to be searchable sources of information that serve as a permanent and public record of all the research done in a specific field, independent of the final results.

The first registries were created in the biomedical sciences[14] and quickly became a requirement for publication in top medical journals (De Angelis, 2005). This has produced a drastic reduction in the observed fraction of positive findings. Kaplan and Irvin (2015) document how this fraction went from 57% to 8%, after requiring registration in clinical trials evaluating drugs and dietary supplements to treat cardiovascular disease. In the biomedical sciences, registries can be extensively detailed and registering a study may require significant effort. In the social sciences, and especially in economics, two terms are used to distinguish between registrations with low and high level of detail. The term *registration* in this setting refers to a document with minimal information on the planned study, whereas the term *pre-analysis plans* refer to documents that contain extensive, detailed information on most of the key

[13] As of end-2018, more than 140 journals have integrated the use of registered reports into their publishing practices. See https://cos.io/rr/

[14] clinicaltrials.gov was created in 2000.

analytical choices that researchers anticipate for a given study. In this document the term registration will refer specifically to this last definition[15].

As explained below, there are several different registries a project team may identify as appropriate for study registration based on the research methodology, location of the research, and the type of research. Regardless of these factors, all registries tend to require the same high-level, minimal information regarding the study. Study registries typically require: title, authors, study country, status, keywords, abstract, start and end dates, outcomes, intervention information, basic research design, whether or not treatments are clustered (when performing an RCT), and Institutional Review Board (IRB)[16] information.

As with any new practice, registrations (and pre-analysis plans) have come up against some resistance. Nosek et al. (2018) provide a list of common objections to registrations, as well as counter points to such objections. These are presented below.

Table 2.1: Challenges about Registrations and its Answers (Nosek et al., 2018)

Challenge	Response
Changes to intervention during study administration	Changes are expected. Providing a record of all changes is compatible with a registration (and much more transparent).
Discovery of assumption violations during analysis	As you learn something about the data, you can update the registry (again with a record). Ideally researchers should have standard operating procedures for deviations (Lin and Green, 2016).
Data are preexisting	If its secondary data that are preexisting, then registration might not apply.
Longitudinal studies and large, multivariate datasets	Use previous results as exploratory and register confirmatory test for upcoming waves.
Many experiments	If experiments are frequent and inexpensive, use these experiments to verify results, publishing all the replications (Coffman and Niederle, 2015).
A program of research	Many studies might lead to some of them to generate positive results by chance. Register each study and present all the results when discussing the overall program of research.
Few a priori expectations	Even after very little research, we all generate priors about hypothesis. Register them as they emerge.
Competing predictions	Register both, and learn.
Narrative inferences and conclusions	No problem in highlighting positive results over null, just provide access to all the hypothesis registered.

[15] In psychology, both terms are commonly known as pre-registration (Nosek et al., 2018).
[16] IRBs are discussed in Chapter 4 Ethical Research

When it comes to registration, unfortunately, researchers' incentives do not always align with those of the wider scientific community. In registering a study, researchers are making a public statement about their future work and constraining their own ability to tell a story that can cleanly explain their results, regardless of whether or not they used selective reporting, p-hacking, or otherwise non-reproducible practices to make such a story (this is precisely the point of research transparency). Additionally, registration has a small cost in time and few clear benefits for individuals. For this reason, registries require only minimal basic information, such that the marginal effort to register a study is almost negligible. An additional concern for researchers is that their idea may be "scooped" by other scientists who see their registration. However, while registrations eventually become public, researchers can choose to embargo portions or the entirety of a registration for several years.

Given the lack of individual incentive for a researcher to register a study, publishers and funders have followed medicine's example by making registration mandatory. For example, in 2018, all nine American Economic Association (AEA) journals provided the following statement:

> "The American Economic Association operates a Registry for Randomized Controlled Trials (RCTs). As of January 2018, **registration in the RCT registry is mandatory for all applicable submissions**. You will be asked to provide your AEARCT identification number in the online submission form. Please include your number in the acknowledgement footnote in your paper, as well." [AEA website, emphasis added].

HOW TO

Registering a study requires a low level of effort as it requires little information. Also registering with one service (eg. AEA Registry) does not exclude the researchers from registering in another website (eg. Clinical Trials). All of the registration websites will make the study public (after an optional embargo period) and researchers will not be able to modify its content without leaving a record.

For project teams, the primary and relevant registries are listed in Table 2.2. To register a study, project teams should create an account in at least one of these services and create a record with the basic information described in the previous section. Registrations are encouraged (and becoming increasingly mandatory) for all prospective studies, not only RCTs.

Table 2.2: Registries

Type of research	Registry	Embargo	Notes
RCTs in non-health-related fields*	AEA Registry (socialscienceregistry.org)	Until completion of the study	Default for RCTs in economics
RCTs in health-related fields*	Clinical Trials or ICTRP (clinicaltrial.gov)	Available but undefined limit	Highly standardized
Impact Evaluation/Dev Economics	3ie (ridie.org)	Unavailable	High level of detail. Accepts RCTs and quasi-experimental
Governance/Political Science	EGAP (egap.org)	Up 18 months after registration	Experiments and observational studies in governance and politics
Any other	OSF (osf.io)	Up to 4 years after registration	Multiple formats: short, long, structured, and open ended

*Health fields are defined by the target journal. For example, if your goal is to publish in a health-related journal, your study would be considered a health-related paper.

For impact evaluations, it is also important to note that most of the information required in a registration should be available in the Monitoring and Evaluation Annex of a POD[17].

[17] It is recommended that project teams inform country partners whenever a registration takes place.

Box 2.1: Example of a registration in the AEA RCT Registry[18]	
Title	Every Child Counts! An “at-scale” test of an early mathematics curriculum
Authors	Esther Duflo, Elizabeth Spelke
URL	https://www.socialscienceregistry.org/trials/3143
Country	Embargoed
Status	In development (as of September 2018)
Keyword	Education
Project Start Date/End Date	2018-07-23 / 2020-12-31
Additional Keywords	Early Childhood Education
Abstract	“The performance of primary schools in developing countries…”
Intervention Start Date/End Date	2018-09-01 / 2019-03-31
Primary Outcomes[19]	Can we harness children’s innate capacities at the foundations of mathematics to give preschool children the skills and confidence to succeed in school? Can we extend our curriculum to enhance children's math learning in primary school? Can we make our interventions “robust” enough to be implemented at scale in pre-schools and in the early grades of primary education?
Experimental Design	Designing and evaluating a modified curriculum, linking the non-symbolic games to the symbol systems of elementary school mathematics. A new RCT would test its effectiveness against both the government's standard preschool and Grade 1 curriculum. This curriculum will be delivered by teachers recruited by the Directorate of Education, Delhi.
Randomization Method	Randomization done through Stata Code
Randomization Unit	Cluster based on the district the school belongs to, number of sections at grade level, school working hours and school gender restrictions.
Sample size/number of clusters	A total of 143 schools will form a part of our study.
Sample size/number of observations	3000 students
Sample size/by treatment arms	70 schools will be randomly selected as treatment schools. The remaining schools will form the control group.
IRB name/num id	MIT/1805377780

[18] Authors granted permission to re-publish in this Technical Note.
[19] Here the authors choose to report the outcomes in the form of question marks.

PRE-ANALYSIS PLANS

DESCRIPTION

As discussed above, while registering a study can help to reduce publication bias, it does not always suffice to prevent p-hacking. Any given study contains many analytical choices still allowing researchers to, consciously or not, choose a set of preferred specifications (and robustness test) that achieve statistical significance. Pre-Analysis Plans (PAP) are an approach to prevent p-hacking and to help researchers clearly identify the hypothesis to be tested in a study[20]. PAPs are extensive methodological descriptions of the analysis to be performed *before* the endline data is collected[21].

Researchers have the freedom to specify any number of hypotheses. However, the tradeoff between number of hypothesis and statistical power should be acknowledge. For this purpose, it is necessary to know how many hypotheses will be tested with their respective specifications. When testing multiple hypothesis, the researchers should specify how to adjust the p-values (with methods like FDR, or FWER) and provide a credible proof that no additional hypothesis were tested. A PAP does precisely this, by requiring researchers to publicly specify before-hand the main hypothesis and how they will be tested.

In addition to providing accurate statistical testing, PAPs have the benefit of shielding researchers from having to hide undesired results. After an impact evaluation takes place, funders of the intervention might have strong incentives not to disclose undesired results (e.g., limited or no impacts of an intervention). A PAP provides a strong resource for researchers to deal with situations like this – publishing the study (or at least its design with the outcomes that it will look at) is already out of their hands.

The universe of studies that can benefit from a PAP is quite large. PAPs are recommended for all prospective studies, where the researchers can provide proof that they have not seen the final data of the project. This means that, in addition to RCTs, PAPs apply to quasi-experimental studies where the data has not been collected or obtained yet. Moreover, even though the usage of PAP (and registrations) is meant to increase the rigor of scientific publications, the IDB could benefit from applying it to studies that are not intended for publication in academic journals. In addition to increased rigor, a large benefit for the bank could come from the protection against the threat of censoring described in the previous paragraph. For a more detailed discussion on how to used PAPs in observational studies, see Burlig (2018).

Unlike a registration, PAP do require a high level of effort. As detailed next in the “How To” subsection, a thorough PAP should resemble a paper without the results sections. However, this does not imply that the researchers do any additional work for a project; researchers must always describe what they will model and why. A PAP only shifts the moment in time when this is done, to before the analysis is conducted. This can be disruptive to the traditional workflow of a researcher, but also provides significant and immediate benefits, including a better research design, a strong signal of rigor, and the satisfaction

[20] Another method to add transparency and prevent p-hacking is to split the data set into two sections: the first section is used to explore the data and formulate hypothesis, and the second sections is used test those hypotheses. For more information on this novel approach see Anderson and Magruder (2017)

[21] The optimal timing for a PAP is still up for debate. The earlier a PAPs is carried out, the least space will be for p-hacking. However, doing a PAP later in the project (after baseline and before endline) can benefit the amount of detail that can be added to the PAP. PAPs however should never be submitted after the endline data has been obtained.

of doing science in much closer accordance to the Mertonian norms describe in the introduction (especially disinterestedness and organized skepticism).

To illustrate how a PAP could improve the research design, consider the following situation: when designing the identification strategy of a study, a researcher might consider using the distance to the point of provision (e.g., healthcare or schooling) as an instrumental variable. As they think about the threats to the validity of the design, a possible criticism might be that people who value the service most (eg. education) may choose to live close to the best schools. In a traditional workflow, the researcher cannot do much more than speculate and hope that this will convince reviewers. In a workflow with a PAP, the researcher will go through such considerations in their in the design stage, gaining the opportunity to collect information on location choices and provide evidence on this plausible correlation.

A final and common concern has to do with the idea that a PAP prevents researchers from performing additional analysis or that it is hard to foresee all possible contingencies before running the analyses. PAP do not stop or discourage researchers from running additional analysis, they only ask for those analysis to distinguish from the original set of hypotheses. Regarding the inability to plan everything ahead, researchers can still do a PAP and record adjustments and deviations as they occur. See the additional resources section for examples and protocols.

Table 2.3: Criticisms to PAPs and response (Christensen et al., Forthcoming)

Critique	Response
PAPs take too much time and are too difficult (Olken 2015)	A PAP does not increase the total amount of work dedicated to a research project, though it does change the timing of the analytic component (which nearly always requires a high level of effort and time).
Scientific discovery often comes from surprises. PAPs stifle discovery (Olken 2015)	PAPs do not prevent researchers from doing exploratory work; they only require researchers to be clear about the objectives of their analyses. Analyses can be either exploratory (with the goal of discovery) or confirmatory (with the goal of testing hypotheses established before looking at the data), but exploratory analyses should not be presented as confirmatory.
If replications are cheap they will rule out false positives, making PAPs irrelevant. (Coffman and Niederle 2015)	Very few experiments are inexpensive as to perform many replications. Moreover, most of the false positives have been identified where experiments are least expensive (lab experiments).

HOW TO

There is no one specific recipe for how to write a PAP, but there are many examples and resources. Glennerster and Takavarasha (2013) and Christensen et al. (Forthcoming) provide checklists for PAPs. Additionally, a template for a PAP was created by Alejandro Ganimian and is available on the OSF here.

PAPs should be posted in a public repository that can provide a digital time stamp. Typically, the chosen repository is the same as the one for the registration discussed above, adding the PAP as a later attachment (with its corresponding time stamp). At the latest a PAP should be submitted before the researchers can access the dataset that contains the final outcomes and treatment status (i.e., endline data for RCTs). Before this critical deadline there is a trade-off for when to submit a PAP. The earlier a PAP is written up makes less likely that the analytical choices where made on the basis of expected results. But at the same time the later a PAP is submitted, there is more contextual information to write a more comprehensive plan.

According to Christensen et al. (forthcoming), PAP should include, at a minimum, the following elements:

1 – Study Design. For RCTs, are there multiple treatments or a single treatment? Detail the randomization process. For quasi-experimental, declare covariates and the estimation method (regression discontinuity design, instrumental variables, difference in difference, propensity score matching) with detailed specification.
2 – Study Sample. Define the sample frame. Describe strategies to deal with non-response, attrition, non-compliance with treatment assignment, and missing data. When performing secondary data analysis (on pre-existing administrative or survey data), the researcher should specify precise file and survey weights to be used.
3 – Outcome Measures. Define in detail the outcomes to be used in the analysis. Distinguish between primary and secondary importance to the main research questions. For each outcome, a clear formula or code should demonstrate how the outputs will be constructed (i.e., exact inputs and transformations).
4 – Mean Effects Families. When combining multiple outcomes into an index, the PAP should pre-specify all the elements behind the index (i.e., variables and weights).
5 – Multiple Hypothesis Testing Adjustment. Declare how to adjust the p-values of multiple tests. This can be done by adjusting the Family-Wise Error Rate (FWER) or choosing a specific False Discovery Rate (FDR) (see Glennerster and Takavarasha (2013)).
6 – Subgroups. Even a few baselines variables can be used to construct a very large number of subgroups. And each of these subgroups can be rationalized ex-post as a relevant group for the analysis. For this reason, declaring in your PAP the groups of interest greatly increase the quality of your analysis. While declaring subgroups, it is also recommended to use multiple hypothesis testing adjustments.
7 – Direction of Effect (optional). When declaring the direction of the effect in the PAP, there is an important gain in statistical power. The rationale for the direction has to originate with one of the mechanisms underlying the causal chain. This point should be included only when there is a strong prior around the expected sign of the results.
8 – Exact Statistical Specification. Define if regression models are linear or generalized linear. List control variables and fixed effects (when appropriate). Specify how standard errors will be computed (e.g., robust, clustered, etc.).
9 – Structural Model (optional). If the study will estimate a specific parameter of a model derived from micro-foundations, then include this information in the PAP. This information should include specific functional forms of utility functions or profit maximization functions, specific constraints, and the underlying rationale.

10 – Timestamp. The main objective of a PAP is to pre-specify analytical choices *before* the final data is available. Without a verifiable timestamp, the PAP loses much of its value.

PAPs are submitted as an additional document to the registration described in the previous section.

EXAMPLES

The first PAP in economics is from Neuwmark (1999, 2001) on the effect of a prospective raise in the minimum wage, an observational study. In more recent years PAPs have become predominant in field experiments, particularly in development economics. A now classic PAP is the work of Casey et al. (2012) on the effects of a Community Driven Development (CDD) intervention on the economic and institutional development of communities in Sierra Leone. Quality of institutions is a complicated construct and there are many ways to measure it. To strengthen the quality of their analysis, the authors created an extensive PAP that outlined how to construct their outcomes, how to assign treatment, and how to estimate the specific regressions. The authors found effects on "hardware" improvements that demonstrate that the program was executed properly, but also found no effect on their pre-specified definition of institutions. To emphasize the importance of PAPs, the authors p-hacked their own data and demonstrate how they could have found significant results in either direction.

Regarding their PAP, Casey et al. (2012) emphasized four key elements for future researchers:

1. *Timing of PAPs.* The authors identify two stages: (1) the general areas of likely impact should be defined before the project is implemented; (2) describe the analysis in detail after implementation, but before the endline data has been collected.
2. *Defining the Hypotheses and Outcomes.* The authors clearly document all their hypotheses in detail. This is illustrative of how a PAP does not preclude the authors from adding additional hypotheses later on or from rearranging their order to create a cohesive story. What is important is that they make clear what changes took place after the PAP was written, their rationale, and both versions are presented.
3. *Choosing the Optimal Level of Detail.* When describing the regressions to be run in the analysis, the authors did list all the specifications and covariates. However, they did not make explicit what their preferred specification was among robustness checks and mentioned that this oversight could have weakened their results (fortunately for them, the estimates did not vary much across robustness checks).
4. *Accounting for Multiple Inference.* The authors tested 12 different hypotheses and adjusted their p-values using the Westfall and Young (1993) free step-down resampling method. This method is an example of FWER correction and involves ranking all p-values (from lowest to highest in value) and rejecting the null of all hypotheses up to rank *k* if $pvalue_k < \frac{k}{R} * \alpha$,where R is the total number of hypotheses (R=12 in this example) and α is the chosen critical value for significance.

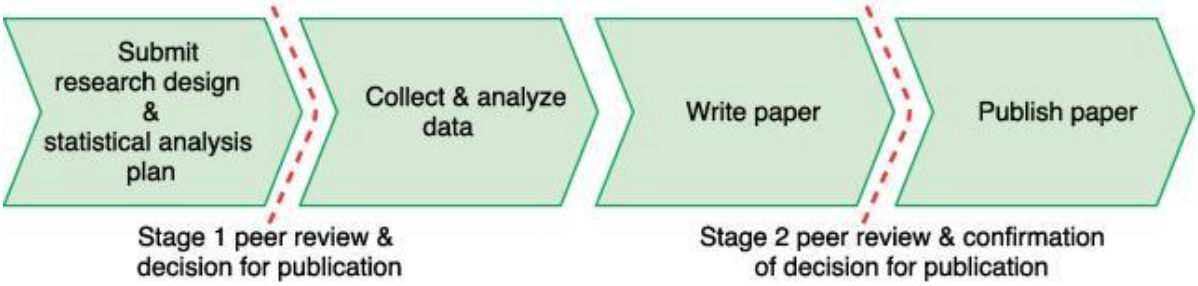
REGISTERED REPORTS

As described in previous sections, though PAPs can help to address p-hacking (and strengthen the quality of research in general), at an individual level, a researcher may not have the incentives to adopt such practices. Registered Reports (RR) are a format of peer review and publication that aims precisely

to address this lack of incentives. An RR is a paper that goes through peer-review *before* the final data is analyzed. Reviewers judge the contribution of the paper on the basis on the relevance of the question and the quality of the research design, providing a decision that is independent of the results obtained.

Once a paper receives "in-principle acceptance," researchers can carry out their protocol without concern for how publishable the results will be. In a second stage of peer review, reviewers only verify that the authors did what they promised in the first stage, and the paper is published in the journal as a scientific article. Figure 2.1 (from a note on the JDE in the BITSS website) illustrates this point.

Figure 2.1: Research timeline and review for a Registered Report (source: JDE website)



Since its first proposal in 2013 in a neuroscience journal, the idea has gained increasing traction among journal editors, and as of November 2018, 142 Journals have adopted the practice either regularly or in a special issue. In March of 2018 the *Journal of Development Economics* (JDE) became the first journal in economics to adopt the format. All prospective research designs are accepted (RCTs and quasi-experimental). The key requirement is that authors provide proof of lacking access to the data by the time the RR was drafted. The JDE is a top journal of high interest to development economists, so the incentives of a large numbers of researchers at the IDB are now aligned with the RR format.

In their Author Guidelines, the JDE makes explicit that papers using the RR format should be judged based on the following dimensions: (i) importance of the research question; (ii) logic, rationale, and plausibility of hypotheses; (iii) methodology and statistical analyses (including power calculations where appropriate); (iv) compliance with the mandatory replication policy; and (v) that pre-specified tests are sufficient and robust to test stated hypothesis[22].

Note that all of these elements are included in the description of what would constitute a good PAP from the previous section. Now writing PAPs is not only good for science, in the sense that it reduces p-hacking, but also good in a very concrete way to each individual researcher. Transforming a highly detailed PAP into a RR should entail mainly formatting, such that the PAP takes the structure of a paper, and emphasizing the previous literature and motivation of the problem.

[22] For the interested reader, Appendix C reproduces a detailed checklist of all components expected in a RR at the JDE.

REPORTING GUIDELINES

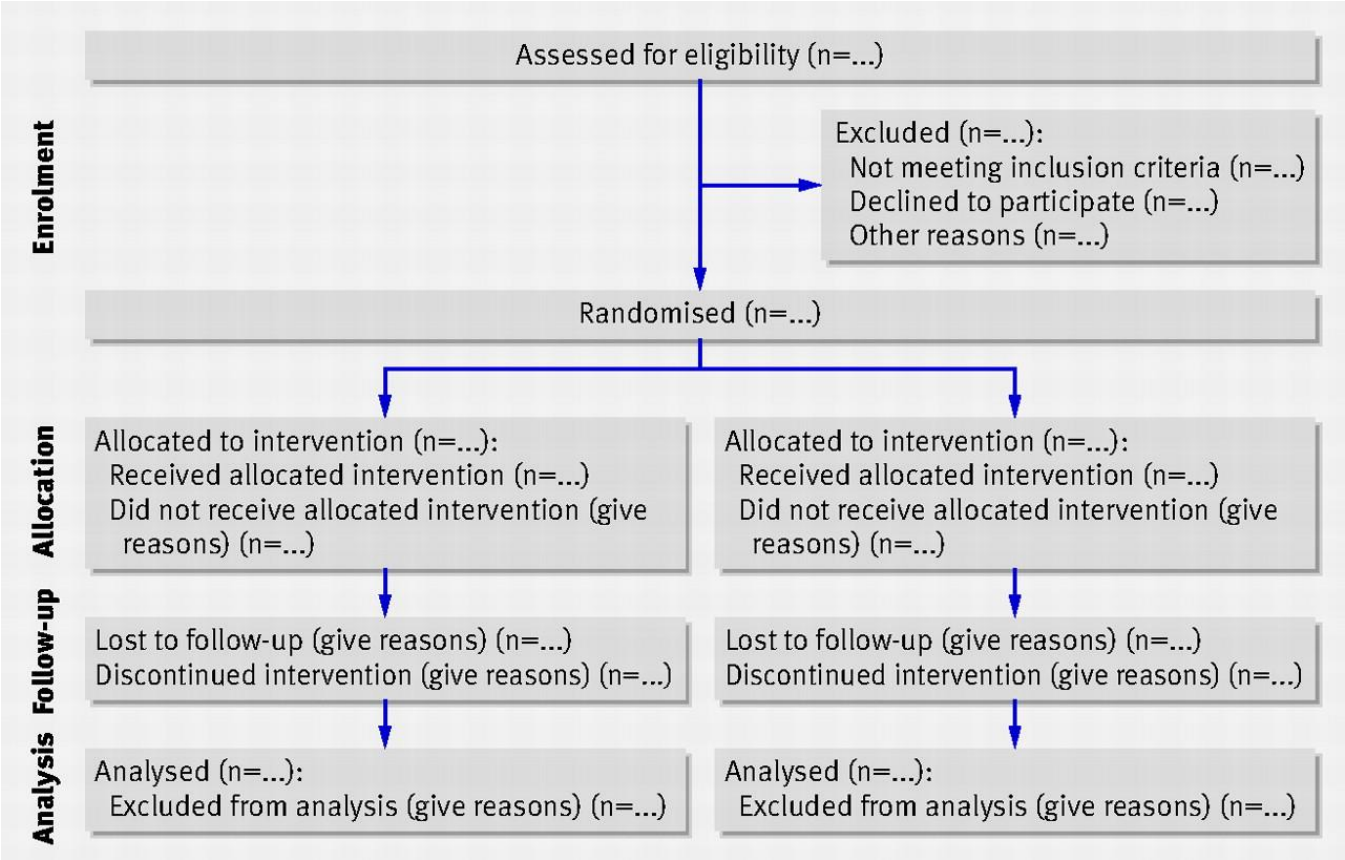
Reporting guidelines are a standardize procedure to report the research output. The evidence generated by any single impact evaluation (or piece of research in general) is meant to contribute to a body of knowledge. For example, the evidence produced by the IDB would ideally be used to inform future policy decisions. It is possible that the hundreds of impact evaluations produced by the bank could eventually feed into the ex-ante economic analyses used to inform future loan decisions. In order to aggregate such information, the relevant studies have to be found first.

When building a body of knowledge, finding all relevant studies can be challenging and resource-intensive. In medicine, this has given birth to a growing branch of research called systematic reviews. Researchers and private firms now specialize in defining the correct universe of potential studies, the specific rules of inclusion and exclusion, and the often-labor-intensive process of screening each study to extract the relevant information. Even when a study is identified, picking out the relevant statistics can be challenging and time-intensive. To address these costs the biomedical sciences have begun using Reporting Guidelines. Reporting Guidelines are designed for specific fields and methods to instruct authors and journals on how to present outputs in a paper. The goal is to standardize as much as possible the format of the results in order to streamline subsequent study aggregations.

The most commonly used guidelines are the CONSORT guidelines (CONSORT Statement, 2010), which are designed for the reporting of randomized control trials. These guidelines require the authors of any RCT (most often in medicine, but also applicable to other contexts) to report two homogenous outputs: (1) a figure representing the main sampling information of the study (Figure 2.2), and (2) a checklist that verifies the existence and page location of required items in the paper.

The CONSORT guidelines are considered general purpose. Some systematic reviews, however, rely on different methodologies and focus on more specific fields. For this purpose, the EQUATOR-Network provides an online catalog of 405 reporting guidelines. These are largely focused on approaches for the biomedical sciences, though guidelines for other disciplines have emerged.

Figure 2.2: CONSORT Flow Diagram
Source: https://www.bmj.com/content/340/bmj.c332



In the social sciences, the field of psychology has taken the lead in developing reporting guidelines. The Journal Article Reporting Standards (JARS) Guidelines provide a CONSORT-type of framework for experimental and observational research in social sciences (Appelbaum et al., 2018). Though they are designed with multiple studies per-paper in mind (a common feature in psychology papers), they are a good starting point, along with CONSORT, for developing reporting guidelines for impact evaluations. Various aspects of JARS could be used to develop reporting guidelines that are specific to impact evaluations. While there is not yet a commonly agreed upon set of reporting guidelines for impact evaluations, the World Bank Development Impact Evaluation (DIME) group provides useful guidance (Evans & Snilstveit, 2016), summarized below in Table 2.4.

Given its position as a large producer of impact evaluations, as well as a potentially large consumer of cost-benefit analyses, the IDB is well-positioned to pioneer the first set of reporting guidelines for economics with a focus on impact evaluations. A good starting point could be a combination of the recommendations outlined above and the checklist provided by the JDE for their RR initiative (see Appendix C).

Table 2.4: Minimum Reporting Guidelines For Impact Evaluations (Based on Evans & Snilstveit, 2016)

Reporting Goal	Reported?	Page #
Compute Effect Sizes		
Outcome data, separately, for treatment and control groups		
Sample standard deviation pooled across treatment and control groups		
Standard errors or confidence intervals of the treatment effect (for cluster RCTs, standard errors should be adjusted for clustering, and the intra-cluster correlation should be provided)		
Sample sizes for treatment and control groups (if clustered, number of clusters and average number of students per cluster), at baseline and at follow up		
Appraise Methodology		
Unit of allocation and unit of analysis		
Type of treatment estimate provided (e.g., ATE, ITT)		
Details about treatment allocation, including how any randomization was implemented and if it was successful (balance on pre-treatment variables)		
Clearly report and justify methods of analysis		
Describe the conditions in the comparison group, including distance to the groups receiving the intervention and any steps to address risks of contamination		
Report results for all primary and secondary outcomes clearly, including results that were not statistically significant or negative		
Quantify Costs		
Describe the intervention design in sufficient enough detail for replication (what was delivered, by whom, for how long)		
Describe what actually happened; document all deviations from original plan		
Provide a description of the context in which the program was delivered		
Report details about resource use and costs to facilitate cost-effectiveness analysis		

ADDITIONAL RESOURCES

Other relevant tools and references not cited directly in the main body of the chapter:

- PAP on Unconditional Cash Transfer (Haushofer and Shapiro, 2016): Paper, registration, initial PAP, and document describing all deviations.
- PAP on Oregon Health Insurance Experiment (Finkelstein et al, 2012): Extensive PAP (120 pages) This closely mirrored the final paper published. They also look at control endline data before writing the final PAP.
- Green and Lin (2016) published a set of Standard Operating Procedures on how to handle deviations from PAP in lab and field experiments for political science.
- Tool for simulating all analyses in R and perform power calculation: DeclareDesign.
- JDE resources on BITSS website: guidelines, FAQs, templates and additional information for the register reports submission format for the JDE.

CHAPTER 3: COMPUTATIONALLY REPRODUCIBLE RESEARCH

"An article about computational results is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result". Jon Clarebout 1992 (quoted in Buckheit and Donoho, 1995)

The quote above refers to the "Clarebout Principle," an idea that clearly articulates the rationale for the importance of complete computational reproducibility. A research project generates a large amount of knowledge, only a small portion of which is codified in the final paper. An important fraction remains only in the code (and notes) that executes the entire analysis. Full computational reproducibility is a collection of tools and best practices for recovering the entire knowledge output from a research project.

This chapter describes the elements required to reproduce the entirety of the results from an empirical analysis of a project (referred from hereon as just the project). The chapter is organized around four key areas of best practice: file management, version control, code readability and dynamic documentation, and the sharing of materials (i.e., code and data). The material in this chapter is primarily based on papers by Wilson et al. (2014), Wilson et al. (2017), Matthew and Shapiro (2014), and Christensen et al. (Forthcoming).

A common theme in the following sections is that best practices should be adopted *before* a project begins. From file management to sharing of materials, all the recommendations outlined below are meant to be considered at the planning stages of a project. Implementing any of these strategies in later stages will be much harder and may face strong resistance from previously established practices that lack reproducibility.

FILE MANAGEMENT

FILE STRUCTURE

Having a well-organized file structure within a project team is perhaps the practice with the highest benefit-to-cost ratio. Missing files are among the main reason projects become irreproducible, and a well-organized file structure can help prevent this problem. The overall message of this section is that researchers should establish one file organization standard and follow it. The specifics of this section provide suggestions on how to organize project files based on best practices identified in the literature.

A common suggestion is the idea of portability. A project folder should be self-contained, such that a complete copy of it should allow other users to re-run the analysis from the beginning. When a project is portable it is possible to set all the directory references in a relative fashion. For example, instead of referring to a file as 'C:\username\documents\project_folder\data\dataset1.dta' the file can be referred to as '~\project_folder\data\dataset1.dta'. This way, there is no need to rename every single directory call in a project folder, facilitating computational reproducibility.

A second common suggestion is to follow a standardized folder structure within the main project folder. Most recommendations focus on having a few high-level folders separating raw data, processed data, code, and documentation. Raw data is any data that was received by the researcher and has not been

processed in any way. Processed data include both intermediary files and final files to be used in the analysis. Code contains any program used to process the data (cleaning and analysis). Documentation should include the final report and all additional inputs required for the project such as questionnaires, bibliography, and raw output (tables and plots).

In the root (main) folder, there should be a readme file in plain text format. This file should detail the contents of the folder and their order of execution. Ideally, project folders should also include a flow chart diagram displaying the workflow as readme files might not make explicit all the dependencies and order of execution of the code. Some tools for this purpose include draw.io and coggle, but diagrams can be drawn even using spreadsheets. Sometimes the construction of the analytic files has sub-steps that require a complex workflow. In these instances, it is better to draw that process in a separate flowchart.

A sample file structure is presented in Figure 3.1.

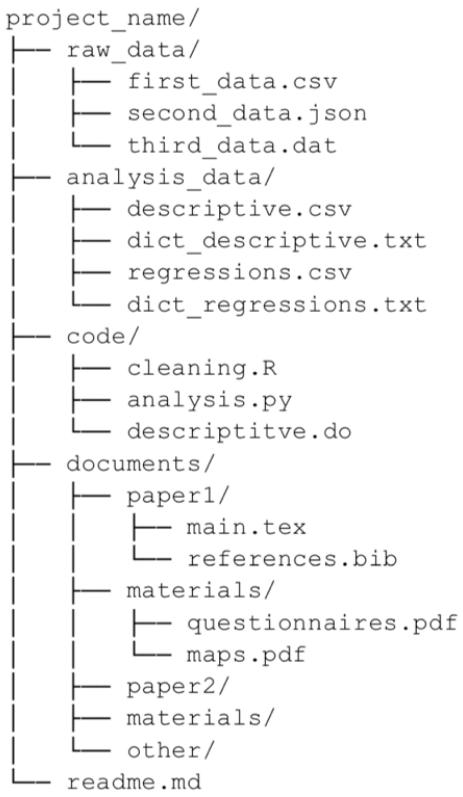


Figure 3.1: Sample standardized folder structure

Another final suggestion, highlighted by Glenskow and Shapiro (2014), is that cleaned data should be *normalized*. The concept of data normalization, from computer science[23], refers to the process of eliminating redundancies from a data set (for example, when the value of state population is repeated for each county) and producing a set of data sets that interact through a set of unique identifiers, or relational databases. By ensuring normalization is a step in the analysis process, data manipulations such as

[23] Not to be confused with the concept of normalization in statistics (subtracting the mean and dividing by standard deviation)

merging and reshaping become more straightforward, and a clear set of primary data sets ease future analyses.

For an extensive treatment and guidance regarding best practices on file structure, see Project TIER specifications.

FILE FORMATS

Plain text files are strongly recommended for a reproducible workflow. These formats allow the user to see exactly the same input that the computer uses to produce the output (plain text formats are usually referred as what-you-see-is-all-there-is, or WYSIATI, formats). Examples of plain text formats are files with extensions like .txt, .tex, or .do. Examples of non-plain text formats are web pages, Word, Excel, or PowerPoint files where the user observes the compiled version already (e.g., the user sees **bold** while the computer "sees" bold in HTML format).

A common practice when coding is to use plain text formats like '.R, .do, .txt,' or '.m'. A less common practice involves using plain text formats to write complete drafts of the paper in a project. The most common syntax for this type of writing is LaTeX, designed specifically for scientific writing. More recently, a strong alternative has emerged in the syntax of Markdown. Both are plain-text languages and have the benefits of facilitating version control and using non-proprietary software. LaTeX allows for many features, but has a steep learning curve. Markdown achieves only the main features of LaTeX or Word, it has a minimal learning curve, but it can be outputted into .tex or .docx files for final edits. This means that now it's fairly easy to write the draft of a paper in Markdown, obtaining all the benefits of a plain text language, and then generate the output into Word or LaTeX for final edits (like institutional headers or more detailed formatting).

VERSION CONTROL STRATEGY

Version control tracks the entire history of a given document. Here we discuss two version control Strategies:

Strategy 1: Naming conventions and protocols for saving files.

The goal of this strategy is to keep track of meaningful changes to a given file. Without any previously agreed upon structure, files tend to be renamed in a "narrative" way. For example, sample-file.do might become something like sample-file-with-state-vars.do, then sample-file-final.do, then sample-file-final-v2.do, etc. Over time, this becomes unmanageable and uninformative for those returning to a file after a time period that could be as short as a few weeks.

The first strategy consists of establishing simple ground rules across a project team. Some of the most common conventions are:

- Save any file that was shared (e.g., over email or published or posted so that others could access it) with a new name. It is advised to rename as often as possible to track the incremental changes. Given storage constraints, the recommendation is to save at least every week and extend to monthly after a few months (this means that after, for example, three months the weekly files should be deleted, keeping only one per month).

- Use the YYYYMMDD date format as a prefix, so that sample-file.R becomes 20181024-sample-file.R. Adding the date prefix ensures that possible future manipulations of the file (like copying or minor edits) do not overwrite the information in the last meaningful edit.
- When working with collaborators, add a suffix with the initials of the last editor, so 20181024-sample-file-JD.R reflects edits by team member Jane Doe.

This strategy has the advantages of easy adoption and, when followed closely, manages to leave a minimal trail of previous work. The disadvantages are that some meaningful changes might still be overwritten with the same name, the naming convention does not provide information about the changes, and at the end of the study, there are dozens of files for each document.

More recently, cloud storage services (like Dropbox and Google Drive) have begun offering a service of version history. This service can be thought as a strategy between 1 and 2. Currently the IDB uses OneDrive for internal purposes which provides a similar service of version history.

Strategy 2: Use version control software.
The problem of keeping a complete history of meaningful changes to code is one that underlies the entire industry of software development, which has responded with version control software (VCS). VCS are programs that are designed to track the entire work history of projects that can involve an unlimited number of lines of code. The most popular VCS to date is Git, a program that runs in the command line (there is no "Git app" develop by the same authors of Git). GitHub is a company that provides several products for using Git. Github.com provides free (public) cloud service and tools for collaboration. The GitHub Desktop App is a graphical user interface (GUI) developed by GitHub that helps user run Git on a computer, and sync the content of a project folder (or repository) in the cloud.

Using Git in the command line is a powerful tool, but has a steep learning curve, which might deter some audiences (a tutorial can be found here: https://swcarpentry.github.io/git-novice/). The GitHub Desktop app provides an easier starting point, and has the complete functionality for storing all changes done to plain text files. The app, however, provides limited tools to explore and access the history of a file.

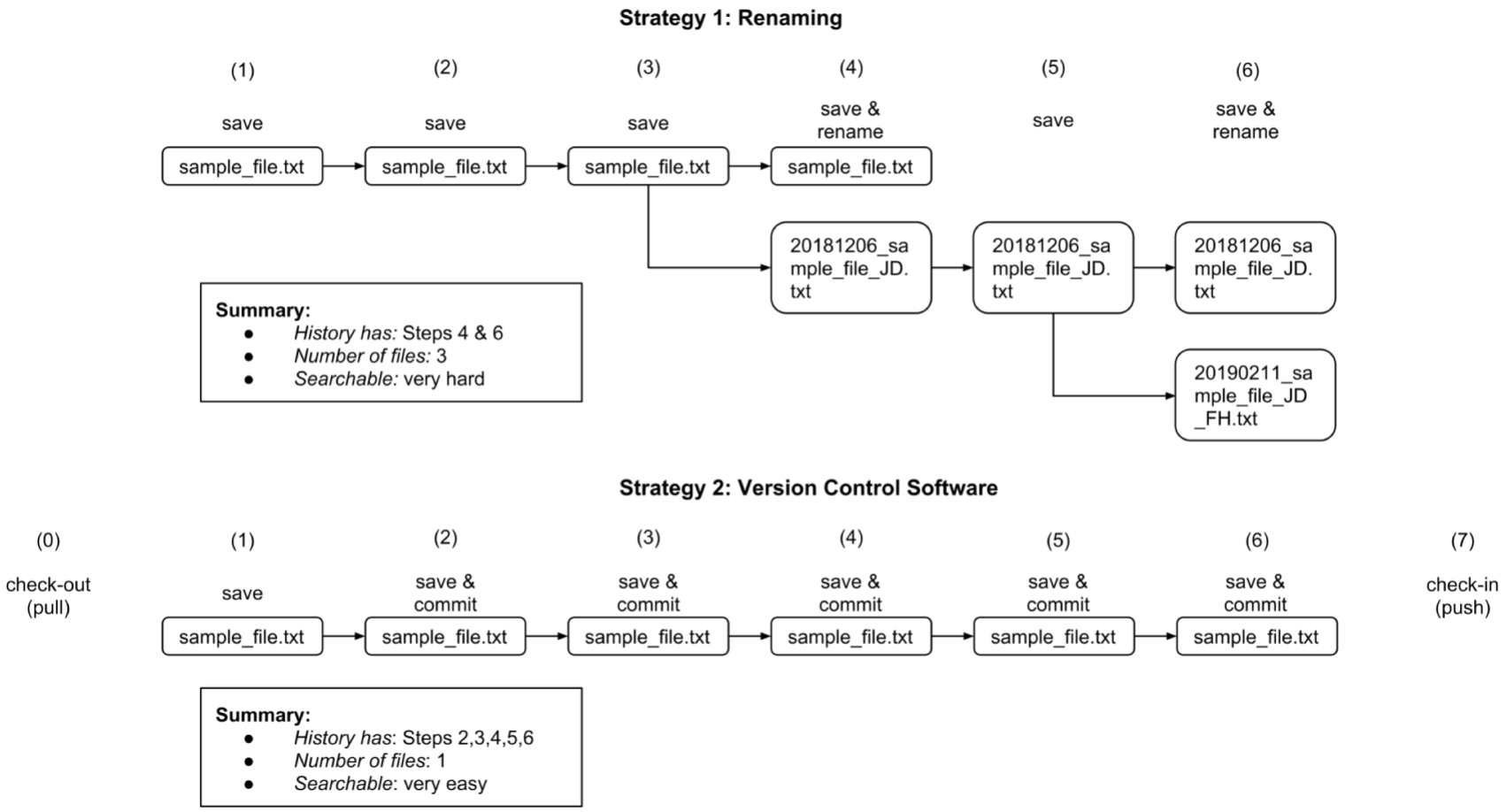
The biggest advantage of using a VCS is that it provides a complete, annotated, and searchable history of the work done on any set of documents, and does not create additional files for the same document. Complete history in this context means every snapshot of the work environment taken by the authors, which happen with daily, if not hourly, frequency (as opposed to the weekly/monthly unstable snapshots taken with the renaming strategy). VCS tracks the version of any file format, but can compare (and search) across versions only for formats with plain text files as defined at the end of the File Formats section.

The main disadvantage of using a VCS is that it requires some changes in the usual workflow of a researcher. At a high level, the biggest changes are the concepts of *checking-out* and *checking in* to a file from a *remote repository*, as well as the concept of *committing* a file. Checking-out (or pulling in VCS lingo) is the action taken at the beginning of every work session and indicates that the analyst will begin to modify the project folder in her own computer (locally). Checking in (or pushing) is the action of returning the newly edited project folder to the main server, an external computing environment located

in the organization or in the cloud. The project folder is known as the *repository* in VCS lingo, and the server is known as the *remote.* Finally, *committing* is the action of taking a snapshot of the project folder (repository) to track all saved changes up to a point in time.

Figure 3.2 compares the workflows with the two different strategies. In the renaming strategy the analyst ends with several files and an incomplete history of the changes made to a file (if followed closely). In the VCS strategy, the analyst adds one step at the beginning and one at the end, and commits the project folder instead of renaming. As a result, it ends with one file and a much more complete and searchable history of the changes made. It is important to highlight that the additional steps of checking-out, checking in, and committing, are at the level of the project folder. This mean that the VCS approach becomes less costly as the number of files in a project folder grow.

Figure 3.2: Comparison of workflow and output of version control strategies



CODE READABILITY & DYNAMIC DOCUMENTATION

Good practices that allow for code readability are essential for (with other colleagues and your future self). Keeping track of an entire workflow depends on the ability to understand what any given piece of code does and how it relates to the final outcome. However, most programming work in empirical research (in the social sciences, at least) is made with ad-hoc programming styles. This lack of standardization is one of the likely culprits behind the alarmingly low rates of computational reproducibility in economics discussed in the introduction.

In this section, we discuss high-level recommendations for better coding and provide a review of a specific approach called dynamic documents.

RECOMMENDATIONS FOR CODING

Each project team should agree on a minimum set of style conventions for coding. This will allow collaborators to be able to quickly assess a script and find the relevant pieces with minimal effort. These style guides should be available to anybody on the team and reviewed periodically.

Examples of style guides include:

- Google's R style guide: https://google.github.io/styleguide/Rguide.xml
- General style guide for Python: https://docs.python-guide.org/writing/style/
- Stata style guide #1 (Stata Journal): https://www.stata-journal.com/sjpdf.html?articlenum=pr0018
- Stata style guide #2 (User-develop): http://www.econometricsbysimulation.com/2013/03/my-not-so-brief-stata-formatting-guide.html

Information regarding standard coding practices can be communicated as part of standard onboarding or training for new project team members. Employees can be presented with the coding style guides mentioned above or with brief sessions of parallel coding with senior analysts. Parallel coding is another practice imported from the software development industry where two analysts/programmers (one senior and one new arrival) sit right next to each other and program the same tasks and compare styles. Usually the practice is recommended up to a two hours of parallel coding. When bringing someone new up to speed, it is particularly helpful to do paired programming on particularly tricky problems (Gentzkow and Shapiro, 2014).

A note on documentation:

Commenting code might seem a somewhat trivial recommendation. However, among the advocates for computationally reproducible research, there are two clearly distinctive camps. One group advocates for extensive commenting and argue that well annotated code is the predecessor of dynamic documents, described in the next section (Christensen, Miguel and Freese, Forthcoming), while others argue for minimal commenting as updating comments is costly and it is better to embed the name of objects (functions, variables, data sets) with as much information as possible (Gentzkow and Shapiro, 2014). This document favors the concept of making names informative (while respecting style guides) and suggest that the "commenting philosophy" should be a choice that is indicated in the project team's style guide. Christensen, Miguel, and Freese (Forthcoming) provide a set of specific coding recommendations

for Stata specifically, and other languages more generally. A version of these recommendations is summarized in Box 3.1.

Box 3.1: General coding suggestions from Christensen, Miguel and Freese (Forthcoming)

General	Stata
1. Project folder and scripts should be self-contained. 2. Add tests to sections of code to verify proper execution (unit test) 3. Comment extensively 4. Indent your code (add spaces or tabs) 5. Rename often or use Version Control Software 6. Separate cleaning data from analysis data 7. Rename files using conventions (e.g., date_filename_intials.tex) 8. Name binary variables after it's 1-value (eg 'female' and not 'gender') 9. Use temp_prefix to delete objects after they are no longer needed (eg tem_var1, temp_data1). 10. Add a label describing each variable (column in data set) 11. Use relative directory paths (e.g., use ~/proj/..; not C:/username/proj/..)	• Use multiple missing values (.a - .z to distinguish, for example, NAs from No response) • Merge on unique identifiers • Use "If" unit tests (eg: *if* _merge != 'expect num' *then* generate error) • Do not use abbreviations (eg. write capture and not cap) • Global macros for paths • Local macros for varlists • Use computer generated locals (e.g., `r(mean)') • Run multiple log files when running master.do • Label data and use notes • Use notes for long information • Use 'datasignature' to verify data (unique identifier of dataset) • Use value labels • Don't use caps • Use 'saveold' (instead of save)

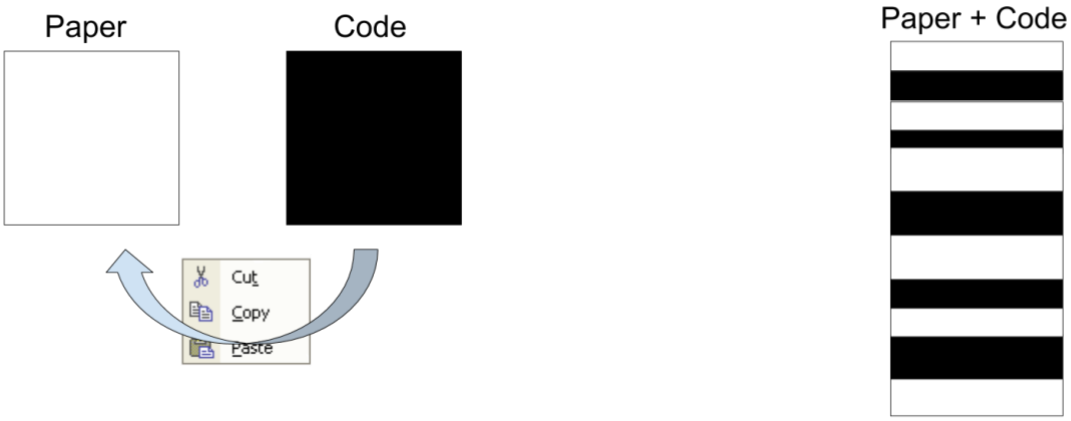
DYNAMIC DOCUMENTS

In 1984, computer scientist Donald Knuth published *Literate Programming*, a book that changed the way programmers think about coding (Knuth, 1984). Knuth's key message was that code should not only be made machine readable (or executable by the computer), but also that code should be human readable. This means that the logical structure of the code should reflect a narrative that resembles that of a paper, as opposed to a set of somewhat ad-hoc instructions that reflect the order in which the analysis took place. To operationalize this concept, Knuth proposed that both code and narrative should be weaved into one single file. More than 30 years after its publication, *Literate Programming* has gained traction with the emergence of easy-to-use tools that implement these concepts and the increasing awareness of computational reproducibility concepts. These easy-to-use tools are generically called *dynamic documents*.

The problem that a dynamic document solves is illustrated in Figure 3.3a. The traditional way in which the narrative components of a project (i.e., the paper) interact with the analytical components is by writing the paper and the code in separate environments. In this format, results produced from the code are usually transported to the paper via copying and pasting. This method is prone to human error, and is also labor intensive, which can make reproducibility more challenging.

The solution offered by dynamic documents is depicted in Figure 3.3b. Dynamic documents (DD) combine both the narrative and the analytical environment in one file (or set of files corresponding to sections).

Figure 3.3: Interaction Between Code and Paper.



(a) Without Dynamic Documents **(b) With Dynamic Documents**

To achieve this integration, the narrative component is written into the code as comments, then the code is run through the given statistical software (ie. R, Stata, Python) and a log file that contains output and narrative is generated (with a specific Markdown format). This log file is then fed into a program called pandoc (a command-line-only type of software whose details are not relevant for this guidelines) that outputs the information from the log file format (.md) into almost any possible format including .docx, .tex, .pdf and .html.

There are two main implementations for dynamic documents: RMarkdown (Figures 3.4 and 3.5) and Jupyter Notebooks (Figure 3.6). RMarkdown is primarily designed to run using the R statistical programming language, and Jupyter Notebooks are meant to be program-agnostic, though it is mainly associated with Python.

Stata 15 also has the capability to build dynamic documents using the command dyndoc, however its development is still in early stages (Tvorak, 2017), and it seems more likely that the academic community will use Stata through Jupyter for dynamic documents (de Kok, 2016). A good example of a dynamic document written for Stata can be found here.

Figure 3.4: Examples of Dynamic Documents in R (RMarkdown)

$$\widehat{\Delta E} = \sum_{g \in \{A,T\}} \left(\widehat{N_g^{final}} \times \widetilde{\eta_{w \leq MW}^{g}} \times \overline{\%\Delta w^g} \right) - \widehat{OF}$$

2.5.1 Code to compute each component

Code

Components of Elasticities

	Adult	Teen
η_{lit}	-0.03	-0.10
$\eta_{w \leq MW'}$	-0.23	-0.13
F_{adj}	4.50	4.50
$\overline{\%\Delta w}$	13.36	16.59
$\widetilde{\eta_{w \leq MW}}$	-0.15	-0.45

Figure 3.5: Examples of Dynamic Documents in R (RMarkdown)

2.5.1 Code to compute each component

Hide

```
#Elasticity from the literature
eta.lit.f <- function(SA.eta.lit = param.eta.lit) - 0.1 * SA.eta.lit

#Extrapolation factor used by CBO
factor.extrap.f <- function(SA.factor.extrap = param.factor.extrap) 1/3 * SA.factor.extrap

#All remaining components required to compute effect on employment
final.other.comp <- function(df.temp = df) {
  eta.lit <- eta.lit.f()
  factor.extrap <- factor.extrap.f()
  rem.comp <-  df.temp  %>%
    filter(pop_of_int == 1) %>%
    mutate("adult" = ifelse(age>=20, "Adult", "Teen") )  %>%
    group_by(adult)  %>%
    summarise("$\\overline{\\%\\Delta w}$" = wtd.mean( ifelse(wages.final <=10.10,
```

Figure 3.6: Examples of Dynamic Documents in Python (Jupyter) (Source: QuanEcon.org)

The Filtering Step

We are now presented with some good news and some bad news

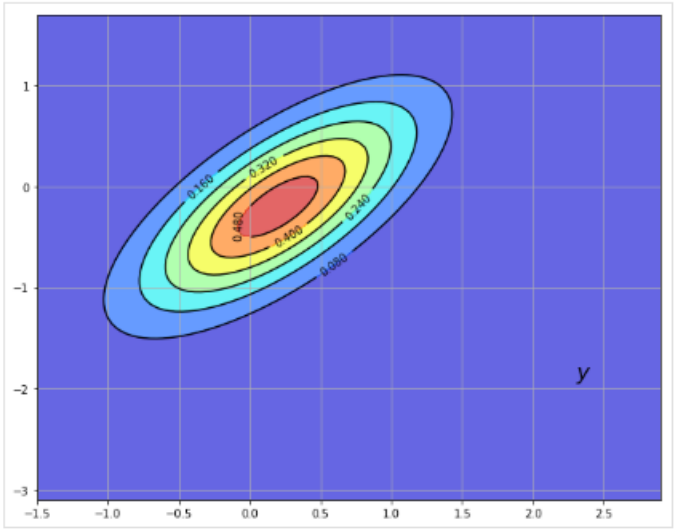
The good news is that the missile has been located by our sensors, which report that the current location is

The next figure shows the original prior $p(x)$ and the new reported location y

```
fig, ax = plt.subplots(figsize=(10, 8))
ax.grid()

Z = gen_gaussian_plot_vals(x_hat, Σ)
ax.contourf(X, Y, Z, 6, alpha=0.6, cmap=cm.jet)
cs = ax.contour(X, Y, Z, 6, colors="black")
ax.clabel(cs, inline=1, fontsize=10)
ax.text(float(y[0]), float(y[1]), "$y$", fontsize=20, color="black")

plt.show()
```



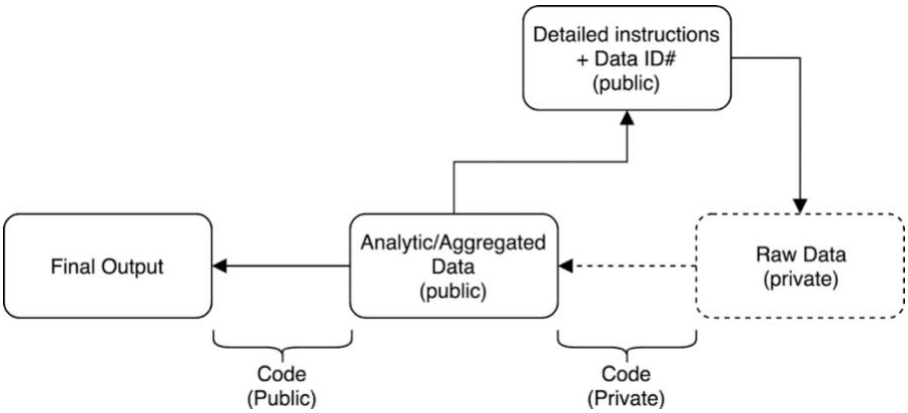
CODE AND DATA SHARING

PRINCIPLES

Another essential step for computational reproducibility requires access to the data and code. Here, it is important to clarify that the data can be broadly characterized into three stages: *raw data* refers to the data as originally obtained (from the field or a secondary source, for example) where the only possible transformation was the deletion of personally identifiable information; *processed data* refers to data that have been modified from raw data but that are not ready for the final analysis; and *analytical data* refers to the data that have been cleaned and is ready to be used to obtain the final statistical output. Each stage should also have corresponding code: cleaning code to transforms raw into processed data, and analytic code to transform processed data into analytical data and generate final outputs. Full computationally reproducibility can only be obtained with access to the raw data such that the final results can be reproduced from the beginning[24].

In many instances, public access to the raw data set will not be possible to protect sensitive information (see section 4.4). In this context, researchers should provide detailed written instructions outlining all of the required steps to obtain the original raw data, including specific contact information and a unique identifier of the dataset as a whole (not of individual observations). The unique identifiers are obtained by applying hash functions (for example, 'datasignature()' in Stata and 'digest()' in R) to the original raw data and obtaining a long alphanumeric string that cannot be reversed. This function maps each data set to a unique alphanumeric key, such that if the function is applied to a data set with the same content (columns, rows and format) it will always produce the same key. If just a single observation is modified in the data, then the hash function will produce a different key. The unique identifier should be posted with the public information for verification purposes. This workflow is presented in Figure 3.7.

Figure 3.7: Reproducible Workflow with Sensitive Data



[24] This section reviews best practices and resources for data sharing assuming that the data has been properly de-identified, for recommendations on de-identification see Chapter 4 of this document.

HOW TO SHARE CODE AND DATA

1. **Comply with Data De-Identification and Management:** As described in Chapter 4 Ethical Research, researchers should carefully consider when in the analysis workflow they can begin de-identifying the data to ensure analysis is conducted on data that closely aligns with data that can be shared (as feasible).
2. **Organize all files with proper documentation.** Follow recommendations from the File Management section of this chapter and produce a self-contained project folder with all components clearly identified. A key step when adding files to an internal project folder is removing any personally-identifying information (PII) as discussed in the Data De-identification section of the next chapter on "Ethical research".
3. **Publish the data in a reputable, DOI-issuing repository.** Once the data is organized and all redundant materials have been removed, submit the replication material to a reputable repository that issues Digital Object Identifiers (DOIs). These provide a unique key to track, share, and cite your data. Once a replication package has been submitted, it cannot be changed without modifying the DOI. Subsequent versions can be submitted, but a record will remain. Table 3.1 summarizes the main repositories that provide DOIs for research[25].

Table 3.1: Data and Code Repositories that issue DOIs

Repository	Space Limits	Web Address
Figshare	100 GB	figshare.com
Dataverse*	10 GB	dataverse.org
OSF	5 GB	osf.io
Zenodo	50 GB	zenodo.org

* Allows for set-up of an institutional version.

4. **Add a link to data (with DOI) to the original paper.** Once the data and code have been posted and the DOI has been recorded, the link of the replication package should be added to the paper.

Frontier developments in code and data sharing for computational reproducibility

One final note is that important progress is being made in computational reproducibility towards eliminating dependencies on specific software requirements to run code. As with version control, solutions to this problem originated as a response to what is common known in computer science as "dependency hell". This term refers to the difficulties encountered when attempting to run software based on an old version, and can range from lack of access to the correct licenses, to compatibility issues across operating systems.

[25] The IDB currently owns a data repository where it stores mainly aggregated data. In Chapter 4 of this document recommendations are made to increase the publication of micro-data in this platform in the near future. The recommendations provided in this section are meant to describe the best practices currently available on data sharing in the academic community and to inform possible future innovations in the bank. Examples of future innovations are: incorporating DOI issuing capabilities to the current platform, or incorporating some of the platforms describe in this section into the workflow of the project teams at the bank.

In response, cloud computing services have been developed to provide a historical record of all the versions associated with a set of software and its libraries. Currently, the two best known implementations of these ideas are project Binder (mybinder.org) in open source, and Code Ocean (codeocean.org) in proprietary format. These are recent developments, and currently there is no standardized way to interact with the other elements described in this chapter. However, it is easy to see how they may eventually provide a platform to easily run a fully reproducible analysis using dynamic documents based on a DOI-issuing repository.

ADDITIONAL RESOURCES

1. IPA's Best Practices for Reproducible Research (2015): Coding recommendations for Stata users.
2. Guidelines for Data Publication also from IPA
3. The World Bank's DIME Wiki: In particular the following sections: Checklist for Data Cleaning, Stata Coding Practices, and Publishing Data
4. Practical Tips for Ethical Data Sharing by Meyer (2018): a tutorial with dos and don'ts for data sharing.
5. BITSS Resource page (https://www.bitss.org/resource-tag/education/) An extensive list of additional resources related to computational reproducibility and research transparency in general
6. Collection of examples of computational reproducible research: https://www.practicereproducibleresearch.org/

CHAPTER 4: ETHICAL RESEARCH

This chapter summarizes the principles and objectives for conducting ethical research. Following this discussion, the chapter presents best known practices for how to operationalize these principles and objectives during the design, implementation, and dissemination phases of a research project, all of which should be considered alongside any necessary local requirements to ensure contextual and cultural sensitivity.

PRINCIPLES AND BEYOND

This document is informed by the three main principles for ethical research conduct presented in the Belmont Report (1979):

Respect for persons incorporates at least two ideas: (i) individuals are treated as autonomous agents and (ii) individuals with diminished autonomy are entitled to protection. In most cases, respect for persons requires that research subjects enter into the research voluntarily and with adequate information. This principle is operationalized through the informed consent process, however operationalizing the "voluntary" component also requires full understanding of subject vulnerabilities and selection, discussed in more detail below. One additional challenge is the increasing availability of "big data" and other data sources that can be used without the research subjects' knowledge of the research – essentially removing both voluntary participation and adequate information requirements.

Beneficence incorporates two ideas: (i) do not harm and (ii) maximize possible benefits and minimize possible harms. As discussed in later sections, harm and/or exploitation of research subjects, as well as project teams, can occur as a result of participating in the research. While some potential harm may be inherent to participation in the research, it is for the project team and partners, such as local experts and Institutional Review and/or Ethics Board expertise, to determine an appropriate balance between potential harm and potential benefits for the research subjects themselves.

Justice in research refers to the just distribution of the risks and burdens of the research and the benefits expected to be produced by the research. A primary way in which this is operationalized is through defining the relevance of the study and through appropriate selection of research subjects. Project teams, along with local experts and Institutional Review Board expertise, should assess whether some subjects (e.g., specific communities, particular racial and ethnic minorities, or persons confined to institutions) are being systematically selected for participation in research simply because of their easy availability, their compromised position, or their manipulability, rather than for reasons directly related to the problem being studied. Justice also means the research should not unduly involve persons from groups unlikely to be among the beneficiaries of subsequent applications of the research. Particularly in evaluation research, a main question for project teams will be around continuation of services and scale-up – if the intervention is found to be successful, will those in the treatment group and in the control group who participated in the study benefit from the intervention after the research is complete? If the intervention is found to be harmful or unsuccessful, will it be terminated? Justice in research requires considering these questions early and often throughout the research.

In addition to the above, this document considers several additional points for ethical research beyond Belmont:

1. Research needs **independent review that is culturally and contextually sensitive** to assure, both in advance and by periodic review, that appropriate steps are taken to protect the rights and welfare of those impacted by the research. This is often done through Institutional Review Boards (IRBs) to review in relation to specific regulations, however IDB research is conducted in dozens of countries with varying laws and regulations and therefore IRBs. For this reason, when local IRB review is not sufficient or even available for some project teams, this document suggests independent review alongside any local requirements.
2. The human subjects involved in the research may not be the only population for which the research poses risks – the project team may need to **consider other bystanders, or even the general population, in accordance with respect for persons, beneficence, and justice**. An example of this is a randomized experiment to "Get out the Vote" with the goal to significantly increase voting for one issue, party or another. The result of the experiment can have real impacts on the result of the election which can have real impacts on the lives of many people. (For discussions on this topic, please refer to Desposato, 2016) This document suggests careful consideration of who should participate in informed consent beyond the research subjects.
3. As also discussed below in the section on vulnerability, risks change over time. Under current US regulation, many research studies may not require IRB "continuing review", or even review at all. However, known risks to those who may be impacted by the research – either as research subjects and/or as bystanders – may evolve over time. Particularly when considering data sharing practices, and specifically identifiable data sharing, this document suggests **continual assessment of risks for those expected to be impacted by the research**. (Please reference Knott, forthcoming for discussion on types of risks that can materialize in dynamic environments.)

OBJECTIVES

In consideration of the principles and other issues described above, the objectives of ethical research are to understand and protect research subjects' vulnerabilities, understand and protect personally identifiable and sensitive data, and ensure proper risk management for the research subjects, research team, and research institution(s).

UNDERSTANDING RESEARCH SUBJECTS' VULNERABILITIES

The first objective for ethical research is to understand the research subject population's vulnerabilities. As governed by its Institutional Strategy (IDB, 2015), the IDB's work is focused on three main development challenges: (i) social exclusion and inequality; (ii) low productivity and innovation, and (ii) lack of regional economic integration. In addition, there are three cross-cutting issues: (i) gender equality and diversity, (ii) climate change and environmental sustainability, and (iii) institutional capacity and the rule of law. Addressing these challenges and cross-cutting issues is complemented with a focus on specific countries, where the IDB has a goal to "*address the needs of small and vulnerable countries*[26]". Since its research activities naturally align with its strategy, project teams must engage with diverse, and often vulnerable, populations for whom the *participation* in the research activities and/or *provision* of

[26] As defined in the IDB's Report on the Ninth General Capital Increase, these borrowing member countries, also known as Group C and D countries, are Bahamas, Barbados, Belize, Bolivia, Costa Rica, the Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Nicaragua, Panama, Paraguay, Suriname, Trinidad and Tobago and Uruguay.

personally identifiable and/or sensitive data by research subjects may pose risks for harm and/or exploitation.

Because some groups and individuals are particularly vulnerable and may have an increased likelihood of being wronged or of incurring additional harm (WMA, 2013), some guidelines define specific groups as vulnerable with specific requirements while others have moved toward broader categorizations of vulnerability. Examples include:

- **Belmont Report (1979):** Vulnerable populations include children, the institutionalized mentally ill, and prisoners (with mention of dependency and compromised capacity for consent);
- **US Federal Regulations 45 CFR 46:** Vulnerable populations include pregnant women, human fetuses, and neonates (Subpart B), prisoners (Subpart C), children (Subpart D), mentally disabled persons, and economically or educationally disadvantaged persons; and
- **Council for International Organizations of Medical Sciences (CIOMS) 2016:** Vulnerable populations may include (i) individuals without capacity to consent, (ii) individuals in hierarchal relationships, (iii) institutionalized persons, (iv) women, (v) pregnant women, and (vi) other potentially vulnerable groups.

Instead of listing specific groups as vulnerable, this document offers two suggestions. First, project teams should consider a research subject's vulnerability as defined as "***a diminished ability to fully safeguard one's own interest in the context of a specific research project. This may be caused by limited decision-making capacity or limited access to social goods, such as rights, opportunities, and power. Individuals or groups may experience vulnerability to different degrees and at different times, depending on their circumstances***."[27] Second, project teams should carefully consider the context - social, economic, cultural, medical, and other conditions - that may contribute to subjects' vulnerability (Levin et al, 2004; Resnik, 2004). To do both, project teams may use the matrix in Table 4.1 (adapted from NBAC, 2001) to assess whether or not their research subjects are vulnerable and how to mitigate risk through ethical research practices.

In addition to the above, it is recommended that project teams carefully consider whether or not participation in research contributes to research subjects' vulnerability, even if research subjects do not initially fit any of the vulnerability categories. For example, if it is generally known in a community that the study is focused on loan recipients, then subjects may become targeted for their financial assets (*this may even be true for "control" survey sample since outside intruders may not know to distinguish between treatment and control individuals/households*).

Regardless of whether or not research subjects are identified as vulnerable at baseline, their **vulnerabilities should be continuously assessed and updated** based on any new vulnerabilities to inform risk mitigation strategies throughout the research life cycle of design, data collection, data management, analysis, and dissemination.

[27] Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada. Tri-Council Policy Statement: ethical conduct for research involving humans, December 2014. (Accessed on July 11, 2018 http://www.pre.ethics.gc.ca/pdf/eng/tcps2/TCPS_2_FINAL_Web.pdf)

TABLE 4.1: Taxonomy of Vulnerability for Research Participants (adapted from NBAC 2001[28])			
Vulnerability	**Definition**	**Potential Causes**	**Ethical Research Practice**
Cognitive Vulnerability	The research subject does not have the capacity to deliberate and decide whether or not to participate in the study	Immaturity (through age, other cause), dementia, certain types of mental illness, disability; educational deficits and unfamiliarity with the language; situational mental distress/crisis	Mitigated through proper **Informed Consent**: plain-language, advance directives (where incapacity is anticipated), supplementary educational measures to ensure comprehension, and the proper use of surrogates and advocates
Juridic Vulnerability	The research subject is liable to the authority of others who may have an independent interest in the research subject's participation	Prisons and the military, where wardens and officers have legal authority over prisoners and enlistees; Children under the authority of their parents, Students subordinated to Professors, Institutionalized persons subject to the authority of custodians, women legally subject to their husbands; Program beneficiaries and their benefactors	Mitigated through proper **Informed Consent**: devise a consent procedure that will adequately insulate the research subject from the hierarchical system to which he or she is subject. This is particularly challenging if the researcher/project team is a part of the hierarchical system (so program beneficiaries who are surveyed by their benefactors).
Deferential Vulnerability	The research subject exhibits patterns of deferential behavior that may mask an underlying unwillingness to participate	May be driven by social and political pressures to follow/defer to others despite own desire to not follow/defer *(often present with juridic vulnerability)*	Mitigated through **Sample Recruitment/Screening** and **Informed Consent**: Inclusion Criteria/Sample Selection may require input of local informants or consultants to devise a process that eliminates as much as possible the social pressures a research subject feels. Informed consent mitigation same as above.
Allocational Vulnerability	The research subject is lacking in important social goods that will be provided as a consequence of participating in the research	When participation in the research can provide research subject a social good - money, housing, medical care, childcare, burial benefits, opportunities to benefit the community, freedom – that they otherwise do not have access to	Mitigated through **Sample Recruitment/Screening** and **Compensation**[29]: The Inclusion Criteria/Sample Selection may require input of local informants to determine whether or not the offering of research participation may coerce certain individuals/groups based on their baseline allocations; Project Teams must also carefully consider Compensation packages to limit their under or over-value and may need to consider not just their research sample, but also neighboring communities/individuals/households that are excluded and may feel resentment for the exclusion.
Infrastructural Vulnerability	The political, organizational, economic, and social context of the research setting does not possess the integrity and resources needed to manage the study	Research subjects have access to research requirements (phone, transport); Project teams have access to research requirements (skills for specific biomarker tests, psychological tests, etc; electricity, transport, safety)	Mitigated through **Study Design**: The study design/protocol should be carefully reviewed for local context and cultural sensitivities.
Medical Vulnerability[30]	The research subject has been selected, in part, because he or she has a serious health-related condition for which there are no satisfactory remedies	When (i) illness is severe and (ii) no safe, effective, and otherwise satisfactory treatments are available, patients can be primarily driven to participate based on false hope for benefits	Mitigated through **Study Design** and **Informed Consent**: Given the interests and aspirations of both parties (and the poor bargaining position of the research subject) work toward fair division of the benefits and burdens of cooperation and design the study to maximize the likelihood of subject benefit based on medical intervention found to be safe and effective;

[28] Original author (Kenneth Kipnis) granted permission for adaptation and use in this document.

[29] This is not taken from NBAC 2001 and is a recommendation of this document. As noted in CSE discussion and Appendix A, project teams should work with community stakeholders and IRB to determine if compensation is appropriate and the value of appropriate compensation to mitigate coercion (allocational vulnerability) and exploitation of research population. For example, if a household interview takes 2 hours to complete, consider if the research subject may be compensated, with financial or in-kind payment, the equivalent of 2-hours labor in the local context.

[30] Limited relevance to IDB research activities.

			communicate possible benefits and their probabilities for success through Informed Consent.

PROTECTING PERSONALLY IDENTIFIABLE AND SENSITIVE DATA

A second objective is the protection of personally identifiable and/or sensitive data to adhere to promises of confidentiality. This objective is independent of direct or indirect contact with research subjects, and independent of possible subject vulnerability. It is important for project teams to define when their research requires collection and management of these data. Whenever these data are required for research, which is often, it can increase risks to research subjects who consent to participate in the research, despite their original vulnerabilities.

Personally Identifiable Information (PII) is information that can be used, on its own or in conjunction with other information that is linked or linkable to a specific individual (or household, community, school, etc.), to determine the identity of an individual or otherwise locate or contact the individual. It includes:

- *Direct Identifiers*: such as full name, date of birth, mailing or home address, email address, telephone number, GPS coordinates, national identification number, physical/biological identifiers (physical appearance, through photo or video data collection, fingerprints, DNA, etc.). Depending on the study and data needs, direct identifiers can also include the name of the school, health facility, community, etc. that directly identify the location of the data collection or extraction; and
- *Quasi (Indirect) Identifiers* are unique, observable or otherwise knowable characteristics that may identify a specific individual (or household, community, school, etc.) even when direct identifiers are removed. Quasi-identifiers may include *visible* assets, loan and credit information, and unique combinations of demographics (such as ethnic minority groups, widow status, or very high or very low education attainment). Quasi-identifiers are also created when there are readily available *linkage documents/data to the study sample*. For example, in program evaluation, treatment status/assignment may become a quasi-identifier if it is well-known who received the treatment in specific areas.

Sensitive data is information that may pose a risk to the individual (or household, community, school, etc.) if it is collected or released in a way that is linkable to the research subjects. This type of data may include income, assets, tax status, health status, but also violence, abuse, mental health information for which the disclosure could lead to harm and/or exploitation to the research subjects. Whether or not data is sensitive is also context-specific – some information may be considered sensitive in some contexts but not in others. Project teams should therefore build in an assessment of the sensitivity of their data, using tools such as the Harvard Information Security Data Classification Table, early in the research life cycle (*see Community and Stakeholder Engagement*).

ENSURING APPROPRIATE RISK MANAGEMENT

There are risks to research subjects, project teams, as well as the IDB and its partners, when engaging in research. A third objective behind the practices of ethical research is to identify and mitigate these risks over the life of the research cycle.

Research subjects (and other bystanders). The main potential risks to research subjects (and other bystanders) when participating in research stem from harm and/or exploitation. The following are examples of risks that may occur:

- **Coercion by project team**. There may be power dynamics at play between the project team and research subject if not carefully considered. Research subjects may feel obligated to participate even if they do not want to because of these power dynamics (juridic and deferential vulnerability). This should be carefully considered in the selection and training of interviewers, as well as the content and delivery of informed consent.
- **Direct harm from improper research management**. This is when harm may befall the research subject as a direct result of the survey. For example, the survey may require women to be interviewed separately from men about sensitive topics, such as domestic abuse or household finances and resource allocation. If the project team does not provide sufficient protection of the woman during the interview and she is overheard discussing sensitive issues by family members or neighbors, she could be harmed as a direct result of improper survey protocol. Another example is inadequate or improper compensation and/or oversampling of certain research subject populations that result in high opportunity costs or other excessive burdens on the research subjects.
- **Direct harm from loss of confidentiality**. If research subjects' PII and/or sensitive data is not sufficiently protected and there is a loss of confidentiality – i.e. intruders or other stakeholders have sensitive information that is linkable to the research subjects – there is risk that this disclosure could be used to harm and/or exploit the research subject. For example, if the survey is on financial inclusion services and survey participants are identified as loan recipients, with the loan amounts linked to their PII, a loss of confidentiality could result in these individuals – or their households, family members, friends – becoming targets for financial extortion.
- **Directly impacted by the study itself**. An example of this is a randomized experiment to "Get out the Vote" with the goal to significantly increase voting for one issue, party or another. The result of the experiment can have real impacts on the result of the election which can have real impacts on the lives of many people. (For discussions on this topic, please refer to Desposato, 2016) Project teams should carefully consider how the research itself – participation in it, or simply the conduct of it – may pose risks in the lives of the population.

Project teams. The main potential risks to project teams for implementing research stem include harm, loss of reputation and funding:

- **Direct harm from improper research management**. Insufficient detail on how project teams will be protected during survey work – *from sexual harassment to road safety to physical safety in less secure neighborhoods to sufficient access to food, water, and breaks during field work* – may result in direct harm to project teams.
- **Loss of reputation**. Survey firms, research assistants, and principal investigators all stand at risk for loss of reputation if their research practices do not adhere to best practices in ethical research.
- **Loss of funding and/or employment**. As a direct result of loss of reputation, project teams may be blacklisted or sidelined from research projects.

Implementing partners. The main risks to the IDB and its clients (i.e. government ministries) or partners stem from loss of reputation and loss of trust in the institutions.

- **Loss of reputation and trust**. Positioning itself as an evidence-driven agency, the IDB and its partner institutions could suffer loss of reputation if its research was considered unethical. This could result in loss of reputation, and loss of trust in the IDB's quality of work by clients and partners.

TRAINING AND OTHER RESOURCES

To support deep-dives on topics discussed above and implementation of best practices described below, project teams – *including principal investigators, research assistants, research managers, etc.* – may strengthen existing knowledge regarding ethical research practices and protection of human subjects through training and use of other existing resources. Training options include both free, online courses as well as paid certificate programs:

- World Health Organization (WHO) - This **free**, online Research Ethics course takes four-seven hours to complete and has a particular focus on international health research.
- United States National Institute for Health (NIH) – The online training program - Protecting Human Research Participants - consists of seven modules; each addressing the principles used to define ethical research using humans and the regulations, policies, and guidance that describe the implementation of those principles. The course takes three hours to complete. Although the free course was discontinued, resources are still available and a fee-based training is expected in 2019.
- Collaborative Institutional Training Initiative (CITI). Thousands of academic and research organizations subscribe to the **fee-based** online training services offered by CITI. CITI also offers a range of resources, as well as supporting organizations in the area of protection of human subjects.
- Public Responsibility in Medicine and Research (PRIM&R). PRIM&R advances the highest ethical standards in the conduct of biomedical, behavioral, and social science research through **fee-based** education, membership services, professional certification, public policy initiatives, and community building.

It is recommended that staff **renew training every two-three years.**

In addition, these Guidelines draw from and complement existing resources, including:

- International Ethical Guidelines for Health-related Research involving Humans by CIOMS
- Inter-University Consortium for Political and Social Research (ICPSR) for data management and sharing
- Handbook of the Modern Development Specialist by Responsible Data

As per the one-pager companion to this document, the following sections follow the research life cycle from design to implementation to dissemination to identify key best practices for ethical research.

RESEARCH DESIGN

Much of the work involved with implementing ethical research begins in the design stage when project teams lay the foundation for proper research implementation and management through (i) community and stakeholder engagement, (ii) research protocol(s), (iii) informed consent(s), (iv) data management plan(s), (v) data use agreement(s), and (vi) conducting independent review (by an IRB and/or other review committee).

Project teams should document their plans in the Monitoring and Evaluation Plan- with milestones, timelines, and budgets - for implementing best practices described here (*as applicable*) to inform their ethical research. Adherence to and deviations from these plans can also be assessed at the Project Completion Report (PCR) stage.

COMMUNITY AND STAKEHOLDER ENGAGEMENT

Community and stakeholder engagement (CSE) can inform project teams of the needs and priorities of the communities they will engage with during the research, support elimination of barriers to research participation, and ensure the implementation and results of research are sensitive to cultural norms. It can be even more critical when the research project requires multiple research sites. Project teams may consider several objectives for CSE across (Gooding et al., 2018; Musesengwa, Chimbari and Makaratirwa, 2018)):

- Define what **ethical research practice means given contextual and cultural sensitivities**. Specifically, CSE should support defining and identifying subjects' vulnerabilities, what is/isn't sensitive, definition of minors, appropriate compensation considering cultural norms (and allocational vulnerabilities), how to avoid harm and/or exploitation of research subjects and staff, and how to ensure appropriately informed consent considering subjects' literacy level, language, and other potential vulnerabilities. In addition, CSE can support defining requirements for ethical treatment and consideration of project team needs, such as road and field safety, appropriate compensation, proper access to food, water, and breaks during field work, etc.
- Assess and address **research-naïve vs. research-overburdened communities**. If the research project requires sampling in communities where little to no research has taken place, the project team may need to build in research awareness training/information campaigns. If it requires sampling in communities that are overburdened by research – meaning multiple data collection/surveys take place on a regular basis – the project team may need to assess how best to re-balance this burden through appropriate compensation or adjusted sampling.
- Determine how the research study will be **relevant to and beneficial for local communities**. Identify the primary users/audience locally for the research findings – such as political, administrative, and traditional authorities - and define how the findings of the research will be used to inform programmatic and/or policies for the research participants (in both treatment and control groups as relevant).
- Assess **local capacity to review, conduct, and inform research**. Identify potential local partners and/or collaborators that can also contribute to ethical research design, review, implementation, analysis, interpretation, and dissemination. Considering the discussion below on independent review, this may also support (i) identifying appropriate IRB(s) and (ii) identifying when a separate independent review is necessary.

LEGAL REQUIREMENTS

With or without a formal CSE activity, project teams must identify and comply with all relevant local laws. Applicable laws include data privacy and protection laws, as well as any national regulations on research and protection of human subjects. There are several resources available to project teams to consult and identify relevant laws, including:

- 2018 International Compilation of Human Research Standards by Health and Human Services is a listing of over 1,000 laws, regulations, and guidelines on human subjects' protections in 130 countries and from many international organizations.
- Data Protection Laws of the World by DLA Piper Law Group and Data Protection around the World by Commission Nationale de l'Informatique et des Libertés (CNIL) allow users to compare laws and regulations between countries.

This information will most likely evolve over time, so project teams should work with local staff, clients, and partners to determine if there are other relevant laws not identified in the resources above.

INDEPENDENT REVIEW – IRB AND OTHERS

As discussed above, ethical research relies on **independent review that is culturally and contextually sensitive** to assure, both in advance and by periodic review, that appropriate steps are taken to protect the rights and welfare of those impacted by the research. This is often done through Institutional Review Boards (IRBs) to review in relation to specific regulations, however IDB research is conducted in dozens of countries with varying laws and regulations and therefore IRBs. For this reason, when local IRB review is not sufficient or even available for some project teams, this document suggests independent review alongside any local requirements.

According to US regulations, IRBs "*assure, both in advance and by periodic review, that appropriate steps are taken to protect the rights and welfare of humans participating as subjects in the research. To accomplish this purpose, IRBs use a group process to review research protocols and related materials (e.g., informed consent documents and investigator brochures) to ensure protection of the rights and welfare of human subjects of research*.[31]" For IDB project teams, there are three main forms of IRBs to consider:

- National IRB – This is a centralized IRB established within a country to review and govern research in that country.
- Academic Institution IRBs – IRBs are based within universities to govern the research produced by university staff.
- Independent IRB firms – There are independent IRBs that may be contracted for academic and non-academic research.

In the early stages of design, project teams should identify the IRB review and clearance standards and requirements for their specific research project. Requirements for IRB review may include:

Local Requirements. Consult with local experts to identify local registered IRBs and if local law or regulation requires IRB (or other regulator) review and clearance prior to research.

[31] https://www.fda.gov/RegulatoryInformation/Guidances/ucm126420.htm

International Standards. If one of the goals of the research is publication, project teams should consider international best practices and standards for ethical research. As per international standards issued by the Committee on Publication Ethics (COPE), published authors should "*ensure that appropriate approval, licensing or registration is obtained before the research begins and details should be provided in the report (e.g. Institutional Review Board, Research Ethics Committee approval, national licensing authorities for the use of animals). In addition, if requested by editors, authors should supply evidence that reported research received the appropriate approval and was carried out ethically (e.g. copies of approvals, licenses, participant consent forms)"* (Wagner and Kleinert 2011).

Journal/Publisher Requirements. In addition to international standards, many publishers (and therefore journals) require authors confirm and/or submit documentation of IRB approval prior to publication (as per COPE standards above). Project teams should assess the journals they are likely to aim to publish in and review policies to identify requirements for their future publications. For one example, Springer[32] Publishing covers more than 2,900 journals and follows the COPE guidelines, with a specific statement on ethical approvals: *All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.*

Other Partner Requirements. Project team members and/or collaborators based at academic institutions may be required to go through their academic IRB review boards in addition to other local requirements. In addition, some funders or other partners may require IRB review. Project teams should therefore assess all other requirements of their project team members and partners to inform their next steps in the research design stage.

Project teams should note that IRBs review research to be in line with required regulations (for example: US-based IRBs are established to ensure research follows 45 CFR 46). For this reason, while IRB review may be required and strongly suggested to meet certain objectives (i.e. ensuring the research meets regulation requirements), it may not be sufficient or appropriate for other objectives. For example, many of the IDB research projects may not be defined as "research" under existing regulation, or may be considered "exempt" research under existing regulation. For this and many other reasons, it is **up to the project team to "own" the ethics of their research and not merely outsource ethical review to IRBs and other review boards. IRB and other independent reviews are often necessary, but not the only means toward ethical conduct in research** (for a far more detailed discussion on this point, reference Desposato, 2016).

IRB PROCESS

As defined above, the project team may identify one or more IRB requirements depending on international standards, local requirements, journal, and partner requirements. The following issues should then be considered regarding the IRB process to inform the project team's planning:

- **Identify IRB –** Depending on requirements above, the project team may need to submit a protocol to multiple IRBs (for example, if a local IRB is required, but it is not accredited). Project teams can

[32] https://www.springer.com/gp/authors-editors/journal-author/journal-author-helpdesk/before-you-start

reference the 2018 International Compilation of Human Research Standards and Office for Human Research Protections (OHRP) Database to identify appropriate contacts and IRBs to ensure their protocol is reviewed by at least one IRB that is HHS registered. However, it should be noted that HHS registration does not include quality control/review. The Association for the Accreditation of Human Research Protection Programs (AAHRPP) provides a list of AAHRPP-accredited IRBs on their website (http://www.aahrpp.org/learn/find-an-accredited-organization) if that is a requirement for the project.

- **Timing** – The project team should identify the schedule(s) for the IRB(s) in advance and the time requirements for submission and review. These may vary significantly across contexts and IRB-type (for-hire vs. academic-based for instance). Given this process can take 1-3 months or more, depending on the IRB process, project teams should both build this into their projected timeline as early as possible, and also explore appropriate methods for quality and timely IRB review, such as real-time review (Spellecy et al 2018).
- **Cost** – The costs of initial and periodic IRB reviews vary by country and by how many years the protocol must be in place. This is because standard IRB review requires both an initial, larger submission and review fee, as well as an annual review fee to maintain the IRB coverage over the course of the research life cycle. For reference on costs for for-hire IRB firms, initial reviews can range from $1300-$1500 USD, with annual review fees between $800-$1500 USD per site.
- **Representation** – Depending on local requirements, the project team may need to submit and/or present the research protocol and documents to the IRB in person. This should be built into the research workplan accordingly.

RESEARCH PROTOCOL

A research protocol is the document that is typically required for review and clearance by an IRB or other ethics review board. Standard research protocols include information on the study that is required to present the research concept and support proposal writing, fundraising, and communication efforts around the research in operations. Therefore, regardless of the project team's decision on how to proceed with engaging IRB(s), a research protocol can be a **useful tool for documenting the planned ethical research design and practices, governing its implementation, and communicating its objectives and expected contributions**.

In addition, **a well-defined research protocol can be essential to study replication** – if the same research team, or a different research team, want to replicate the study at a different time and/or in a different context.

To maximize usability of this documentation, project teams may consider (i) requirements of the IRB(s) they will submit to and (ii) reporting guidelines and standards that they will be using for subsequent publications and reporting. To the extent these can be aligned, efforts to generate the Research Protocol will contribute to subsequent efforts related to development of Pre-Analysis Plan, and sufficient reporting and disclosure following international standards (see Chapter 2 Transparency). In general, research protocols follow a standard format and required content, an example of which is detailed in Appendix A, Table 2 (adapted from WHO).

DATA MANAGEMENT PLAN

As noted in the standard format of the research protocol, it is recommended that the project team have a clearly defined Data Management Plan (DMP) in place that governs who will have access to what data and when. Project teams will need to carefully consider how to manage the data in preparation for future public, limited, and/or restricted-access use upon completion of the research project. While dissemination of data is discussed in Data Sharing and Dissemination, the project team needs to lay the foundation for data sharing in the design stage.

There are many DMP templates and guidance available (ICPSR, NSF, Dataverse), including online tools such as the Data Management Tool. Regardless of what template is used, the goal is the same – to define who has access to the data and why, what type of data is available, when, and how.

1. **Who** will have access to the data and **Why**? There are many stakeholders involved in any given research project. The project team should define who the stakeholders are and their objectives for accessing data:
    a. **Data collection and entry staff**. These stakeholders will have access to all the data. When the study requires use and/or collection of PII – both direct and indirect identifiers – these stakeholders hold direct responsibility with how it is managed to mitigate disclosure risk.
    b. **Primary project team**. These stakeholders may have access to all the data. The project team should define in advance whether or not they require *direct* identifiers for analysis and/or study management. Often the main reason for holding direct identifiers is to maintain the study sample contact information for panel surveys. If the study is not a panel survey, the project team should define what direct identifiers they should have access to and why. When the study requires use and/or collection of PII – both direct and indirect identifiers – these stakeholders hold direct responsibility with how data is managed to mitigate disclosure risk. Additionally, as discussed in Privacy Protection, the primary project team should carefully consider what data is required for analysis and therefore what must be collected/extracted – and how they may or may not facilitate access to the data that underlies analysis for transparency and reproducibility objectives.
    c. **Other researcher(s)**. These stakeholders may access all the data. It will depend on what is approved through informed consent, but other project teams may require access to both *direct and indirect identifiers* – for example if they wish to extend the timeline of a panel survey – or semi-de-identified (*direct identifiers removed*) or de-identified (*direct and indirect identifiers removed*). What is possible will be defined by the informed consent process and the objectives of the research. Depending on what is shared, these stakeholders will also hold direct responsibility for how data is managed to mitigate disclosure risk.
    d. **Other Partner(s).** These stakeholders may have access to some of the data. It is typically unnecessary for non-research partners – such as government ministries, project teams – to require and be allowed to hold *direct* identifiers, but it may be necessary for them to access *indirect* identifiers, particularly depending on their objectives for accessing and using the data. One risk here is that some partners may have a motivation for re-identifying the data – for example, if the study is an evaluation of the quality of services provided by a government agency. In this case, the project team has a direct responsibility for mitigating risk to the

research subjects and ensuring these stakeholders have access only to de-identified (*direct or direct and indirect identifiers removed*) data that suits their specific needs.

e. **Funder(s) and Journal(s), as well as general public**. These stakeholders may have access to some of the data for the specific purposes of accountability, transparency, and reproducibility of the research they are funding or publishing. This means that *direct* identifiers are (typically) universally removed for these stakeholders, and inclusion of *indirect* identifiers are carefully considered as per the discussion in Privacy Protection.

2. **What** data should each stakeholder have access to in terms of *identifiable* (*direct identifiers included*) and/or semi-de-identified (*direct identifiers removed*) and/or de-identified data (*direct and indirect identifiers removed*) data? As described above, determine what data will be provided to who.
3. **When** will data be shared (if ever)? Some research studies require one round of data, some require multiple rounds of data. There are trade-offs between timely release of data (for example, sharing baseline data prior to completion of the study), full knowledge/ability to assess re-identification risks to inform de-identification efforts, and allowing primary project teams sufficient windows of opportunity for "sole access" to the data for the primary research requirements. For this reason, the project team should work with stakeholders to not only identify who has access to what, but when, in a way that strikes a balance between timely release, informed de-identification efforts, and aligning with project team incentives for sole access.
4. **How** will the data be shared? As per discussions in Data Sharing and Dissemination, the project team should determine if data will be shared through **public, limited, and/or restricted-access data platform(s)** that allow for direct download, password protected virtual data enclave or other access, in what **format** (XML, CVS, .dta, etc), and with what protocols for **security, version control, documentation**, etc. Identifying the mechanism(s) by which data will be shared is useful for laying the foundation for that sharing in an ethical way, particularly if those platforms are not in place yet. The platform(s) used by project teams is discussed in Data Sharing and Dissemination. In addition, clearly defining the format in which the data will be shared to maximize use early can mitigate issues in the future – such as some stakeholders not having access to certain formats and thereby limiting the usefulness of the data sharing due to format restrictions.

Defining these points early will inform both the required information for an informed consent statement, as well as ethical data management during implementation.

DATA TRANSFER AND USE AGREEMENT(S)

With a proper Research Protocol and DMP in place, the project team will have sufficiently identified all required data sources for the study. After defining the data requirements, the project team should also define who "owns" each of the data sources, and therefore who governs how the data can and will be shared in the future – for computational reproducibility, for broad usability - pending informed consent by research subjects. This may require establishing Data Transfer and/or Use Agreement(s) with stakeholders in advance to define who will have access to what, and ultimately what the project team may or may not be able to disseminate through public, limited, and/or restricted-access use data. Having these agreements in place early not only lays a strong foundation, it also supports the future efforts even when the research timelines are long and (inevitable) staff turnover across stakeholders creates a different landscape than when the study was designed and/or implemented.

INFORMED CONSENT

Informed consent is where respect for persons is operationalized, and where research teams operationalize objectives related to understanding the research subject's vulnerabilities, how their personally identifiable and sensitive data will be shared or not, and what risks may exist and how they will be mitigated. It is therefore recommended that project teams carefully consider the informed consent and/or assent process for all research to determine:

1. **Research subject(s).** Define who needs to grant consent, regardless of whether or not the subject is considered vulnerable or not. Multiple layers of required (or recommended) consent may exist even for one study, regardless of vulnerabilities. For example:
    a. **Community**. Through CSE, the project team should determine if/when a community-level consent for a survey team conducting research in a community may be useful, or when it may actually serve to increase subject vulnerabilities, such as deferential vulnerability (Brear, 2018). An example of this may be in a place with little to no gender equality and limited rights for women. If the main research subjects are pregnant women and mothers and the (male) village leader grants consent for the research to take place in the community and demands all women to participate, the research subjects' deferential vulnerability will need to be assessed with additional consent, and possibly compensation, requirements.
    b. **Facility/Firm/Agency**. When conducting interviews, or extracting data, for a facility-level effort – such as health facility survey and medical record extraction – the project team should determine who needs to consent for facility-level information, as well as who will also provide individual-level data and therefore require individual-level consent. For example, interviewing management on facility-level data such as financials, staffing, and aggregate patient numbers per service may require a manager's consent, but if the survey also interviews health care providers and patients, then each additional research subject should also be considered for their own informed consent process.
    c. **Household**. In some settings, even if the primary research subject is a specific individual – such as women 15-49 years old - the project team may need to obtain consent and/or assent from the Head of Household or similar household decision-maker. The issue of deferential vulnerability of the research subject(s) should be considered in these cases, similar to Community-level consent.
    d. **Individual**. The individual research subject(s) who is being interviewed should go through the necessary informed consent process. However, this is insufficient if the research subject(s) is a minor.
    e. **Child/Minor**. Minor is defined as a person "under the legal age of responsibility". Therefore, first the project team must identify the legal age for consent/legal definition of adult and/or minor in the research setting. For example, in certain countries, a minor is a person under the age of 18. If the research subject(s) is a minor, he/she cannot provide legal consent and must go through the informed assent process, along with Parental Consent. The project team should work with the IRB and CSE to define the protocol for if a child/minor does NOT assent, but Parental Consent is obtained (i.e. obtaining height and weight measurements for children under five).
    f. **Parent**. If the research subject(s) is a minor, Parental Consent is often a requirement, along with Minor Assent.

2. **Content**[33]. Regardless of length, the informed consent will be context-specific and built to inform research subjects on:
    a. Statement that this is research and participation is voluntary
    b. Purpose, duration, and description of specific procedures
    c. Reasonable expected risks
    d. Reasonable expected benefits
    e. Alternative procedures (if applicable)
    f. Promises of Confidentiality and Data Sharing – There is a strong argument for the default position of the project team to protect the confidentiality and anonymity of the research subject and that any deviations from this should be clearly defined in the informed consent. The project team should first determine if (i) PII data needs to be collected (for specific study purposes) AND (ii) if confidentiality promises are required. If PII data is not needed – it should not be collected, thereby limiting specific risks to confidentiality. If the data collected is public – directly observable and not sensitive – promises of confidentiality should be carefully considered as they may be unnecessary. The following statements in the informed consent lay the foundation for (future) ethical open data and data sharing:
        i. Who will have access to what data, particularly when there is a distinction between identifiable dataset and de-identified dataset. If computational reproducibility requires access to identifiable data, then a statement on who will have access to identifiable data for the purpose of reproducibility should be included. If de-identified data will be made public or otherwise shared, then the statement should include this[34].
        ii. Statement on how data will be de-identified (as discussed in Privacy Protection) as applicable;
        iii. Broad consent – The project team may also consider obtaining broad consent for identifiable data to be shared with other researchers for unknown research purposes[35]. Seeking broad consent should be discussed during CSE and the Institutional and/or Ethics Board Review. However, even if broad consent is obtained, future use of the identifiable data should be carefully considered to avoid improper or inappropriate research (for example, see Sterling, 2011 regarding genetic research on Havasupai).
3. **Language**. There are several issues to consider regarding language:
    a. Literacy and comprehension levels. A recommendation is to develop consent forms that are readable and understandable at a 5th grade reading level (or level otherwise defined through CSE). Researchers may also consider visual information such as symbols, pictures or diagrams on the consent form to enhance understanding.
    b. Primary language vs. translation. Project teams will need to consider when they require (i) English version informed consent (for IRB, journal submissions, other requirements); (ii) Spanish/French/Portuguese version informed consent; and (iii) minority language(s) version informed consent given the research subjects primary language. Translation and back-

[33] There are many possible references for this, particularly the Health and Human Services (HHS) site. However, many academic-based research groups have great references for operationalizing the **Common Rule and Revised Common Rule**. University of Michigan's site was specifically referenced for this section - http://research-compliance.umich.edu/human-subjects/common-rule-other-changes/u-m-implementation-informed-consent-changes

[34] This is an explicit requirement in the Revised Common Rule under ***subpart A 45 CFR part 46.116(b)(9)***

[35] This is an explicit requirement in the Revised Common Rule under ***subpart A 45 CFR part 46.116(b)(9)***

translation may be necessary, but require additional time and financial resources, as well as pilot testing.

4. **Comprehension**. Ultimately, how will the project team know that the research subject has full comprehension of the content of the informed consent? Will the interviewer require the subject to confirm their comprehension by repeating the study objectives and completing a comprehension test? Understanding that the goal of informed consent is to mitigate certain vulnerabilities, such as cognitive and deferential, and not just obtain an affirmative response, is critical for the project team to practice ethical research. There should be some discussion in the research protocol and with the IRB, CSE stakeholders on how comprehension can be confirmed.
5. **Documenting consent and/or assent and/or withdrawal**. Once comprehension is confirmed and the research subject(s) is willing to grant consent and/or assent, the interviewer will need to document the consent/assent. How this is done may vary by context. In some settings, having the research subject sign the consent form is appropriate, however this may not be appropriate in settings where the research subject is illiterate or otherwise unable to sign. In such cases, the project team should have a protocol in place with the IRB for verifying consent/assent, such as fingerprinting, having a neighbor or other household member sign, or other appropriate method for documenting consent. The project team will also need to determine how the documentation will be shared with the research subject and stored with the project team. An additional process should be in place for documenting the subject's withdrawal if necessary.

As per US regulation, it is possible to request a waiver for informed consent when certain conditions are met (see 46.116(e) and (f)). As with any regulation, it is for the project teams to determine, preferably with an IRB or other external ethics review board, how respect for persons, beneficence, and justice will be adhered to in the event a waiver of informed consent is requested (and accepted).

In addition to the above, NIH offers some suggestions for consent strategies if the standard written form approach is not sufficient or may be insufficient given the research subjects' needs and vulnerabilities:

- Consider videotapes that show the process or performance of treatment or research.
- Consider the provision of videotapes for use by interpreters and subjects to explain consent forms. (Remember to protect the confidentiality of subjects if videotapes are used to obtain consent.)
- Consider developing interactive computer programs to assist in informed consent process.
- Allow prior research subjects to interact with new subjects.

One important question is whether and how to obtain informed consent when the project team does not have direct engagement with the research subject. For example, if the project team obtains data from a third-party source that has collected (whether individual or aggregate-level) data on the research population and agrees to share the data with the project team. While access is not an issue – the third-party may agree to share the data with the project team – the question still remains whether or not the research subject is aware her data is being used for the purpose of the specific research project. As with other issues surrounding ethics, this document recommends the project team still engage with a sub-sample of the research population to assess whether or not there are concerns/issues with the third-party data being used for the defined research objectives. This should inform discussions with the IRB on the issue of waivers of consent.

FACILITATING (FUTURE) DATA SHARING

When documenting the study design through a research protocol, DMP, or other design document(s), the project team should carefully consider how elements of the study design or plan may inform (or even prevent) future efforts for data de-identification and dissemination, as discussed in Privacy Protection. These elements include:

- **Informed consent –** How will the project team manage documentation of informed consent to share identifiable and/or de-identified data? If some research subjects consent to this data sharing and some do not, how with the project team ensure appropriate differentiation and what does this mean for meeting transparency and reproducibility objectives?
- **Linkage documentation –** For future de-identification efforts, the project team will need to be well-aware of linkage documentation that may support re-identification efforts or at least mitigate de-identification efforts. For example, for a public-use data set, the project team may determine village names need to be de-identified and removed from the data. This may be done in the data, but if the names of the villages in the sample are disseminated elsewhere, such as in a Design or Baseline Report in earlier stages of the research life cycle, this information could be used to re-identify the village names in the dataset and increase household/individual re-identification risk. For this reason, the project team should carefully consider what information is and will be available and when about the study sample that may pose a re-identification risk and limit its dissemination.
- **Sample frame** – What is the source for the sample frame, how available is this source to others, and what percentage of the sample frame will be selected for the study? These are important questions that will inform data de-identification efforts, specifically focused on understanding the extent to which outliers in the study sample may be outliers in the population, and therefore potentially useful for re-identification of individuals, households, communities, etc.
- **Knowledge of Treatment** – If the research is a program evaluation, how well-known will "treatment" status be? For example, will random selection of communities/villages/schools/facilities/etc receiving the treatment be publicized? This should be carefully considered as it is a form of linkage that may support re-identification efforts if the treatment status, or other information about the treatment group, is known and can help to re-identify individuals. Project teams may therefore consider carefully how treatment status, program beneficiary lists, etc, may be managed to mitigate future re-identification risk.
- **Create flags for identifying and sensitive data** – Beginning with questionnaire design and data entry, the project team should consider creating flags – such as a specific suffix in the variable number or name – to create an easy reference in data analysis, de-identification, and dissemination for variables which should be carefully considered. These variables may then easily be removed from the data for ethical data sharing following the promises of confidentiality and who should have access to what according to the DMP.

RESEARCH IMPLEMENTATION

After setting a strong foundation for ethical research in the research design stage, during data collection and/or extraction, storage, and transfer activities, the primary goal is to follow the protocol and plan. However, during implementation there are three big issues for concern:

Develop a protocol/plan if you don't have one yet. Even if a project team is halfway through a research project and well past the design stage, if there is no research protocol and/or data management plan in place, putting one in place sooner rather than later is better than never. In this case, refer to the above section regardless of which stage the research is in.

Maintain fidelity/compliance to the plan. The project team should determine how it will monitor implementation to ensure compliance with the ethical research practices put forth in the research protocol and/or DMP. Ochieng et al. (2013) offers some insight on research implementation in low-resource settings. In their study of research sites in Uganda, 28 site monitoring visits covering 40 research projects were reviewed for a four-year period. 25% of the site monitoring reports revealed violation of the regulatory requirement for valid ethical approval. 36% of the site reports showed some instances of informed consent violation, including not obtaining consent and not having documentation of consent. 28% showed violation of the rights and welfare of research participants, including not compensating participants for their time or providing inadequate compensation. For example, in some sites, participants waited for long periods without a snack or a meal and were not compensated for time and work lost. The practice of ethical research is certainly in the details of how it is implemented, and project teams should have a careful plan in place for how to monitor and ensure compliance with the methods prescribed in the research protocol and DMP.

Ensure effective breach management and risk mitigation. The project team should determine how it will respond if there is a disclosure risk – for example, a box of completed paper questionnaires with PII and sensitive data are stolen during transit - during implementation. The project team should have a plan for what documentation, reporting, and course correction is in place for such an event. In the unfortunate event of a breach, it will be beneficial for the project team to already have assessed and documented the sensitivity and risk of required data, for example using tools such as the Harvard Information Security Data Classification Table. This assessment can provide a guide to the team in terms of the risks of the disclosure and inform necessary steps for the breach response. This may include:

- Full documentation of the breach and potential risks to the human subjects and/or others
- Notification to the IRB(s)
- Notification to the human subject(s) affected by the breach

COLLECTION

During data collection and/or extraction period, the project team should consider the following for compliance, risk monitoring, and course corrections:

- Sufficient training and testing for survey team
- Sufficient pilot testing for survey team
- Sufficient pilot testing for course corrections
- Field visits and spot checks

- Real-time data quality reviews
- Secure transport

DATA STORAGE, TRANSFER, AND DISPOSAL

As defined in the DMP and depending on the flow of raw data from the field, the following should be considered for ethical study materials storage, transfer, and disposal of survey materials:

Table 4.2: Mapping Data Types, Storage, Transfer, and Disposal Guidelines

Type	Purpose	Storage	Transfer	Disposal	
				Timeline	Requirements
Paper – *All Data*	Verification of survey responses	Locked box and/or cabinet	Paper-based questionnaires should only be transferred from the field to centralized office for data entry. Transfer should be in secured storage and securely handled by defined data handlers.	Immediately following double-data entry	Shredded or burned depending on local resources and local requirements
Digital – *All Data*	Data analysis	Encrypted data files; Password protection on data systems and data encryption; **end point encryption software should meet** AES-256 encryption **standards or above**.	Communication channels are encrypted, especially Wi-Fi connections; File transfers should occur only through **https** connections; Use of hyperlinks for connections should be prohibited; instead, users should only connect to trusted sites by manually starting a new web-browsing session; Password protect and encrypt all PDFs or other document types if there are no other solutions available for secure file transfers; Send passwords via a separate email or phone the recipient.	As defined per data handler in the DMP	Delete/wipe from hard drive and/or change password(s) for managed access to cloud-based platforms
Digital – *De-identified data*	Data analysis				
Digital – *Direct Identifiers only*	Re-contact for panel surveys				

Once data is transferred to the project team from the field, digital data should be stored in the centralized, secured folder.

While this is a high-level summary, a more detailed guidance on Data Security Procedures for Researchers, including storage, access, transfer, and erasing data is found in O'Toole, et al (2018).

PRIVACY PROTECTION

Privacy protection is a particularly relevant component of ethical research and often sits at the intersection of transparent, reproducible, and ethical research. Data privacy protection requires attention at all stages – design, implementation, and dissemination. As discussed above, project teams should be carefully considering privacy protection of the human subjects involved in their research throughout the research life cycle, regardless of whether their research requires direct or indirect engagement with human subjects. The ability of a project team to be fully transparent and share the data underlying the research - publicly or through a limited or restricted-access mechanism - relies on maintaining this privacy protection, when applicable.

As discussed above, some research may have minimal privacy concerns – for example if the data collected or extracted is already publicly available and well-known with minimal risk to human subjects. Some research may pose significant privacy concerns – for example if the data collected or extracted is sensitive and disclosure of which may pose risk to human subjects for harm and/or exploitation. Understanding how the project team may de-identify the study documentation and/or data, as well as the risks for re-identification, is critical for considering how transparent and computationally reproducible the research can be while still providing adequate privacy protection.

DEFINITIONS

For issues related to privacy protection, this document uses the following definitions (from NIST, 2016):

- **Direct identifying data**: Data that directly identifies a single individual.
- **Quasi-identifier**: Data that can be used to identify an individual through association with another variable. Also referred to as *indirect identifiers* as defined in Personally Identifiable and Sensitive Data.
- **Perturbation-based methods**: Perturbation-based methods falsify the data before publication by introducing an element of error purposely for confidentiality reasons. This error can be inserted in the cell values after the table is created, which means the error is introduced to the output of the data and will therefore be referred to as output perturbation, or the error can be inserted in the original data on the microdata level, which is the input of the tables one wants to create.
- **De-identification** is the general term for any process of removing the association between a set of identifying data and the data subject. De-identification includes all techniques that provide researchers with access to microdata while simultaneously limiting the opportunity for disclosure. De-identification takes an original dataset and produces a de-identified dataset.
- **Re-identification** is the general term for any process that restores the association between a set of de-identified data and the data subject.
- **Disclosure limitation** refers to statistical methods used to hinder anyone from identifying an individual respondent or establishment by analyzing published data, especially by manipulating mathematical and arithmetical relationships among the data.

Based on the above definitions, **data de-identification is the process of applying data perturbation-based methods to appropriately manage direct identifying data and quasi-identifiers to remove or significantly reduce risk of re-identification and ensure disclosure limitation**.

DE-IDENTIFICATION - DOCUMENTS

This section provides guidance for project teams on what to consider for de-identification of study documentation and data. As discussed above in Design, a first step to data de-identification is project teams considering re-identification risk with regards to available documentation about the study sample. Certain information that may support re-identification includes:

- In program evaluation, lists of program beneficiaries available through the program website, newspaper stories, other dissemination materials about the program beneficiaries. This documentation can provide linkage between the research sample data and already publicly available data.
- Study design documents that detail the sample frame, including information regarding sample units and number of observations. In cases where de-identification efforts may require de-identifying these geographic units, the provision of earlier information on the sample frame can facilitate re-identification of sample units.

With this in mind, project teams should do a careful sweep of documentation they are developing and disseminating, as well as a sweep of all other available documentation regarding their study sample prior to considering data de-identification.

DE-IDENTIFICATION – DATA

Prior to conducting data de-identification actions, project teams should consider:

- Early and often about what data is required for study analysis and determine which data perturbations and techniques can be applied BEFORE study analysis to **maintain a better link between the data that underlies analysis and the data that can be shared through public, limited, and/or restricted-access use**.
- Data de-identification relates to **how high the probability is for re-identification** and how to mitigate this risk. Issues to consider are presented below in Table 4.3.
- There is a **balancing act between applying data perturbation-based methods and techniques to de-identify data and the quality, usability, and relevance of the data**. In many cases, significant de-identification efforts may result in data that is less useful and/or relevant, even for replication of original study analysis. This is particularly true for **qualitative data**[36], where the data may not only be sensitive, but very difficult to de-identify in a way that retains its usefulness.
- Project teams may have to **carefully consider combinations of variables**, even when individual variables do not pose a re-identification risk. For example, age alone, gender alone, and marital status alone may not pose re-identification risk, but when combined there may be only one 20-year-old female widow in the sample AND in the general population the study sample was drawn from. This combination results in a re-identification risk where de-identification efforts may be required in one or more of the individual variables.
- When balancing data de-identification and usability for computational reproducibility and other analysis, the project team should carefully consider **different data access-levels**. While significantly mitigating re-identification risk should be the goal for direct download, publicly accessible data, the

[36] For data transparency discussions related to qualitative data, please reference The Qualitative Transparency Deliberations (QTD) available here - https://www.qualtd.net/

project team may consider other access-levels, such as limited access and restricted-access data enclaves. This is discussed further in Data Sharing.

Table 4.3 – Initial Risk Assessment by Risk Factors		
Risk Factor for re-identification	**Lower probability**	**Higher probability**
Sample representation: Are outliers in the data outliers in the general population?	When the sample is a small percentage of the general population, visible and known characteristics that are outliers in the sample may not pose a re-identification risk	When the sample is a large percentage of the general population, visible and known characteristics that are outliers in the sample may pose a stronger re-identification risk
Linkage documentation: What documentation about the sample exists outside the research data but can link to it?	If little to no documentation exists about the study sample then linkage documentation may not pose a re-identification risk	If documentation exists about the study sample then linkage documentation may not pose a re-identification risk (examples: loan information obtained on study sample mirrors loan information at bank)
Timing and population characteristics: How closely does the data reflect current and future state for the sample population?	If significant time has passed and the study population is transient or nomadic, there is lower re-identification risk	If the data was recently collected and the study population is more permanent, there is higher re-identification risk

Once the above has been considered, project teams may consider the following high-level data perturbation techniques for standard data de-identification:

- **Removal of all direct identifiers.** This is the first condition of privacy protection. Removal of direct identifiers may not be as simple as removing the specific variables where known direct identifiers were recorded by the survey team. For example, the written response within “Other” responses may include detailed information and direct identifiers.
- **Geographic units**. Consider the highest geographic level that should remain identifiable for specific analytic purposes and de-identify all lower geographic units through use of a randomized numeric identifier. Similar to the discussion above on sample representation, the higher the geographic unit that is identifiable, the lower the risk for re-identification at individual, household, and other sample unit levels. When geographic units remain identified at lower levels, such as village, there remains a higher risk for re-identification of the individuals, households, other sample units within that geographic unit. If lower geographic units can remain de-identified, there is less risk for re-identification at the individual variable level, and therefore less need for significant data perturbation on a variable by variable basis.
- **Top and Bottom Coding**. When specific continuous variables are visible and/or known characteristics about the research subject (i.e. visible asset holdings, age, years of education), outliers may need to be considered for top and bottom coding. There is no specific rule (top and/or bottom 2%, 5%, etc) given the decision on where to cut outliers should be made based on the data and what is known about the study sample population. Once a threshold is identified, to retain data values and avoid lost data, researchers can send outlier values to the median.
- **Re-categorization**. When specific categorical variables are visible and/or known characteristics about the research subject (i.e. ethnicity, religion, language spoken, education level), minority groups may need to be considered for re-categorization. To retain the value of the data, it’s preferable to re-categorize into meaningful groups, combining categories, rather than collapsing into an unknown “Other” category. However, this is dependent on context, data, and risk.

- **Removal**. When specific variables cannot be retained given potential re-identification risk, the variable(s) should be removed from public-use datasets (and clearly documented as such).

For project teams that deal with sensitive data with significant risk for re-identification, there are other de-identification tools that may be considered if the tools above are considered insufficient. One such tool is Differential Privacy. From the Harvard's Differential Privacy research group: "*Differential privacy is a rigorous mathematical definition of privacy. In the simplest setting, consider an algorithm that analyzes a dataset and computes statistics about it (such as the data's mean, variance, median, mode, etc.). Such an algorithm is said to be differentially private if by looking at the output, one cannot tell whether any individual's data was included in the original dataset or not. In other words, the guarantee of a differentially private algorithm is that its behavior hardly changes when a single individual joins or leaves the dataset -- anything the algorithm might output on a database containing some individual's information is almost as likely to have come from a database without that individual's information. Most notably, this guarantee holds for any individual and any dataset. Therefore, regardless of how eccentric any single individual's details are, and regardless of the details of anyone else in the database, the guarantee of differential privacy still holds. This gives a formal guarantee that individual-level information about participants in the database is not leaked.*" For more information on differential privacy, project teams are referred to Harvard's Differential Privacy research group.

DE-IDENTIFICATION – PROCESS, REVIEW, AND CLEARANCE

Project teams are advised to ensure proper documentation of data de-identification actions. This is essential for transparency and accountability, and ensuring that new users of the data are aware of how data perturbations may affect usability of the data – for computational reproducibility and/or broader analysis.

Project teams should therefore document their de-identification actions in the Data De-Identification Worksheet, and ensure the following is answered (adapted from NIST 2016):

- What is the access-level for the data?
- How many rounds of data are required for the research and when in the research life cycle is de-identification performed? Will de-identification occur for individual rounds of data or all together as one package?
- Are there specific datasets that can be used to re-identify the de-identified data? If so, what controls are in place to prevent intentional or unintentional re-identification?
- What data perturbation(s) were applied and for what variables? Can the original study analysis be reproduced from the de-identified dataset?
- What is the risk to research subjects if re-identified occurs?
- Is there a mechanism that will inform the project team if there is an attempt to re-identify the de-identified dataset? Is there a mechanism that will inform the IDB if the attempt is successful?
- Will the original dataset be retained after de-identification? If so, where? Is it accessible? Or is there a key or map retained, so that specific data elements can be re-identified later?

Once the project team has completed the necessary data de-identification and documentation, they should prepare the Data Package for review and dissemination, including the (i) Data De-Identification

Worksheet, (ii) Informed Consent statements, (iii) De-Identified Data, (iv) Analysis Code (for replication of analysis), (v) Questionnaires, (vi) Codebook (exported from the de-identified data).

RESEARCH DISSEMINATION

There are several motivating factors for project teams to share their data:

Intrinsic Motivation	Extrinsic Motivation
The data is a public good	Someone may find an error that prevents incorrect conclusions/recommendations
Sharing fosters collaboration with other researchers	Data sharing can increase impact of research by making it more credible
	Data sharing can increase impact for researchers by increasing citations on data, as well as research
Data can contribute to broader meta-analysis efforts	Researchers with outside funding may be required by other funders
	Researchers who want to publish in certain journals may be required by journals (See AEA journal policy for example)

In addition to these motivating factors, working toward open data – public and/or restricted-access – aligns with the IDB's commitments to research transparency and reproducibility, as codified in its 2010 Access to Information Policy, where it is stated: "*The Bank reaffirms its commitment to transparency in all aspects of its operations as a means of aligning itself with international best practice, especially among the countries of Latin America and the Caribbean, and as a matter of enhancing its accountability and development effectiveness. Through implementation of this policy the Bank seeks to demonstrate its transparent use of public funds, and by deepening its engagement with stakeholders, to improve the quality of its operations and knowledge and capacity building activities*."

In addition, project teams should determine how best to manage their data sharing to maximize use. This includes considering the following:

- Ensure there is contact information posted with the data so any new users know who to contact with questions.
- Develop and deliver data dissemination workshops to train new users on how the data is built and may be used. These can also be filmed and posted with the data.
- Develop and share data dissemination training or instruction videos with the data files.

IDB DISSEMINATION PLATFORMS

At the IDB, there are two main mechanisms for dissemination:

- Publications – This is the platform for disseminating all documentation related to the research projects. This platform can be considered the "backbone" for disseminating research products since project teams can directly link the underlying data that they are publishing in the Open Data platform to the Publications platform.
- Open Data – This is the platform for data dissemination. Data can be posted here to be searchable in the Open Data platform, but ideally also linked back to the respective research project's Publications entry to link to the Documentation related to the research.

An example of an IDB research project that demonstrates the link between the Publications and Open Data platforms is: Serving Citizens: A Decade of Civil Service Reforms in Latin America (2004-13)

When preparing for dissemination, project teams should consider the following documentation in addition to the Working Papers or other published Articles:

Table 4.5: Documentation Requirements

Document	Requested Format	Description
Metadata File (for both Publications and Open Data platforms)	IDB requirements	The metadata can be updated/revised as necessary over the course of the research life cycle and should provide the basic elements of the study: Title, Project team, Timeline, Methods, Sample, while pointing interested users to the full documentation available.
Study Design Materials	Word, searchable PDF	Depending on what the project team has produced, documentation related to the research design (Design Report, Pre-Analysis Plan, etc) should be posted as complementary documentation for any Working Paper or published Article.
Informed Consent Statement	Word, searchable PDF	The informed consent statement should be published, either independently or as part of the questionnaire(s).
Questionnaires and/or other survey materials	Original editable source and searchable PDF	All survey questionnaires – baseline, interim, final - should be shared in a way that enables reuse by sharing the original editable source file. Project teams may also submit a searchable PDF. Related documentation may also include sampling, field operations and interviewer manuals when needed for complete documentation of survey protocols. For qualitative data, this documentation may include de-identified codebooks, field notes, researcher journals, etc. that would enable replication of the study.

APPENDIX

APPENDIX A: RESEARCH DETAILS REQUIRED FOR M&E REPORT, RESEARCH PROTOCOLS FOR IRB, AND PRE-ANALYSIS PLANS.

Click here to download file with hyperlinks.

Table A0: One Page Summary of Activities and Resources for Transparency, Reproducibility and Ethics Across the Research Cycle

Make it:		Design/Project Preparation			
	Transparent	Registration (section 2.1 of this document)	Create a public record of the study's main hypothesis and methods. Recorded ideally before implementing the study.	AEA Registry: registry for RCTs ClinicalTrials.gov: Health related RTCs	OSF Registry: All methods RIDIE registry of impact evaluations in development
		Pre-analysis plan (section 2.2)	Create an extensive methodological descriptions of the analysis to be performed. Should be recorded before implementing the study.	PAP Template IPA guide to PAPs	JPAL Examples Declare Design
	Reproducible	File Management & File Structure (3.1)	Coordinate with team members on uniform file structure. Create a project folder and pre-populate with chosen structure.	TIER protocol	TIER specifications
			Decide format for report writing (Word, Tex, md). Create a readme.md[.txt] in the root folder of the project	TIER readme file	Markdown guide
		Version Control (3.2)	Define version control strategy: (i) systematize renaming (ii) version control software.	Tutorial on Git Renaming Guidelines	Git/Github + R Guidebook
		Code Readability (3.3)	Adopt/define style guides for coding.	Style guides for R Stytle guides for Stata v1	Style guides for Python Stytle guides for Stata v2
	Ethical	Training	To fully own ethics of research, project teams may conduct additional training on protection of human subjects and other ethical research considerations.	World Health Organization (WHO) United States National Institute for Health (NIH) Public Responsibility in Medicine and Research (PRIM&R)	
		Community Stakeholder Engagement (CSE)	Assess vulnerabilities and identify sensitive data based on local context; as well as needs and priorities of the communities involved in research.	Reference Harvard Information Security Data Classification Tool to determine sensitivity/risk of data	
		Legal requirements	Understand laws, regulations, and guidelines on human subjects' protections in 130 countries and from many international organizations.	2018 International Compilation of Human Research Standards by Health and Human Service	
		Independent Review	Understand (i) local, partner, funder, journal requirements; (ii) process for identifying IRB, documentation requirements, costs, and timelines.	AAHHRP-accredited IRBs	
		Research Protocol	Define the planned ethical research design and practices to inform the implementation.	Follow outline in Appendix A1.	
		Data Management Plan	Define who will have access to what data and when.	DMP Tool ICPSR NSF	
		Informed Consent	Determine who needs to grant consent and/or assent and in what language and the necessary information regarding study objectives and data management plan.		

Table A0: One Page Summary of Activities and Resources for Transparency, Reproducibility and Ethics Across the Research Cycle

Make it:

		Implementation		
Transparent	Registered Reports (2.3)	Detailed pre-analysis plan to be submitted to a journal before analysing the final data.		JDE RR
	Pre-analysis plan (2.2)	Document deviations from PAP.		Standard Operating Procedures
	Additional Analysis (2.2)	Perform additional exploratory analysis.		Example (Hausofer UCT)
Reproducible	File Management (3.1)	Track file workflow with diagram. If the workflow diagram becomes too complex, use different levels of abstraction.	draw.io Examples	coggle.it
	Data management plan (3.4)	Execute Data Magagement Plan (see Ethics/Design) and perform data cleaning and storing protocols.	Data cleaning checklist from DIME	
	Code (3.3)	Perform parallel coding session to verify similar style. Run unit tests. Create a master code that reproduces all the analysis.	Section 3.3 of guidelines	
	Dynamic Documents (3.3)	Write your results using an implementation of dynamic documents, eg Rmarkdown,	DD in R DD in Stata	DD in Python DD in Stata example
Ethical	Collection, Storage and Transfer	Locked cabinents; Encrypted data files; Password protection on data systems and data encryption; End point encryption software should meet AES-256 encryption standards or above.	Data Security Procedures for Researchers (JPAL) AES-256 encryption standards.	
	Disposal	Shred/burn paper versions; Delete/wipe from hard drive and/or change password(s) for managed access to cloud-based platforms.		
	Breach Management Plan	Define how to ensure risk mitigation in the event of a data breach (during interview; lost questionnaires; hacked/stolen computers; etc) or other breach in protocol (harm or safety issues with research team) and follow if necessary.		

Table A0: One Page Summary of Activities and Resources for Transparency, Reproducibility and Ethics Across the Research Cycle

Make it:

	Dissemination			
Transparent	Reporting Guidelines (2.4)	Declare which guidelines are followed to present the results to maximize visibility in future meta-analysis/systematic reviews.	CONSORT (RCTs in Medicine) JARS (RCT and Observational in psychology)	Evans & Snilsvert in Development JDE Checklist (for register reports component)
Reproducible	Code and Data Sharing (2.4)	Review project folder and remove personal identifiable information; If PII is needed to reproduce results create two version and provide instructions to access private version.		Dataverse OSF Figshare Zenodo
	Transparency and Reproducibility	Use the TOP guidelines to disclose degree of	Attach a Transparency and Reproducibility checklist to the final report	
Ethical	De-identification of Documents	Ensure privacy protection and adherence to promises of confidentiality of human subjects as per the Informed Consent process by reviewing all study documentation prior to publication to mitigate risks for re-identification of human subjects. This includes careful consideration of sample frame definitions and documentation as well as keys related to the sample.		
	De-identification of Data	Ensure privacy protection and adherence to promises of confidentiality of human subjects as per the Informed Consent process by identifying direct and indirect identifiers and considering necessary data perturbations (data removal, top/bottom coding; re-categorization; other methods).		
	Managed data sharing	Ensure privacy protection and adherence to promises of confidentiality of human subjects as per the informed consent process by utilizing appropriate data dissemination platforms for restricted-access use (such as virtual data enclave) if necessary.		

APPENDIX

APPENDIX A: RESEARCH DETAILS REQUIRED FOR M&E REPORT, RESEARCH PROTOCOLS FOR IRB, AND PRE-ANALYSIS PLANS

The main document describes four key documents involve in transparent, reproducible and ethical research: A study registration (section 2.1), an IRB research protocol (section 4.2.5), a pre-analysis plan (section 2.2) and a register report (section 2.3). This appendix compares the content required for each component and suggests where to align some of these components in the IDB's Monitoring and Evaluation (M&E) report, at the beginning of a project, and in the final report (or paper) at the completion of a project.

Table A2 (in the accompanying excel file) lists all the components required in each document, and identifies common elements across them. For a high-level comparison, table A1 compares the content and level of detail for all of these documents. Each document represents a column, sorted according to their role in a project timeline. The M&E report contains, among many other elements, the first elements of the research design, methods and a plan specifying the project teams' approach regarding computational reproducibility (it can also contain IRB information but is not required at this point in time). Per section 2.1 the registration is minimal requirements and should be a subset of the M&E report. The IRB research protocol requires a more extensive explanation of the project combined with extensive proof of compliance with ethics protocols. The PAP requires additional detail in the methodology but only suggest proof of IRB compliance. The register report is essentially the final report except for the results section. Table A2 (in the companion excel file) uses the framework for an IRB research protocol and identifies which components are necessary for the M&E report, for the registration and for the PAP. The details required for a RR are more extensive and are presented in Appendix C.

Table A1: Research Details Required per Research Product. Summary

	M & E Report	Registration	IRB Protocol	PAP	RR	Final Report / Paper
Relevancy + Methods/ Research Design						
Reproducibility (comp)						
Ethics Reporting						
Results						

Little detail
High detail

Appendix A2: Research Details required per Research Product adapted from WHO					
Section	**Description**	**M&E Report**	**Registration**	**Research Protocol for IRB**	**Pre-Analysis Plan**
Summary	Should be no more than 300-500 words. It should summarize all the central elements of the study: rationale, objective(s), method(s), population(s), time frame(s), and outcome(s).	Yes	Yes	Yes	
General Information	Title, Identification/Registration Number (if any), and date.	Yes	Yes	Yes	
	Name and address of the sponsor/funder(s).			Yes	
	Name and title of the investigator(s) who is (are) responsible for conducting the research, and the address and telephone number(s) of the research site(s), including responsibilities of each.	Yes	Yes	Yes	
	Name(s) and address(es) of the data collection firms; laboratories; other relevant research collaborators.			Yes	
Rationale and Relevance	Defines the reasons for conducting the research given current evidence and knowledge. It should include a **well-documented statement of the need/problem that is the basis of the research, the cause of this problem, and possible solutions. It should answer the question of why and what: why the research needs to be done and what will be its relevance**. This should be followed by a brief description of the most relevant studies that serve as the foundation for and inform the proposed study. This section should also describe the generalizability/external validity of the study. Reference for the literature should be included as an Annex.	Yes		Yes	Yes
Description of Intervention	**IF APPLICABLE**, this section should link the rationale and problem diagnostic to the study objectives by describing WHAT the study will be observing. This is directly relevant to **program evaluation studies, where there is a specific intervention proposed to address the problem(s) and the study is being used to test the effectiveness/impact of that intervention**.	Yes	Yes	Yes	Yes
Research Questions or Objectives and Hypothesis	This should detail the Research Queestions or Objectives (specific objectives are statements of the research question(s)). Questions/Objectives should be simple (not complex), specific (not vague), and stated in advance (not after the research is done). This section should also detail the researcher(s)' hypothesis for each Question/Objective *(i.e. What is the impact of the intervention on school enrollment rates? We hypothesis an increase in enrollment by at least 5%).*	Yes		Yes	Yes
Design	This section should include information on the type of study – for example if it is an impact evaluation if it will be an experimental or quasi-experimental design. Additionally, this section should describe the research population, power calculations and required sample size and sampling frame, inclusion and exclusion criteria, withdrawal criteria etc. There should also be some discussion on required exposure period (*how long the population must be exposed to the treatment for expected impact on outcomes to be achieved*) as applicable.	Yes	Yes	Yes	Yes

Appendix A2: Research Details required per Research Product adapted from WHO					
Section	**Description**	**M&E Report**	**Registration**	**Research Protocol for IRB**	**Pre-Analysis Plan**
Method(s)	This section is the detailed information on how the study design will be operationalized. For example, if it is experimental design and a randomized control trial - additional information on the process of randomization (specifying unit of randomization and unit of observation as necessary) and blinding, description of stopping rules for individuals, for part of the study or entire study, the procedures and conditions for breaking the codes etc. should be described. Additionally, a graphic outline of the study design and procedures using a flow diagram should be provided. Procedures for biomedical data collection (collection of blood or sputum samples to develop a diagnostic test), and interviews (doing a questionnaire survey, carrying out a focus group discussion as part of formative research, observation of the participant's environment, etc.) should be described. Instruments which are to be used to collect information (questionnaires, FGD guides, observation recording form, case report forms etc.) should also be provided.	Yes		Yes	Yes
Duration of Study	The protocol should specify the time that each phase of the project is likely to take, along with a detailed month by month timeline for each activity to be undertaken.	Yes	Yes	Yes	Yes
Expected benefits	This section should define who is expected to directly and/or indirectly benefit from the research findings and what the benefit(s) of the research findings are expected to be. If compensation will be provided, this may be included, in consultation with the IRB.			Yes	
Potential for harm and/or exploitation	This section should define who may be potentially harmed and/or exploited, what the potential harm and/or exploitation may be, and steps to mitigate.			Yes	
Informed Consent(s)	The protocol should include all copies of necessary informed consent and/or assent.			Yes	
Risks and Challenges	This section should discuss the difficulties that the investigators anticipate in successfully completing their projects within the time frame stipulated and the funding requested. It should also offer possible solutions to deal with these difficulties.			Yes	
Safety Considerations	This may be a sub-section of the Harm and/or Exploitation section. The safety of research participants AND research teams should be considered in the research protocol, with appropriate procedures on how harm and/or exploitation will be mitigated given the context and study requirements.			Yes	
Quality Assurance	This section should describe the quality control and quality assurance system for the conduct of the study, including data quality control, sample attrition management, etc.			Yes	
Disclosure and Reporting	This section should include clarification on what the interview and/or research team will do if the data collection results in obtaining information regarding direct risk to the research subject. This can include:			Yes	
	Field based biomarker tests – If research subjects consent to be tested for anemia, malaria, other biomarkers in the field, how will the research team manage disclosure/reporting to the research subjects if they found to have a serious health risk (such as testing positive for malaria, testing positive for anemia). Will subjects be treated in the field? Will they be referred to a clinic? Does the research team need to facilitate this treatment and/or referral and how?			Yes	
	Mental health crisis – If the research subjects report levels of severe depression and anxiety, what is the research team's responsibility for facilitating treatment?			Yes	
	Abuse, violence, other direct harm – If the research subjects report specific incidence of abuse, violence, or other direct harm, what is the research team's role in responding to this and/or facilitating a response?			Yes	

Appendix A2: Research Details required per Research Product adapted from WHO					
Section	**Description**	**M&E Report**	**Registration**	**Research Protocol for IRB**	**Pre-Analysis Plan**
Data Management and Statistical Analysis	This section should describe how the data will be managed, including data handling and coding for analysis, monitoring and verification. The statistical methods proposed to be used for the analysis of data should be clearly outlined, procedures for accounting for any missing or spurious data etc. For projects involving qualitative approaches, specify in sufficient detail how the data will be analyzed. This is the foundation for a Pre-Analysis Plan,	Yes		Yes	Yes
Workflow Management*	This section should describe how all files from the project will be archived. It should specify the specific folder structure	Yes			
Version Control Strategy*	This section should specify what type of version control strategy was chosen to keep track of the code for the project. If a rename-and-save strategy is chosen, this section should include the patterns, protocols and examples. If a version control software is chosen, this section should name it.	Yes			
Code readability & Dynamic Documentation*	This section should discuss the coding style used through the project. It can mention a third-party style guide or describe an ad-hoc guide develop for the project. In this section researchers should announce if the they will be using dynamic documentation.	Yes			
Code and Data Sharing*	Describe where is the all the project workflow folder being shared. Specify chosen repository and Digital Object Identifier. If public and private project folders differ for privacy considerations, explain and provide instructions on how to access the complete data sets.	Yes			
Reporting Guidelines*	This section should provide information about the reporting guideline used if any. Examples include CONSORT, Equator Network, and Minimal recommendations provided in the IDB guidelines.	Yes			
Dissemination Plan	This section should describe how results will be disseminated across relevant stakeholders, with particular emphasis on ensuring the benefits of the research are made available to those who participated as feasible. This is also discussed in more detail in the Dissemination Chapter of these Guidelines.			Yes	
Roles, and Responsibilities	The CVs of all key research team members should be available and this section should define the roles and responsibilities of each.			Yes	
Budget	The budget section should contain a detailed item-wise breakdown of the funds requested for, along with a justification for each item.	Yes		Yes	
* These components where not in the WHO original and where added to reflect the elements required for computational reproducibility					

APPENDIX B: CONTENT TO INCLUDE IN MONITORING AND EVALUATION PLANS

I. Considerations of Transparency, Reproducibility and Ethics

To meet the "transparency and credibility" requirements of the DEM, monitoring and evaluation plans should follow the procedures outlined in the Technical Note on Best Practices in Transparent, Reproducible, and Ethical Research, as applicable for the context of each project. The monitoring and evaluation plans should indicate the specific activities to be undertaken for the impact evaluation, prospective economic analysis and other empirical research associated to the project such as surveys financed through the operation. The following text can be adapted by project teams for inclusion in the monitoring and evaluation annex of the POD:

I.1 Transparency and Reproducibility

On transparency and reproducibility of the [research activity], the project will comply with the three key principles of research transparency (following Miguel et al., 2014): (i) disclosure, (ii) registration and pre-analysis plans, and (iii) open data and materials. These high-level principles will be operationalized into four specific deliverables outlined below.

1. ***Registration of the [research activity]**: the [research activity] will be registered in in [clinicaltrials.gov if there is an expectation of publishing result in health-related journal] [socialscienceregistry.org if it is an RCT[37]] [in osf.io or ridie.org for other cases]. The registration will include: title, authors, country, status, keyword, abstract, start and end dates, outcomes, intervention, basic research design, whether treatment clustered, and IRB information. The registration is anticipated to be finalized before [date] when the final data is collected for analysis.*
2. ***Pre-analysis plan (PAPs):** a detailed plan specifying all the planned analysis will be attached to the registration describing in extensive detail, all they hypothesis to be tested. PAPs will be submitted before the final data is available for analysis. The contents of the PAP will help to identify the hypothesis to be tested in a confirmatory fashion. This does not preclude the project from running additional analyses as long as they are properly distinguished from the ones in the PAP (this latter part is should be labeled "exploratory analysis").*
3. ***A computationally reproducible workflow:** all the components required to reproduce the final analysis from the original data will be provided in a self-contained folder, named 'eval_name_of_the_program', with a well-defined folder structures (see section 3.1 of guidelines), and with clear instructions on how to execute the different files to produce intermediary and final output.*
4. ***Code and data sharing plan:** in addition to the DMP (below) the project will share the final replication package using [dissemination channels]. Prior to releasing information for public use, the project will ensure compliance with protection of privately identifiable information.*

[37] If it is an RCT and expected to publish in health-related journal register in clinicaltrials.gov.

I.2 Ethics

For the ethics component of the impact evaluation, the proposal for the [research activity] will take the following steps. The [research activity] will be reviewed by an Institutional Review Board, certified in [the country where the study takes place and/or other location]. Moreover, given that the [research activity] will make use of [administrative records and/or individual surveys], it will follow standard protocols for obtain informed consent and protect human subjects and protect private and sensitive information.

The following activities will be taken to guarantee the protection of participants (control and treatment):

1. ***Community Stakeholder Engagement (CSE):*** *Assess vulnerabilities and identify sensitive data based on local context; as well as needs and priorities of the communities involved in research.*
2. ***Legal Requirements:*** *Understand laws, regulations, and guidelines on human subjects' protections in [country of study] and from any related international organizations.*
3. ***Informed Consent:*** *Define who needs to grant consent and/or assent and in what language and the necessary information regarding study objectives and data management plan.*
4. ***IRB*** *[submission/information]: The [research activity] will be submitted for the approval of an IRB. Potential IRBs include [add possible IRBs the project could use].*
5. ***Data Management Plan:*** *the [research activity] will outline a plan for data management, including the person/institutions who have access to the data, what type of data will be available, when each type of data will become available and how (in which format and repository), and measures for mitigating risks inherent in the collection, transmission, storage and dissemination of data.*

Most of these activities should be recorded as an item into the research protocol (section 4.2.7 of the guidelines). The research protocol can be used as a master checklist to track all the components of the final impact evaluation. For the stage of the M&E report the research protocol should describe all the completed to that date and name the expected components to be produced by the end of the project.

Table B1 below lists different the activities required for transparency, reproducible and ethical research, their associated products for the M & E report (when applicable), and their respective reference in the IDB guidelines.

Dissemination Plan

The results from the impact evaluation will be reported and published through reports, seminar presentations, conferences and other media. In a first stage, and prior to its publication, the results from the evaluation will be presented within the [implementing organization/government counterparts] for feedback. Other organizations associated with the evaluation will also be able to provide corrections and comments to reports, papers, working papers and other products from the evaluation within a previously stablish period, and before the materials go into the public domain. Every publication produced using the data from this study will acknowledge the [implementing organization] and other institutions that had provided financial or other type of support to the evaluation.

In a second stage the results will be presented in academic and policy audiences to disseminate the results and obtain feedback from outsider perspectives. Finally, the evaluation may be published in the website of participating institutions and in one or more academic journals. Accompanying the papers with the results from this study will be the materials for transparency and reproducibility: the registration of the study, the pre-analysis plan, the replication package and the TOP checklist with the planned and effective levels of transparency and reproducibility.

The authorship rights will be agreed between the [implementing organization], the IDB, and any other partner involved in funding or technical support of the impact evaluation, before the beginning of the preparation of the analysis and reports, with the expectations that authorship will be attribute to members of the team that provide original contributions to the theory or empirical formulations of the evaluation, following standard criteria for publication in academic journals[38]. As part of an initiative based on evidence and with high potential to contribute to an international body of knowledge, it is of particular interest that the results from the impact evaluation comply with the standards of transparency and reproducibility presented here.

Table B1: Items for Transparent, Reproducible and Ethical Research to be Conducted as part or [Research Activity]

Activity	Products
Registration of the impact evaluation (Section 2.1 of IDB guidelines)	Number of completed registry (ie: AEA00002334)
Pre-analysis plan (PAPs) (Section 2.2 of IDB guidelines)	Attached an initial draft [or template] of a PAP with expected date for submission in specific repository.
A computationally reproducible workflow (Sections 3.1–3.3 of guidelines)	(i) Clear and pre-stablished file structure (with readme.md/.txt on root). (ii) Declare choice of Version Control Strategy (renaming protocol or name of version control software). (iii) Declare choice of literate programming strategy (code style guidelines or dynamic documents).
Code and data sharing plan (Section 3.4 and 4.2.6 of guidelines)	Name of data repository. Description of procedure to protect personal identifiable information. Description of procedures to access all the data.
Community Stakeholder Engagement (Section 4.2.1 of guidelines)	Summary in IRB submission
Legal Requirements (Section 4.2.2 of guidelines)	Summary in IRB submission
Informed consent forms (Section 4.2.5 of guidelines)	Content [when applicable]: (i) Statement on how PII will be managed/removed as applicable; (ii) Statement on how data will be de-identified as applicable; and (iii) Statement on who will have access to what data, particularly when there is a distinction between identifiable dataset and de-identified dataset.
IRB approval/submission form (Section 4.2.4 of guidelines)	IRB submission number or draft of submission package with expected date for submission to a specific IRB.
Data Management Plan – DMP (Section 4.2.6 of guidelines)	Define who will have access to what data and when.

[38] See for example the criteria stablished by the ICMJE: http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html

APPENDIX C: REPORTING CHECKLIST FOR REGISTERED REPORTS AT THE JDE

Section	Item	Description and details to report	Reported?	Page(s)
Cover page *(required)*	*Title*	Informative title specifying the study design, population, and interventions		
	Date of latest draft	Date of when the prospective review article was last edited.		
	Study registration status	Link, registration identifier and registry name (or intended registry if not yet registered)		
	Keywords	Up to six keywords, to be used for indexing purposes.		
	JEL codes	Up to six codes.		
Abstract (*required)*	*Abstract*	Summarize research question, outcome variables, methodological framework and contribution in less than 150 words.		
Timeline *(required)*	*Expected completion date*	Expected date for completion of the pre-specified research design.		
Introduction	*Background and relevance of the study*	Brief overview of previous research, and relevance of the research question(s) for the field of economic development		
	Research question(s)			
Research design	*Basic methodological framework*	Outline of the identification strategy in your study (experimental/non-experimental)		
	Hypotheses	Pre-specified hypotheses to be tested in the study and reported as primary findings in the Stage 2 full manuscript		
	Outcome variable(s)	Definition of the main outcome variable(s) and (if applicable) secondary outcome variable(s)		
		Specification of how outcome(s) will be constructed from the dataset		
	Intervention(s)	Details of the intervention (when, where, how, by whom)		
		Number of treatment arms and whether they are exclusive or overlapping		
		Randomization strategy		
		Blinding strategy (if applicable)		
		Instructions and supporting materials for administering the intervention		
		Source(s) of exogenous variation		
	Theory of change	How and why the intervention is predicted to lead to certain effects		

	Sample	Specification of unit of analysis (individuals, organizations, countries, etc.)		
		Data source(s)		
		Projected sample size and statistical power calculations		
	Variations from the intended sample	Specification of the degree of attrition that may threaten the robustness of the study		
		Strategies to deal with attrition, non-compliance with the assigned treatment, etc.		
	Data collection and processing	Type of data, collection method/data source(s), and timeline for collection		
		Rule for terminating data collection / stopping rule		
		Data management plan		
		Pilot data and experiments run in preparation of the Stage 1 submission		
Empirical analysis	*Statistical method(s)*	Main evaluation method(s) and underlying assumptions		
		Rules for handling missing values		
		Definition and rules for handling outliers		
	Multiple hypothesis testing	Strategies to prevent false positives		
	Heterogeneous effects	Anticipated heterogeneous effects and theoretical justification		
	Statistical model	A functional (mathematical) form of the causal mechanism explored in the study		
		Specification if regression model is linear, generalized linear, or other		
		How will standard errors be calculated		
Limitations and challenges	*Challenges in the study implementation*	Potential objective circumstances that might jeopardize the implementation of the proposed study design		
Administrative information *(required)*	*Ethics approval*	Statement confirming that all necessary ethics approvals are in place.		
	Funding	Funding sources in the suggested format		
	Acknowledgments	List of (non-author) individuals who provided help to the research project.		
Bibliography	*Bibliography*	References can be in any style or format as long as the style is consistent.		
Other items	*Appendices*	Tables and figures		

Source: Linked at the author guidelines for Register Reports in the Journal for Development Economics

GLOSSARY

Beneficence incorporates two ideas: (i) do not harm and (ii) maximize possible benefits and minimize possible harms.

Commit: (in VCS) whenever a researcher takes a local snapshot of their *saved* work, they *commit*. It is recommended generate the habit of committing early and often.

Community and stakeholder engagement (CSE) can inform project teams of the needs and priorities of the communities they will engage with during the research..

Computational Reproducibility: ability to reproduce all the output of a study using the same code and data.

Data Management Plan (DMP) defines who will have access to what data and when.

Data de-identification is the process of applying data perturbation-based methods to appropriately manage direct identifying data and quasi-identifiers to remove or significantly reduce risk of re-identification and ensure disclosure limitation.

Direct Identifiers may include an individual's full name, date of birth, mailing or home address, email address, telephone number, GPS coordinates, national identification number, physical/biological identifiers (physical appearance, through photo or video data collection, fingerprints, DNA, etc.). Depending on the study and data needs, direct identifiers can also include the name of the school, health facility, community, etc. that directly identify the location of the data collection or extraction.

Disclosure limitation is statistical methods used to hinder anyone from identifying an individual respondent or establishment by analyzing published data, especially by manipulating mathematical and arithmetical relationships among the data.

File Drawer Problem: tendency in research to not complete (write up and publish) studies that have shown "uninteresting results" (typically null results).

File Management: Best practices regarding the organization of the files, and choice of file formats, in a research workflow.

Garden of Forking Paths: popular metaphor to illustrate the large number of analytical choices that a researcher has when performing a study. Its main purpose is to emphasis that robustness checks are not sufficient to prevent p-hacking.

Git: software that does tracks changes across your files. All happens under the hood (there is no "Git app")

GitHub: implementation of Git that is easier to use, provides free (public) cloud service, and tools for collaboration.

GitHub Desktop App: software developed by GitHub that helps you run Git in a specific computer and access all the project work on the web.

Informed consent is the action required in research to operationalize respect for persons, where research subjects are informed on the objectives, duration, and description of the research, its expected

benefits and risks, promises of confidentiality, how and who data will be shared with, and that their participation is voluntary.

Institutional Review Board (IRB) review is "*to assure, both in advance and by periodic review, that appropriate steps are taken to protect the rights and welfare of humans participating as subjects in the research. To accomplish this purpose, IRBs use a group process to review research protocols and related materials (e.g., informed consent documents and investigator brochures) to ensure protection of the rights and welfare of human subjects of research*."

Justice in research refers to the just distribution of the risks and burdens of the research and the benefits expected to be produced by the research.

Linkage documentation may support re-identification efforts or at least mitigate de-identification efforts.

Local: (in VCS) term used to describe when changes are made to files in a specific computer, individual users working in their own computers are said to be working locally.

Mertonian Norms: set of norms that define good scientific behavior. They include: disinterestedness, organized skepticism, communality, and universalism.

Minor is defined as a person "under the legal age of responsibility". Therefore, first the project team must identify the legal age for consent/legal definition of adult and/or minor in the research setting.

Personally Identifiable Information (PII) is information that can be used, on its own or in conjunction with other information that is linked or linkable to a specific individual (or household, community, school, etc.), to determine the identity of an individual or otherwise locate or contact the individual. PII includes both direct and quasi (or indirect) identifiers.

Perturbation-based methods: Perturbation-based methods falsify the data before publication by introducing an element of error purposely for confidentiality reasons.

P-hacking: known also as "data-mining" or "specification search" defines all the analytical alternatives that a research might test in order to obtain a statistically significant result. Examples include: restrict the sample, test subgroups or redefine variable *after* looking at the final data.

Pre-Analysis Plans (PAPs): An extensive description of a study before data is available for analysis. As per Appendix A, the content required for a PAP also aligns with requirements for registration, research protocol, and registered report.

Publication Bias: systematic difference between conducted and reported research. Occurs when results in published studies differ in a predictable direction from all conducted studies in a topic.

Project Team is the collection of individuals involved in commissioning, designing, implementing, analyzing and publishing a research study

Quasi (Indirect) Identifiers are unique, observable or other characteristics that may identify a specific individual (or household, community, school, etc.) even when direct identifiers are removed.

Re-identification is the general term for any process that restores the association between a set of de-identified data and the data subject.

Registered Report: A format for peer review and publication whereby reviewers assess a manuscript for which the analysis has not yet been conducted. As per Appendix A, the content required for a Registered Report also aligns with requirements for registration, pre-analysis plan (PAP), and research protocol.

Registration: A brief description of a study before data is available for analysis. As per Appendix A, the minimal content required for a Registration also aligns with requirements for pre-analysis plan (PAP), research protocol, and registered report.

Remote: (in VCS) whenever you make changes to the files in the cloud/server you are working remotely.

Replicability: (as used in this document) ability to obtain qualitatively similar results when repeating a study in a similar population.

Repository/repo: (in VCS) a repository or is a master folder that contains all the project work.

Reporting Guidelines: Reporting standards laying out which content is to be presented in a manuscript and designed to facilitate systematic reviews and meta-analysis.

Reproducibility/Credibility Crisis: general term to describe findings in the last decade that describe several problems across scientific fields. These include: low rates of replicability, low rates of computational reproducibility, high prevalence of publication bias and of p-hacking.

Research Protocol can be a useful tool for documenting the planned ethical research design and practices, governing its implementation, and communicating its objectives and expected contributions.

Researcher is an individual member of a Project Team

Respect for persons incorporates at least two ideas: (i) individuals are treated as autonomous agents and (ii) individuals with diminished autonomy are entitled to protection. In most cases, respect for persons requires that research subjects enter into the research voluntarily and with adequate information.

Sensitive data is information that may pose a risk to the individual (or household, community, school, etc.) if it is collected or released in a way that is linkable to the individual.

Scientific Misconduct: direct and intentional violation of any of the Mertonian norms.

Version Control Software (VCS): programs that are designed to track the entire work history of projects that can involve an unlimited number of lines of code.

Vulnerability refers to a diminished ability to fully safeguard one's own interest in the context of a specific research project. This may be caused by limited decision-making capacity or limited access to social goods, such as rights, opportunities, and power. Individuals or groups may experience vulnerability to different degrees and at different times, depending on their circumstances.

REFERENCES

Andrews, Isaiah, and Maximilian Kasy. 2017. “Identification of and Correction for Publication Bias.” https://doi.org/10.3386/w23298.

Anderson, Michael L., and Jeremy Magruder. *Split-sample strategies for avoiding false discoveries*. No. w23544. National Bureau of Economic Research, 2017.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. “The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics.” The Journal of Economic Perspectives: A Journal of the American Economic Association 24 (2): 3–30.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). “Journal Article Reporting Standards for Quantitative Research in Psychology: The APA Publications and Communications Board Task Force report”. American Psychologist, 73, 3–25. http://dx.doi.org/10.1037/amp0000191

Baker, Monya. 2016. “1,500 Scientists Lift the Lid on Reproducibility.” Nature 533 (7604): 452–54.

Beecher HK. Ethics and clinical research. The New England Journal of Medicine, 1966, 274: 1354–1360.

Brandt, Allan M. 1978. "Racism and research: The case of the Tuskegee Syphilis study." The Hastings Center Report 8(6): 21-29. http://nrs.harvard.edu/urn-3:HUL.InstRepos:3372911

Brear, Michelle. March 13, 2018. Ethical Research Practice or Undue Influence? Symbolic Power in Community- and Individual-Level Informed Consent Processes in Community-Based Participatory Research in Swaziland, Journal of Empirical Research on Human Research Ethics, DOI: 10.1177/1556264618761268, PMID 29529932

Broockman, David, Kalla, Joshua, and Aronow, Peter (19 May 2015). "Irregularities in LaCour (2014)" (PDF). Retrieved 23 January 2019.

Brodeur, Abel, Lé Mathias, Marc Sangnier, and Yanos Zylberberg. 2012. “Star Wars: The Empirics Strike Back.” SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2089580.

Buckheit, Jonathan B., and David L. Donoho. 1995. “WaveLab and Reproducible Research.” In Lecture Notes in Statistics, 55–81.

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

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al. 2016. “Evaluating Replicability of Laboratory Experiments in Economics.” Science 351 (6280): 1433–36.

Camerer, Colin F., et al. "Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015." *Nature Human Behaviour* 2.9 (2018): 637.

Card, David, and Alan B. Krueger. 1995. “Time-Series Minimum-Wage Studies: A Meta-Analysis.” The American Economic Review 85 (2): 238–43.
Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2011. “Reshaping Institutions: Evidence on Aid Impacts Using a Pre-Analysis Plan.” https://doi.org/10.3386/w17012.

Chang, Andrew C., Board of Governors of the Federal Reserve System, and Phillip Li. 2015. “Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say ‘Usually Not.’” Finance and Economics Discussion Series 2015 (83): 1–26.

Christensen G, Freese J, and Miguel E. “Transparent and Reproducible Social Science Research” *University of California Press* (Forthcoming 2019).

Clemens, Michael A. 2015. The Meaning of Failed Replications: A Review and Proposal.

Coffman, Lucas C., and Muriel Niederle. 2015. “Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible.” The Journal of Economic Perspectives: A Journal of the American Economic Association 29 (3): 81–98.

Collaboration, Open Science. 2015. “Estimating the Reproducibility of Psychological Science.” Science 349 (6251): aac4716.

Council for the International Organization of Medical Sciences (CIOMS) in collaboration with World Health Organization (WHO). 2016. International ethical guidelines for health-related research involving human subjects. (Accessed July 11, 2018 https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf)

Crespo, Anna Risi Vianna; Azuara Herrera, Oliver, October 2017, *IDB’s Impact Evaluations: Production, Use and Influence*, Inter-American Development Bank, Office of Evaluation and Oversight, http://dx.doi.org/10.18235/0000852

De Angelis, C. D. 2005. “Is This Clinical Trial Fully Registered? A Statement from the International Committee of Medical Journal Editors.” Canadian Medical Association Journal 172 (13): 1700–1702.

Desposato, S. (Ed.). (2016). Ethics and Experiments. New York: Routledge.

Dewald, William G., Jerry G. Thursby, and Richard G. Anderson. "Replication in empirical economics: The journal of money, credit and banking project." *The American Economic Review* (1986): 587-603.

Duflo, Esther and Elizabeth Spelke. 2018. "Every Child Counts! An “at-scale” test of an early mathematics curriculum" AEA RCT Registry. July 24. https://www.socialscienceregistry.org/trials/3143/history/32119

Enders, Walter, and Gary A. Hoover. 2004. “Whose Line Is It? Plagiarism in Economics.” Journal of Economic Literature 42 (2): 487–93.

Federal Regulations on the Protection of Human Subjects (subpart A 45 CFR 46) Online at https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=83cd09e1c0f5c6937cd9d7513160fc3f&pitd=20180719&n=pt45.1.46&r=PART&ty=HTML

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group. 2012. “THE OREGON HEALTH INSURANCE EXPERIMENT: EVIDENCE FROM THE FIRST YEAR.” The Quarterly Journal of Economics 127 (3): 1057–1106.

Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2014. “Social Science. Publication Bias in the Social Sciences: Unlocking the File Drawer.” Science 345 (6203): 1502–5.

Gelman, Andrew, and Eric Loken. "The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time." *Department of Statistics, Columbia University* (2013).

Gerber, Alan. 2008. “Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals.” Quarterly Journal of Political Science 3 (3): 313–26.

Gertler, Paul, Sebastian Galiani, and Mauricio Romero. 2018. “How to Make Replication the Norm.” Nature 554 (7693): 417–19.

Gentzkow, Matthew, and Jesse M. Shapiro. "Code and data for the social sciences: A practitioner’s guide." *Chicago, IL: University of Chicago* (2014).

Glennerster, Rachel, and Kudzai Takavarasha. 2013. Running Randomized Evaluations: A Practical Guide. Princeton University Press.

Gooding K, Makwinja R, Nyirenda D et al. Using theories of change to design monitoring and evaluation of community engagement in research: experiences from a research institute in Malawi [version 1]. Wellcome Open Res 2018, 3:8 (doi: 10.12688/wellcomeopenres.13790.1)

Haushofer, Johannes, and Jeremy Shapiro. 2016. “The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya.” The Quarterly Journal of Economics 131 (4): 1973–2042.

Inter-American Development Bank, 2015, *Summary: Update to the Institutional Strategy 2016-2019 Partnering with Latin America and the Caribbean to Save Lives* (accessed on July 9, 2018 http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=39933723)

IDB. 2017. “IDB Key Facts 2017.” 2017. http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=40862380.

Ioannidis, John P. A. 2007. “Why Most Published Research Findings Are False: Author’s Reply to Goodman and Greenland.” PLoS Medicine 4 (6): e215.

Ioannidis, John PA, Tom D. Stanley, and Hristos Doucouliagos. "The power of bias in economics research." *The Economic Journal* 127.605 (2017): F236-F265.

John, Leslie K., George Loewenstein, and Drazen Prelec. "Measuring the prevalence of questionable research practices with incentives for truth telling." *Psychological science* 23.5 (2012): 524-532.

Kaplan, Robert M., and Veronica L. Irvin. 2015. “Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time.” PloS One 10 (8): e0132382.

Klein, Richard A., et al. "Many Labs 2: Investigating variation in replicability across samples and settings." *Advances in Methods and Practices in Psychological Science* 1.4 (2018): 443-490.

Knott, Eleanor (Forthcoming) ‘Beyond the Field: Ethics after Fieldwork in Politically Dynamic Contexts’, Perspectives on Politics. (conditionally accepted) Accessed on February 1, 2019 (http://eprints.lse.ac.uk/87445/1/Knott_Beyond%20the%20Field.pdf)

Knuth, Donald Ervin. 1992. Literate Programming. Stanford Univ Center for the Study.

LaCour, Michael J.; Green, Donald P. (2014). "When contact changes minds: An experiment on transmission of support for gay equality" (PDF). *Science*. 346 (6215): 1366. doi:10.1126/science.1256151. PMID 25504721.

Leamer, E. E. (1983). Let's take the con out of econometrics. *The American Economic Review*, *73*(1), 31-43.

Lin, Winston, and Donald P. Green. 2016. “Standard Operating Procedures: A Safety Net for Pre-Analysis Plans.” PS, Political Science & Politics 49 (03): 495–500.

Levine C, Ruth Faden , Christine Grady , Dale Hammerschmidt , Lisa Eckenwiler & Jeremy Sugarman (2004) The Limitations of “Vulnerability” as a Protection for Human Research Participants, The American Journal of Bioethics, 4:3, 44-49, DOI: 10.1080/15265160490497083

McCloskey, Deirdre N., and Stephen T. Ziliak. "The standard error of regressions." Journal of economic literature 34.1 (1996): 97-114.

Merton, Robert (1942) “A Note on Science and Democracy,” Journal of Legal and Political Sociology: 1: 115-26. Reprinted as “Science and Democratic Social Structure” in Social Theory and Social Structure ([1949, 1957]1968), pp. 604-615.

Meyer, Michelle N. "Practical tips for ethical data sharing." *Advances in Methods and Practices in Psychological Science*1.1 (2018): 131-144.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., ... & Laitin, D. (2014). Promoting transparency in social science research. Science, 343(6166), 30-31.

Musesengwa, Rosemary, Moses J. Chimbari, Samson Mukaratirwa, April 2018, A Framework for Community and Stakeholder Engagement: Experiences from a Multicenter Study in Southern Africa, Journal of Empirical Research on Human Research Ethics, https://doi.org/10.1177/1556264618769002

National Bioethics Advisory Commission [NBAC]. Ethical and Policy Issues in Research Involving Human Participants. Volume II: Commissioned Papers. Rockville, MD: National Bioethics Advisory Commission [NBAC]; 2001: G1-G13 (Accessed July 11, 2018 http://www.onlineethics.org/cms/8087.aspx)

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. [Bethesda, Md.]: The Commission.

National Institute of Standards and Technology (NIST) Special Publication 800-188 68 Natl. Inst. Stand. Technol. Spec. Publ. 800-188, 85 pages (December 2016)

Neumark, David. 1999. The Employment Effects of Recent Minimum Wage Increases: Evidence from a Pre-Specified Research Design.

Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. “The Preregistration Revolution.” Proceedings of the National Academy of Sciences of the United States of America 115 (11): 2600–2606.

Ochieng, J., Ecuru, J., Nakwagala, F., & Kutyabami, P. (2013). Research site monitoring for compliance with ethics regulatory standards: review of experience from Uganda. BMC Medical Ethics, 14, 23. http://doi.org/10.1186/1472-6939-14-23

Olken, Benjamin A. "Promises and perils of pre-analysis plans." *Journal of Economic Perspectives* 29.3 (2015): 61-80.

O'Toole, Elisabeth; Laura Feeney, Kenya Heard, Rohit Naimpally. August 2018. "Data Security Procedures for Researchers". JPAL North America. Accessed online on January 24, 2019 at https://www.povertyactionlab.org/sites/default/files/documents/Data_Security_Procedures_December.pdf

Spellecy, Ryan; Eve, Ann Marie; Connors, Emily; Shaker, Reza; Clark, David C. June 2018. Real-Time IRB: A Collaborative Innovation to Decrease IRB Review Time. Journal of Empirical Research on Human Research Ethics. DOI 10.1177/1556264618780803; PMID 29902956

Sterling, Robyn. Genetic Research among Havasupai: A Cautionary Tale. Virtual Mentor. 2011;13(2):113-117. doi: 10.1001/virtualmentor.2011.13.2.hlaw1-1102.

Resnik DB. Research subjects in developing nations and vulnerability. Am J Bioeth. 2004 Summer;4(3):63-4; discussion W32. PubMed PMID: 16192146.

Rosenthal, Robert. 1979. "The File Drawer Problem and Tolerance for Null Results." Psychological Bulletin 86 (3): 638–41.

Turner, Erick H., Annette M. Matthews, Eftihia Linardatos, Robert A. Tell, and Robert Rosenthal. 2008. "Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy." *New England Journal of Medicine* 358 (3):252–60. https://doi.org/10.1056/NEJMsa065779.

Wager E & Kleinert S (2011) Responsible research publication: international standards for authors. A position statement developed at the 2nd World Conference on Research Integrity, Singapore, July 22-24, 2010. Chapter 50 in: Mayer T & Steneck N (eds) Promoting Research Integrity in a Global Environment. Imperial College Press / World Scientific Publishing, Singapore (pp 309-16). (ISBN 978-981-4340-97-7)

Wilson, Greg, Jennifer Bryan, Karen Cranston, Justin Kitzes, Lex Nederbragt, and Tracy K. Teal. 2017. "Good Enough Practices in Scientific Computing." PLoS Computational Biology 13 (6): e1005510.

Wilson, Greg, et al. "Best practices for scientific computing." PLoS biology 12.1 (2014): e1001745.
de Kok, Ties. "Combine Stata with Python using the Jupyter Notebook." *2016 Stata Conference*. No. 2. Stata Users Group, 2016.

World Medical Association. World Medical Association Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects. JAMA. 2013;310(20):2191–2194. doi:10.1001/jama.2013.281053