**IRB Protocol Template**

By: Jeffrey A. Shero & Sara A. Hart

Florida State University

*This document aims to lay out the sections in a typical US IRB Research Protocol (HRP-503), highlighting those sections (in red font) that may be different or require special attention due to making data publicly available.*

**Study Summary**

This section should contain Information on study title, design, primary. Additionally, secondary objectives, interventions, populations, sample size, duration, and terminology specifically related to the study should be included here.

**Objectives**

This section should contain the purpose or aims/objectives of the study. Also, should address any specific hypotheses that will be tested.

**Background**

This section should provide a scholarly and research based background for the study including information on what gaps are missing in the research and how your study fits in/aims to advance the field/topic.

**Study endpoints**

This section should list the primary and secondary endpoints of the study.

**Study intervention/investigational agent**

Provide a description of the intervention that is being assessed.

**Procedures involved**

This section should lay out and explain exactly what the participant will be asked to do/done to the participant if they participate in the study. This should also include how data will be collected, and if there will be any pot-research follow up to collect further data.

**Data and specimen banking**

This section should explain how data or specimens will be stored for future use. This mainly pertains to the physical records or identifying information for the study and how it will be stored.

[“Written records will be stored in a locked-cabinet/secured location/on a secure computer that only the researchers will have access to.”]

If planning on making data publicly available should state that here, but specify that only the deidentified data will be made available.

[“Data with no identifying information will be made open by [METHOD FOR MAKING DATA AVAILABLE]. Any data made open will be made deidentified and made so that it cannot be linked to the participant in anyway….”]

* [“Data will be placed on a secure data repository [LINK TO SITE], where only deidentified data with no liking information will be available.”]
* [“Information regarding the project and data variables will be made available at [LINK], where data can be requested/applied for. It will then be up to the discretion of the PI [OR OTHERS] to decide whether or not the deidentified dataset will be shared with the requestors.”]

**Sharing of results with subjects**

This section should explain how any sharing of data between researchers and subjects, researchers and other institutions, researchers and the individual, etc. will be carried out.

For data being made open, should specify how data will be deidentified prior to being made available.

[“Any identifying information including names, addresses, school ID’s, etc. will be removed prior to the sharing of data online or with other researchers by request/application. Through doing so, this will ensure that no one is able to find out that the participant was a part of the study, or link their results to them in anyway.”]

This should also include specific data-repositories or other means of making data available to the public that are anticipated.

[“The deidentified data/information about the project will be uploaded to the [REPOSITORY NAME] data repository, and as such many questions about the repository/open data can be answered here [FAQ/ABOUT PAGE URL].”]

**Study timelines**

This section should describe the duration of the study, frequency of testing, and generally the amount of time a participant can expect to devote to the study.

**Inclusion and exclusion criteria**

This should describe the types of individuals who will be included in the study. Any desirable participant traits should be included and listed as inclusionary criteria (i.e. study seeking students with identified learning disabilities, in grades K-3, in a given geographic region. This should also include any exclusionary criteria that may limit an individual from participating (i.e. only interested in students for whom English is not the first language).

Additionally, this should list any special populations that will or will not be considered for study participations, specifically those who are more at-risk for harm when participating (i.e. Non-adults, pregnant women, prisoners, or individuals unable to consent on their own).

**Vulnerable populations**

This section should list any populations that will be studies who are at increased-risk due to participating in the study, and list the additional safeguards, procedures, and steps that will be taken to further ensure their rights and welfare are protected.

**Local number of subjects**

This section should give an estimate of the number of participants that are expected to partake in the study.

**Recruitment methods**

This section should give information on how recruiting of participants will take place and address the following questions:

* What population will be targeted when recruiting participants?
* How, when, and where will recruitment take place?
* How will potential subjects be identified?
* Will subjects be compensated? If so, how much and how often? How will compensation be carried out?
* Will fliers or materials be handed out in attempts to recruit individuals? If so, include copies of these here.

**Withdrawal of subjects**

This should discuss the circumstances under which an individual will be removed from the project by the researchers without their consent, and should describe what will happen when a participant removes themselves from the study.

For data being made publicly available, what will happen to this individual’s data? Will it still be made available?

**Risks to subjects**

This section should list any discomfort that may result from participation in this study. This should list any immediate or delayed harms, as well as those that are the direct or indirect result of participating in the study.

For parts of the study that have foreseeable risks, should list these specific sections and associated risks here. Specifically list any procedures that may be of increased risk to the vulnerable populations that take part in the study.

If outlined prior that data is expected to be used for open science or made publicly available, include a quick statement that making data available poses the minimum amount of risk to participants since all data will be deidentified to the point that no one outside of the researchers will be able to link an individual to their data or to the study as a whole to the extent possible.

[“Any data made publicly available will be made deidentified and made so that it cannot be linked to the participant in anyway as much as possible. It is impossible to guarantee that the participant’s data will in no way be linked back to them, but through extensive deidentification making the data available to the public will involve only the minimal amount of harm or risk to the participant.”]

It is important to note here that no matter what level of deidentification is performed for the dataset, you cannot guarantee that an individual will not be reidentified.

[“Due to the nature of the data, it is possible individuals can be reidentified but any such estimates about the likelihood of this reidentification are not possible to compute. Rather, through (ensuring larger minimum cell sizes/aggregating data/blurring data/etc.) we will use techniques to limit the possibility of reidentification as much as possible.”]

**Potential benefits to subjects**

This section should describe the potential benefits to being in the study. This should include direct benefits to the individual participant, as well as benefit to others. Compensation is not a form of benefit and as such should not be included in this section.

**Data management and confidentiality**

This should include a plan for data-analysis, including specific statistical techniques or power analyses, and should include specific transformations or tests to ensure that data will be deidentified.

[“Names, addresses, school ID’s, etc. will be removed from all data prior to analysis and data being made publicly available/shared with other researchers by request. Statistical techniques will be used to ensure that other variables such as race, gender, and age will not be able to link the individual to the study. Any records including this information will be stored in a locked location that only the researchers have access to/destroyed/stored on a secure computer/etc.”]

[“For instances in which information about the entire population is known and demographic or quasi-identifying (sex, race, etc.) variables pose a risk to identifying individuals, cross-tabs of quasi-identifying variables will be made into broader categories using blanking and imputing/aggregating/other techniques to ensure a minimum bin size of 5. This is to ensure that if only few individuals meet a certain classification (i.e. a female Hispanic with a specific learning disability), that they cannot be identified by this information.”]

This should also include a statement on how any confidential or potentially identifiable information will be treated or stored, including long term plans for the data after the project has ended. Additionally, this should include specific information regarding any written or physical records of the data. This includes plans for destroying the data when necessary.

This section is highly important to address when planning to make data publicly available.

Should again address that the data will be made publicly available, and some description of the plan to do so (specific data-repositories, open to share with other researchers by request, etc.), and explain that data will be deidentified and what that means.

[“Any information that can link the participant to this study will be removed prior to any data being made publicly available/shared with outside researchers by request. There will be no information that can link the individual’s participation to the study, and as such steps will be taken so that no one outside of those directly involved with the research will know the participant took place in this study, to the extent possible by current deidentification techniques.”]

**Provisions to monitor the data to ensure the safety of subjects**

This section is important when the study has more than minimal risk to the participant. This section should outline how participants will be monitored to ensure that no more than minimal harm is done to the participants. This should include specific timelines, frequency, and methods for assessing risks/collecting this data.

**Provisions to protect the privacy interests of subjects**

This section is similar, but different to the confidentiality and risk sections. Here, this discusses how the study will protect the participant’s “sense of privacy”. Include here specifics on how many researchers will have access to the identified data, and what will be done to minimize this number.

It is important to reemphasize here that only the original research staff will be the ones who have access to any of the identifying information.

[“Although the data from this study will be made available to the public/to other researchers by request, only the members of the research team will ever see or have access to any identifying information. Any data that will be made available to the public will have no information linking the participant to the study to the extent possible. We cannot guarantee that reidentification is impossible, but can ensure that the risk will be as minimal as possible.”]

**Compensation for research-related injury**

If the study involves more than the minimal amount of risk to the subject, this section should include specifics on how individuals will be compensated if injury or harm does occur.

**Economic burden to subjects**

This should describe any costs that the participant may be personally responsible for related to the study.

**Consent process**

This should include the plan for obtaining consent from participants, including where and when obtaining consent will take place. When possible, should include copies of any informed consent or assent forms.

This should also address what extra steps are being taken for subjects whose first language is not English, or plans for obtaining consent/assent from vulnerable populations.

For data being made publicly available or available to outside researchers by request, should include the extra steps that are being taken to inform the participants on what deidentification, open science, and publicly available data mean, as well as the plan for informing and getting consent for making the data available to the public.

[“Since data will be made available to the public/for request, an additional section in our informed consent documents was included describing that nature of the deidentified data that will be uploaded. If participants are not comfortable with their deidentified data being made publicly available/requestable, we invite the participants to still take part in the study with the condition that their individual data will not be made available to the public.”]

**Process to document consent in writing**

If not included in the prior section, this should include the specific informed consent and assent documents that will be used for the study.

For data being made available to the public/requestable, can specifically highlight or point out the additional sections that are included to address this, if not included in prior sections.

**Setting**

This should describe the settings for all aspects related to the study.

* Describe where recruitment will take place.
* Describe the settings and locations in which the research/data collection will take place.
* When applicable, if the research is being conducted at a third-party site outside of the institution specify the customs or regulations related to that site that are relevant to the study.

**Resources available**

This section should describe the resources, funding, and time available for the completion of the study. When applicable, describe resources for any additional safeguards needed during the research process (i.e. medical or counseling staff, data-scientists for data-deidentification, etc.). This should also include any training materials or procedures for training staff and ensuring they understand all protocol.

**Multisite research**

If study is conducted across multiple sites, this section should include specifics on the number of other sites, names and contacts for other sites, total sample sizes, etc. This should include information on any of the above sections that are relevant to the study as a whole (i.e. procedures to obtain consent for the whole study rather than just one site specifically, recruitment process across multiple-sites and coordinating this between sites, how and when sites will interact, compensation/benefits/risk differences between multiple sites, and how the study will deal with differences in local laws or policies governing research/consent between sites.