**Protocol Template for Retrospective Chart Review Studies/ Research Limited to Existing Data**

This template is appropriate for retrospective chart review studies where the investigators wish to extract information that already exists in the medical records or for use in studies using existing data sets. If participants have not yet been seen or data has not yet been collected, the study design is not retrospective and this template should not be used.

# Title page

Title

Principal Investigator

Co-investigators

Chart abstractors

Statistician/methodologist

Study sites

Funding source

# Specific aims

Provide a description of the overall purpose/rationale of the research. Include a list of the main aims of the study/project. Whether in the text or as a component of the aims, there should be a very clear statement of the main hypothesis and any exploratory hypotheses.

# Background and significance

Describe the scientific basis for the study. There should be sufficient reference to the literature to demonstrate the study will contribute new and generalizable knowledge and the significance of the study in relation to human health. If there are any preliminary data, these should be described here.

# Approach

## Source of records

Explain where the existing records are held. This should include the following information:

* From what source you are obtaining the existing records
* Who are the records maintained by: CCHMC, UCHealth, outside hospital, or databases maintained by another affiliated institution?
* Whether the existing records are electronic or paper

Certain activities related to the collection of data may be subject to FDA regulations. Does the research involve the collection of data or other results from individuals that will be submitted to, or held for inspection by, the FDA? In general this would include research that involves any data that will be provided (in any form) to a pharmaceutical, medical device or biotech company.

\*Of note if the research is FDA regulated it must be registered on clinicaltrials.gov\*

**If this study is being done at multiple sites this information needs to be given for each site, listing out all sites. Please list if the PI on this protocol is responsible for all sites or just the research being conducted at their own institution. Also list the IRBs that will govern the study at external sites, if any, and how those submissions will be handled.**

## Identification of eligible participants/records

### Potential participants

Describe how the pool of records will be identified. This might be by querying a database for all patients with a certain procedure code, DRG code, ICD-9 code, or it might be by exporting data from all patients during a time period, or data that was collected in a previous or ongoing research study or by some other means. Be sure to mention the data range or collection range of the records of interest. Also include any additional data that will be used to augment the records.

**Identify whether data will be pulled by individuals at the site where the data is held that are not involved in the study and then sent to the study team, or whether data will be pulled by study staff listed in the protocol.**

**Indicate whether there will there be any contact with participants, patients, their families, or their physicians.**

### **Screening/Inclusion and exclusion criteria**

List the specific criteria that will be used to decide which records will be included and which will not.

Describe how the records will be reviewed to make sure only eligible records are included. It might be possible to do this by an automated query, or it might be by manual review. If it is by manual review, describe how you will ensure consistency and accuracy of the screen, such as by having two people decide on eligibility.

## Gender and minority inclusion:

If the chart review will include vulnerable populations (prisoners, pregnant women, fetuses, neonates, individuals who are cognitively impaired or minors), information must be given for justification as to why these individuals are being included. The UC IRB has a checklist for vulnerable populations that should be used to ensure all answers to questions on the forms are laid out in this section.

For chart review studies, it is not always possible to exclude vulnerable subjects. If individuals from vulnerable population cannot be removed from the data set, then those individuals must be considered as enrolled and therefore explained here.

Example wording when vulnerable subjects will not be included: *Adults (age >18) of any gender or race and ethnicity are potentially eligible for inclusion. Pregnant women and children may be treated differently to adults and their inclusion would add unnecessary variability. Prisoners will not be included; the results will not generalize to this patient population.*

## Request for Waiver of Consent/Authorization (HIPAA)

To be eligible for a waiver of consent and a waiver of authorization for release of medical record information, the following must be provided:

* An explanation as to why obtaining consent or HIPAA authorization would prevent the research from being completed.
* An explanation as to whether it is possible or feasible to obtain informed consent, given the scope of the research, including whether the research could still occur if consent was obtained.
* An explanation as to how the removal of the consent process from the conduct of this research will not adversely affect the rights or welfare of participants. (Note: "participants" in this case is human data/specimens. The use of human data/specimens for research purposes constitutes human subjects research.) An example of adverse effects might include a person's right to decide whether or not they want to be a participant in a research study.
* Methods for destroying the information on completion of the work, and methods to maintain confidentiality.

Example wording: *This is a request to use identifiable information in the conduct of this research study under a waiver of authorization. The identifiable information being requested is: [Insert here a description of the data to be used, “in a specific and meaningful fashion.” The description should be understandable; not a mere recitation of data elements understandable only to the research team. The description should be specific and the request should be limited to that information necessary to the research protocol. Examples of specific and meaningful descriptions include "Lab tests," "clinic visit data," "X-ray readings," etc.]*

*The identifiable information will be used or disclosed only by members of the research team and the following persons: [identify with specificity and justify the need to disclose the information to anyone outside the covered entity (e.g. UC Health)].*

***The proposed study poses minimal risk to the privacy of the subjects because:***

1. *The identifiable information will be protected from improper use or disclosure by: [detail how this will be accomplished including limitations of physical or electronic access to the information and other protections]*
2. *The identifiers will be destroyed at the earliest opportunity consistent with the research [discuss the timeframe or the reasons the identifiers must be retained, including health or research justifications or any legal requirement to retain them]*
3. *The identifiable information will not be reused or disclosed to any other person or entity outside the covered entity other than those identified in the protocol, except as required by law, for authorized oversight of this research study, or as specifically approved for use in another study by an IRB.*

***The proposed study cannot be practicably conducted without a waiver of authorization because:* [discuss reasons why it would not be possible to obtain authorization from individual subjects]**

***The proposed study cannot be done without the specified identifiable information because:* [discuss reasons why it would not be possible to conduct the research without the identifiable information being requested]**

## Data collection

**Provide a DETAILED description of the study procedures. This section must describe how each piece of research data will be obtained from the records, including a detailed plan of how records will be accessed, from where the records will be accessed and how data will be collected.**

# 4.6 Timeline

Briefly describe the expected length of time it will take for you to extract the data you wish you use from the data set. Estimate the length of time for primary analysis and final report writing/publication. Since studies can be delayed for multiple reasons, wording can be designed as such. If the intent is to have an open-ended registry with no set end date, his should be made clear here.

Example wording: *Collection will begin within a month of IRB approval. It is expected that collection of the data will take up to 6 months. Analysis of the data is expected to take 6 months to complete. Our goal is to submit a manuscript within 4 months of the end of data analysis.*

### **Data collection form and data dictionary**

Create a data collection instrument (a case report form, data abstraction form, or electronic database) in advance of the study. Your data collection sheet will need to be submitted to the IRB so that it can be verified that data collected in the form is not outside the parameters of this protocol. Your data collection sheet can be part of the protocol or submitted as a separate document. Each data element should have a definition that describes where in the medical record the information will come from, how missing data will be recorded, and what to do if there is discordant data in the chart. If data does not fit a simple format, for example when applying new methods to review and quantification of imaging records, describe how you intend to record findings.

### **Data collection team**

The persons abstracting the data from the existing data set should be described, including their level of knowledge related to the material and how they will be trained to review the data consistently.

### **Data verification**

The method used to ensure the abstracted data are accurate should be described. The following options, among others, might be considered

* Dual abstraction of each chart with a third reviewer to arbitrate discordance
* Dual abstraction with joint agreement on discordant data elements
* Single abstraction with dual abstraction of a proportion of charts to verify accuracy, and methods for corrective action if accuracy is not established
* Single abstraction with data queries and consistency checks done on the final dataset

## Data management

Describe whether the data will be written on paper forms and then entered into an analysis package, or if they will be entered directly into electronic format. If the data collection form is electronic, describe how the data will be verified and secured. If data are written on paper forms and then entered into electronic datasets, describe whether this will be dual data entry.

For all electronic data, state where the servers are housed, and whether there is user-level access control to the information.

Indicate whether direct identifiers (e.g. name and address), indirect identifiers (e.g. data that could reasonably be used to identify a subject, or a coding system with a key), or no identifiers will be written on case report forms or stored with data.

Of note, Institutional Review Boards are not in favor of Microsoft Excel (or similar) products for data collection and management. REDCap is a secure, web-based, electronic data capture system that is freely available at many institutions, including UC. This software, or similar, should be used wherever possible.

## ****Data safety management, security and monitoring****

Taking steps to ensure that the collected data will be adequately secured and stored in a manner that minimizes the risk of loss is an essential step in protecting the rights and welfare of research participants. Indicate how the research team will ensure that only authorized individuals are able to access research data (both hard copy and electronic copy). If anyone not listed as a part of the research team (i.e. not listed as study staff on this IRB application) will have e access to the study data, provide their names and brief explanation of their role in the research.

**In order to ensure the collected research data is being monitored over the course of a research study for any signals that participation in the research may be placing participants or their data at increased risk, ALL research protocols must contain some provisions for the review/monitoring of the data for safety. As such provide a description of the plan to monitor the collected data during the course of the research. This might be limited to how the team will review the security of records.**

## Potential risks and protection against risks

Example wording: *The risks of this research are limited to risks associated with confidentiality of the data. We will ensure that paper forms are stored securely under lock and key when not in the direct custody of an investigator. Forms may have direct identifiers for the purposes of linking data from different medical records. Electronic data will not be entered with direct identifiers. A subject number will be used to link back to the identifier. The key linking identifier to subject number will not be stored with the electronic data. Electronic data will be managed using a secure, web-based system that has user-level access control. RedCAP satisfies these requirements and is the system of choice. Analytical datasets will be stored on secure servers that also limit access to the investigator team. Should results of the study be published or reported, individual names or other identifying information will not be used.*

## Potential benefits

There is no direct benefit to the study patients. However there may be a benefit to future patients with the condition of interest by answering the study question.

## 5. DATA ANAYLSIS

**5.1 Analyses Methods**

Describe how the data will be analyzed to answer the study question, **including the methods to be used and a basis for how you have determined the number of records that are needed to answer the research question (including as appropriate a power analysis or other sample size calculation).** A reasonable statistical approach must be provided. The approach might be descriptive, it might involve statistical testing to compare groups, it might involve statistical modeling, or it might involve some other approach or combination of approaches. Help for statistical analysis can be obtained from the Center for Clinical and Translational Science and Training.

Indicate whether the research team will disclose any data that is individually identifiable as part of the research. In most cases, collected research data is coded and aggregated, or at the very least de-identified prior to reporting to sponsors, funding agencies, etc. If you intend to share any identifiable data with these groups, describe exactly what identifiable information will be shared and for what purpose.

## 5.2 Data archiving

Describe where and how any paper records and electronic files will be securely stored and archived. If the data will be destroyed, the timeline and method for ensuring their destruction should be provided. If electronic records will be anonymized, the method to ensure they cannot be re-identified should be described. If records are to be kept as identifiable please be sure to list out what identifiable information will be kept and how confidentiality will be maintained, including if sharing data.

Be sure the following questions are answered, and that they are consistent with the section requesting a waiver of authorization (Section 4.4):

1. What will happen to the collected data at the end of this currently described research study, e.g.
	1. Data will be retained per sponsor requirements for a pre-determined period of time. No future research is planned for the collected data.
	2. Data will be retained in an identifiable (including coded) state for potential future currently unspecified research.
	3. Data will be anonymized (i.e. no link back to an identifiable person) and used for future currently unspecified research.
	4. Following data analysis and publication/dissemination of results collected data will be destroyed.
2. If data will be retained following completion of the currently described research study, describe how/where the data will be retained, e.g.
	1. Data will continue to be maintained in the format and location as described in the "Data Management and Security" section of this protocol.
	2. Data will be transferred to another CCHMC/UC IRB approved data registry/repository.
	3. Data will be transferred to an outside data registry/repository.
3. Describe possible future uses, e.g.
	1. Describe what data will be retained in the repository (all or just some data points).
	2. Describe the length of time the data will be held in the repository.
	3. Describe what the stored data will/may be used for in the future.

Example wording: *The data will be retained for future studies that have yet to be design. Possible analyses may include:*

1. *Descriptions of patient cohorts with conditions of interest*
2. *Analysis of care practices for patient cohorts with conditions of interest*
3. *Analyses exploring predictors of outcomes, or factors associated with illness, or disease*
4. Describe the way in which requests for use of the information in the data repository will be handled, e.g.

*Researchers who are interested in using this study’s information must complete a request with specific information about their study, including a description of how they will use the information. Requests will be reviewed by a group of experienced scientists who will judge each request for its scientific merit, its potential contribution to the prevention or cure of disease, and for the qualifications of the research team. If a plan is approved, the information will only be provided to the researcher once IRB approval has been obtained.*