**IRB PROTOCOL TEMPLATE – Healthcare Cost and Utilization Project (HCUP)**

*To implement a research project with HCUP data you will need to consult with the WMed Data Analytic Services Unit. Please complete the* [*WMed Project Request Triage Form*](https://edc.wmed.edu/surveys/?s=8KLWR3YFLW) *to initiate this process.*

***Guidance for using this template:***

*The template contains some sample text and instructions as to the type of information that needs to be included in a study protocol on a project searching a deidentified, clinical database.*

* *Language in italics should be used as an example for development of your protocol and should be replaced.*
* *Remove any template language prior to submitting to the Data Analytics Services Unit if applicable.*

1. **Protocol Title:**

1. **Investigators/Study Personnel:**
2. **Version Date: (mm/dd/yyyy Version 1)**

1. **Introduction:**

*Three to four paragraph section outlining the research study. Paragraph one begins with a broad introductory sentence. Last sentence outlines the GAP or NEED in research.*

*Paragraph one and two provides scientific or scholarly background and rationale for the research based on the existing literature (include references). Paragraph 3 (optional) describes any relevant preliminary data.*

*The last paragraph outlines the significance of the research. This paragraph should contain a clear significance statement: This research is significant because …*

1. **Purpose of Study:**

*Overall Goal: (State the overall goal of the study.)*

*List the specific research question for the study. The questions should be related to the overall goal of the study. Questions should be simple, specific (stating variables of interest), and stated in advance (before data analysis starts). As many questions that can be asked before data analysis starts is key; adding questions after will result in a need for an updated protocol and potentially a re-analysis. Having the research questions fully flushed out beforehand saves significant time with analysis.*

*Example:*

* *Primary Question 1:* 
  + *What is the rate of in-hospital post-procedure comorbidities such as pneumothorax, pneumonia, hemoptysis, COPD exacerbation, Endobronchial valve removal (procedure code), and acute respiratory infection?*
* *Primary Question 2:* 
  + *What is the in-hospital mortality rate for patients with EBV?*
* *Secondary Question 1:* 
  + *What is the trend of utilization of EBV in the USA?*
* *Secondary Question 2:* 
  + *What is the length of stay?*
* *Secondary Question 3:* 
  + *Is there a difference in the length of stay between males and females?*
* *Secondary Question 4:* 
  + *What are the predictors of in-hospital mortality?*

*Questions to consider for the analysis & include in protocol:*

* *Will the project only use a targeted sample? Or will it at some point* *compare to the full dataset? (Such as p-values in a demographics table to determine significant differences in sample versus full population – i.e., comparing EBV cases to other hospitalizations)*
* *Will any comparisons be made, such as between demographics? Please list these. (Ex: Compare between age groups, between gender, etc.) Please note what would be compared (i.e., rate of hypertension between male and female)*
* *Will any modeling be completed? What is the outcome (Ex: predict in-hospital mortality)? What are the possible variables in the model (demographics or comorbidities)?*

1. **Research Design:**

*Describe the study procedures (Describe what target population will be accessed, from where it will be collected (which HCUP database), what time frame is being included, and what other variables will be reviewed to answer the study questions). Include the ICD 10 diagnosis/procedure codes that will define your patient population (CPT codes are not searchable in the HCUP database).*

*Example*

*This study will be a retrospective cross-sectional study using the Healthcare Utilization Project – National Inpatient Sample (NIS). This database includes a 20% sample of the US hospital admissions in participating states.*

*We will query the NIS database using the International Classification of Diseases 10th revision (ICD-10-PCS) procedure codes for Endobronchial valve implant in primary and secondary procedure fields to extract study population.*

*We will then analyze the outcomes during the index admission of post procedure pneumothorax, acute respiratory failure, pulmonary infection, all-cause in-hospital mortality rate, length of stay (LOS). The trend of utilization of EBV and the trend of complications since FDA approval will also be analyzed.*

***NOTE:*** *A separate Excel file for how the biostatisticians collect ICD/PCS codes from the study team is available for teams to fill in and submit along with a completed protocol. This ensures efficiency and accuracy for the biostatisticians while coding. This template is available on the Data Analytics Service Unit* [*website*](https://wmed.edu/dataanalytics) *or can be available by emailing the Research Navigator,* [*research.navigator@wmed.edu*](mailto:research.navigator@wmed.edu)*.*

1. **Inclusion and Exclusion Criteria (if applicable)**

*Describe any factors that need to be taken into account when defining the target study population.*

*Example*

*This study will include individuals who meet the following inclusion criteria:*

* *Endobronchial valve implant procedure code (ICD-10 procedure codes), and*
* *The timing of the procedure should be no longer than 1 day after hospitalization. The reason for that is because endobronchial valves also can be used in the setting of persistent bronchial pleural leak, refractory to conventional treatment (chest tube drainage), and*
* *Diagnosis of COPD, accordingly to ICD-10 diagnostic codes*

*Patients whose endobronchial valve implant procedure was done after day 1 of hospitalization will be excluded.*

*Patients who do not have COPD as comorbidity on the primary or secondary diagnosis will be excluded.*

*Patients with age < 18 years, with missing data for age, gender, or mortality will be excluded.*

***NOTE:*** *A separate Excel file for how the biostatisticians collect ICD/PCS codes from the study team is available for teams to fill in and submit along with a completed protocol. This ensures efficiency and accuracy for the biostatisticians while coding.*

1. **Data Variables of Interest**

*Describe the variables of interest, including demographic or institution-specific variables available in the HCUP databases as well as comorbidities/procedures of interest.*

*A full list of variables in the HCUP website can be found on the* [*HCUP website.*](https://hcup-us.ahrq.gov/databases.jsp)*The biostatisticians are available to discuss options for each specific database during the project initiation meeting.*

*Example:*

* *Demographic variables*
  + *Age*
  + *Race*
  + *Insurance*
  + *Hospital Location/Teaching status*
  + *Etc.*
* *Comorbidities*
  + *Hypertension*
  + *Diabetes*
* *Procedures*
  + *Mechanical Valve Implant*

1. **Disposition and Confidentiality of Data (location/length of data storage)**

*Describe how the data will be accessed and stored for study purposes. Please note that in most cases, the example language below should be used.*

*Example: HCUP data are provided in a de-identified format. The database is publicly available with written agreement and purchase through the Agency for Healthcare Research and Quality (AHRQ). The database is updated annually and stored indefinitely on WMed institutional servers. WMed study team members, including the data manager and biostatistician, who have completed the HCUP training and data use agreement will be granted access to the raw data files. Otherwise, only summarized and aggregate data will be shared with study investigators. Project specific data will be maintained for 2 years following study completion.*

1. **Statistical Plan\*:**

*Describe how the data will be summarized (i.e., means and standard deviations, medians and ranges, percentages). Identify the statistical test for the analysis of the primary outcome variable. Define the tests for the analysis of the secondary outcome variables. Set the level of significance (i.e., significance will be assessed at p < 0.05). If no statistical tests are planned, denote that only summary/descriptive statistics will be used.*

*\*The Data Analytics Services Unit at WMed can help you with setting up your plan for analysis.*

1. **Risks:**

*Risks: A confidentiality breach is a risk associated with data review research.*

*Describe the risks involved in the study activity. (i.e. physical, psychological, social, or economic). As applicable, describe why the risks to participants, if any, are reasonable in relation to the anticipated benefits and/or knowledge that might reasonably be expected from the results. Describe what will be done to minimize any known risks.*

*Example*

*This is a retrospective cross-sectional study. There will be no interventions in this study mitigating direct patient risk. Patients in retrospective studies may be at risk for breach of privacy which will be lessened by raw data only being accessed and analyzed on a private WMed server with access given only to those who have completed the HCUP data use agreement. Any data reported from the project will be done in aggregate to prohibit the direct or indirect identification of a person or institution in the database.  Release and disclosure of observed values less than ten will be avoided.  For this project, data will only be reported for research purposes and will not be used for commercial or competitive interests.  Any reports or publications of the data will acknowledge the data source as the [HCUP database name].*

1. **Benefits:**

*Benefits: Participants are not likely to receive any benefit from the proposed research; however, others may benefit from the knowledge gained.*

*Describe potential benefits for participants that are involved in the study. It is acceptable to state that there are no anticipated benefits to the subject.*

*Example:*

*It is unlikely that any patient in the study will have a direct or indirect benefit, however, the community may benefit if there is a better understanding of the disease, risk factors, complications, and trends.*

**10. References:**

*Identify any literature cited for any information referenced in the protocol. Organize this information like that found in a medical journal.*