**Use of 3rd Party (Honest Broker)**

**Definition of Roles**

**3rd Party Honest Broker for Specimen De-Identification** – A 3rd Party Honest Brokeris an individual, organization or system acting for, or on behalf of, the covered entity to collect and provide information to research investigators in such a manner whereby it would not be reasonably possible for the investigators or others to identify the corresponding research subjects directly or indirectly.

Individuals serving as an honest broker must have legitimate access to the data requested by the investigator and they must be completely independent of the research team.

The 3rd Party Honest Broker will be the <specify person’s role, and name if you have it> who is not otherwise associated with this study.

<Delete if not needed> **Backup for 3rd Party Honest Broker** - In the event the responsibilities of the Honest Broker are transferred to another individual, the current Honest Broker will transfer the file associated with this study to his/her replacement in a secure manner upon IRB approval of the amendment designating the replacement. The new 3rd Party Honest Broker will maintain the file as noted in this SOP.

<Specify a person’s role and name if you have it> in the <Department> office is not associated with this study and will function as a backup to the 3rd Party Honest Broker.

**Study Team** – No member of the study team can serve as the 3rd Party Honest Broker.

The Study Team will remain blinded to the code linking <data/specimens> to their <data key, i.e., study ID number, survey record number> (ID #), accession number, or any other identifiers.

The study team will receive from the 3rd Party Honest Broker only the de-identified data set.

**Procedure for De-identifying Data**

**Primary Objectives:** <*provide the objective in collecting the data; why was it done?*

Example: Primary objective is to explore the underlying cause of high cesarean rate in the U.S. population, describe contemporary labor progression, and determine when is the more appropriate time to perform a cesarean delivery in women with labor protraction and arrest.>

**Study Population:** <*Describe the study population inclusion criteria>*

Study Population: <Example: Individuals with <X, list inclusion criteria>. Dates: <list date range. Example: 2002 – present>

**Database Description**: <provide detailed description of the data set, i.e., Collected detailed information from <source, i.e. 19 U.S. hospitals>.

* **De-Identified Information**  o De-identifiedmeans information which does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual. o The study team does not have actual knowledge that the identifiable information accessed by the 3rd party.
  + De-identified received from the 3rd party honest broker could be used alone or in combination with other information and not identify the individual who is the subject of the information.
  + The honest broker must adhere to all of the terms and conditions specified by the IRB for any research study for which the honest broker will perform de-identification services.

**Data Abstraction by 3rd Party Honest Broker**:

* 3rd Party Honest Broker will review data and remove <describe how data will be abstratced>.
* After data abstraction, the 3rd Party Honest Broker will assign a random Study ID linked to the

subject’s data.

**Data De-Identification Methodology:** <Describe in detail the methodology to de-identify the data>

All databases have been stripped of identifying subject information. There are no dates in the research database, original site numbers have been re-coded, and any information that could identify a <study location, i.e., school, community center, hospital> has been removed. Specific information on each deidentified dataset is provided below.

Each record is uniquely identified by <describe unique study identifier>.

The following describes changes to the data to de-identify the information:

Example

* We have 10 sites with one hospital, 1 site with 5 hospitals, and 1 site with 4 hospitals, for 19 hospitals total.
* Scrambled site numbers 81 to 99 and assigned to each of the existing hospitals (use of 81-99 was used as to not confuse by participating sites who know current numbering scheme).
* Replaced 1st 2 digits of momid with new site number in all the databases
* Dropped following variables from the CSLlinkedbypreg dataset
* ACOG district (ACOGd)
* Physician characteristic provider ID (DocID)
* Hospital name (Hospname)
* Safe Labor institution (CSLsitename)
* City, state of hospital (Hosplocat)
* Date of delivery
  + 1. Delete variable
    2. Extract following from date of delivery variable to include in database
       - Year of delivery
       - Day of week (Monday, Tuesday….)
       - Hour of delivery (if time is available) (on a 24-hour military time scale)
* For all the following date and time variables, calculate in minutes (drop seconds) from the date/time of delivery. Negative minutes indicate prior to delivery, positive minutes are after delivery. The labels of these times will note that it these are calculated from date/time of delivery, which was set as time 0)
  + 1. Admission to L&D: Date/time of Admission
    2. Admission to L&D: Date/time of spontaneous onset of labor
    3. Admission to L&D: Date/time of arriving at triage
    4. Admission to L&D: Date/time of rupture of membranes
    5. Labor and Delivery summary: Date/time of complete dilation
    6. Labor and Delivery summary: Date/time of placental delivery
    7. Labor and Delivery summary: Date/Time of Induction
    8. Labor and Delivery summary: Date/Time of Cervical Ripening
    9. Labor and Delivery summary: Date/Time of ROM
    10. Labor and Delivery summary: Date/Time of Epidural Placement
    11. Labor and Delivery summary: antibiotic date
* For each baby, removed birthday and replaced with:
  + 1. Year of delivery
    2. Day of week (Monday, Tuesday….)
    3. Hour of delivery (on a 24-hour military time scale)
    4. Duration of time in minutes between deliverydt\_B# and Deliverydt of first child born for this pregnancy. For example, if this pregnancy resulted in twins, then Deliverytime1 would equal zero and Deliverytime2 will be difference in minutes between the delivery of the first baby and the second.
* For each baby replaced the date of death, with days from date of the birth of baby # to the date of death of baby #.
* Replaced date of first ultrasound and LMP date with the days between the date of ultrasound or LMP to the date of delivery.
* Repeated measures
  + 1. Removed Vaginal Exam Date/Time, and calculate minutes between vaginal exam or full dilation and delivery (CSLRepeateddata.sas7bdat).
    2. Removed date/time of oxytocin dosage and replace with difference between dosage date/time and delivery (CSLOxtocinData.SAS7bdat.

Example 2:

**Participant Identifier:** The original identification variable has been removed and replaced with a new randomly generated variable, ID, with values between 1000 and 2000.

**Clinic Identifier**: The original clinic identifier has been removed and replaced with a new randomly generated variable, ClinicID, with values between 1 and 7. All participants from a given clinic have the same value of ClinicID.

**Other identifying variables**: Variables identifying the person who administered a test and the person who entered data have been removed.

**Date variables**: Most dates were removed from the databases: date of birth, data-entry dates, dates of testing, date of enrollment. Remaining date variables were removed and replaced with the number of days since the date of enrollment. The names of the deidentified date variables are the same as the names of the original date variables except that they have been modified by having a leading underscore added (for example, the name of the deidentified version of a date variable named VisitDate would be \_VisitDate.) The labels of the deidentified variables are the same as the labels of the original variables with except that “(deidentified)” has been added to the end of the label. Components of the dates (day, month, year) have been removed.

**Height and Weight at Enrollment**: Height and weight have been replace by body mass index (BMI.) BMI values less than 20 have been set to 19; BMI values greater than 40 have been set to 40.

**Age at Enrollment**: Age at enrollment has been replaced with a categorical variable, AgeGroup, with the first category being <= 45, the last category being > 75, and increments of five years in between. **Education**: Years of education has been replaced with a categorical education-level variable, EducLevel **Race**: Race categories other than White and African-American were combined into Other.

**Free-text fields and Comments**: These have been removed from the data sets.

**Re-Identification Procedures:** <State if there is any method in which subjects could possibly be reidentified.

**Data Delivery:**

* After de-identification, the 3rd Party Honest Broker will provide de-identified data to study team for analysis.
* Data delivered will include <specify what is given>.
* The 3rd Party Honest Broker will establish a code linking <data/specimens> to their <data key,

i.e., medical record (MR) number, survey record number> (ID #) and data via a study number.

* Serving as the custodian of identifiable information, the 3rd Party Honest Broker will protect the confidentiality of subject with a Master list which will be kept on a password- protected electronic file on the network drive and not be accessible to any members of the study team.
* The 3rd Party Honest Broker will delete the ID # from the master list at the end of the study.