NYC
Health

SECTION F: Vital Data Matching or Linkage
To be completed only if your request requires data linkage and/or matching.

Vital data matching or linkage is performed by and at the New York City Department of Health and Mental Hygiene (DOHMH). If a link/match is to be made to New York State Department of Health (SDOH) administrative databases, matching and linkage may be performed by SDOH Office of Quality and Patient Safety (OQPS) with DOHMH permission.

The purpose of this section is to determine which identifiers you will provide for matching. This information will be used to determine how and where matching or linkage will be conducted. Please provide the following information.

1. **Vital Data Type/Years for linkage/matching**
   You are requesting a record linkage/match to which of the following NYC vital data registries:

Vital Data Type	Years/Dates Needed
☐ Birth	
☐ Death	
☐ Matched births to infant deaths (Composite)	
☐ Spontaneous Termination of Pregnancy	
☐ Other, (Describe)	

2. **Database/Registries to which you want to match Vital Data**
   Indicate the database and/or disease specific registry to which you want to link or match NYC vital data:
   ☐ New York State's Statewide Planning and Research Cooperative System (SPARCS)
   　　What is the (anticipated) approval date of the SPARCS application to NYS?
   　　Click here to enter text.

   　　Please provide SPARCS Data Protection Review Board approval documentation, once available. Click here to enter text.

   　　Application will be forwarded to NYSDOH for NYC vital stats - SPARCS match/linkage
   　　when all approvals and documents are received.

   ☐ Other Database 1. Describe Click here to enter text.
   ☐ Other Database 2. Describe Click here to enter text.
   ☐ Other Database 3. Describe Click here to enter text.
   ☐ Other Database 4. Describe Click here to enter text.
   ☐ Other Database 5. Describe Click here to enter text.

   If there are additional datasets you wish to link, please submit on separate page.

Needed for Data Match:
All applicants requesting vital data matches, regardless of center performing the match, must complete the applicable Data Matching Variable List at the end of this section.

Post Linkage/Match Dataset(s) returned to requester:
To protect confidentiality of the NYC vital records, all identifiers (including those provided to conduct the match) are stripped from the final matched dataset unless specifically listed and justified in Section C. If your request involves matching and you intend to have Case ID returned along with matched vital data, you must indicate this in Section C of this application and provide justification.

3. **Anticipated MATCHING/LINKAGE PROTOCOL**
Provide a succinct step-by-step narrative of the procedures and methods you anticipate to link NYC vital statistics data file(s) with each additional dataset. Consult the Data Use Office (**vsdata@health.nyc.gov**) if you have questions. Include all registries/databases that will ultimately be matched to vital data-including those that will be matched or linked to the final analytic dataset. For requests involving release of birth and death *paper/pdf* certificates, please specify data items to be redacted (e.g. informant's name, address, parent(s) name(s), SSN) and provide a rationale for HIPAA-sensitive items that are to be retained. *(1500 character maximum)*

Protocol:

GUIDELINES FOR DATA PREPARATION AND SUBMISSION

It is the responsibility of the data match requester to prepare a clean, ready to use dataset for match.

Minimum data needed to conduct a <u>linkage</u>:
A linkage is when you have a *common key variable* that links same individuals in two or more datasets: For example,

1. Vital records have been linked to medical records using the medical record number that is listed on birth certificate.
2. The National Death Index will provide a researcher with select certificate numbers and year of certificate for matched data. Then the researcher can obtain the corresponding certificate data when they provide the DOHMH with the certificate number list with NDI approval documentation.
3. The New York City Office of the Chief Medical Examiner (OCME) unique identifier is also listed on the NYC death certificate and may be used to link OCME records to death records.

Minimum data needed to conduct a <u>match</u>:
A match is required when you do <u>not</u> have a *common key variable* for same individuals in two or more datasets, however you need to determine whether same individuals reside in two or more datasets. Using the direct identifiers provided by the requestor, DOHMH will compare them to the vital record identifiers. Though matches may not be exact, probabilities of matches are determined and used to accept or reject matches.

Full name and *full date of birth* are the minimum *required* data elements for DOHMH to conduct a match. Note, if you harbor other data elements (e.g. any of those indicated on the matching grids below) and are permitted to disclose them to DOHMH for matching purposes, the match rate will likely improve as will the matching accuracy.

Condition of data (data cleaning and formatting in preparation for match)
Data errors: Examine and correct data with common formatting and validity errors. For example, all dates should adhere to identical formats.

Separate data elements: Separate each variable into respective parts. For example separate name into: prefix, first name, middle name, last name, and name suffix. Similarly, separate: house number, street name, apartment number, city, zip code, etc. into separate data strings.

Address fields: Run data through address correction software and provide both the original and corrected address; this will enhance matching feasibility and accuracy.

Social Security Number: Remove dashes and include leading 0 when applicable to create a 9-digit data string.

Dates: Provide date in a recognized date format as opposed to a text string.

Duplicate data records: Ensure your dataset does not include multiple records for the same individual.

Data Submission
Please submit data in a SAS or other database format that provides data element formats and nomenclature. Comma Separated Values (csv) files are also acceptable. If you must submit an excel file, verify that dates are in date format.

Identifiable data submission to DOHMH must be conducted according to DOHMH Division of Informatics and Information Technology and Telecommunications (DIITT) policy. DO NOT submit identifiable data by email, fax, mobile drives or U.S. Mail. Contact the Data Use Office (**vsdata@health.nyc.gov**) for information on proper submission. Note, the DOHMH Data Delivery System (BISCOM) is recommended for all transfers of identifiable data involving 50+ records.

Birth Data Matching Variable List

Instructions: Review the variables in Column A. Check off the variables you are providing for the match in Column B. In Column C, indicate the percentage of your records that contain the specified variable.

Indicate File Name for Dataset(s) Submitted for Match (if more than one dataset, separate with comma and fill out separate sheet for each dataset): Click here to enter text.
Indicate the estimated number of records per dataset submitted for match. If you anticipate multiple submissions, please indicate desired frequency and estimate number of records submitted with each match:
Click here to enter text.

A. Review these variable names & identify those that you will submit for matching. Enter any additional variables that will be included in your match.	B. Indicate by checkmark each variable listed in Column A that will be used for matching.	C. Indicate the % of your records that contain the variable checked in Column B.
Case ID (if you are requesting Case ID to be returned subsequent to match, please indicate justification in Section D of application)	☐	**100%**
Infant's Last Name	☐	Click here to enter text.
Infant's First Name	☐	Click here to enter text.
Infant's Middle Name	☐	Click here to enter text.
Infant's Gender	☐	Click here to enter text.
Infant's Full Date of Birth	☐	Click here to enter text.
Mother's Last Name	☐	Click here to enter text.
Mother's First Name	☐	Click here to enter text.
Mother's Middle Name	☐	Click here to enter text.
Mother's Maiden Name	☐	Click here to enter text.
Mother's Full Date of Birth	☐	Click here to enter text.
Mother's Age	☐	Click here to enter text.
Mother's Social Security Number (only last 4 digits may be used for matching purposes)	☐	Click here to enter text.
Mother's House Number	☐	Click here to enter text.
Mother's Street Name (include directional, if part of street name)	☐	Click here to enter text.
Mother's Apartment Number	☐	Click here to enter text.
Mother's City	☐	Click here to enter text.
Mother's Zip Code	☐	Click here to enter text.
Mother's Telephone Number(s) (e.g. home, cell, whatever is available)	☐	Click here to enter text.
Father's Full Date of Birth	☐	Click here to enter text.
Father's Age	☐	Click here to enter text.
Father's Social Security Number (only last 4 digits may be used for matching purposes)	☐	Click here to enter text.
Click here to enter text.	☐	Click here to enter text.
Click here to enter text.	☐	Click here to enter text.
Click here to enter text.	☐	Click here to enter text.
Click here to enter text.	☐	Click here to enter text.

Death Data Matching Variable List

Instructions: Review the variables in Column A. Check off the variables you are providing for the match in Column B. In Column C, indicate the percentage of your records that contain the specified variable.

Indicate File Name for Dataset(s) Submitted for Match (if more than one dataset, separate with comma and fill out separate sheet for each dataset): Click here to enter text.
Indicate the estimated number of records per dataset submitted for match. If you anticipate multiple submissions, please indicate desired frequency and estimate number of records submitted with each match:
Click here to enter text.

A. Review these variable names & identify those that you will submit for matching. Enter any additional variables that will be included in your match.	B. Indicate by checkmark each variable listed in Column A that will be used for matching.	C. Indicate the % of your records that contain the variable checked in Column B.
Case ID (if you are requesting Case ID to be returned subsequent to match, please indicate justification in Section D of application)	☐	**100%**
Decedent's Last Name	☐	Click here to enter text.
Decedent's First Name	☐	Click here to enter text.
Decedent's Middle Name	☐	Click here to enter text.
Decedent's Alias Name	☐	Click here to enter text.
Decedent's Gender	☐	Click here to enter text.
Decedent's Full Date of Birth	☐	Click here to enter text.
Decedent's Age	☐	Click here to enter text.
Decedent's Social Security Number	☐	Click here to enter text.
Decedent's House Number	☐	Click here to enter text.
Decedent's Street Name (include directional, if part of street name)	☐	Click here to enter text.
Decedent's Apartment Number	☐	Click here to enter text.
Decedent's City	☐	Click here to enter text.
Decedent's Zip Code	☐	Click here to enter text.
Decedent's Date of Last Contact (used for QI for Date of Death)	☐	Click here to enter text.
Click here to enter text.	☐	Click here to enter text.
Click here to enter text.	☐	Click here to enter text.
Click here to enter text.	☐	Click here to enter text.
Click here to enter text.	☐	Click here to enter text.
Click here to enter text.	☐	Click here to enter text.
Click here to enter text.	☐	Click here to enter text.
Click here to enter text.	☐	Click here to enter text.
Click here to enter text.	☐	Click here to enter text.