

Request Overview

Data Use, Privacy, Security

What data elements are you seeking? This should also specify the timeframe of data you are seeking.	For each person in the NYeC sMPI with a NY address between ages 2 and 19 as of when the report is run. We will be querying for the most recent CCD (raw or consolidated) on file with each QE.		
	Below is a list of data elements anticipated to be provided to the NYSDOH. Additional metadata may be captured by NYeC for troubleshooting purposes, such as CCD identifiers and/or timestamps, but are not relevant to the use case and therefore will not be supplied to DOH.		
	Data Element	Definition	Expected Source
	Patient Census-tract	Census tract where the patient resides	NYeC
	County	County, derived from resident address	NYeC
	QE Source	QE generating the data	NYeC/QE
	Race	Patient's race	SHIN-NY Data Lake/CCD
	Ethnicity	Patient's ethnicity	SHIN-NY Data Lake/CCD
	Sex assigned at birth	Can be used in analysis of the results	SHIN-NY Data Lake/CCD
	Facility Name	Name of the facility where data was recorded	CCD
	age-in-months	age-in-months at time of the measurements.	CCD
	Weight	Weight of the patient	CCD
	Weight Unit	Unit of measure for weight	CCD
	Date Weight Recorded	Date of the encounter where weight was recorded	CCD
	Height	Height of the patient	CCD
	Height Unit	Unit of measure for height	CCD
	Date Height Recorded	Date of the encounter where height was recorded	CCD
	BMI	Body Mass Index of the patient	CCD
	Date BMI Recorded	Date of the encounter where BMI was recorded	CCD
	Obesity Diagnosis	Obesity diagnosis code system, code, description	CCD

Is the data being requested Identified, Limited, or De-Identified?	Limited
Do the data elements requested meet the standard of Minimum Necessary ¹ ? If not, please explain.	Yes. Based on discussions with the SHIN-NY, the key information is mainly exchanged in CCDAs and would be a new source of data into the Data Lake. NYeC will be using the list of children 2-19 from the state MPI to query for CCDAs. The key data elements will be extracted from the CCDAs and passed to the Data Lake for storage. The remainder of the CCDAs will be discarded, keeping only the minimum necessary data.
Proposed delivery method to NYSDOH	Text file sent via SFTP
Who will be the data recipient(s)? Please list all.	Staff within the Division of Chronic Disease Prevention, Bureau of Chronic Disease Evaluation and Research will have access to data files provided by NYeC. The NYeC technical teams will be receiving the data within the NYeC infrastructure.
Who will be accessing the data? Please list all.	Staff within the Division of Chronic Disease Prevention, Bureau of Chronic Disease Evaluation and Research will have access to data files provided by NYeC. The NYeC technical teams and Data Team will have access to the data within the NYeC infrastructure.
What safeguards are in place / will you put in place to protect the data?	The NYSDOH will take the steps outline below to ensure data are stored properly. NYeC will ensure the data will be shared and stored securely within the NYeC infrastructure. The data will be parsed using a vendor tool and stored in the NYeC Data Lake, which is in the AWS cloud and follows all necessary data protections and protocols required under HITRUST.
1. SECURE WORKSTATIONS	NYSDOH staff will store data shared by NYeC on a dedicated secure NYSDOH local area network (LAN) drive where access will only be granted to project staff working with the data.
2. TRAINING OF STAFF ON PROPER DATA HANDLING	NYSDOH staff will comply with training requirements related to proper data handling and security.
3. QE-SPECIFIC TRAINING AND PROCEDURES	NYSDOH staff accessing the data will participate and complete any initial or supplemental trainings required by the QE.
4. NYSDOH, NEW YORK STATE OFFICE OF INFORMATION TECHNOLOGY SERVICES (ITS) AND NYS RULES FOR PROTECTING DATA	Project staff accessing the data will review and comply with all NYSDOH, ITS and NYS rules for protecting data.

¹ <https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/minimum-necessary-requirement/index.html>

De-Identification Methodology

NYeC will only share the data elements listed above.

Attestation(s) for De-Identified Data:

- ☒ DOH will not, and will not permit any third parties to, use de-identified data for any purposes other than those permitted in an Approved Data Use Request or otherwise required by law
- ☒ DOH will not re-identify or attempt to re-identify the de-identified data
- ☒ DOH will report to NYeC if DOH has knowledge that de-identified data has been re-identified

Background

Obesity and overweight-related conditions, such as heart disease, stroke, and type 2 diabetes, are the second leading preventable causes of death in the United States. Over the past three decades, obesity prevalence among children and adolescents has tripled, reaching epidemic proportions in New York State, particularly affecting specific racial and ethnic minority groups. The current Student Weight Status Category Reporting System relies on paper-based surveys collected by school nurses and aggregated for the NYSDOH over a two-year period. By leveraging SHIN-NY data, the Department of Health aims to alleviate the administrative burden on school-based health centers and county school systems, enhancing efficiency and timeliness in reporting and addressing childhood obesity.

The pilot, conducted by Rochester, HealtheLink, and HealthiX, tested the SHIN-NY's ability to calculate BMI for age, obtain record-level data from Qualified Entities (QEs), and assess the infrastructure's capability to support population measurement of childhood obesity. QEs submitted flat files with record-level data on the age and demographic information of children and adolescents ages 2-19 from clinical records. Pilot specifications included inclusion and exclusion criteria, a standard set of codes for defining height, weight, BMI, and BMI-for-age, and the file format for the record-level files transmitted.

The pilot findings from NYS DOH indicated significant potential for SHIN-NY to function as a comprehensive statewide data source for monitoring childhood obesity. The strategy involves utilizing the existing Statewide Master Patient Index (SMPI) and the Secure Patient Record Locator (SPRL), along with insights gained from the pilot. NYeC will query the Qualified Entities (QEs) using the current systems, extract the specified data elements, and store them in a data lake. Upon completion of the transactions, the C-CDAs will be deleted from NYeC's systems.

Data Exchange

The table below details all the data elements that will be included in the extract to NYSDOH. The source is indicated per data element, as some of the data will originate from the QE CCDA and others will be added by NYeC.

Data Element	Definition	Format/Value (if applicable)	Required
Patient Census-tract	Census tract where the patient resides	String	Yes
County	Patient county as derived from the address.	String	Yes
QE Source	Source of the data	String	Yes
Facility Name	Name of the facility where data was recorded	String	No
Race	Patient's race	String	No
Ethnicity	Patient's ethnicity	String	No
Sex Assigned at Birth	Patient's sex assigned at birth	String	No
Age-In-Months	Months from date of birth, when the measurement was taken. ²	Integer	Yes
Weight	Weight of the patient	Numeric	Height and Weight or BMI must be provided
Weight Unit	Unit of measure for weight	Follow USCDI Kilograms (kg) Pounds (lb)	If weight provided
Date Weight Recorded	Date of the encounter where weight was recorded	YYYY-MM-DD	If weight provided
Height	Height of the patient	Numeric	Height and Weight or BMI must be provided
Height Unit	Unit of measure for height	Follow USCDI Centimeters (cm) Inches (in)	If height provided
Date Height Recorded	Date of the encounter where height was recorded	YYYY-MM-DD	If height provided
BMI	Body Mass Index of the patient	Numeric	Height and Weight or BMI must be provided
Date BMI Recorded	Date of the encounter where BMI was recorded	YYYY-MM-DD	If BMI provided
Obesity Diagnosis Code	Clinical codes (such as ICD-10or SNOMED) related to Obesity	Code(s)	Yes, if provided

² $\text{agemos} = \text{intck}(\text{'month'}, \text{patient_dob}, \text{encounter_date}) - (\text{day}(\text{encounter_date}) < \text{day}(\text{patient_dob}))$;

The first part counts the number of times a month boundary is crossed between the two dates and the second part corrects for when the end date is less than the start date.

- Ex 1: DOB of 2/14/18 and an encounter date of 7/19/21 was calculated as 41 months (41 months 5 days old...)
- Ex 2: DOB of 5/29/09 and encounter date of 12/20/21 was calculated as 150 months (151 months rounded down one because 20 is less than 29)

Inclusion Criteria

The criteria for NYeC to include sending data to the DOH is the following:

1. List extracted from sMPI of patients between 2 and 19 at the time the data is pulled
2. Where the 'golden record' address is in NY State
3. Where the patient has a CCDA at least one QE
4. Where the CCDA has at least height and weight, or BMI, or both.

Exclusion Criteria

An individual in sMPI of a different age (outside 2-19 years) or any address outside NY State.

File Transmission

Start Date:

End Date:

Transmission:

☐ Email

☒ SFTP Site

☐ Health Commerce System

How will files be handled/stored/deleted at the end of the project?

All identified data elements originating from the CCDA will be stored in the Data Lake and DOH on an ongoing basis to allow for state trending over time. The original CCDA will be destroyed immediately after the data elements defined in this use case as being sourced from the CCDA are extracted from the CCDA.

NYSDOH staff will document and track the datasets received from NYeC. Project staff will develop a metadata worksheet for each dataset received from the QEs which will include information about the file (e.g., data source, description of data, QE contact, date downloaded, etc.) and a data dictionary.

Data will be stored on a dedicated secure NYSDOH LAN drive where access will only be granted to project staff working with the data. Project staff will utilize data analytics tools including SAS, ArcGIS, R and SQL to analyze and evaluate data which will be operated on NYSDOH secure networks, and all files associated with QE data will be saved and archived on a dedicated secure NYSDOH LAN drive. Deletion of source data files from the QEs will adhere to QE, NYSDOH, and ITS policies and timeline for data and records retention.

Analysis

NYeC and/or DOH will do basic analysis including:

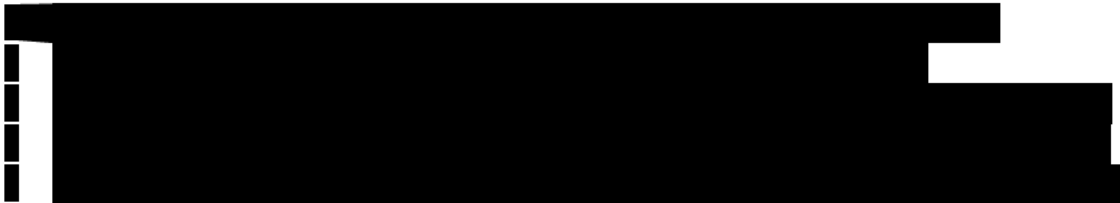
- Trends over time
- Mapping
- Correlations between data elements
- Aggregate counts

Project Contact

Primary project contacts are listed above.

Other Project Team Members

The requested data will be accessible to staff in the Bureau of Chronic Disease Evaluation and Research involved in public health surveillance for childhood obesity.



File Template

The file will be provided to DOH in CSV format. Below is a sample file template:

Census-tract	Age-in-months	QE Source	County	Date of encounter for BMI	Race	Date of encounter for Height	Date of encounter for Weight	Ethnicity	Facility Name	Weight	Weight Unit of Measure	Height	Height Measure	BMI	Obesity Diagnosis
123456	120	Source A	West Chester	2023-04-01	White	2023-04-01	2023-04-01	Hispanic	Facility X	60	Pounds	48	Inches	21.8	True
123457	110	Source B	Clinton	2023-03-15	Asian	2023-03-15	2023-03-15	Non-Hispanic	Facility Y	50	Pounds	45	Inches	20.5	False
123458	132	Source C	Nassau	2023-02-28	Black	2023-02-28	2023-02-28	Hispanic	Facility Z	55	Pounds	46	Inches	22.6	True
123459	204	Source D	New York	2023-01-20	White	2023-01-20	2023-01-20	Non-Hispanic	Facility W	58	Pounds	47	Inches	23.1	True
123460	201	Source E	Niagara	2023-05-05	Asian	2023-05-05	2023-05-05	Hispanic	Facility X	52	Pounds	44	Inches	21.2	False
123461	195	Source F	Putnam	2023-04-18	Black	2023-04-18	2023-04-18	Non-Hispanic	Facility Y	60	Pounds	49	Inches	22.4	True