Facilitating Research through the Childhood Obesity Data Initiative

Alyson B. Goodman, MD/MPH¹, Ray King, PhD/MSc¹, Peter Mork, PhD², Andrew Gregorowicz, MS², Matthew Daley, MD³
¹CDC, Atlanta, GA; ²MITRE, Atlanta, GA; ³Kaiser Permanente, Denver, CO

Abstract

Childhood obesity impacts almost 14 million U.S. children and is associated with serious and costly health risks. Research that assesses childhood obesity interventions is hindered by an inability to link patient records across information systems where pediatric health-related data are stored. The Childhood Obesity Data Initiative (CODI) aims to improve data capacity for childhood obesity research by facilitating access to patient-level, linked, longitudinal data that include health outcomes, information about weight management interventions, and risk factors. This panel will include a federal project lead, technical experts, and a clinical researcher who will present:

1) The studies that CODI enables—childhood obesity comparative effectiveness research, program evaluation, and population health surveillance.
2) The CODI data model; architecture for contributing data; process for assembling longitudinal data; and means by which research queries are processed.
3) The perspective of a CODI participant, including research interests and the steps taken to participate in CODI.

Learning Objectives:

- Learn how CODI supports childhood obesity research and
- Gain insight into the informatics requirements of research data networks.

Introduction

Childhood obesity impacts almost 14 million U.S. children and is associated with serious and costly immediate and future health risks, school absenteeism, social issues such as bullying, and poor mental health. Research that assesses and compares childhood obesity interventions is hindered, however, by limitations in researchers’ ability to link patient records across information systems where pediatric health-related data are stored.

To address this challenge and to improve data capacity for childhood obesity research, program evaluation and surveillance, the US Centers for Disease Control and Prevention (CDC), funded by the Patient-Centered Outcomes Research Trust Fund, is leading the Childhood Obesity Data Initiative (CODI). CODI leverages existing IT tools in innovative ways to facilitate access to individual-level, linked, longitudinal data that includes clinical outcomes, clinical and community settings’ weight management interventions, and risk factors across health systems and sectors. Access to patient-level linked data through CODI will:

1) Enable multisector patient-centered outcomes research on interventions,
2) Align evidence-based interventions with the needs of patients & communities,
3) Improve the timeliness and geographic specificity of public health surveillance using local multi-sector data within distributed networks, and
4) Enhance data capacity for healthcare quality improvement and program evaluation.

To achieve improved data capacity for childhood obesity research, CODI solved several informatics challenges, which the presenters will describe in this panel:

- Researchers defined childhood obesity research use cases and identified priority research questions for CODI; answers to these questions can inform implementation and efficacy of childhood obesity interventions
- Informaticists analyzed the research questions to determine the extent to which the PCORNet Common Data Model (CDM) can answer those questions. When this analysis identified research questions that could not be
answered using CDM, the informaticists augmented CDM with ancillary data tables that contain the information needed to close these gaps.

- Systems engineers developed a query architecture for CODI that allows pediatric weight management programs—both clinical and non-clinical—to contribute data to CODI, a trusted third party to assemble longitudinal patient data, and researchers to receive answers to their queries.

CODI developed these solutions within the PCORnet infrastructure. In CODI, a researcher submits their query to a trusted third party. The trusted third party determines what data must be retrieved from the participating CODI data contributors. Those participants review the requests to ensure compliance with data use agreements and institutional policies. The trusted third party assembles the data into a longitudinal record for each child that includes clinical and non-clinical data. Finally, the researcher receives the answer to their research query.

In this panel presentation, we will delve into the technical details of how CODI provides value to the research community. The intended audience for this panel comprises public health researchers and implementors of research data networks and surveillance systems.

**Moderator**

Dr. Alyson Goodman is the CODI Director and will moderate the panel. Dr. Goodman is a board-certified pediatrician and medical epidemiologist in the Division of Nutrition, Physical Activity, & Obesity at CDC. Dr. Goodman is the Division lead for electronic health information, including child obesity data integration and interoperability standards, clinical decision supports, and leveraging novel technology to spread best practice and harmonize data elements for surveillance, program evaluation and research. Her role as the Obesity Prevention & Control Branch lead for health services research on childhood obesity includes augmenting public health support for implementation, spread and scale of best-practices for childhood obesity prevention & treatment. She received her MD and MPH from Emory University in 2007, completed pediatrics residency training at Boston Children’s Hospital/Boston Medical Center, and a post-doctoral CDC fellowship in applied epidemiology with the Epidemic Intelligence Service.

**Research Questions**

Dr. Ray King is the CODI Co-Director. Dr. King is an epidemiologist and informaticist in the Division of Nutrition Physical Activity and Obesity. He leads projects investigating the utility of EHR data for obesity population health and surveillance and co-leads the CODI effort developing methods, tools and services for integrating multisector data for childhood obesity population health decision support. He has 18 years of experience at CDC with an interest in spatial epidemiology and a systems approach to understanding the distribution and determinants of disease and the development of multisector solutions for population health prevention and control. Dr. King is a graduate of CDC’s Public Health Informatics Fellowship and Emerging Infectious Diseases Fellowship. He received his PhD and MSc from the London School of Hygiene and Tropical Medicine and BA from Emory University.

Dr. King will provide an overview of the Childhood Obesity Data Initiative, including how it will improve data capacity for health services researchers and public health surveillance. Dr. King will review how CODI leverages existing infrastructure to improve access to data that can answer questions around the prevalence of childhood obesity within communities, childhood obesity screening and actions, weight management intervention “dose” and characteristics associated with weight management programs, and cost effectiveness of those interventions. He will describe the innovative elements of the CODI data models. Finally, he will describe the significant level of stakeholder engagement in the CODI decision-making process.

**Data Modeling**

Dr. Peter Mork is the chief data engineer for CODI. Dr. Mork is a Principal Health Systems Engineer with MITRE. They have nearly two decades of experience in biomedical & health informatics, with an emphasis on information management. They provide technical guidance to a wide range of government agencies. Their research activities have resulted in dozens of publications and patents in data integration, cost estimation, object-relational modeling, intrusion detection, privacy policies, and semantic web knowledge sharing.

Dr. Mork will first describe the CODI data model. The CODI team extracted data concepts from the high-priority research questions that would shed light on the information gaps that exist in order to evaluate the effectiveness of childhood obesity programs. We then conducted a technical environmental scan to determine the extent to which CODI participants would be able to provide data related to those data concepts, and for the record linkage to develop
a patient-level longitudinal data file. This environmental scan also revealed potential gaps in data quality and concepts for which insufficient data standardization has transpired.

Based on the research questions and the identified gaps, we developed ancillary tables that augment CDM with data elements specific to CODI. For example, CODI needs one set of ancillary tables to support record linkage and another set of tables to support weight management programs or services or interventions. In the presentation, we will demonstrate how the CODI data model aligns with the researchers’ questions.

**Building and Querying Longitudinal Records**

Andrew Gregorowicz is the technical lead for CODI. Mr. Gregorowicz is a Principal Software Systems Engineer with MITRE. He has been an active member of the Health IT standards and interoperability community for over 10 years. In his role, Mr. Gregorowicz has developed system architectures and prototypes for several government agencies. He is an active member of the open source software community, with contributions spanning infrastructure tools to clinical applications. Mr. Gregorowicz received his MS from Boston University.

Mr. Gregorowicz will describe how CODI processes data. He will begin by providing an overview of the extensions the CODI team made to the PCORnet architecture described above. These extensions provide support for privacy-preserving record linkage used to link children across sites and for iterative queries necessary to identify study cohorts based on the cohort definition provided by a researcher.

Finally, he will describe the techniques CODI uses to assemble longitudinal data. Decision rules describe how to handle situations in which conflicting information exists for a patient. For example, different sources may have different values for a patient’s race or date of birth. These rules also flag biologically implausible values for height, weight, and BMI based on all available data, as illustrated in Figure 1. Finally, he will describe the end-to-end sequence of events that transpire to answer a researcher’s query.

**Participating in CODI**

Dr. Matthew Daley is a pediatric clinician-researcher and site lead for CODI. He is a pediatric health services researcher and practicing pediatrician. He has worked extensively in national and regional distributed data networks.

Dr. Daley will provide the perspective of a CODI participant. He will begin by describing the specific research interests from his organization’s perspective and summarize the 28 use case questions that CODI will support. These use cases fit into three general themes, such as “What dose and characteristics of weight management interventions are associated with effectiveness?”

Next, he will share the steps that have been taken to make clinical data available for research within PCORnet. He will share lessons learned supporting the CODI data model.

**Discussion Questions**

- What other conditions are well suited for the application of CODI infrastructure and technology?
- How might CODI benefit from other investments in data interoperability?
- How might CODI handle missing data when building a longitudinal record for a child?

**Participation**

All participants have agreed to take part on the panel.