

# The Childhood Obesity Data Initiative: A Case Study in Implementing Clinical-Community Infrastructure Enhancements to Support Health Services Research and Public Health

Raymond J. King, PhD, MSc; Dawn M. Heisey-Grove, PhD, MPH; Nedra Garrett, MS; Kenneth A. Scott, PhD, MPH; Matthew F. Daley, MD; Matthew A. Haemer, MD; Pradeep Podila, PhD, MHA; Jason P. Block, MD; Tom Carton, PhD, MS; Andrew J. Gregorowicz, MS; K. Peter Mork, PhD; Renee M. Porter, DNP; Daniel L. Chudnov, MS; Jim Jellison, MPH, MS; Emily M. Kraus, PhD, MPH; Megan R. Harrison, MPH; Marissa Scalia Sucusky, MPH; Sarah Armstrong, MD; Alyson B. Goodman, MD, MPH

## ABSTRACT

**Context:** We describe a participatory framework that enhanced and implemented innovative changes to an existing distributed health data network (DHDN) infrastructure to support linkage across sectors and systems. Our processes and lessons learned provide a potential framework for other multidisciplinary infrastructure development projects that engage in a participatory decision-making process.

**Program:** The Childhood Obesity Data Initiative (CODI) provides a potential framework for local and national stakeholders with public health, clinical, health services research, community intervention, and information technology expertise to collaboratively develop a DHDN infrastructure that enhances data capacity for patient-centered outcomes research and public health surveillance. CODI utilizes a participatory approach to guide decision making among clinical and community partners.

**Implementation:** CODI's multidisciplinary group of public health and clinical scientists and information technology experts collectively defined key components of CODI's infrastructure and selected and enhanced existing tools and data models.

**Author Affiliations:** Obesity Prevention and Control Branch, Division of Nutrition, Physical Activity, and Obesity, National Center for Chronic Disease Prevention and Health Promotion (Drs King and Goodman and Mss Harrison and Sucusky), and Center for Surveillance Epidemiology and Laboratory Services (Ms Garret), Centers for Disease Control and Prevention, Atlanta, Georgia; Health Technical Center, The MITRE Corporation, McLean, Virginia (Drs Heisey-Grove and Mork and Messrs Gregorowicz, Chudnov, and Jellison); Denver Public Health, Denver, Colorado (Dr Scott); Institute for Health Research, Kaiser Permanente Colorado, Aurora, Colorado (Dr Daley); University of Colorado Department of Pediatrics and Children's Hospital Colorado, Aurora, Colorado (Dr Haemer); Department of Population Medicine, Harvard Medical School/Harvard Pilgrim Health Care Institute, Boston, Massachusetts (Dr Block); Louisiana Public Health Institute, New Orleans, Louisiana (Dr Carton); McKing Consulting Corporation, Atlanta, Georgia (Dr Porter); Public Health Informatics Institute, Decatur, Georgia (Dr Kraus); and Duke University School of Medicine, Durham, North Carolina (Dr Armstrong).

The authors thank the following people for their contributions in developing and implementing the Childhood Obesity Data Initiative: Heidi M. Blanck, Marnie Bloom, Matt Boyas, Arthur Davidson, Carrie Daymont, Sara J. Deakney Davies, Cynthia Dittmar, Seth Foldy, David Freedman, Brooke von Fricken, Suzianne Garner, Mark Gray, Robert Grundmeier, Lisa Johnson, Ruth Petersen, John Rose, Teresa Kinley, Adrean Mabry, Toan Ong, Drew Pellitieri, Katie Redfield, Brett Reeder, Mona Sharifi, Steffani Silva, Asheley Skinner, Carmen Smiley, and Rachel Zucker.

This work was supported by the Office of the Secretary Patient-Centered Outcomes Research Trust Fund through an interagency agreement between

the Office of the Assistant Secretary for Planning and Evaluation and the Centers for Disease Control and Prevention's Division of Nutrition Physical Activity and Obesity (IAA #750118PE090005).

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Office of the Assistant Secretary for Planning and Evaluation, the Centers for Disease Control and Prevention, or the Department of Health and Human Services.

The authors have indicated no potential conflicts of interest to disclose.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (<http://www.JPHMP.com>).

Written work prepared by employees of the Federal Government as part of their official duties is, under the U.S. Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the United States Code is not available. As such, copyright does not extend to the contributions of employees of the Federal Government.

**Correspondence:** Raymond J. King, PhD, MSc, Obesity Prevention and Control Branch, Division of Nutrition, Physical Activity, and Obesity, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, 1600 Clifton Rd, NE, MS F77, Atlanta, GA 30333 ([rjking@cdc.gov](mailto:rjking@cdc.gov)).

DOI: 10.1097/PHH.0000000000001419

We conducted a pilot implementation with 3 health care systems and 2 community partners in the greater Denver Metro Area during 2018-2020.

**Evaluation:** We developed an evaluation plan based primarily on the Good Evaluation Practice in Health Informatics guideline. An independent third party implemented the evaluation plan for the CODI development phase by conducting interviews to identify lessons learned from the participatory decision-making processes.

**Discussion:** We demonstrate the feasibility of rapid innovation based upon an iterative and collaborative process and existing infrastructure. Collaborative engagement of stakeholders early and iteratively was critical to ensure a common understanding of the research and project objectives, current state of technological capacity, intended use, and the desired future state of CODI architecture. Integration of community partners' data with clinical data may require the use of a trusted third party's infrastructure. Lessons learned from our process may help others develop or improve similar DHDNs.

**KEY WORDS:** clinical-community linkage, pediatric obesity, record linkage

Childhood obesity affects almost 14 million US children<sup>1,2</sup> and is associated with serious and costly immediate and future health risks, poor academic and coping skills, and poorer mental health.<sup>3-8</sup> Research, program evaluation, and public health surveillance that assesses childhood obesity risk factors, interventions, and social determinants of health (SDOH) services are limited because of the inability to link, longitudinally arrange, and securely share individual health-related information stored across different sectors and information systems. Child health data, including clinical information, SDOH, weight management programs, and geographic markers are maintained throughout communities in many separate information systems operated by hospitals, provider networks, and clinical and community-based programs. Current distributed health data networks (DHDNs; also referred to as clinical data research networks or learning networks) such as the health system-based Clinical Research Networks of the National Patient-Centered Clinical Outcomes Research Network (PCORnet)<sup>9</sup> provide a foundational infrastructure that can be extended for clinical and community organizational interoperability.<sup>10</sup>

To improve data capacity for childhood obesity research, program evaluation, and surveillance, the US Centers for Disease Control and Prevention (CDC) is leading the Childhood Obesity Data Initiative (CODI).<sup>11</sup> On the basis of a participatory framework of local and national clinical, community, and public health representatives, CODI collaboratively developed and implemented an information technology infrastructure to standardize and link individual health data across clinical and community sectors while preserving patient privacy and confidentiality. CODI leverages an existing DHDN, the Colorado Health Observation Regional Data Service (CHORDS) network,<sup>12</sup> and other open-source technology solutions in an end user-driven development process.

CHORDS and other DHDNs currently provide infrastructure allowing research on populations spread across health care systems by distributing and executing queries against locally stored data.<sup>10</sup> This distributed query process is possible because participating organizations map their data to a common data model (CDM), thereby standardizing data concept tables and elements within the network. Organizations participating in DHDNs maintain individuals' personally identifiable information (PII) locally behind their firewalls. Typically, only aggregate or limited data sets are released to investigators by the DHDN. Most DHDNs lack unique patient data linkage across organizations, making it difficult to follow individuals over time and across health systems and community settings. They also lack information on clinical and community weight management interventions, community health programs, for example, after-school physical activity programs or faith-based health initiatives, and navigation to benefits or services that may address SDOH.

Our objective in this article is to share the CODI framework and lessons learned to help others develop or improve similar DHDNs. To achieve this, we describe the implementation of and lessons learned from CODI in Denver, Colorado, with clinical, public health, and community partners. We also describe the CODI framework for collectively enhancing and implementing innovative changes to an existing DHDN infrastructure to support clinical-community linkages for childhood obesity research, surveillance, and evaluation by enabling (1) access to standardized clinical and community data; (2) individual cross-sector privacy-preserving record linkage (PPRL); (3) distributed clinical-community longitudinal queries; and (4) governance for data sharing across sectors and information systems. We outline an iterative and participatory process for collaborative engagement that structures and balances infrastructure development across 3 critical and interdependent components of an information system: people, process, and

technology. We present the people component in the CODI Program section and the technology aspects in the CODI Implementation section; our processes are described across each of these sections. We describe lessons learned in the CODI Evaluation section.

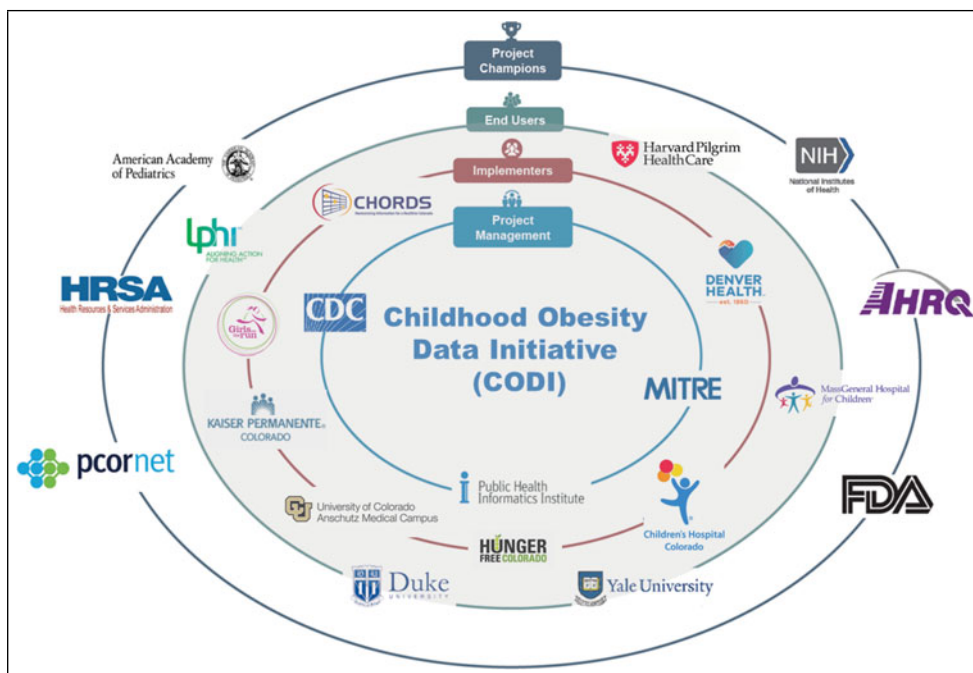
## CODI Program

CODI leverages the people-process-technology approach for organizational change.<sup>13,14</sup> This approach ensures that all 3 components—people, process, and technology—are integrated throughout the development process to facilitate organizational interoperability. This section describes the people and the processes used to engage those people, which included multisector stakeholders and CODI's iterative and participatory approach based upon the community health record framework designed to ensure end user needs are addressed.<sup>15</sup>

CODI's goal was to ensure that the network infrastructure was both useful and usable by community partners, health services researchers, and public health scientists, in government, health care, and academia. To achieve this, we established several collaborative groups. The largest of these was the CODI Collaborative (Collaborative; Figure 1).

Collaborative membership included staff from federal, state, and local public health and health care agencies, national organizations representing stakeholder groups, academicians, and practicing clinicians. The Collaborative was united by a common interest in improving data capacity for child obesity research, surveillance, and evaluation; and innovative reuse of existing technologies to conduct research, public health surveillance, and program evaluation. This group met quarterly to develop a shared vision of CODI's goals and achieve trust-building among stakeholders.

A subset of Collaborative members leading the CODI pilot—project managers, implementers, and end users—constituted the CODI Collaborative Working Group (CCWG; Figure 1). CCWG membership included technical and informatics experts and subject matter experts (SMEs) in clinical care, public health, DHDN research methods, data governance, and community programs. The project managers included staff from the CDC and the lead technical entity, The MITRE Corporation. These members coordinated all activities, managed subgroups, ensured adherence to timelines and completion of deliverables, and developed the technical infrastructure in partnership with CCWG members. The implementers



**FIGURE 1** The CODI Collaborative<sup>a</sup>

Abbreviation: CODI, Childhood Obesity Data Initiative.

<sup>a</sup>The CODI Collaborative membership included staff from federal, state, and local public health and health care agencies, national organizations representing stakeholder groups, academicians, and practicing clinicians. Members were grouped by role: project management, implementers, end users, and project champions. The CODI Collaborative workgroup members consist of the 3 inner rings (gray shading) of project management, implementers, and end users.

provided technical and research-based insights into the decisions around the infrastructure and included representatives from the organizations participating in the CODI pilot. The end users included researchers, informaticists, and public health scientists. They informed CODI infrastructure development in alignment with child obesity research and public health surveillance questions and provided domain expertise and research and informatics perspectives to ensure that final products appealed to the target audience.

The CCWG met in person for a 2-day project kickoff meeting to develop a shared vision and common understanding of CODI's scope and solutions. It provided an early opportunity to develop relationships and understanding of partner organizations; reach consensus on CODI's vision, objectives, and scope; select a preliminary governance approach; and collectively define solutions and an agenda to achieve success within the time and resources allowed. Deliverables included the project charter, communication plan, and preliminary decisions regarding CODI scope, research question domains, and potential technical solutions.

For the duration of the project, the CCWG met at regular intervals to review progress, make key decisions, discuss findings, identify next steps, and form subgroups to execute the work. Participating organizations were financially supported by the CDC.

CCWG subgroups—small teams of SMEs—met regularly to inform and conduct specific tasks (Table 1). Subgroups were designed to leverage members' specific expertise and limited availability. Subgroup leads performed much of the work; used recurring meetings to validate the process, information, and recommendations with members; and documented all activities. The CCWG vetted all subgroup recommendations before implementation.

## Implementation of CODI Developmental Phase

For CODI to develop solutions that add value to participating organizations and other end users, it was important to understand the information needs, the current information technology processes, and the desired future state of the Denver partners. Here, process refers to the human and technical actions and tasks performed by individuals, institutions, and systems in support of current and proposed future childhood obesity business practices<sup>13</sup> and the decision-making steps and agreements achieved by those organizations to achieve a shared and contractual understanding of an achievable future state. Activities included a scoping exercise to define priority research questions and information gathering to characterize (1) the

landscape of clinical and community organizations, (2) technical capacity of clinical and community organizations, and (3) options for the record linkage solution. Findings from these activities were used to design and develop the CODI technical infrastructure and governance policy.

All CODI technology is open-source and stand-alone with accompanying implementation guidance. Solutions were collectively identified, iteratively developed, and locally implemented with end user involvement at every stage (see Supplemental Digital Content Appendix A, available at <http://links.lww.com/JPHMP/A841>). To accelerate development and optimize resources, CODI leverages and enhances existing technology where available, allowing for rapid implementation. All CODI technology solutions highlighted in the following sections were available for implementation within 18 months of the project launch.

### *Priority research questions*

CODI-specific research questions were developed and used to inform, scope, and prioritize CODI infrastructure development and implementation. The CCWG Research Subgroup of child obesity health services researchers, public health scientists, and SMEs met over 2 months to formulate a comprehensive list of child obesity research questions. The subgroup categorized questions by type and prioritized them. To ensure relevance to all stakeholders, the CCWG reviewed, provided input, and approved the final questions.

Priority research questions scoped the CODI infrastructure to address gaps in childhood obesity comparative effectiveness and public health research. Table 2 presents these research questions, which address the concepts of (1) screening for child obesity, (2) intervention efficacy, (3) cost and cost-effectiveness of interventions, and (4) surveillance of child obesity in local communities.

### *Programmatic environmental scan*

The CCWG completed a programmatic environmental scan (PES) of 9 potential clinical and community organizations for the Denver pilot. Organizations selected for the PES had to provide 1 or more services, including (1) nutrition or physical activity counseling, (2) multidisciplinary clinical interventions to treat overweight and obesity (pediatric weight management interventions [PWMIs] only), (3) education and training on nutrition practices and physical activity skills, (4) navigation services for publicly provided resources and assets, or (5) social assets such as food assistance and insurance. PES discussions solicited

**TABLE 1**  
**The CCWG Subgroups Established to Conduct Project Task<sup>a</sup>**

Subgroup	Tasks	Membership Composition
<b>Research question</b>	<ul style="list-style-type: none"> <li>• Draft CODI research questions</li> <li>• Provide subject matter expertise to inform the gaps analysis between the PCORnet CDM and CODI research questions</li> <li>• Draft CODI pilot use cases</li> </ul>	<ul style="list-style-type: none"> <li>• Project management</li> <li>• Health services researchers</li> <li>• Childhood obesity subject matter experts</li> <li>• Public health scientists and informaticians</li> <li>• DHDN experts</li> <li>• Clinicians</li> </ul>
<b>Programmatic environmental scan</b>	<ul style="list-style-type: none"> <li>• Identified organizations to include in information-gathering activities</li> <li>• Conducted outreach to organizations</li> <li>• Summarized and presented information for CCWG decision making</li> </ul>	<ul style="list-style-type: none"> <li>• Project management</li> <li>• Childhood obesity intervention experts</li> <li>• Pilot site implementation leads</li> <li>• Evaluation expert</li> </ul>
<b>Technical environmental scan</b>	<ul style="list-style-type: none"> <li>• Conducted crosswalk between concepts in the PCORnet CDM and CODI research questions</li> <li>• Developed questions and conducted outreach to organizations for the scan</li> <li>• Summarized and presented information for CCWG decision making</li> </ul>	<ul style="list-style-type: none"> <li>• Project management</li> <li>• Data science experts</li> <li>• Pilot site implementation leads</li> </ul>
<b>Record linkage</b>	<ul style="list-style-type: none"> <li>• Evaluated and summarized record linkage strategies to inform CCWG decision making</li> <li>• Identified and evaluated open-source PPRL tools</li> <li>• Developed evaluation criteria to assess performance of the PPRL tools</li> </ul>	<ul style="list-style-type: none"> <li>• Project management</li> <li>• Identity management experts</li> <li>• Data science experts</li> <li>• DHDN experts</li> <li>• Pilot site implementation leads</li> </ul>
<b>Governance</b>	<ul style="list-style-type: none"> <li>• Conducted gaps analysis of CHORDS governance functions and CODI governance needs and made recommendations to the CCWG about mitigation strategies</li> <li>• Evaluated different data sharing and use agreement frameworks and recommended strategies to the CCWG</li> <li>• Drafted the data sharing and use agreement</li> <li>• Facilitated approvals by each organization's legal authorities</li> </ul>	<ul style="list-style-type: none"> <li>• Project management</li> <li>• Technical leads from each pilot organization</li> <li>• Governance leads from each pilot organization</li> <li>• DHDN governance experts</li> </ul>
<b>Evaluation plan</b>	<ul style="list-style-type: none"> <li>• Identified key evaluation questions, goals, indicators, and metrics</li> <li>• Recommended a prioritized implementation of the evaluation plan for the CODI pilot</li> <li>• Developed data collection tools based upon the evaluation plan</li> </ul>	<ul style="list-style-type: none"> <li>• Project management</li> <li>• Evaluation experts</li> <li>• Pilot site implementers</li> </ul>
<b>Implementation</b>	<ul style="list-style-type: none"> <li>• Provided input and feedback on CODI solutions and implementation guides</li> <li>• Tested and implemented CODI solutions</li> </ul>	<ul style="list-style-type: none"> <li>• Project management</li> <li>• Pilot site implementers</li> </ul>

Abbreviations: CCWG, CODI Collaborative Work Group; CDM, common data model; CHORDS, Colorado Health Observation Regional Data Service; CODI, Childhood Obesity Data Initiative; DHDN, distributed health data network; PCORnet, National Patient-Centered Clinical Outcomes Research Network; PPRL, privacy-preserving record linkage.

<sup>a</sup>The project management team was responsible for facilitating meetings, developing draft materials for review and discussion, maintaining membership, and documenting decisions and recommendations. Subgroup members were expected to attend and actively participate in all meetings, share their expertise, and provide occasional review and offline feedback of materials in preparation for meetings.

information about participants' program description and goals, delivery mechanisms, and technical capacity. The PES captured the estimated number of annual participants, geographic catchment area, program and intervention sustainability and intervention types, the availability of PII to support record linkage, and other relevant content. The CCWG used PES findings to select PWMIs and community programs for the technical environmental scan (TES).

Participation by any organization in the CODI pilot required the following: (1) electronic documentation of individual participation; (2) sufficient number of participants ( $n \geq 200$  annually) within the Denver metro area for linkages with clinical organizations; (3) a willingness to contribute data; and (4) an active clinical PWMI or community program. Current participation in CHORDS was required of all clinical organizations.

**TABLE 2**  
**Prioritized CODI Research Questions Defined by the CODI Research Question Subgroup to Inform, Scope, and Prioritize CODI Infrastructure Development and Implementation**

Concept(s) Included	Priority Research Question	Data Elements
Screening	When is obesity screening (measuring BMI) occurring? In what settings? What actions does it trigger, including comorbidity screening?	<ul style="list-style-type: none"> <li>• Demographics: Age, sex, race/ethnicity, etc</li> <li>• Geographic area; neighborhood; Census Tract</li> <li>• Health factors: Weight category, presence/absence of comorbidity</li> </ul>
Intervention efficacy based on intervention dose and characteristics	What “dose” and characteristics of weight management interventions are associated with effectiveness?	<ul style="list-style-type: none"> <li>• Payer</li> <li>• Social determinants: Socioeconomic status (ie, poverty and level of poverty); food security</li> <li>• Parent characteristics (eg, weight/height)</li> </ul>
Cost and cost-effectiveness	What are the cost and cost-effectiveness of weight management intervention?	<ul style="list-style-type: none"> <li>• Intervention characteristics</li> <li>• Cost of care</li> </ul>
Surveillance	What is the prevalence of obesity (and weight categories) among children aged 2-19 y and has the prevalence changed (trends) overtime within small geographic areas (in Denver CODI catchment)?	<ul style="list-style-type: none"> <li>• Weight categories: underweight, healthy weight, overweight, obesity, and severe obesity classes</li> <li>• Age, sex, race, ethnicity, and other demographics</li> <li>• Neighborhood, geographic area, rural vs urban</li> <li>• Social determinants</li> </ul>

Abbreviations: BMI, body mass index; CODI, Childhood Obesity Data Initiative.

Two community programs were selected for participation in CODI: Girls on the Run of the Rockies, which offers physical activity and self-confidence building activities; and Hunger Free Colorado, which provides navigation to safety net programs such as the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC). Four community programs were eliminated from consideration because they had insufficient participants (n = 3) or lacked resources to participate in CODI (n = 1). Three multidisciplinary clinical intervention programs were also selected for inclusion: Children’s Hospital Colorado’s Lifestyle Medicine Clinic; Denver Health’s Healthy Lifestyle Clinic; and Denver Health’s Mind, Exercise, Nutrition, Do it! (MEND) program.

**Technical environmental scan and business process analysis**

A TES was conducted within 4 clinical and 2 community organizations to gain an in-depth understanding of their data assets, software platforms, and information-sharing capacity relevant to identified research questions and patient-level record linkage capabilities. We designed and used a table shell to guide discussions with each organization (see Supplemental Digital Content Appendix B, available at <http://links.lww.com/JPHMP/A842>) around whether and how organizations collected important data elements in their information systems. In addition, general technical capabilities associated with data sharing and record linkage were assessed.

We conducted a business process analysis with these organizations to identify stakeholders’ objectives, information transactions, tasks, functional requirements for information systems supporting childhood obesity interventions, and data quality and standardization issues.<sup>16</sup> This analysis explored the delivery of child obesity health services related to screening of children for overweight and obesity, provision of nutrition and physical activity counseling, and navigation of patients to resources and assets that would promote healthy lifestyles and access to clinical services.

Through these efforts, we gained an understanding of participating organizations’ current technical infrastructure, data availability, and identified data and technology gaps relative to CODI requirements. We identified important research question data elements that were inaccessible for this pilot. For example, there was no consistency in how SDOH were captured across the implementing organizations’ electronic health record (EHR) systems, nor were there systematic processes that ensured that available fields were consistently and reliably collected. We also learned that implementers could not easily or reliably extract concepts for alerts notifications, referrals, or costs from their information systems. Tables containing data elements relevant to these concepts were therefore not prioritized for the Denver pilot.

**CODI data models**

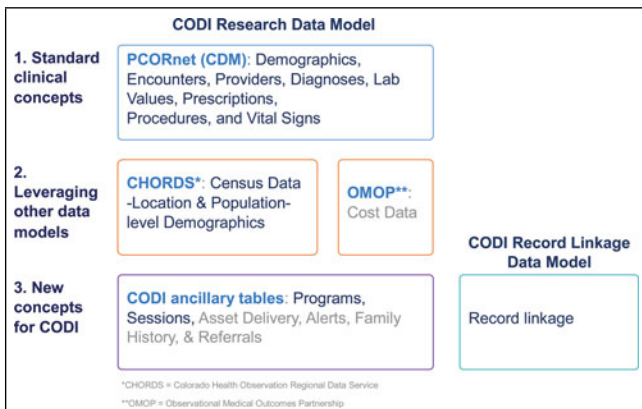
CODI data models were designed to standardize and accommodate all data concepts and elements

required to answer the CODI priority research questions and perform record linkage. Whenever possible, the CODI data models leveraged existing standards and concepts from other CDMs, adding supplemental elements and tables only when needed to address identified gaps. The CODI data models built primarily upon the PCORnet CDM v5.1.<sup>17</sup> A gaps analysis compared the CDM concepts and data elements with concepts and data elements needed for CODI. We learned that the CDM did not include data concepts for weight-related interventions such as dose (time and exposure), nor could it accommodate data from community programs. It also did not include concepts for cost, referrals between PWMIs and community programs, SDOH, clinical decision support alerts, identity management to support record linkage, and social asset delivery (eg, food benefits).

On the basis of the gap analysis findings, the project's technical lead collaborated with the Research Subgroup to develop the CODI Research and Record Linkage data models (Figure 2).<sup>18</sup> Ancillary data tables were created to capture concepts absent in the PCORnet CDM. Because of the limitations of existing SDOH data at the individual level, we supplemented the CODI data models with population-level proxies for SDOH (eg, education level, family income, poverty level, language spoken, employment rates, and occupancy rates). We identified these proxies based on individuals' geocoded census tract location of residence and a corresponding static reference table derived from the CHORDS Virtual Data Warehouse model<sup>12</sup> derived from 2010–2014 American Community Survey data. Cost data concepts were standardized using relevant OMOP CDM tables.<sup>19</sup>

### Record linkage solution

Traditionally, DHDNs do not include a record linkage function because of their desire to limit the PII



**FIGURE 2** Overview of CODI Data Models

that leaves institutional firewalls. As a result, patients may be counted multiple times across institutions or appear to be lost to follow up if they switch their care to another system. The CCWG Record Linkage Subgroup conducted a landscape analysis of open-source deidentified record linkage solutions using publicly available information (see Supplemental Digital Content Appendix C, available at <http://links.lww.com/JPHMP/A843>). Those findings informed the selection of tools to undergo detailed testing to compare performance and ability to support pediatric record linkage.

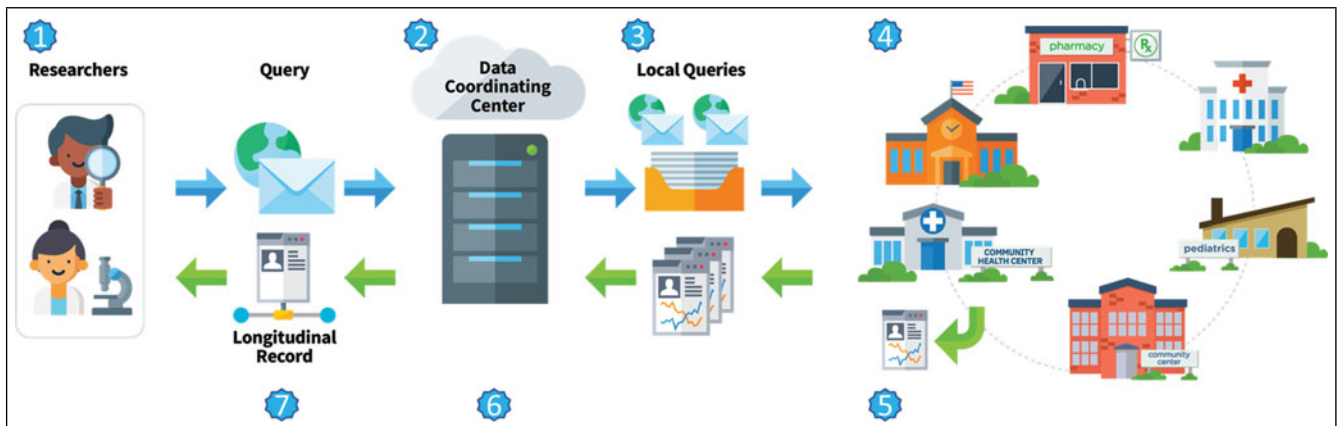
PPRL was selected for CODI because it maintains patient privacy while enabling record linkage across sectors and settings. In PPRL, PII remains secure behind the source institution's firewall. PPRL relies on complex encoding of a patient's data, called hashing, which prevents access to an individual's PII but allows records to be linked across organizations using a unique encoded identifier. An open-source tool, anonlink,<sup>20</sup> was selected for pediatric record linkage because of superior performance in tests of sensitivity and specificity in matching synthetic data.<sup>21</sup> These tests included matching sibling records, a difficult challenge for pediatric record linkage. We modified anonlink so that data from more than 2 organizations could be matched at once, a typical scenario within DHDNs.<sup>17,22</sup>

### Query architecture

We assessed different query architecture strategies to identify the best fits for CODI's query distribution and longitudinal result generation needs. Options included (1) a central data repository hosted by a trusted third party or data coordinating center (DCC) that compiled longitudinal records with individuals' PII and protect health information (PHI) in a physical database; or (2) a logical data warehouse in which participating organizations run queries behind their firewalls and share only limited data sets with a DCC that reconciles data received from all organizations into a longitudinal data set. The CCWG selected a virtual data warehouse model that leveraged existing query infrastructure to support the CODI distributed data query process.<sup>23</sup> The process steps are illustrated in Figure 3.

### Longitudinal data reconciliation, cleaning, and analysis

The CCWG also evaluated distributed longitudinal data reconciliation approaches (eg, how to resolve data conflicts on time-invariant data elements such as birth date and patient sex at birth), growth data



**FIGURE 3** The CODI Query Process Steps<sup>a</sup>

Abbreviations: CODI, Childhood Obesity Data Initiative; DCC, data coordinating center; PPRL, privacy-preserving record linkage.

<sup>a</sup>The CODI query process steps included (1) a researcher(s) submits a research request to the DCC that includes the research question of interest with specifications for data elements, study population definitions, and other details necessary for data extraction; (2) the DCC finalizes the request details and receives appropriate data use approvals; (3) the DCC develops and distributes a series of analytic queries to participating organizations using PopMedNet,<sup>17</sup> an open-source software package that supports distribution of research queries and data; (4) data partners execute queries within their information systems; (5) data partners return patient-level deidentified output to the DCC with the unique linkage identifier derived through PPRL; (6) the DCC uses the unique identifier to perform optional analyses, including data reconciliation and cleaning, and to assemble individuals' records received from multiple systems or settings into a longitudinal patient-level record; and (7) the DCC returns the longitudinal record to the researcher(s).

cleaning strategies, and analytic methods to ensure information output reliability and validity. From our Research Subgroup, we learned that end users preferred to specify data reconciliation criteria on a project-by-project basis.

To clean longitudinal child growth data by detecting height and weight measurement errors, we augmented an existing open-source EHR anthropometric data cleaning algorithm, *growthcleanr*,<sup>24</sup> based on feedback from a panel of experts. This tool flags biologically implausible longitudinal height and weight measures by assessing the difference between modified *z*-scores for weight and height at each visit and the expected values based on an exponentially weighted moving average for each child. We enhanced the algorithm to better detect biologically implausible height values by integrating logic for the World Health Organization height-velocity curves,<sup>25</sup> streamlined the tool's functionality, and improved supporting documentation.<sup>26</sup> We developed an accompanying novel open-source tool, *GrowthViz*, that facilitates visualization of *growthcleanr* output and data postprocessing.<sup>27</sup> Finally, the CCWG identified methods and developed an open-source algorithm to generate weighted prevalence estimates based upon nonprobability (ie, convenience) samples derived from EHR data to develop population estimates for childhood obesity.<sup>28,29</sup>

## Governance

Governance is essential in establishing policies, processes, and formal agreements among organizations

that allow information sharing necessary to facilitate research. The CCWG Governance Subgroup collaboratively identified a data governance strategy that leveraged existing local DHDN governance infrastructure.<sup>30</sup> The subgroup considered end user needs for efficient project review and approval and information-sharing needs of clinical and community organizations. End users requested a streamlined governance process for requesting research data and the pilot sites wanted to minimize burden for developing data sharing agreements to support CODI. The subgroup identified governance gaps and based upon these considerations and requirements of the technical solution opted for a Master Sharing and Use Agreement (MSUA). The subgroup then developed processes, drafted governance documents, and facilitated approvals by each organization's legal authorities (Kraus et al, in this issue).

The CODI MSUA (1) eliminated the need for data sharing agreements between researchers from participating organizations; (2) expedited the data sharing agreement process for researchers from non-participating organizations by appointing the DCC to coordinate agreements and sign on behalf of partners; and (3) streamlined research project review and approval through the DCC. For each research project, participating organizations opt in following review and approval by the DCC.

## Evaluation of CODI Developmental Phase

We developed a CODI evaluation plan based on the Good Evaluation Practice in Health Informatics



guidelines, the Reach Effectiveness Adoption Implementation Maintenance Framework, and the Consolidated Framework for Implementation Research.<sup>31–33</sup> These frameworks provided the context for (1) identifying essential aspects of the development phase; (2) assessing project elements that improve sustainable adoption and implementation of effective interventions; and (3) identifying barriers to and facilitators for successful implementation. The CCWG Evaluation Subgroup developed the evaluation during the initial phase of the project. Supplemental Digital Content Appendix D (available at <http://links.lww.com/JPHMP/A844>) presents the CODI evaluation plan for all phases.

Evaluation results presented in this article focus on lessons learned during the development phase of CODI, when CCWG members made critical decisions about the final state of CODI infrastructure. For this phase of the evaluation, an independent analyst captured lessons learned through structured interviews of the 9 subgroup leads and coleads. The CCWG reviewed and validated these findings.

The evaluation identified several lessons learned regarding working with individuals as part of CODI. A key to workgroup success was engaging the correct SMEs from each organization early and often. Overall, CCWG members reported feeling engaged in CODI decision making and that the resulting infrastructure reflected their needs. However, failure to engage selected technical and governance leads in the CCWG at the project's outset contributed to delays in establishing the Governance Subgroup and achieving a common understanding of governance goals and needs.

For the Research Subgroup, a broad range of childhood obesity research and public health science expertise were vital in ensuring a comprehensive scope. We learned, however, that as subgroup size increased group discussion decreased. We determined that discussion decreased within the group when membership exceeded 15 and therefore strived to limit membership to those individuals necessary to scope the research questions in an engaged and participatory fashion.

Aside from the selected Denver pilot sites, other Denver organizations with an interest in childhood obesity were not initially included in the Collaborative. We learned that organizations not directly involved in the local implementation were still interested in being updated on the progress and potential future participation. We therefore created a CODI Local Collaborative that provided quarterly updates on CODI progress to those organizations.

Finally, the evaluation findings stressed the importance of translating technical content into plain

language to ensure informed and shared decision making occurred throughout the CODI process. Identifying and defining key terms helped ensure a common understanding of CODI goals and activities. For example, we transformed references of “trusted third party” to “data coordinating center” and we replaced some technical terms with commonly used terms (eg, encryption key in lieu of “salt”).

## Discussion and Conclusion

CODI success is based on an iterative participatory approach that depends on the continued engagement of the right *people* and expertise; *processes* that elucidate and address the needs of public health, health care, community partners, and health services researchers; and open *technologies* that can scale for

### Implications for Policy & Practice

- When developing a clinical-community–based infrastructure across sectors and systems, our pilot found it important to engage a multidisciplinary group of stakeholders in a participatory process during all stages of development and to conduct work collaboratively based on a shared understanding of the problem, solution, and mutually beneficial objectives, using an iterative process with agreed-upon incremental steps. Others developing similar infrastructures can consider this approach.
- Rapid innovative technical implementations are possible through modifications of existing resources and tools. By avoiding a “build new” approach, collaborative clinical-community-public health partnerships may be able to conserve and optimize human and technical resources while facilitating answers to previously hard-to-answer questions. Leveraging existing technical, governance, and organizational infrastructure supported by trusted public health partners may simplify and streamline the process for incorporating community partners’ data into the DHDN.
- Clinical-community data sharing in a DHDN is reliant on a trust and governance framework that minimizes participation burden for data owners and researchers. An MSUA can ensure a common approach and promotes efficiency within the network. It can also streamline the process for researchers while allowing data providers autonomy over their data.
- Longitudinal health services and public health research and surveillance across an integrated clinical and community DHDN may benefit from linking patient-level data across systems and settings. PPRL enables patient-level clinical-community linkages at scale without exposing PII beyond organizations’ firewalls.

widespread adoption. CODI brings together clinical and community data to create longitudinal data sets that may improve the ability to conduct research, program evaluation, and public health surveillance on the significant public health problem of childhood obesity. Throughout the process, our CCWG members reported feeling engaged in CODI decision making and that the resulting infrastructure reflects their needs.

CODI infrastructure innovated by standardizing how intervention program data concepts were collected, using PPRL to link data within a DHDN and from both clinical and community partners, developed a streamlined governance process, and creating a longitudinal data set consisting of both clinical and community data from a DHDN. By leveraging existing open-source tools and resources, we were able to innovate rapidly to develop an accessible open-source set of tools and resources that can be applied by other communities and to other health conditions.

CODI is not limited to one-use case. In fact, the participatory process and infrastructure are being implemented in a second community, and other entities are exploring use of its open-source tools for other applications and use cases (eg, deidentified record linkage for federal government programs and across community health centers; public health surveillance of national data sets; expansion to other chronic conditions and infectious diseases such as COVID-19; expansion to adults).

## References

- Ogden CL, Fryar CD, Martin CB, et al. Trends in obesity prevalence by race and Hispanic origin—1999–2000 to 2017–2018. *JAMA*. 2020; 324(12):1208–1210.
- Hales C, Fryar CD. QuickStats: prevalence of obesity and severe obesity among person aged 2–19 years—National Health and Nutrition Examination Survey 1999–2000 through 2017–2018. *MMWR Morb Mortal Wkly Rep*. 2020;69(13):390.
- Gill N, Gjelsvik A, Mercurio LY, Amanullah S. Childhood obesity is associated with poor academic skills and coping mechanisms. *J Pediatr*. 2021;228:278–284.
- Anderson EL, Howe LD, Jones HE, Higgins JP, Lawlor DA, Fraser A. The prevalence of non-alcoholic fatty liver disease in children and adolescents: a systematic review and meta-analysis. *PLoS One*. 2015;10(10):e0140908.
- Egan KB, Ettinger AS, Bracken MB. Childhood body mass index and subsequent physician-diagnosed asthma: a systematic review and meta-analysis of prospective cohort studies. *BMC Pediatr*. 2013;13:121.
- Friedemann C, Heneghan C, Mahtani K, Thompson M, Perera R, Ward AM. Cardiovascular disease risk in healthy children and its association with body mass index: systematic review and meta-analysis. *BMJ*. 2012;345:e4759.
- Quek YH, Tam WWS, Zhang MWB, Ho RCM. Exploring the association between childhood and adolescent obesity and depression: a meta-analysis. *Obes Rev*. 2017;18(7):742–754.
- van Geel M, Vedder P, Tanilon J. Are overweight and obese youths more often bullied by their peers? A meta-analysis on the correlation between weight status and bullying. *Int J Obes (Lond)*. 2014; 38(10):1263–1267.
- Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national Patient-Centered Clinical Research Network. *J Am Med Inform Assoc*. 2014;21(4):578–582.
- Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med Care*. 2010;48(6)(suppl):S45–S51.
- Centers for Disease Control and Prevention. Childhood Obesity Data Initiative (CODI). <https://www.cdc.gov/obesity/initiatives/codi/childhood-obesity-data-initiative.html>. Accessed October 13, 2020.
- Colorado Health Institute. Colorado Health Observation Regional Data Service (CHORDS) governance plan, version 1.4. [https://www.coloradohealthinstitute.org/sites/default/files/file\\_attachments/CHORDS%20Governance%20Plan%20Version%201.4\\_August2017.pdf](https://www.coloradohealthinstitute.org/sites/default/files/file_attachments/CHORDS%20Governance%20Plan%20Version%201.4_August2017.pdf). Accessed October 13, 2020.
- Leavitt HG. Handbook of organizations. In: March JG, ed. *Handbook of Organizations*. Chicago, IL: Rand McNally; 1965:1144–1170.
- Sharma A. Collaborative product innovation: integrating elements of CPI via PLM framework. *Comput Aided Des*. 2005;37(13):1425–1434.
- King RJ, Garrett N, Kriseman J, et al. A community health record: improving health through multisector collaboration, information sharing, and technology. *Prev Chronic Dis*. 2016;13:E122.
- Public Health Informatics Institute. Business process analysis. <https://phii.org/business-process-analysis>. Accessed October 13, 2020.
- The MITRE Corporation. Tools for the Childhood Obesity Data Initiative (CODI) data owners and partners to use in record linkage. <https://github.com/mitre/data-owner-tools>. Accessed October 13, 2020.
- The MITRE Corporation. CODI data model implementation guide. <https://github.com/mitre/codi/blob/main/CODI%20Data%20Model%20Implementation%20Guide.pdf>. Accessed October 13, 2020.
- Observation Medical Outcomes Partnership (OMOP). OMOP Common Data Model. <https://www.ohdsi.org/data-standardization/the-common-data-model/>. Accessed July 31, 2021.
- Schnell R, Bachteler TJR. Python implementation of anonymous linkage using cryptographic linkage keys. GitHub. <https://github.com/data61/anonlink>. Accessed October 13, 2020.
- The MITRE Corporation. Synthetic Denver. <https://synthea.mitre.org/downloads>. Accessed October 13, 2020.
- The MITRE Corporation. Tools for data coordinating centers in the CODI project to accept and match inputs from data owners. <https://github.com/mitre/dccools>. Accessed October 13, 2020.
- Davies M, Erickson K, Wyner Z, Malenfant J, Rosen R, Brown J. Software-enabled distributed network governance: the PopMedNet experience. *EGEMS (Wash DC)*. 2016;4(2):1213.
- Daymont C, Ross ME, Russell Localio A, Fiks AG, Wasserman RC, Grundmeier RW. Automated identification of implausible values in growth data from pediatric electronic health records. *J Am Med Inform Assoc*. 2017;24(6):1080–1087.
- World Health Organization. *WHO Child Growth Standards and the Identification of Severe Acute Malnutrition in Infants and Children: A Joint Statement by the World Health Organization and the United Nations Children's Fund*. Geneva, Switzerland: World Health Organization; 2009. <https://www.ncbi.nlm.nih.gov/pubmed/24809116>. Accessed October 13, 2020.
- The MITRE Corporation. mitre/growthcleanr. GitHub. <https://github.com/mitre/growthcleanr>. Accessed October 13, 2020.
- The MITRE Corporation. Visualization tools for working with growth data processed by growthcleanr. <https://github.com/mitre/GrowthViz>. Accessed October 13, 2020.
- NORC. ZCTA3 ZIP CODE tabulation area 3 pre-processing and prevalence. <https://github.com/NORC-UChicago/CODI-PQ/tree/master/ZCTA3%20ZIP%20CODE%20TABULATION%20AREA%203%20Pre-Processing%20and%20Prevalence%20Query>. Accessed January 10, 2021.

29. NORC. County pre-processing and prevalence query. <https://github.com/NORC-UChicago/CODI-PQ/tree/master/COUNTY%20Pre-Processing%20and%20Prevalence%20Query>. Accessed January 20, 2021.
30. Bacon E, Budney G, Bondy J, et al. Developing a regional distributed data network for surveillance of chronic health conditions: the Colorado Health Observation Regional Data Service. *J Public Health Manag Pract*. 2019;25(5):498-507.
31. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci*. 2009;4:50.
32. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health*. 1999;89(9):1322-1327.
33. Nykanen P, Brender J, Talmon J, et al. Guideline for Good Evaluation Practice in Health Informatics (GEP-HI). *Int J Med Inform*. 2011; 80(12):815-827.