PCORnet Obesity Observational Study:
Short- and Long-term Effects of Antibiotics on Childhood Growth

Updated: March 2018

A. Study Design and Approach

A.1. Specific Aims

The chief **objectives** are to assess the comparative effects of different types, timing, and amount of antibiotic use in the first 2 years of life with (*Aim 1*) body mass index (BMI) and obesity at ages 5 and 10 years and (*Aim 2*) growth trajectories to age 5 years. In *Aim 3*, we will examine effect modification (~heterogeneity of treatment effect) according to several *a priori* specified socio-demographic, clinical, and maternal variables. We will account for missing data, and control for confounding, following PCORI Methodology Standards. Primary analysis will occur via distributed queries run against the Common Data Model. In addition, using individual patient-level data on the whole study population, we will evaluate the distributed data system. Further, within subgroups of the study population with appropriate data, we will explore the extent to which observed associations are confounded by maternal and perinatal characteristics, and the extent to which using dispensing instead of prescribing data for medication use alters the findings. Through this use of patient-level data we will gauge the adequacy of the aggregate-level data from distributed queries to obtain valid answers to the study questions, an important comparison that addresses Overall Goal II, to examine how well we can address questions like these in an observational study within PCORnet. In our *Secondary Aim,* we will employ focus groups of parents and in-depth interviews of clinicians to explore how best to put the epidemiologic findings into everyday practice.

Specific Aim 1: To evaluate the comparative effects of different types, timing, and amount of antibiotics used during the first two years of life on body mass index and risk of obesity at ages 5 (primary outcome age) and 10 (secondary) years.

- *Hypothesis:* There will be a "dose-response" relationship between the number of antibiotic courses given during the first 2 years of life and both higher BMI and the probability of obesity at ages 5 and 10 years. This relationship will be strongest for broad-spectrum antibiotics prescribed in the first 6 months of life.

Specific Aim 2: To assess the comparative effects of different types, timing, and amount of antibiotics used during the first two years of life on the rates and patterns of childhood growth during the first 5 years of life.

- *Hypothesis:* There will be a "dose-response" relationship between the number of antibiotic courses given during the first 2 years of life and subsequent growth trajectories of children in a pattern that increases children's risk of later overweight and obesity. This relationship will be strongest for broad-spectrum antibiotics prescribed in the first 6 months of life.

Specific Aim 3: To explore how the effects of different types, timing, and amount of antibiotics on childhood BMI, obesity risk and growth (Aims 1 and 2) vary according to patient socio-demographic, clinical, and maternal characteristics, including

1. Socio-demographic
    - Child sex
    - Child race/ethnicity
    - Geography, based on location of clinical facility
2. Clinical
    - Prescription of medications that also cause obesity, esp. corticosteroids
    - Low birth weight or macrosomia in term infants
3. Maternal (for seven node sites that have linked maternal and child records)
    - BMI (before or early pregnancy)
    - Gestational weight gain
    - Gestational diabetes
    - Maternal receipt of antibiotics during pregnancy
    - Type of delivery, i.e., Cesarean v. vaginal
    - Birth weight of child
    - Gestational age at birth
    - Maternal infections
    - Maternal smoking status

- Maternal demographic information: age (in years), ethnicity, and race

Hypotheses:

• The antibiotic effects will not vary by socio-demographic or maternal characteristics.

• Long-term corticosteroid use will potentiate the effect of antibiotics on childhood obesity.

Secondary Aim (Section C): Through focus groups and in-depth interviews, to explore how parents and other caregivers and their providers assess information related to current and future benefits and risks, particularly for treatments such as antibiotics in early childhood, which can have substantial near-term benefits along with moderate long-term risks. We will also explore how clinicians, health care organizations, and policy makers should best present study findings to help parents understand its strengths and limitations in the context of shared clinical decision-making.

A.2. Study Population for Aims 1-3.

A.2.1. Data sources and infrastructure. The source of data will be PCORnet's Clinical Data Research Networks (CDRNs). PCORnet funded 11 CDRNs during Phase I, 13 CDRNs in Phase II. Beginning in Phase I in January 2014, these CDRNs have been developing the infrastructure to "create a large, highly representative, national network for conducting clinical outcomes research."[45] The CDRNs enter their electronic clinical data in a Common Data Model that is consistent within and across CDRNs.[46] The Common Data Model ensures reliable data across a variety of data systems, thereby allowing centralized queries rather than *ad hoc* programming at each data site. It is not a centralized data repository; rather it is "virtual," consisting of parallel databases set up identically at each CDRN site to facilitate interoperability across sites. Each of the CDRNs has several subnetworks of one or more clinical institutions contributing data to the CDRN; we refer to each of these subnetworks as a node site. For this study, 10 of the 13 CDRNs will participate, with 39 node sites providing data (**Table 1**).

Table 1: Participating CDRNs and their 37 node sites

CDRN Name	Node sites	
	Number	**Names**
ADVANCE	2	OCHIN, Health Choice Network of Florida
CAPriCORN	5	Northshore UHS, University of Chicago, Loyola Medicine, Rush University Medical Center, Lurie Children's Hospital
GPC	4	University of Texas Health Science Center – San Antonio, Medical College of Wisconsin, University of Iowa Health Care, Marshfield Clinic
Mid-South	3	Vanderbilt University, Greenway Health, University of North Carolina
NYC-CDRN	2	Montefiore/Einstein, Mt. Sanai Health System
OneFlorida	3	University of Florida, Orlando Health, Tallahassee Memorial Health System
PEDSnet	8	Children's Hospital of Philadelphia, Seattle Children's Hospital, Children's Hospital Colorado, Nemours Children's Hospital System, St. Louis Children's Hospital, Cincinnati Children's Hospital, Nationwide Children's Hospital
PORTAL	6	Kaiser Permanente Colorado, Denver Health, Kaiser Permanente Washington Health Research Institute, Health Partners Research Foundation, Kaiser Permanente Mid-Atlantic, Kaiser Permanente Northwest
REACHnet	3	Tulane University, Ochsner Health System, Baylor Scott and White Health
SCILHS	2	Boston Health Net/BMC, Wake Forest Baptist Hospital

The PCORnet Common Data Model will be the source for almost all of the data for this study. Because medication data are needed to achieve our Aims, we will rely on v3.0 (**Table 2**), which requires prescription data. The majority of CDRN data are from electronic medical records rather than claims (e.g., pharmacy benefit managers), and thus, the dispensing field may be sparse for many node sites.

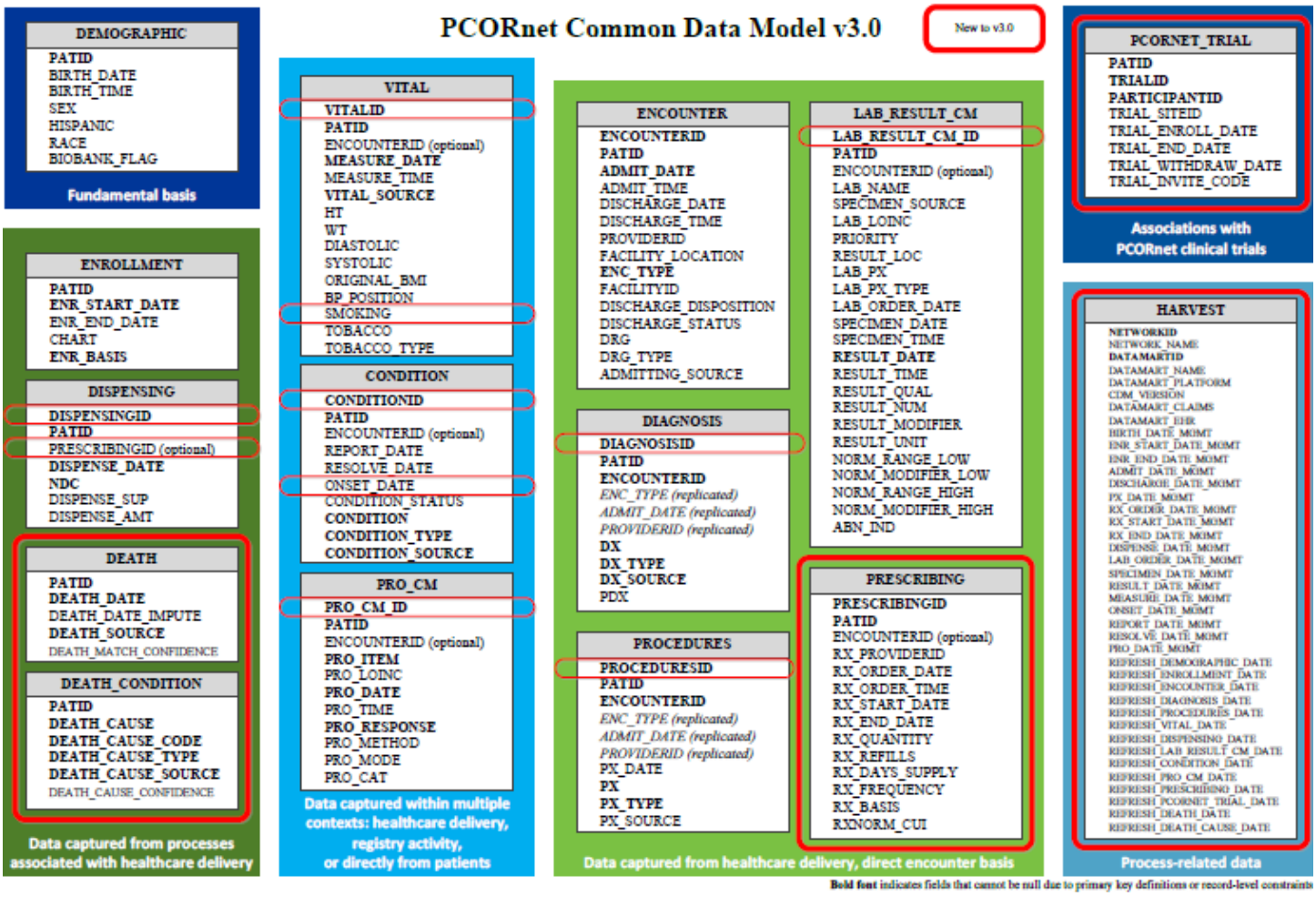
Table 2: Common Data Model (CDM) v3.0: Data tables that are applicable to the PCORnet Antibiotics Study

CDM Table	Description	Applicability for PCORnet Antibiotics Study
Demographic	Contains 1 record per patient that includes key demographic variables	Age at encounter, sex, Hispanic (y/n), and race. Eligibility criteria (age) and covariates

Encounter	Contains 1 record for each time a patient sees a provider in ambulatory setting or is hospitalized; multiple encounters per day are possible if they occur with different providers or in different care settings	For Aim 1 we focus on well-child visits for the outcome, which are coded in visit type. For Aim 2, we will include all outpatient visits.
Diagnosis	Contains all uniquely recorded diagnoses for all encounters. Each diagnosis is associated with a specific patient and encounter.	We will use diagnosis codes and associated encounter dates to establish eligibility criteria and covariates.
Procedure	Contains all uniquely recorded procedures for all encounters. Each procedure is associated with a specific patient and encounter.	N/A
Vital	Contains one record per height or weight result/entry. Multiple measurements per encounter are recorded as separate measures.	Part of eligibility criteria and main outcomes: Height (length for 0-23 months assumed under this variable name) and weight for calculation of BMI.
Prescribing	Contains RxNORM code for each medication prescribed, date of prescription	For main exposure variables (antibiotics), plus other medication as covariates.
Dispensing	Contains NDC code for each medication dispensed, date dispensed.	Available for a subset of node sites.

The PCORnet Common Data Model will be the source for almost all of the data for this study. Because we need medication data to achieve our Aims, we will rely on Common Data Model v3.0 (**Figure 1**), which requires prescription data. While v2.0 has a field for dispensed medications, the majority of CDRN data are from electronic medical records rather than claims (e.g., pharmacy benefit managers), and thus the dispensing field may be sparse for many node sites; this is why we use dispensing as a sensitivity analysis.

Figure 1: Common Data Model v3.0; updates from v2.0 in red



In addition to data retrieved from the CDM, analyses including maternal factors will utilize additional data obtained by participating CDRN node sites that have previously been able to link maternal and child records. These data will be extracted and de-identified by each node site individually, and the de-identified data aggregated with CDM-derived data for that child to create analytic datasets.

While the CDM supports semantic interoperability by creating a common structure for clinical data, the specification provides wide latitude in the use of terminologies for specific facts. The intent is to permit sites to record information in the terminology used for clinical operations (e.g. diagnoses as ICD9-CM, ICD10-CM, or SNOMED-CT). As a result, we expect there to be semantic variability for the data of interest across node sites participating in the study. For simplicity, we refer in this protocol to a single terminology for each data type. In practice, the data characterization phase of the study will include an assessment of terminologies and coding practices in use across participating node sites. The Methods Core will use the results of this assessment in their design of study analyses, to insure that analyses appropriately account for different coding systems in use at node sites. These results will also inform the Coordinating Center's development of resources to facilitate this task in future studies.

Data Network Infrastructure

PopMedNet™ is the query management platform for the PCORnet Distributed Research Network. In PCORnet's distributed data environment, code is developed centrally and distributed to each partner to execute against data that are stored in a common format. The Coordinating Center, in collaboration with the Study Team, will undertake several steps

to examine query functionality and data quality prior to internal testing, piloting with 1 CDRN (PEDSnet) and sending full queries to the CDRNs and datamarts. The study team will send a detailed set of functional and technical requirements to the CC lead programmer. The programmer, working in conjunction with study programmer will write the code. That code will then undergo a series of tests, including an initial test by an outside vendor, who will determine whether the code can run across software programs (multiple versions of SAS) and operating systems (Unix, Windows, Linux). After feedback from the vendor, the code will be further refined and then tested on the PEDSnet CDRN. Further refinement will precede query distribution to the CDRNs via PopMedNet. The CDRNs execute these programs against their own data, and return the results via PopMedNet to the Coordinating Center, whence it is passed on to the study team.

The CDM includes some of the 18 elements that define PHI under the HIPAA Safe Harbor guidance, including encounter dates and date of birth. The necessary "cross-walks" between other identifiers included in the CDM and their originating data are not included in the scope of the CDM, but are expected to be maintained by each CDRN. This study will not require that sites transmit any of the Safe Harbor identifying information to the CC or the study team.

A.2.2 Identifying patients who meet the eligibility criteria

To allow wide generalization of results, we specify relatively broad eligibility criteria. Conceptually, because early infancy may be a critical period for the obesogenic effect of antibiotic use, we are interested in following an inception cohort of children from birth until at least age 5 years (and secondarily to 10 years). In practice, entry into the cohort may occur after birth because children represented in many CDRN node sites deliver in non-CDRN hospitals and enter primary care at several weeks or months of age. To be eligible, patients must have entered care at the node site before 6 months of age, and adhere to the following criteria:

Inclusion criteria:

1. Data contained in PCORnet Common Data Model 3.0, including prescribed medications in RxNorm coding.
2. $\geq$ 1 encounter with length and weight measured in each of the following age intervals: 0-5 m, 6-11 m, 12-23 m, and
3. $\geq$ 1 encounter with height and weight measured in either or both of the following age intervals: 4.0 to 5.9 y ("age 5 years"), 9.0 to 10.9 y ("age 10 years"), or eligible to be followed to these ages for use in multiple imputation to account for missing data (Section B.4).

Exclusion criteria:

1. Obesity with identified secondary cause (see below for ICD-10 codes and diagnoses)
2. Other clinical conditions that substantially alter growth (see below for ICD-10 codes and diagnoses)
3. Biologically implausible length/height or weight measurements (Section B.4.5)

Members of the study team have used similar approaches in the past to identify participants in repeated cross-sectional and longitudinal studies that employ large numbers of length/height and weight measures from electronic medical records. These include HPHCI studies of trends in obesity in young children, how crossing percentile lines on growth charts in the first 2 years of life predicts obesity at age 5 years, and the CHOP/PEDSnet study of antibiotics in the first 2 years of life and obesity at ages 3-5.[47-51] We will use these approaches to identify and characterize participating children.

ICD-10 Codes and Diagnoses for Exclusion Criteria

a. Obesity with identified secondary cause, including 2 codes on separate visits (to minimize rule outs) for Rare genetic syndromes

Q87.1 Prader-Willi (Congenital malformation syndromes primarily involving short stature)
Q87.3 Beckwith-Wiedemann (Congenital malformation syndromes involving early overgrowth)
Q87.89 Biedl-Bardet (Other specified congenital malformation syndromes)
Q87.8 Cohen, Alstrom
Q78.0 Carpenter
Q98.4 Klinefelter
Q96.8 Turner

Endocrine disorders
E23 Hypofunction and other disorders of the pituitary gland (includes disorders of hypothalamus)
E24 Cushing syndrome
E03 Hypothyroidism

b. Other disorders affecting growth

E84 Cystic fibrosis
E30.1 Precocious puberty

P07.2 Extreme immaturity of newborn (< 28 weeks)
P07.0 Extremely low birth weight newborn (< 1 kg)

c. For sensitivity analyses:

P07.3 Preterm [premature] newborn [other] (<37 weeks, can also code by specific weeks)
P07.1 Other low birth weight newborn (1000-2499 g, can also code by 250 g increments)

P08.0 Exceptionally large newborn baby (>4500 g)
P08.1 Other heavy for gestational age newborn (4000-4499 g)

A.2.3. Study sample: size, characteristics

In our prep-to-research survey, we asked each participating CDRN to provide estimates of the number of children who would fulfill the inclusion criteria, with the caveats that we did not ask them to subdivide the 1st 2 years of life into 3 groups, nor did we ask them to apply exclusion criteria. Because most CDRNs are still putting the Common Data Model into place, many of the data came from pre-existing virtual data warehouses or more direct queries of electronic records at the node sites. They ran these initial queries using SAS® (version unspecified) or SQL. However, participation in Phase II of the CDRN infrastructure development requires CDRNs to be able to implement queries via SAS version 9.3 or higher. Thus, we anticipate completing our entire query work with SAS v9.3 or higher.

Table 3. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from 38 of 42 node sites of 9 participating CDRNs. More detail in Protocol.

	A	B	C	D
Years	*Total 0-17.9 yr*	*Subset of A 0-23 mo*	*Subset of B 4.0-5.9-yr follow-up*	*Subset of B 9.0-10.9-yr follow-up*
2005-2008	994231	311822	180645	16877
2009-2013 (observed)	2427940	1316840	131044	1
2009-2015 (projected)			447505	
Total (observed)	3422171	1628662	311689	16878
Total (projected)			**628150**	

We asked CDRNs to report on data availability in 2005-2008 and then each year from 2009 through 2013. Based on these data, provided by 38 node sites at the 9 of the 10 participating CDRNs, the years in which the CDRNs have (and will have) data in Common Data Model format, and our projections of additional follow-up data available as of 2015, we estimate that relevant data will be available among over 1.6 million children 0 to 23 months of age, among whom we project more than 600,000 will have an observed encounter with measures at 4.0 to 5.9 years of age (**Table 3**). The observed ~17,000 at 9.0 to 10.9 years of age is an underestimate due to lack of follow-up after 2013, the last year of the survey; we are unable to project an estimate. The CDRNs report that approximately 90% of the 0-23-month-old children have data on race/ethnicity. While we do not yet have data on the breakdown by race/ethnicity, we expect to have substantial diversity because of the wide geographic span of node sites, including some that care for underserved and

uninsured patients.

Assumptions for Projected Sample Size

To calculate our projected sample size for this study, we used prep-to-research data provided by 38 of the original 43 node sites, as shown in Table 3 of the proposal (above) and Expanded Table 3 for each node site (below). We separated counts from two time periods, 2005 to 2008 and 2009 to 2013. Children in the earlier time period should have an *observed* height and weight between 4 and 5.9 years of age, if in fact they were followed up in the system (actual observed 180,645 children, 58%). For the 2009 to 2013 time period, 131,044 (10%) children had an *observed* height and weight at 4 to 5.9 years by the end of 2013, but many more will be eligible for this measurement by the end of 2015, the time period available for our study. Thus, we calculated a *projected* follow-up for these children. We conservatively estimated that 70% of children in this time period would be 4 to 5.9 by the end of 2015 and that 50% of those eligible would have follow-up (447,505, 34%). This led us to a *projected* overall follow-up at 4 to 5.9 years of 628,150. Power for the main comparisons in Aim 1 and Aim 3 is based on this number, of whom about 69% will have received an antibiotic prescription in the first 2 years of life, assuming rates similar to those in the CHOP/PEDSnet study.

Below, in expanded tables, we show the numbers of children contributed by each of the 38 node sites (or data mart for CDRNs who report overall numbers). While some entries are depicted as "not in CDM," these sites intend to populate the Common Data Model for those years, and they provided data outside the CDM for our projected sample size. We used these values to estimate the entries in Table 3 (above and in Section B.2.3), and thus for our power calculations.

These prep-to-research data show a very large sample size. Along with preliminary data from the PEDSnet CDRN on proportions of children who receive antibiotics of different types and at different ages, adequate statistical power for main and subgroup effects is ensured (Section B.5). However, the data also show large loss to follow-up, especially at 10 years, which is expected in a clinical system in which children may move or change care delivery system and because some CDRNs do not have Common Data Models before 2009 or 2010. Nevertheless, we are interested in exploring the extent to which follow-up data until age 10 as well as age 5 reveals insights into both data characterization and longer-term effects of antibiotics on excess weight gain. We address missing data in Section B.4.4, including use of multiple imputation for outcomes as well as covariates.

Observed Numbers of Children from Each Node Site
(CDRNs with single data marts reported over all node sites)

ADVANCE

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from ADVANCE; Oregon Community Health Information Network, Fenway Health, Health Choice Network

	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	127611	8365	7361	330
2009-2013	149803	255164	15937	1
Total (observed)	277414	263529	23298	331

CAPriCORN

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from CAPriCORN; The Alliance of Chicago

Community Health Services, Loyola University Health System, NorthShore University Health System, Northwestern Medicine, Rush University Medical Center, The University of Chicago Medicine, University of Illinois Hospital & Health Sciences System, Edward Hines, Jr.				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	not in CDM	not in CDM	not in CDM	not in CDM
2009-2013	94015	87448	330	0
Total (observed)	94015	87448	330	0

GPC

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from GPC; University of Iowa				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	20708	6929	2416	112
2009-2013	17262	13693	602	0
Total (observed)	37970	20622	3018	112

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from GPC; UT NW				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	not in CDM	not in CDM	not in CDM	not in CDM
2009-2013	746	710	not in CDM	not in CDM
Total (observed)	746	710	0	0

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from GPC; San Antonio, Medical College of Wisconsin, Mercy-Kansas City				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	304293	72743	23539	6112
2009-2013	1206158	136599	20778	0
Total (observed)	1510451	209342	44317	6112

Mid South

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from Mid- South: Vanderbilt, Vanderbilt Health Affiliated Network, Greenway				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	53313	33772	16437	2385
2009-2013	30828	28673	374	0
Total (observed)	84141	62445	16811	2385

NYC

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from NYC-CDRN; Mt Sinai				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	not in CDM	not in CDM	not in CDM	not in CDM
2009-2013	31529	26191	2436	0
Total (observed)	31529	26191	2436	0

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from SCILHS; Montefiore Einstein				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	not in CDM	not in CDM	not in CDM	not in CDM
2009-2013	15372	45260	8158	0
Total (observed)	15372	45260	8158	0

PEDSnet

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from PEDSnet; Boston Children's Hospital Cincinnati Children's Hospital Medical Center, Children's Hospital of Philadelphia (CHOP), Colorado Children's Hospital, Nationwide Children's Hospital, Nemours Health System, Seattle Children's Hospital, St. Louis Children's Hospital				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	Not in CDM	Not in CDM	Not in CDM	Not in CDM
2009-2013	585606	455203	54711	0
Total (observed)	585606	455203	54711	0

PORTAL

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from PORTAL; KPCO				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	not in CDM	not in CDM	not in CDM	not in CDM
2009-2013	38140	33511	4364	0
Total (observed)	38140	33511	4364	0

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from PORTAL; Denver Health				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	not in CDM	not in CDM	not in CDM	not in CDM
2009-2013	24466	21315	4702	0
Total (observed)	24466	21315	4702	0

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from PORTAL; KPWHRI				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	not in CDM	not in CDM	not in CDM	not in CDM
2009-2013	28135	25579	2979	0
Total (observed)	28135	25579	2979	0

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from PORTAL; Health Partner				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	not in CDM	not in CDM	not in CDM	not in CDM
2009-2013	16310	15196	1087	0
Total (observed)	16310	15196	1087	0

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from PORTAL; KP MAS				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	not in CDM	not in CDM	not in CDM	not in CDM
2009-2013	34113	28338	5394	0
Total (observed)	34113	28338	5394	0

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from PORTAL; KP NW				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	32207	8108	5455	0
2009-2013	35888	31214	5905	0
Total (observed)	68095	39322	11360	0

REACHnet

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from REACHnet (formerly La-CDRN); LSU, Ochsner, PATH, Tulane				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	not in CDM	not in CDM	not in CDM	not in CDM
2009-2013	32424	25298	1120	0
Total (observed)	32424	25298	1120	0

SCILHS

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from SCILHS; BMC				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	69125	15483	11345	4726
2009-2013	97820	19396	9656	0
Total (observed)	166945	34879	21001	4726

Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from SCILHS; Wake Forest				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	24542	4557	1154	23
2009-2013	31827	20398	569	0
Total (observed)	56369	24955	1723	23

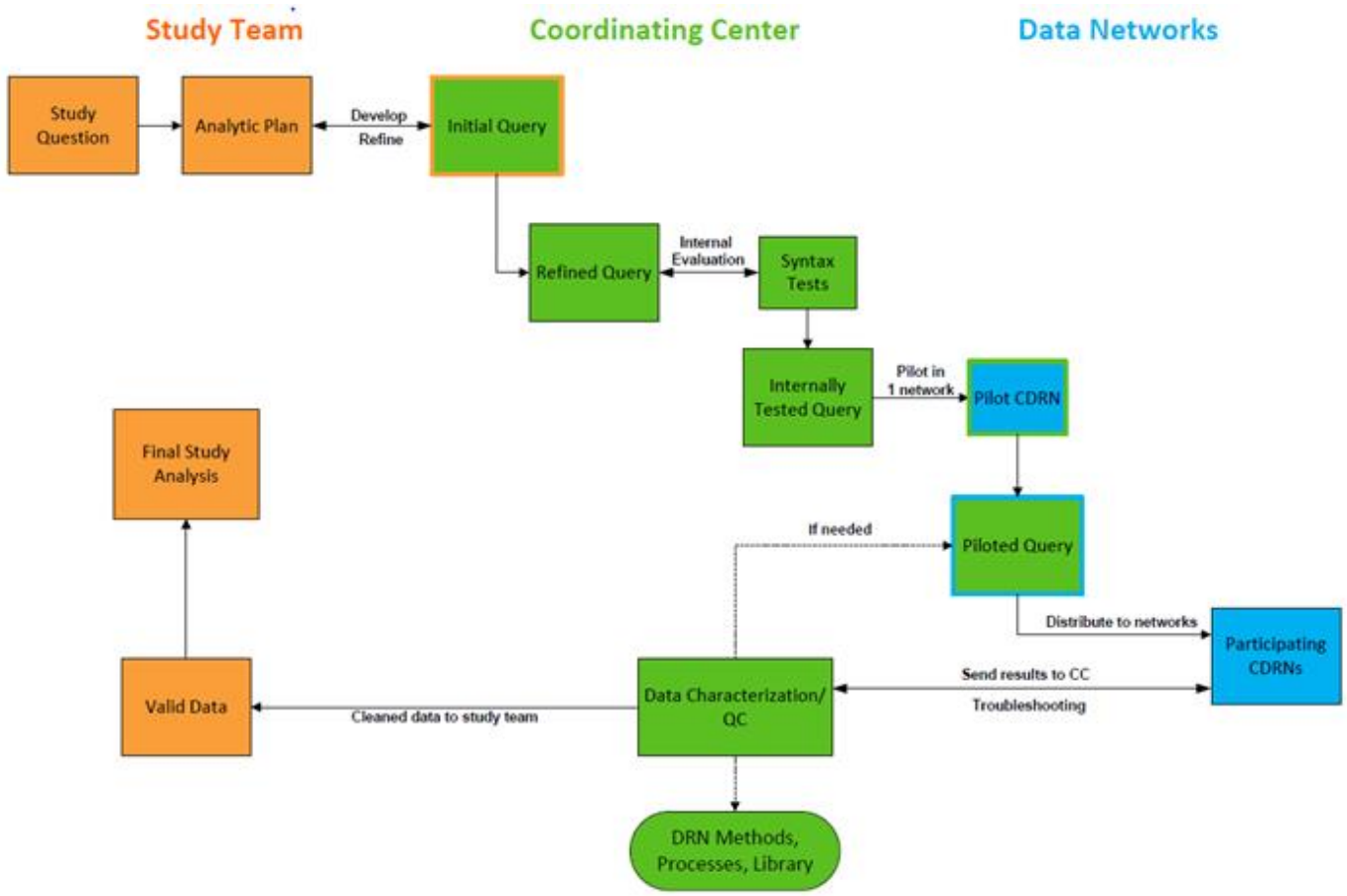
Expanded Table 3 for each node site. Numbers of children with encounters that include height/length and weight, according to age. Prep-to-research data from SCILHS; Partners				
	A	B	C	D
Years	Total 0-17.9 yr	Subset of A 0-23 mo	Subset of B 4.0-5.9-yr follow-up	Subset of B9.0-10.9-yr follow-up
2005-2008	not in CDM	not in CDM	not in CDM	not in CDM
2009-2013	21558	20259	0	0
Total (observed)	21558	20259	0	0

A.3. Study design and methodology

A.3.1. Using the PCORnet CDM infrastructure for distributed research queries. PCORnet has set up a distributed research model in which a main operating principle is to "send questions to the data."[52, 53] Under this principle, the PCORnet Coordinating Center sends honed programming code ("queries") to each CDRN node site, who run them, unaltered, against the Common Data Model. The study team combines the aggregate-level data from the multiple sites for analysis. Advantages over traditional pooled analysis are that data are kept in the hands of the original data holders, which decreases proprietary concerns and data breaches. Data holders transfer only minimum necessary data to a central site for analysis. In addition, data holders know their data well, facilitating cleaning and troubleshooting. In the proposed study, we will adhere to these principles while at the same time comparing findings from a fully distributed programming approach that yields only aggregate data with an approach that employs (de-identified) individual-level data. We propose a data flow system that takes advantage of, and integrates, the knowledge and expertise of the 3 main components of the study: The Study Team, the PCORnet Coordinating Center (CC), and the Data Networks (the CDRNs) (**Figure 2**). In summary, the Study Team will refine the research questions and develop a detailed analytic plan. Working with the Study Team and CC scientists, a programmer at the CC experienced in distributed programming will develop and refine initial queries. The CC will send these queries to a third party to test them across a range of statistical software syntaxes. In the next step, the CC will send the tested queries to a pilot CDRN to identify further issues. In the case of this study, the pilot CDRN will be PEDSnet. Upon remedying any problems or errors, the CC will distribute the queries to the full set of Data Networks, who will return the aggregated data to the CC for characterization and quality control. Given that this is the first PCORnet experience with this process, we anticipate the need for troubleshooting with node sites before the CC can provide the Study Team finalized analytic datasets for statistical analyses. These final analytic datasets will include fully aggregated results, as well as some individual patient-level information to 1.) address Aim 2 on growth trajectories, for which distributed programs will not available, and 2.) compare results from the two approaches—fully aggregated and individual-level—for Aim 1. In our prep-to-research survey, all node sites expressed willingness to share de-identified patient-level data with appropriate safeguards.

A.3.2. Expected data quality and missing data concerns [IR-1, IR-2]. As outlined above and depicted in Figure 1, the close collaboration between the CC, Study Team, and CDRN investigators and analysts will allow for rapid diagnosis and resolution of data quality problems that arise during the course of the study, primarily before distribution of queries but also in the troubleshooting phase after CDRNs produce initial results. Based on our preliminary data and on conversations with the CC and each of the CDRNs, we anticipate several potential challenges to data quality and/or completeness including: 1), defining and accounting for loss to follow-up in a setting of medical records rather than insurance enrollment, 2) linking to maternal records for potential confounding factors, and 3) using dispensed instead of prescribed medications, and 4.) additional threats to validity in interpretation of findings. We have a clear plan for addressing each of these challenges (Sections B.4.3 and B.4.4).

Figure 2. Distributed research network data flow across the 3 main components of the study



A.3.3. Data Characterization Process [IR-1]. This study will be one of the first within PCORnet, which comprises a federated group of regional networks, each with its own mix of data provenance and transformation. Although the PCORnet Common Data Model creates a shared structure and overall semantics for data elements, the semantic interoperability of the CDRNs has not yet been tested at a large scale. Further, even positing ideal agreement on intended semantics, we expect that data from different sources will have variations in operating characteristics, reflecting differences in data capture, informatics resources, and clinical practices. Effective formulation of the main study analyses will therefore require both detailed characterization of the available data and preliminary examination of potential correlations to guide the main analyses. Prior to addressing our study questions, we will conduct an in-depth data characterization process that will include both descriptive and cross-source correlative ("benchmarking") statistics.

<u>All Aims will require data available in Common Data Model (CDM) v3.0</u>
Classification of exposure will require access to prescribed or dispensed medication data. Outcome assessment will require demographic (age, sex) and anthropometric (weight, height/length) data. Covariates will be available from Procedures, Demographics, Encounters, and Diagnoses tables. The availability of information from the Common Data Model will allow us to control for covariates in Aims 1 and 2 and also to stratify by these different covariates, as we specify in Aim 3.

<u>Limited Need for Non-CDM 3.0 data</u>
Maternal Linkage Variables. Childhood BMI is highly associated with maternal BMI, so having access to data on pre- or early pregnancy maternal BMI will be important for controlling confounders in this study. We also believe that data on maternal use of antibiotics during pregnancy will be useful, as well as cesarean v. vaginal delivery, both of which are related to changes in microbiota. Birth weight and gestational age at birth are also typically available and may be effect modifiers. Based on prep-to-research survey responses and discussions with CDRNs, maternal-child linkages are already done in 7 node sites (5 sites in PORTAL, ADVANCE's OCHIN network and the Vanderbilt site of Mid-South CDRN).

We will do subset analyses involving these covariates in these 7 node sites. As described above, these data will be extracted by sites outside the CDM and linked to CDM data in the creation of de-identified analytic datasets.

Data characterization will be necessary to construct robust analyses

The initial queries to CDRNs will address data characterization and cohort definition for Aims 1-3. We will examine data quality and missing data at this point, (cf proposal Section B.4.4). Only after this data is examined and viewed satisfactory will development further queries be undertaken. This data characterization will include both descriptive and cross-source correlative ("benchmarking") statistics:

- *Continuous variables*: This includes variables such as age, time of entry into cohort, and anthropometric measurements. Characterization will include standard descriptive statistics (range, mean, median, SD, and IQR) as well as tabulation of out-of-range, implausible, and missing values.
- *Categorical values, small value sets*: This includes variables such as sex and race/ethnicity. Characterization will include tabulation across the entire value set, and of illegal and missing values.
- *Categorical values, large value sets*: This includes variables such as coded diagnoses and drug utilization. Characterization will include tabulation of the most common 100 values from each source, and of illegal and missing values.
- *Trends*: Dates of medical service will be used to assess temporal trends. Characterization will include number of visits per month (overall and by visit type) and number of prescriptions per month.

In each case, results will be compared to *a priori* expected distributions and to current literature addressing practice in areas related to study questions. In addition, they will be used for "benchmarking" across CDRNs to detect differences in data characteristics that may influence study analyses; these comparisons will be particularly valuable in assessing usage of large codesets such as diagnostic terminologies or drug formularies.

1. Demographic and Clinical Baseline Characteristics
    a. Variables

Variable	Provenance
Age	CDM
Sex	CDM
Race	CDM
Ethnicity	CDM
Visits 0-5 mo	CDM
Visits 6-11 mo	CDM
Visits 12-23 mo	CDM
Maternal prepregnancy BMI	Non-CDM
Maternal gestational antibiotic exposure	Non-CDM
Maternal GDM	Non-CDM
Delivery type	Non-CDM

    b. Methods
        i. Summary statistics
        ii. Cross-site benchmarking
        iii. National samples
        iv. Benchmarking against EPSDT/AAP
        v. Tabulation of maternal variables
    c. Results
        i. Population description
        ii. Estimates of potential outlier sites/stratification
        iii. Demographic inclusion/exclusion criteria
2. Exposure
    a. Variables

Variable	Provenance

Antibiotics prescribed	CDM
Antibiotics	CDM

b. Methods
   i. Cross-site benchmarking
   ii. Drug categorization (ATC)
   iii. Episode construction

c. Results
   i. Medication usage description
   ii. Estimates of potential outlier sites/stratification
   iii. Exposure category definitions

3. Outcomes
   a. Variables

Variable	Provenance
Height	CDM
Weight	CDM
Measured BMI	CDM

   b. Methods
      i. Height imputation
      ii. BMI computation
      iii. Normalization (NHANES > WHO)
      iv. Cross-site benchmarking
      v. Benchmarking to historical NHANES data and published reports
      vi. Duration of follow-up distribution
   c. Results
      i. Estimates of potential outlier sites/stratification
      ii. Outcome category definitions

4. Longitudinal covariates (assessed through end of exposure period)
   a. Variables

Variable	Provenance
Outpatient visits	CDM
Inpatient admissions	CDM
Non-antibiotic medications prescribed	CDM
Non-antibiotic medications dispensed	CDM
Diagnoses	CDM

   b. Methods
      i. Diagnosis categorization (SNOMED-CT)
      ii. Cross-site benchmarking
      iii. Duration of follow-up distribution
   c. Results
      i. Estimates of potential outlier sites/stratification
      ii. Outcome category definitions

<u>Unique challenges might be present for medications</u>

The core of this study require access to medication data. Because of different naming conventions and different coding process for medications, we anticipate particular challenges in this domain. Part of the function of CC will be to work with CDRNs to refine their medication data within the CDM.

This study will primarily capture antibiotic prescriptions, but we also anticipate need for medications such as a

corticosteroids, which also might influence weight gain, and anti-reflux medications, which we presume will not. The length of treatment with antibiotics should be reasonably straightforward to manage because these prescriptions are short-term and typically for less than one month. Chronic medications such as reflux medications and possibly steroids will present additional challenges. These medications may be prescribed for up to three months per prescription with a year's worth of refills, as required by insurance company and pharmaceutical benefits managers. We may not be able to determine if a patient stopped a medication during this year-long period. Therefore, these non-antibiotic medications may categorized as any v. none, or using a simplified multilevel categorization that distinguishes between identifiable short term exposures (e.g. for asthma) and all longer-term exposures as a group.

Prescribed vs. Dispensed Medications

When a provider writes for a prescription medication, there is no guarantee that the prescription will be filled at a pharmacy or that a patient will take it. In this study, we anticipate primarily having data on whether a medication was prescribed. For those CDRNs which have both prescribed and dispensed medication data – 4 CDRNs (PEDSnet, ADVANCE, PORTAL, Medical College of Wisconsin of GPC)– we will do sensitivity analyses to compare the approaches, thus providing validation data for prescribing information.

As part of PCORnet's Health Plan Linkage Demonstration project, we will also examine validation for prescribing information using health plan claims data from HealthCore/Anthem and Humana. HealthCore and Humana will each work with one CDRN (PEDSnet and REACHnet respectively) to link health plan claims data to CDRN HER data using individual linkage methods. The CDRNs will then incorporate the linked claims data into their CDM providing a dataset which the study team will be able to query. As with other study queries, query responses will contain de-identified individual level data without HIPAA identifiers. Ages are reported in whole months, and multiple events of a given type within a month are sequenced as intervals.

CDM 3.0 will include medication prescribing data in a standardized terminology called RxNorm

Institutions use different naming conventions for medications. To ensure that data across node sites and CDRNs are similar, the PCORnet DSSNI Task Force has called for the standardized use of RxNorm in CDM3.0.
Based on 15 source terminologies, including the FDA's National Drug Code directory, the National Library of Medicine (NLM) (http://www.nlm.nih.gov/research/umls/RxNorm/overview.html) created this system for the precise purpose of interoperability and consistency across systems, and the system covers all prescription and many over-the-counter medications available in the United States. The system provides coding at the ingredient, strength, and drug form, and brand name levels. For example, an oral capsule of azithromycin 250 mg has a RxNorm concept identifier (RXCUI) of 141962, which maps to a standardized name of "Azithromycin 250 MG Oral Capsule".

For those node sites that have dispensing data available, CDM 3.0 calls for the use of FDA National Drug Classification terminology. NDC codes can be mapped to RxNorm, allowing for comparison. In some cases, as mentioned earlier, dispensing data may be more helpful. For typically short term prescriptions, like antibiotics, the prescribing data should be sufficient to calculate an end date. Where dispensing data are used the quantity of drug ("RX_QUANTITY" IN CDM 3.0) and the days supply ("RX_DAYS_SUPPLY") should enable recreating this information.

For node sites and CDRNs that are not able to map their medications to RxNorm, source prescribing data should still be available in CDM 3.0. In these cases, depending on the local coding used, a centralized mechanism for transforming medications to RxNorm could allow for use of these data.

Although a small number of active ingredients account for the majority of antibiotic exposures in young children, a large number of formulations exist for these, and less common exposures are also likely to be biologically relevant. We will therefore take as complete an approach to capturing antibiotic exposure as is feasible. Following are selected systemic antibiotics identified in a prior study as available for use in children, with summary counts NDC and RxNorm codes seen in the local electronic medical record implementation for each:

Product Name	NDC	RxNorm
Alatrofloxacin Mesylate IV Conc 5 MG/ML (Base Equiv)	4	1

Product Name	NDC	RxNorm
Amikacin Sulfate Inj 1 GM/4ML (250 MG/ML)	1	1
Amikacin Sulfate Inj 62.5 MG/ML	1	1
Amoxicillin		2
Amoxicillin (Trihydrate) Cap 250 MG	171	5
Amoxicillin (Trihydrate) Cap 500 MG	349	4
Amoxicillin (Trihydrate) Chew Tab 125 MG	19	2
Amoxicillin (Trihydrate) Chew Tab 200 MG	1	2
Amoxicillin (Trihydrate) Chew Tab 250 MG	95	3
Amoxicillin (Trihydrate) Chew Tab 400 MG	2	2
Amoxicillin (Trihydrate) For Susp 125 MG/5ML	58	2
Amoxicillin (Trihydrate) For Susp 200 MG/5ML	33	1
Amoxicillin (Trihydrate) For Susp 250 MG/5ML	54	2
Amoxicillin (Trihydrate) For Susp 400 MG/5ML	48	2
Amoxicillin (Trihydrate) For Susp 50 MG/ML	1	1
Amoxicillin (Trihydrate) Tab 500 MG	9	1
Amoxicillin (Trihydrate) Tab 875 MG	85	1
Amoxicillin (Trihydrate) Tab SR 24HR 775 MG	4	1
Amoxicillin & K Clavulanate Chew Tab 125-31.25 MG	3	2
Amoxicillin & K Clavulanate Chew Tab 200-28.5 MG	7	2
Amoxicillin & K Clavulanate Chew Tab 250-62.5 MG		2
Amoxicillin & K Clavulanate Chew Tab 400-57 MG	8	2
Amoxicillin & K Clavulanate For Susp 125-31.25 MG/5ML	7	2
Amoxicillin & K Clavulanate For Susp 200-28.5 MG/5ML	29	3
Amoxicillin & K Clavulanate For Susp 250-62.5 MG/5ML	20	2
Amoxicillin & K Clavulanate For Susp 400-57 MG/5ML	33	2
Amoxicillin & K Clavulanate For Susp 600-42.9 MG/5ML	37	2
Amoxicillin & K Clavulanate Tab 250-125 MG	6	2
Amoxicillin & K Clavulanate Tab 500-125 MG	109	2
Amoxicillin & K Clavulanate Tab 875-125 MG	127	2
Amoxicillin & K Clavulanate Tab SR 12HR 1000-62.5 MG	16	2
AMOXICILLIN 200 MG OR TBSO		1
Amoxicillin-Pot Clavulanate		3
Ampicillin & Sulbactam Sodium For IV Soln 1-0.5 GM	4	6
Ampicillin & Sulbactam Sodium For IV Soln 10-5 GM	13	3
Ampicillin & Sulbactam Sodium For IV Soln 2-1 GM	4	6
Ampicillin Cap 250 MG	66	5
Ampicillin Cap 500 MG	71	3
Ampicillin For Susp 125 MG/5ML	6	1
Ampicillin For Susp 250 MG/5ML	4	1
AMPICILLIN SODIUM 500 MG IV SOLR	1	1
Ampicillin Sodium For IV Soln 1 GM	6	4
Ampicillin Sodium For IV Soln 10 GM	5	1

Product Name	NDC	RxNorm
Ampicillin Sodium For IV Soln 2 GM	5	4
Atovaquone Susp 750 MG/5ML	3	1
Azithromycin		8
Azithromycin Cap 250 MG	12	4
Azithromycin Extended Release For Oral Susp 2 GM	2	2
Azithromycin For Susp 100 MG/5ML	22	2
Azithromycin For Susp 200 MG/5ML	62	2
Azithromycin Hydrogencitrate IV For Soln 2.5 GM (Base Equiv)	1	2
Azithromycin Hydrogencitrate IV For Soln 500 MG (Base Equiv)	1	2
Azithromycin IV For Soln 500 MG	16	4
Azithromycin Powd Pack for Susp 1 GM	13	4
Azithromycin Tab 250 MG	157	12
Azithromycin Tab 500 MG	76	9
Azithromycin Tab 600 MG	14	2
Aztreonam in Dextrose Inj 1 GM/50 ML	1	1
Aztreonam in Dextrose Inj 2 GM/50 ML	1	1
Bacitracin Intramuscular For Soln 50000 Unit	15	2
BIAXIN 187.5 MG/5ML OR SUSR	1	1
BICILLIN L-A 300000 UNIT/ML IM SUSP	2	2
CECLOR 187 MG/5ML OR SUSR	2	1
CECLOR CD 375 MG OR TB12	1	1
CEFACLOR 187 MG/5ML OR SUSR	11	1
Cefaclor Cap 250 MG	67	2
Cefaclor Cap 500 MG	58	2
CEFACLOR CR 375 MG OR TB12	1	1
Cefaclor For Susp 125 MG/5ML	6	2
Cefaclor For Susp 250 MG/5ML	10	2
Cefaclor For Susp 375 MG/5ML	5	2
Cefaclor Monohydrate Tab SR 12HR 500 MG	5	2
Cefadroxil Cap 500 MG	186	4
Cefadroxil For Susp 125 MG/5ML	3	1
Cefadroxil For Susp 250 MG/5ML	18	2
Cefadroxil For Susp 500 MG/5ML	18	2
Cefadroxil Tab 1 GM	9	4
Cefazolin in D5W Inj 1 GM/50ML	1	1
Cefazolin Sodium For Inj 1 GM	1	1
Cefazolin Sodium For IV Soln 1 GM	1	1
Cefazolin Sodium for IV Soln 1 GM and Dextrose 4%	1	1
Cefazolin Sodium for IV Soln 2 GM and Dextrose 3%	1	1
Cefdinir		4
Cefdinir Cap 300 MG	56	4
Cefdinir For Susp 125 MG/5ML	24	2

Product Name	NDC	RxNorm
Cefdinir For Susp 250 MG/5ML	32	2
Cefditoren Pivoxil Tab 200 MG (Base Equivalent)	3	2
Cefditoren Pivoxil Tab 400 MG (Base Equivalent)	6	2
Cefepime HCl and Dextrose 5% For IV Soln 1 GM/50ML	1	1
Cefepime HCl and Dextrose 5% For IV Soln 2 GM/50ML	1	1
Cefepime HCl For IV Soln 1 GM	1	2
Cefepime HCl For IV Soln 2 GM	1	3
Cefepime HCl IV Soln 1 GM/50ML	1	1
Cefepime HCl IV Soln 2 GM/100ML	1	1
Cefixime		1
Cefixime Cap 400 MG	3	1
Cefixime Chew Tab 100 MG	1	1
Cefixime Chew Tab 200 MG	1	1
Cefixime For Susp 100 MG/5ML	4	2
Cefixime For Susp 200 MG/5ML	2	2
Cefixime For Susp 500 MG/5ML	2	1
Cefixime Tab 400 MG	3	2
CEFOBID 2 GM IV SOLR	1	1
Cefotaxime Sodium For IV Soln 1 GM	5	1
Cefotaxime Sodium For IV Soln 2 GM	4	1
Cefotaxime Sodium in D5W IV Soln 1 GM/50ML	1	1
Cefotaxime Sodium in D5W IV Soln 2 GM/50ML	1	1
Cefotetan Disodium For IV Soln 1 GM	1	2
Cefotetan Disodium for IV Soln 1 GM and Dextrose 3.58%	1	1
Cefotetan Disodium For IV Soln 2 GM	1	2
Cefotetan Disodium for IV Soln 2 GM and Dextrose 2.08%	1	1
Cefoxitin Sodium For IV Soln 1 GM	7	1
Cefoxitin Sodium For IV Soln 2 GM	7	2
Cefoxitin Sodium IV For Soln 1 GM and Dextrose 4%	1	1
Cefoxitin Sodium IV For Soln 2 GM and Dextrose 2.2%	1	3
Cefoxitin Sodium IV Soln 1 GM/50ML in Dextrose 2 GM/50ML	1	1
Cefoxitin Sodium IV Soln 2 GM/50ML in Dextrose 1.1 GM/50ML	2	1
Cefpodoxime Proxetil		1
Cefpodoxime Proxetil For Susp 100 MG/5ML	9	2
Cefpodoxime Proxetil For Susp 50 MG/5ML	8	2
Cefpodoxime Proxetil Tab 100 MG	7	1
Cefpodoxime Proxetil Tab 200 MG	16	2
Cefprozil		2
Cefprozil For Susp 125 MG/5ML	29	2
Cefprozil For Susp 250 MG/5ML	35	2
Cefprozil Tab 250 MG	16	1
Cefprozil Tab 500 MG	26	1

Product Name	NDC	RxNorm
Ceftaroline Fosamil for IV Soln 400 MG	2	1
Ceftaroline Fosamil for IV Soln 600 MG	2	1
Ceftazidime For IV Soln 1 GM	7	12
Ceftazidime For IV Soln 1 GM/50ML and Dextrose 5%	1	1
Ceftazidime For IV Soln 2 GM	7	8
Ceftazidime For IV Soln 2 GM/50ML and Dextrose 5%	1	1
Ceftazidime Sodium in D5W Inj 1 GM/50ML	2	1
Ceftazidime Sodium in D5W Inj 2 GM/50ML	2	1
Ceftazidime Sodium in Dextrose Inj 1 GM/50ML	1	1
Ceftibuten Cap 400 MG	2	2
Ceftibuten For Susp 180 MG/5ML	3	2
Ceftibuten For Susp 90 MG/5ML	3	1
CEFTIN 125 MG OR TABS	4	1
Ceftriaxone Sodium For Inj 1 GM	41	2
Ceftriaxone Sodium For Inj 10 GM	14	1
Ceftriaxone Sodium For Inj 250 MG	34	2
Ceftriaxone Sodium For Inj 500 MG	34	2
Ceftriaxone Sodium For IV Soln 1 GM	4	2
Ceftriaxone Sodium for IV Soln 1 GM and Dextrose 3.74%	1	1
Ceftriaxone Sodium For IV Soln 2 GM	5	4
Ceftriaxone Sodium for IV Soln 2 GM and Dextrose 2.22%	1	1
Ceftriaxone Sodium in Dextrose Inj 20 MG/ML	1	1
Ceftriaxone Sodium in Dextrose Inj 40 MG/ML	1	1
Ceftriaxone Sodium w/ Lidocaine IM Inj Kit 1 GM	2	
Cefuroxime Axetil		1
Cefuroxime Axetil For Susp 125 MG/5ML	3	2
Cefuroxime Axetil For Susp 250 MG/5ML	4	2
Cefuroxime Axetil Tab 250 MG	94	2
Cefuroxime Axetil Tab 500 MG	64	2
Cefuroxime in Sterile Water Inj 1.5 GM/50ML	2	1
Cefuroxime Sodium 1.5 GM and Dextrose 2.9% For IV Soln	1	2
Cefuroxime Sodium 750 MG and Dextrose 4.1% For IV Soln	1	1
Cefuroxime Sodium For IV Soln 1.5 GM	5	3
Cefuroxime Sodium For IV Soln 7.5 GM	1	1
Cefuroxime Sodium For IV Soln 750 MG	2	1
Cefuroxime Sodium in D5W Inj 15 MG/ML	1	1
Cephalexin		1
Cephalexin Cap 250 MG	269	4
Cephalexin Cap 500 MG	525	4
Cephalexin Cap 750 MG	7	2
Cephalexin For Susp 125 MG/5ML	35	3
Cephalexin For Susp 250 MG/5ML	51	3

Product Name	NDC	RxNorm
Cephalexin Tab 250 MG	1	1
Cephalexin Tab 500 MG	6	2
CEPHRADINE 250 MG OR CAPS	20	1
CEPHRADINE 500 MG OR CAPS	31	1
Chloramphenicol Sodium Succinate For IV Inj 1 GM	1	1
Ciprofloxacin 200 MG/100ML in D5W	11	2
Ciprofloxacin 400 MG/200ML in D5W	21	3
Ciprofloxacin For Oral Susp 250 MG/5ML (5%) (5 GM/100ML)	1	1
Ciprofloxacin For Oral Susp 500 MG/5ML (10%) (10 GM/100ML)	1	1
Ciprofloxacin HCl Tab 100 MG (Base Equiv)	5	4
Ciprofloxacin HCl Tab 250 MG (Base Equiv)	173	2
Ciprofloxacin HCl Tab 500 MG (Base Equiv)	379	2
Ciprofloxacin HCl Tab 750 MG (Base Equiv)	77	2
Ciprofloxacin HCl Tab SR 24HR 500 MG (Base Equiv)	3	1
Ciprofloxacin IV Soln 1200 MG/120ML (1%)	1	1
Ciprofloxacin IV Soln 200 MG/20ML (1%)	6	2
Ciprofloxacin IV Soln 400 MG/40ML (1%)	7	1
Ciprofloxacin-Ciprofloxacin HCl Tab SR 24HR 1000 MG(Base Eq)	4	2
Ciprofloxacin-Ciprofloxacin HCl Tab SR 24HR 500 MG (Base Eq)	5	2
Clarithromycin		1
Clarithromycin For Susp 125 MG/5ML	12	2
Clarithromycin For Susp 250 MG/5ML	15	2
Clarithromycin Tab 250 MG	37	2
Clarithromycin Tab 500 MG	139	2
Clarithromycin Tab SR 24HR 500 MG	30	18
Clindamycin HCl		2
Clindamycin HCl Cap 150 MG	439	3
Clindamycin HCl Cap 300 MG	126	2
Clindamycin HCl Cap 75 MG	2	2
Clindamycin Palmitate HCl For Soln 75 MG/5ML (Base Equiv)	7	2
Clindamycin Phosphate in D5W IV Soln 300 MG/50ML	7	4
Clindamycin Phosphate in D5W IV Soln 600 MG/50ML	8	4
Clindamycin Phosphate in D5W IV Soln 900 MG/50ML	7	4
Clindamycin Phosphate Inj 9 GM/60ML	1	1
Clindamycin Phosphate IV Soln 300 MG/2ML	1	1
Clindamycin Phosphate IV Soln 600 MG/4ML	3	2
Clindamycin Phosphate IV Soln 900 MG/6ML	3	2
CLOXACILLIN SODIUM 250 MG OR CAPS	4	1
CLOXACILLIN SODIUM 500 MG OR CAPS	4	1
Dapsone Tab 100 MG	4	1
Dapsone Tab 25 MG	5	2
Daptomycin For IV Soln 500 MG	1	1

Product Name	NDC	RxNorm
Demeclocycline HCl Tab 150 MG	7	2
Demeclocycline HCl Tab 300 MG	7	1
Dicloxacillin Sodium Cap 250 MG	68	1
Dicloxacillin Sodium Cap 500 MG	82	1
Doripenem For IV Infusion 250 MG	4	1
Doripenem For IV Infusion 500 MG	4	1
Doxycycline Calcium Syrup 50 MG/5ML	1	1
Doxycycline Hyclate Cap 100 MG	265	3
Doxycycline Hyclate Cap 20 MG	3	2
Doxycycline Hyclate Cap 50 MG	35	2
Doxycycline Hyclate Cap DR Particles 100 MG	1	1
Doxycycline Hyclate Cap DR Particles 75 MG	1	2
Doxycycline Hyclate For Inj 100 MG	12	3
Doxycycline Hyclate Tab 100 MG	242	2
Doxycycline Hyclate Tab 20 MG	19	2
Doxycycline Hyclate Tab Delayed Release 100 MG	5	2
Doxycycline Hyclate Tab Delayed Release 150 MG	5	2
Doxycycline Hyclate Tab Delayed Release 200 MG	2	1
Doxycycline Hyclate Tab Delayed Release 75 MG	5	2
Doxycycline Monohyd Cap 75 MG w/ Omega 3-Vit E Cap Kit	1	
Doxycycline Monohyd Tab 100 MG &Multivit/Minerals Tab Kit	2	
Doxycycline Monohydrate Cap 100 MG	14	2
Doxycycline Monohydrate Cap 150 MG	5	4
Doxycycline Monohydrate Cap 50 MG	4	2
Doxycycline Monohydrate Cap 75 MG	4	2
Doxycycline Monohydrate For Susp 25 MG/5ML	3	2
Doxycycline Monohydrate Tab 100 MG	21	15
Doxycycline Monohydrate Tab 150 MG	6	6
Doxycycline Monohydrate Tab 50 MG	7	2
Doxycycline Monohydrate Tab 75 MG	7	4
Drotrecogin alfa (Activated) For IV Soln 20 MG	1	1
Drotrecogin alfa (Activated) For IV Soln 5 MG	1	1
DYNAPEN 62.5 MG/5ML OR SUSR	4	1
Ertapenem Sodium For IV Inj 1 GM (Base Equivalent)	1	2
ERYPED 200 MG OR CHEW	6	1
ERYTHROCIN 1000 MG IV KIT	1	4
ERYTHROMYCIN ESTOLATE 125 MG/5ML OR SUSP	11	2
ERYTHROMYCIN ESTOLATE 250 MG/5ML OR SUSP	13	2
Erythromycin Ethylsuccinate For Susp 100 MG/2.5ML		1
Erythromycin Ethylsuccinate For Susp 200 MG/5ML	10	3
Erythromycin Ethylsuccinate For Susp 400 MG/5ML	5	1
Erythromycin Ethylsuccinate Susp 200 MG/5ML		1

Product Name	NDC	RxNorm
Erythromycin Ethylsuccinate Tab 400 MG	84	2
Erythromycin Lactobionate For Inj 1000 MG	3	8
Erythromycin Lactobionate For Inj 500 MG	4	8
Erythromycin Lactobionate For Inj Kit 500 MG	1	4
Erythromycin Powder	10	
Erythromycin Stearate Tab 250 MG	82	6
Erythromycin Stearate Tab 500 MG		4
Erythromycin Tab 250 MG	96	3
Erythromycin Tab 500 MG	53	1
Erythromycin Tab Delayed Release 250 MG	43	2
Erythromycin Tab Delayed Release 333 MG	102	3
Erythromycin Tab Delayed Release 500 MG	41	1
Erythromycin w/ Delayed Release Particles Cap 250 MG	57	2
Erythromycin w/ Enteric Coated Particles Tab 333 MG	3	1
Erythromycin w/ Enteric Coated Particles Tab 500 MG	5	1
Erythromycin-Sulfisoxazole		1
Erythromycin-Sulfisoxazole For Susp 200-600 MG/5ML	27	4
Fidaxomicin Tab 200 MG	1	1
Gatifloxacin Inj 10 MG/ML	1	1
Gatifloxacin-D5W IV Soln 2 MG/ML	1	1
Gemifloxacin Mesylate Tab 320 MG (Base Equiv)	5	1
Gentamicin in Saline Inj 0.6 MG/ML		1
Gentamicin in Saline Inj 0.8 MG/ML	3	1
Gentamicin in Saline Inj 0.9 MG/ML	1	1
Gentamicin in Saline Inj 1 MG/ML	3	1
Gentamicin in Saline Inj 1.2 MG/ML	4	1
Gentamicin in Saline Inj 1.4 MG/ML	1	1
Gentamicin in Saline Inj 1.6 MG/ML	3	1
Gentamicin in Saline Inj 2 MG/ML	1	1
Gentamicin Sulfate Inj 40 MG/ML	3	1
Gentamicin Sulfate IV Soln 10 MG/ML	3	1
Imipenem-Cilastatin Intramuscular For Soln 500 MG	1	2
Imipenem-Cilastatin Intravenous For Soln 250 MG	7	2
Imipenem-Cilastatin Intravenous For Soln 500 MG	7	2
KANTREX 500 MG OR CAPS	1	1
KEFTAB 500 MG OR TABS	10	1
LAMPRENE 50 MG OR CAPS	2	1
Levofloxacin in D5W IV Soln 250 MG/50ML	8	2
Levofloxacin in D5W IV Soln 500 MG/100ML	9	2
Levofloxacin in D5W IV Soln 750 MG/150ML	9	2
Levofloxacin IV Soln 25 MG/ML	11	2
Levofloxacin Oral Soln 25 MG/ML	9	2

Product Name	NDC	RxNorm
Levofloxacin Tab 250 MG	88	2
Levofloxacin Tab 500 MG	180	2
Levofloxacin Tab 750 MG	98	6
Linezolid For Susp 100 MG/5ML	1	1
Linezolid IV Soln 2 MG/ML	2	1
Linezolid Tab 600 MG	6	1
Loracarbef Cap 200 MG	7	1
Loracarbef Cap 400 MG	3	1
Loracarbef For Susp 100 MG/5ML	7	1
Loracarbef For Susp 200 MG/5ML	10	1
MAXAQUIN 400 MG OR TABS	2	1
Meropenem IV For Soln 1 GM	8	2
Meropenem IV For Soln 500 MG	8	2
MetroNIDAZOLE		1
Metronidazole Cap 375 MG	4	2
Metronidazole HCl IV For Soln 500 MG	2	1
Metronidazole in NaCl 0.74% IV Soln 500 MG/100ML	1	1
Metronidazole in NaCl 0.79% IV Soln 500 MG/100ML	6	2
Metronidazole Tab 250 MG	163	3
Metronidazole Tab 500 MG	222	2
Metronidazole Tab SR 24HR 750 MG	1	1
MINOCIN 50 MG/5ML OR SUSP	1	1
Minocycline HCl Cap 100 MG	77	4
Minocycline HCl Cap 50 MG	31	4
Minocycline HCl Cap 75 MG	11	2
Minocycline HCl IV For Soln 100 MG	2	1
Minocycline HCl Tab 100 MG	7	2
Minocycline HCl Tab 50 MG	6	2
Minocycline HCl Tab 75 MG	6	2
Minocycline HCl Tab SR 24HR 105 MG	1	1
Minocycline HCl Tab SR 24HR 115 MG	1	1
Minocycline HCl Tab SR 24HR 135 MG	9	2
Minocycline HCl Tab SR 24HR 45 MG	9	2
Minocycline HCl Tab SR 24HR 55 MG	1	1
Minocycline HCl Tab SR 24HR 65 MG	1	1
Minocycline HCl Tab SR 24HR 80 MG	1	1
Minocycline HCl Tab SR 24HR 90 MG	9	2
Moxifloxacin HCl 400 MG/250ML in Sodium Chloride 0.8% Inj	1	1
Moxifloxacin HCl Tab 400 MG (Base Equiv)	46	4
Nafcillin Sodium For IV Soln 1 GM	5	2
Nafcillin Sodium For IV Soln 2 GM	5	1
Nafcillin Sodium in Dextrose Inj 1 GM/50ML	1	1

Product Name	NDC	RxNorm
Nafcillin Sodium in Dextrose Inj 2 GM/100ML	1	1
Nafcillin Sodium in Dextrose Inj 2 GM/50ML		1
NEBCIN 60 MG IV SOLR	1	1
Neomycin Sulfate Soln 25 MG/ML	1	1
Neomycin Sulfate Tab 500 MG	20	2
Nitazoxanide For Susp 100 MG/5ML	2	1
Nitazoxanide Tab 500 MG	2	1
Norfloxacin Tab 400 MG	5	1
Ofloxacin Tab 200 MG	11	2
Ofloxacin Tab 300 MG	6	2
Ofloxacin Tab 400 MG	10	2
OXACILLIN SODIUM 2 GM IV SOLR	2	1
OXACILLIN SODIUM 250 MG/5ML OR SOLR	1	1
Oxacillin Sodium For IV Soln 1 GM	4	1
Oxacillin Sodium in Dextrose Inj 1 GM/50ML	1	1
Oxacillin Sodium in Dextrose Inj 2 GM/50ML	1	1
Paromomycin Sulfate Cap 250 MG	5	1
PENETREX 200 MG OR TABS	1	1
PENETREX 400 MG OR TABS	1	1
Penicillin G Benzathine & Procaine Inj 900000-300000 Unt/2ML	7	6
Penicillin G Benzathine & Procaine Inj Susp 1200000 Unit/2ML	8	2
Penicillin G Benzathine & Procaine Inj Susp 2400000 Unit/4ML	2	2
Penicillin G Benzathine Intramuscular Susp 1200000 Unit/2ML	7	2
Penicillin G Benzathine Intramuscular Susp 2400000 Unit/4ML	4	2
Penicillin G Benzathine Intramuscular Susp 600000 Unit/ML	5	12
Penicillin G Potassium Inj 20000 Unit/ML in Dextrose	1	1
Penicillin G Potassium Inj 40000 Unit/ML in Dextrose	1	1
Penicillin G Potassium Inj 60000 Unit/ML in Dextrose	1	1
Penicillin G Procaine & Benzathine Inj Susp 300000 Unit/ML	2	3
Penicillin G Procaine & Benzathine Inj Susp 600000 Unit/ML	2	4
Penicillin G Procaine Intramuscular Susp 600000 Unit/ML	10	6
PENICILLIN V 125 MG/5ML OR SOLR	2	
PENICILLIN V 250 MG/5ML OR SOLR	1	
Penicillin V Potassium		1
Penicillin V Potassium For Soln 125 MG/5ML	11	2
Penicillin V Potassium For Soln 250 MG/5ML	14	1
Penicillin V Potassium Tab 250 MG	149	1
Penicillin V Potassium Tab 500 MG	182	3
Piperacillin Sod-Tazobactam Sod in Dex IV Sol 3-0.375GM/50ML	4	1
Piperacillin Sod-Tazobactam Sod in Dex IV Soln 2-0.25GM/50ML	4	1
Piperacillin Sod-Tazobactam Sod in Dex IV Soln 4-0.5GM/100ML	4	1
Piperacillin Sodium For IV Soln 3 GM	1	1

Product Name	NDC	RxNorm
Piperacillin Sodium For IV Soln 4 GM	1	1
Piperacillin Sodium For IV Soln 40 GM		1
Piperacillin Sodium-Tazobactam Sodium For Inj 2-0.25 GM	16	2
Piperacillin Sodium-Tazobactam Sodium For Inj 3-0.375 GM	18	2
Piperacillin Sodium-Tazobactam Sodium For Inj 36-4.5 GM	6	2
Piperacillin Sodium-Tazobactam Sodium For Inj 4-0.5 GM	17	4
PROSTAPHLIN 250 MG/5ML OR SOLR	1	1
Quinupristin-Dalfopristin for Inj 500 MG (150-350 MG)	2	1
Rifaximin Tab 200 MG	10	1
Rifaximin Tab 550 MG	8	1
ROCEPHIN IM CONVENIENCE 500-1 MG-% IM KIT	1	
Streptomycin Sulfate For Inj 1 GM	2	2
Sulfadiazine Tab 500 MG	2	1
Sulfamethoxazole-Trimethoprim		5
Sulfamethoxazole-Trimethoprim IV Soln 400-80 MG/5ML	9	3
Sulfamethoxazole-Trimethoprim Susp 200-40 MG/5ML	50	7
Sulfamethoxazole-Trimethoprim Tab 400-80 MG	48	4
Sulfamethoxazole-Trimethoprim Tab 800-160 MG	354	6
Sulfisoxazole Tab 500 MG	6	1
SUPRAX 200 MG OR TABS	1	1
Telavancin HCl For IV Soln 250 MG (Base Equiv)	2	1
Telavancin HCl For IV Soln 750 MG (Base Equiv)	2	1
Telithromycin Tab 300 MG	1	1
Telithromycin Tab 400 MG	7	1
TEQUIN 400 MG OR TABS	4	1
Tetracycline HCl Cap 250 MG	108	5
Tetracycline HCl Cap 500 MG	129	2
Tetracycline HCl Tab 250 MG	1	1
Tetracycline HCl Tab 500 MG	1	1
Ticarcillin & K Clavulanate For Inj 3.1 GM	2	1
Ticarcillin & K Clavulanate For Inj 31 GM	1	1
Ticarcillin & K Clavulanate Inj 3.1 GM/100ML	1	1
Tigecycline For IV Soln 50 MG	2	1
TOBRAMYCIN 10 MG/ML (D5W) INJECTION CUSTOM PD	1	
Tobramycin Sulfate For IV Soln 80 MG	2	1
TOBRAMYCIN SULFATE IN SALINE 1.6-0.9 MG/ML-% IV SOLN	1	1
Tobramycin Sulfate Inj 0.8 MG/ML in Saline	1	1
Tobramycin Sulfate Inj 1.2 MG/ML in Saline	1	1
Tobramycin Sulfate Inj 40 MG/ML	7	1
Tobramycin Sulfate IV Soln 10 MG/ML		1
Trimethoprim HCl Oral Soln 50 MG/5ML (Base Equiv)	1	2
Trimethoprim Tab 100 MG	17	2

Product Name	NDC	RxNorm
Trimethoprim Tab 200 MG	3	2
Trimetrexate Glucuronate For Inj 25 MG	1	1
TROBICIN 2 GM IM SUSR	1	1
Troleandomycin Cap 250 MG	1	1
Trovafloxacin Mesylate Tab 100 MG (Base Equiv)	2	1
Trovafloxacin Mesylate Tab 200 MG (Base Equiv)	3	1
Vancomycin HCl Cap 125 MG	11	2
Vancomycin HCl Cap 250 MG	12	2
Vancomycin HCl For Inj 10 GM	9	3
Vancomycin HCl For Inj 1000 MG	19	3
Vancomycin HCl For Inj 500 MG	30	4
Vancomycin HCl For Inj 5000 MG	7	1
Vancomycin HCl For Inj 750 MG	2	1
Vancomycin HCl For Oral Soln 250 MG/5ML	2	1
Vancomycin HCl For Oral Soln 500 MG/6ML	1	1
Vancomycin HCl in Dextrose Inj 1 GM/200ML	1	1
Vancomycin HCl in Dextrose Inj 500 MG/100ML	1	1
Vancomycin HCl in Dextrose Inj 750 MG/150ML	1	1
VELOSEF 250 MG/5ML OR SUSR	2	1
ZAGAM 200 MG OR TABS	2	1

A.3.4. Maintaining patient confidentiality in Common Data Model queries and datasets. The Common Data Model contains some of the 18 elements that are defined as personal health information (PHI) under HIPAA, including encounter dates and date of birth. The necessary "cross-walks" between the arbitrary identifiers included in the Common Data Model and their originating data are not specified in the scope of the Common Data Model, but for the purposes of this study, will be maintained by each node site. Furthermore, much of our data analyses will only require aggregated data that has no patient-level information. For the patient-level data analyses that we perform, we will use only de-identified data, without requiring any PHI be released by node sites. This is made possible by calculating age in days or months at the time of encounters and eliminating the date of birth variable.

Maternal-Child Linkage Data in 7 Node Sites

We received prep-to-research information for 6 of the 7 node sites which we will employ for maternal-child linkage, 5 from PORTAL and from ADVANCE (its OCHIN network), as shown below. Mid-South CDRN (Vanderbilt network) has also successfully linked maternal and child records. Examples of published studies attesting to the feasibility of these linkages are available from each of these CDRNs:[123-125]

PORTAL—Kaiser NW). Hillier TA, Pedula KL, Schmidt MM, Mullen JA, Charles MA, Pettitt DJ. Childhood obesity and metabolic imprinting: the ongoing effects of maternal hyperglycemia. Diabetes Care. 2007 Sep;30(9):2287-92.

(ADVANCE—OCHIN) Angier H, Gold R, Crawford C, P O'Malley J, J Tillotson C, Marino M, DeVoe JE. Linkage methods for connecting children with parents in electronic health record and state public health insurance data. Matern Child Health J. 2014 Nov;18(9):2025-33.

(Mid-South--Vanderbilt) Heerman WJ, Bian A, Shintani A, Barkin SL. Interaction between maternal prepregnancy body mass index and gestational weight gain shapes infant growth. Acad Pediatr. 2014 Sep-Oct;14(5):463-70.

PORTAL: Node Sites' Experience and Ability to Create Maternal/Child Linkage

Participating sites:

- Kaiser Permanente Colorado (KPCO - lead site for the PORTAL Network)
- Denver Health and Hospital Authority (DHHA)
- Kaiser Permanente Washington Health Research Institute (KPWHRI)
- HealthPartners Institute for Education and Research (HPIER)
- Kaiser Permanente Mid-Atlantic States (KPMAS)
- Kaiser Permanente Northwest (KPNW)

Each of these node sites were asked the following questions:

1. Does your site have experience creating a mother/child linkage?
2. For what purpose did your site create originally create this linkage?
3. For which calendar years does this linkage currently exist?
4. Does your site use state birth certificates to create the mother/child linkage when necessary (i.e. when you haven't already made the linkage using other data sources)?
5. If the mother/child linkages at your site are not current through 2014, what would it take to update this linkage at your site? (for example, would you need to update and re-run an existing algorithm?; would you need a new agreement with state health to obtain birth certificates?)
6. Any additional comments to include relevant to creating mother/child linkages?

Responses to the above questions, by site.

PORTAL Site	**Experience creating a maternal/child linkage**	**Original purpose of linkage**	**Years the linkage covers**	**Use state birth certificates to create the mother/child linkage when necessary**	**If linkage is not current through 2014, what would it take to update it**	**Any additional comments**
A	No	N/A	N/A	N/A	N/A	Creating the linkage is possible based on delivery records, in our system. We have used the guarantor in the payment field to make linkages between family members in the past.
B	Yes	Medication Exposure in Pregnancy Risk Evaluation	1996-2014 (linkage rate is higher from 2001-2013 where	Yes (2001-2013)	2014 deliveries are linked but not yet enriched with birth certificate	Plan to update the linkage in the next 1-2 yrs with separate funding. We

		Program (MEPREP) and EIPO	we have birth certificate information available to enrich the linking process)		information. The linkage rate (from potential delivery to a known infant) could be raised from 68% to ~90% by requesting the name and address information from the State Department of Health.	will approach this as a modification to our existing IRB with the state department of health.
C	Yes	Vaccine Safety Datalink (VSD) and Medication Exposure in Pregnancy Risk Evaluation Program (MEPREP)	2001-2012	Yes, we use birth certificates from the state department of health	To update state linkage we would need the resources to pay the state the contracted fee of $2000/year. A research programmer would also need to update and run our existing linkage algorithms.	The state department of health requires that separate contracts are executed for each purpose/use of the birth certificate data. We could make use of the mother/child linkage for PORTAL purposes but would be unable to use any of the specific birth certificate data fields for PORTAL purposes without a separate contract.
D	Yes	Medication Exposure in Pregnancy Risk Evaluation Program (MEPREP);	2000-2014	No	N/A (and will be updated through 3/2015 very soon	We are confident in the data we have for mothers that have delivered babies in our

		also for an existing "Perinatal database"				system. We are less confident about linking maternal/child information for new members or members with stepchildren
E	Yes	Several ongoing projects require this linkage	1991-2014	No, we do not have state birth certificate data; we link through our membership database.	N/A (already includes 2014)	Of 200,215 unique MRNs there are 170,730 unique MRNs that could link up to their mother (85%). Of randomly picked 20 MRNs for children born in 2014 all could link to their mother perfectly.
F	Yes	Built for a particular research project	1989 - present	No	N/A – updates run weekly	For most linkages, we use four different databases from different eras that contain both the mom's and babies health record number in the same record. For live birth deliveries and newborns that are still unlinked, we link deliveries from the mom's record to the newborn's date of birth if they are both in the same subscriber group (family account) and if

						the delivery date and the newborn date of birth are within 14 days.

Availability of covariates

Maternal antibiotic prescribing in pregnancy. Maternal BMI (pre- or early preg), Gestational weight gain, Gestational diabetes:
Available for mothers who receive care within the various health care systems of participating sites (which is most mothers).
Birth weight, Gestational age at birth:
Likely to be available in most circumstances for most participating sites. If it is only in the hospital nursery records, and the mother delivers at a hospital outside of the routine health care delivery system, it may not be available. This is likely to vary by site.
C-section v. vaginal
Likely to be available in most circumstances for most participating sites. Even if the mother delivered outside of the health care system, procedure codes for delivery (vaginal versus C-section) are likely to be available in claims information (as opposed to information from the electronic health record).

Table. Linkage data available from ADVANCE CDRN.

	OCHIN		
	Denom	Num	%
Total Women with at least one Pregnancy visit in CDM	56167		
Able to link mother and child (# of children)	46887	8976	19%
Data on c-section v. vaginal (# of mothers)	46887	6907	15%
Data on maternal BMI (pre- or early preg) (# of mothers)	46887	33246	71%
Data on mother gestational weight gain (# of mothers)	46887	36702	78%
Data on gestational diabetes in mother (# of mothers) (1)	46887	4292	9%
Data on child birth weight (# of children) (2)	8976	4305	48%
Data on gestational age (# of children) (2)	8976	4605	51%
Maternal antibiotic prescribing in pregnancy (3)	46887	11936	25%
(1) Encounter dx available now; problem list dx available ADVANCE-wide in CDM v2			
(2) This denominator is the number of children whom we currently link directly to the mother.			
(3) Prescribing data will be available ADVANCE-wide in CDM v3			

A.4. Analysis Plan
A.4.1. Overview. This section depicts how we address Overall Goals 1 and 2 of the Funding Announcement: (1) answer important patient-centered questions about obesity using PCORnet, and (2) rigorously test and evaluate the capability and functionalities of the PCORnet distributed data system for supporting large-scale, multi-center observational studies. Wherever possible and without jeopardizing the ability to answer the proposed patient-centered questions, we will test and evaluate methods that share as little information as possible. We will develop a systematic approach to assess the validity and efficiency of these data-sharing and analytic methods for future use in PCORnet studies (Evaluation Section).

A.4.2. Descriptive analysis

We will develop a distributed program to create the study cohort based on the eligibility criteria described above. The distributed program will also compare the baseline characteristics of the study cohort with patients who are otherwise eligible for the study but have missing information in key baseline covariates, at each CDRN site and across all CDRN sites. Within the overall cohort of children who meet the eligibility criteria, we will further identify those who have longitudinal follow-up data at 5 and 10 years and compare the overall cohort with these sub-cohorts. These comparisons will allow us to assess the representativeness of the study cohort and sub-cohorts and help explore possible selection bias that can later be accounted for in the analysis and interpretation of the study results. We will characterize and compare eligible patients who receive antibiotics and those who don't in detail. Specifically, we will compare the distributions and missingness of the baseline covariates, the length of follow-up and missingness after receipt of antibiotics **[IR-1]**.

All of the proposed descriptive analyses can be done with distributed programs developed centrally by the study team in collaboration with the Coordinating Center (CC) in accordance with the PCORnet Common Data Model v3.0. The CC will beta-test and distribute the program. Each node site will execute the program locally with minimal to no modification to the code. In the fully aggregated analysis, only summary-level information will be sent back to the study team via the CC. The process we develop will be a critical building block for the PCORnet infrastructure as it will allow robust descriptive analysis to be conducted efficiently without sharing individual-level information in future studies.

A.4.3 Causal inference approaches [CI-1, CI-2, CI-3, CI-4, CI-5]; Threats to validity

The conceptual causal model underlying the proposed research is that certain types, timing, and amount of antibiotic use in the first 2 years of life can set children on adverse trajectories of excess weight gain, resulting in higher amounts of adiposity (BMI as proxy) in later childhood. The mostly likely mediating mechanism, as summarized in Section A, is via alterations in gut microbiota. These are questions of comparative harm in which the effect sizes are likely to be small. Valid estimation of the magnitude and precision of effects (in Aims 1-3) is critical to translating into an exploration of how the findings would influence parents' and clinicians' shared decision-making when faced with potentially bacterial infections in infancy and toddlerhood (the topic of our Secondary Aim). In all observational studies, causal inference is limited by potential confounding and both selection and information biases, perhaps especially when anticipated effect sizes are modest. **Table 4** summarizes our view of key threats to validity and how we will address them. Details follow in subsequent sections.

Our approach to causal inference will comprise the following steps:

- Account for missing data (Issues 1 and 2)
- Specify confounding variables
    - Specify *a priori* likely confounding variables
    - Identify additional potentially confounding factors from Common Data Model in pilot steps (Issue 3).
- Develop and distribute query to CDRNs
    - Use multiple imputation procedure to impute missing data and generate multiple (e.g., 40) complete datasets. Steps b-f apply to each imputed dataset except step d which we will apply to a pilot site and a smaller number of datasets.
    - Build propensity score model using logistic regression; estimate probability of exposure from this model
    - Use propensity scores to limit the dataset to the area of covariate overlap
    - Check covariate balance between exposure groups after propensity score adjustment to assess fitness of the imposed propensity score model.
    - For the continuous outcome on BMI Z-score, perform multivariate linear regression adjusting for propensity score as a categorical variable
    - For the discrete outcome on obesity status, return aggregate-level counts stratifying on site, propensity score strata, exposure status, and outcome variable for the central site to conduct a joint (generalized) logistic regression to estimate association between antibiotics use and obesity outcomes.
- Combine the results of the analyses of the imputed datasets and generate valid statistical inference taking into account both within- and between-imputation variation.
- Perform sensitivity analyses (**IR-5**)
    - Adjust for additional covariates from linkage with maternal records (Issue 4)
        - Compare results from a fully aggregated analysis to partially aggregated or fully individual patient-level analyses (Issues 4, 10).

        - Implement propensity score calibration method with patient-level data to assess the impact of the additional covariates from linkage with maternal records.
    - Compare overall associations with those within conditions that are/are not related to obesity and are/are not indications for antibiotics (Issue 6)
    - Compare results using dispensed v. prescribed medications (Issue 7).
- Include prescribing at ages 2-5 years in models (Issue 8).

We will repeat this process for each of the exposures in Aim 1: use of narrow vs. broad-spectrum antibiotics; number of antibiotic uses (0, 1, 2, 3, 4 or more in a generalized logistic regression); and age of first use (0-6 months, 7-12 months, 13-24 months, or not in the first two years in a generalized logistic regression).

Table 4. Ten key threats to validity and how we plan to combat them.

	Issue	*Threat*	*Our Approach*	*Strengths*	*Remaining Issues*
1	Loss to follow-up by age 5 years	Selection bias from informative censoring, could be systematic	Multiple imputation (Section B.4.4). Sensitivity analyses after covariate adjustment	Robust method	Sens. analyses impose untestable parametric assumptions on missing data mechanism
2	Missing covariate data	Selection bias, likely non-systematic	Multiple imputation (Section B.4.4)	Robust method	Less likely to violate missing at random, but possible
3	*A priori* list of Common Data Model covariates may miss important ones	Confounding	Preliminary bivariate analyses in pilot CDRN	Include potentially important confounders in distributed query	Fishing expedition? (Avoid by choosing sensible candidates)
4	Common Data Model lacks some relevant data, e.g., maternal variables	Confounding	Link to maternal variables in 7 node sites with experience	Adjust for type of delivery, maternal BMI, intra-partum antibiotics, etc.	Aggregate v. patient-level comparison only on subset
5	EMR typically does not have relevant data, e.g,, breastfeeding	Confounding	Cannot include those variables	-	Reflects current limitations of many big data studies
6	Even after adjusting for relevant covariates, confounding can still exist	Confounding, including by indication	Comparisons within condition: --otitis media --bronchiolitis --asthma --G-E reflux	Comparing within conditions related/ unrelated to obesity, with/without indication for antibiotics	Residual confounding still possible
7	Prescribing, not dispensing, data	Misclassification of exposure	Analysis using dispensing data	Compare results	Only on subset
8	Prescription exposure in 1st 2 years, outcome at age 5	Misclassification of exposure	Include prescribing at ages 2-5 as one step in modeling	May distinguish "critical period" v. "accumulation of risk" conceptual models	Analytic challenge—mediation, interaction
9	No inpatient medications	Misclassification of exposure	Will not include even though available on subset	Focus on decision-making for mild-moderate outpatient infections	May miss strong exposure for very small number of kids
10	Distributed query approach precludes iterative model building	Any of above	Query refinement, testing, and piloting before distribution	CC and Study Team have experience in distributed queries, incl. PS methods	How viable is this approach in PCORnet?

EMR, electronic medical record; CC, Coordinating Center; PS, propensity score

A.4.4. Approaches to missing data and confounding

A.4.4.1. Methods for handling missing data [MD-1, MD-2, MD-3, MD-5]

We expect to encounter missing information of various degree in both the baseline covariates and outcomes. Our analyses will use primarily the multiple imputation (MI) approach to handle missing covariates and outcome information [**MD-2, MD-3, MD-5]**. MI is a statistical technique for approaching incomplete data which has gained increasing popularity in the past decades.[54-56] It uses an imputation algorithm to fill in missing values *m* times and thus create *m* different "completed" data sets. Each completed data set is analyzed using standard procedures and the results from the *m* analyses are combined in a structured fashion to make inferences. There exist multiple imputation algorithms. We will consider employing the fully conditional specification method because it can handle arbitrary missing patterns and impute missing values for all variables, assuming the existence of a joint distribution for these variables. In particular, the fully conditional specification predictive mean matching method ensures that imputed values are plausible and may be more appropriate when the normality assumption is violated and thus is recommended for imputing missing values for a continuous variable, such as BMI. The analysis can be used for both one-time and repeated outcome measures.

MI assumes missing at random (MAR) which might be violated in this longitudinal study. For instance, it is possible that patients who "drop out" of the system are different than those that stay, for reasons related to their unobserved outcomes, even after we account for all the measured risk factors. In all analyses involving MI, we propose to use time-varying covariates available to account for the selection bias introduced by missing data. In the secondary analysis with individual-level data, we will also explore modifying the imputation algorithm to allow missing data to be non-MAR.

In the primary analysis with individual-level data, we will also explore modifying the MI approach to allow missing data to be non-MAR. Specifically, we propose to conduct sensitivity analysis using the following approach. We will adjust the imputed continuous variable values by adding a constant and we will adjust the imputed categorical variable values by shifting the odds of certain class levels compared to the reference level. The shift parameter can be constant across subjects, or can be simulated from a given distribution; it can apply to the entire study population or, as in the pattern-mixture model, only a specified subset. For example, we will shift imputed BMI by a positive constant (e.g., +0.1 BMI-z) since patients with missing BMI in follow-up will may tend to have higher BMI compared to those with BMI observed, but will also explore the opposite, i.e., negative constant to give plausible bounds to estimates. We can determine plausible constants based on partial follow-up of participants. We will conduct a comprehensive sensitivity analysis varying these parameters over a reasonable range to fully assess the robustness of analyses results.[128, 129]

Other missing data methods, such as the inverse probability censoring weighting (IPCW)[57-59] approach, originated from survey sampling literature. It weights patients with complete data to represent not only themselves but also other comparable individuals with missing data, and thus creates a weighted population in which there is no missing data or bias introduced by non-random missingness under the assumption that patients with and without missing data are comparable conditional on the measured variables. However, IPCW requires a monotone missing data pattern (roughly, once a patient is censored due to a missing variable, all subsequent information is discarded). In our analysis, we expect to have missing data in baseline covariates with an arbitrary missing data pattern. Thus IPCW does not fit well in this setting.

A.4.4.2. Notes on controlling for confounding. See Section 4.5.1 for approach to multivariable modeling.
***A priori* list of Common Data Model covariates may miss important ones (Issue 3)[IR-1]**. As an observational study situated in data from routine delivery of health care, the analyses must make efforts to account for factors across the spectrum of care that may influence exposures and outcomes. We have identified a number of these factors *a priori* (Section B.4.5.1 below), but the universe of available data in the Common Data Model is large, and the entire range of factors that influence both child growth and clinicians' use of antibiotics will probably not be evident before initial data collection. Therefore, data characterization will include a set of simplified preliminary analyses that screen for correlations among sociodemographic factors, diagnoses, other medication types, other aspects of clinical utilization, as well as the main exposures and outcomes. These analyses will be undertaken primarily within the PEDSnet "pilot" CDRN. Results will be used to guide selection of covariates, particularly diagnosis clusters and medication classes, for use in main study analyses. This exploration of the data differs conceptually from the technical feasibility testing described in Section B.3, where the intent will be to assure that final queries formulated by the Coordinating Center execute as expected in the CDRN, but we will undertake them in the same step of the process.

Common Data Model lacks some relevant data, e.g., maternal variables (Issue 4)[IR-2]. While the Common Data Model will provide data for many potential confounding factors in addition to the main exposures and outcomes, it does not include some that may be important, particularly those that reside in the mother's medical record. In our prep-to-

research survey and follow-up telephone calls, we identified 7 node sites that have already performed linkage to maternal data; they represent 26% of participating children (see below for site information). Using data from those node sites, we will perform a sensitivity analysis in which we adjust for one or more of the following variables: maternal pre-pregnancy or early pregnancy body mass index, gestational weight gain, gestational diabetes, use of antibiotics during pregnancy, type of delivery (vaginal, cesarean section), gestational age at birth, and birth weight.

Even after adjusting for relevant covariates, confounding can still exist (Issue 6). Observational studies of medication use are subject to confounding by indication.[60] The reasons for prescribing an antibiotic, whether they be the infection or some features of clinical decision making, may also be related to the outcome. While our PS approach can control for observed confounding, neither that approach nor others such as instrumental variables[61] are sufficient for eliminating confounding by indication. No one approach is perfect; in this study our strategy will be to examine associations of antibiotic use with excess weight gain among children with a particular diagnoses which are relatively common and likely to be accurate when present in the medical record. Here we focus on otitis media (antibiotics, may be related to obesity), asthma (some antibiotics, related to obesity), gastrointestinal (G-E) reflux (few antibiotics, not related to obesity) and bronchiolitis (some antibiotics, probably not related to obesity).

A.4.4.3. Issues of misclassification of exposure.

Prescribing, not dispensing, data (Issue 7)[IR-2]. Some patients are non-adherent with prescribed medications, especially for chronic conditions.[62-65] In one study of over 75,000 patients (mostly adults) enrolled in 3 health plans, patients filled 75% of all prescriptions sent electronically to pharmacies.[66] Among children 0 to 18 years old, 87% of prescriptions were filled. Thus, while dispensings are closer to actual taking of medication (because they are confirmed to be available to the patient), and are a more valid indicator of adherence in chronic disease among adults, we propose that for antibiotics in young children, the prescription will be a reasonable alternative. In our prep-to-research survey, all of the participating CDRNs pledged to have the prescription data in their Common Data Models by January 2016, i.e., in time for analyses of the Aims. In addition, there are several formats for prescription data. PCORnet has chosen RxNorm as the terminology for prescribing. All CDRNs anticipate having their prescription data as RxNorm codes in CDM v3.0. While we will use prescription data for our primary approach, in sensitivity analyses we will compare findings using dispensings for 5 CDRNs that already have those data available or will have them available by January 2016. They comprise a majority of the participating children (node sites with these data represent 80% of the population, but we do not yet know the completeness of the data field). In any case, prescriptions probably overestimate dispensings, and the overestimate is likely not related to excess weight gain. Thus this would be a case of non-differential misclassification, which would bias estimates toward the null, a conservative bias.

Prescription exposure in 1st 2 years, outcome at age 5 (Issue 8). If our conceptual model is correct that during a "critical period" of early and repeated exposure to broad spectrum antibiotics, the microbiota semi-permanently changes in a way that causes obesity, then later antibiotic exposure may make no difference. On the other hand, repeated exposure over age ("accumulation of risk")[67] may be the more important type of exposure. To disentangle these theories, in one regression model we will include antibiotic prescriptions beyond the first 2 years of life. How to include them is not straightforward because early antibiotic use may be related to later use through a number of pathways including repeated infection diathesis or decision-making inertia among clinicians and parents. Thus later antibiotic use could take the form of confounder, mediator, or interactor (~effect modifier). Our approach will be to stratify early use by characteristics (mainly type and amount) of later use. If there are differences across strata, then we depict them. If not, then we discuss mediation v. confounding based on theory, the literature, and other study findings.

A.4.4.4 Overarching issue

Distributed query approach precludes iterative model building (Issue 10). Our main programming approach will result in "fully aggregated" data because 1.) the future of PCORnet depends on its ability, at least in part, to execute distributed programming code, and 2.) all 10 CDRNs will have data that contribute to this analysis. In this approach the CC will send refined, tested, and piloted computer programs (queries) to each node site. These programs will include routines for accounting for missing data and multivariable modeling that includes covariates within the Common Data Model. Programmers at the node sites will run these queries unaltered against the Common Data Model and return de-identified tabulated data to CC for data characterization and quality control before the CC sends them to the Study Team for final analyses, as described below (Section B.4.5).

Additionally, we will perform individual patient-level analyses, for 3 reasons. The first is to compare results, on the same study sample, with the fully aggregated approach. The only difference will be using repeated outcome measures (within the age groups 4.0 to 5.9 and 9.0 to 10.9 y). The second is to further explore confounding by incorporating maternal data outside the Common Data Model on a subset of node sites; here we will compare effect size and precision with the fully aggregated results on the same subset. The third reason to collect patient-level data is that Aim 2 incorporates growth trajectories, distributed programming for which will not be available. Nevertheless, within the patient-level analysis approach, we will explore using distributed programs to obtain propensity-score adjusted exposure variables, which will reduce dimensionality and limit the number of data elements that we will ask node sites to transmit outside of their local data environments. In our prep-to-research survey, all node sites have expressed willingness to share individual-level data with appropriate governance and privacy safeguards.

A.4.5 Analysis plans for Specific Aims [**IR-3**]

A.4.5.1. Analysis Plan for Aim 1 -- *To evaluate the comparative effects of different types, timing, and amount of antibiotics used during the first two years of life on body mass index and risk of overweight and obesity at ages 5 (primary outcome age) and 10 (secondary) years.*

Main exposures

Antibiotic use in 1st 2 years of life. We define a prescription episode as first prescription after a 14-day free interval.

a. Number of prescriptions. Because actual dose is challenging to obtain for liquid prescriptions, and dosing for a common indications is relatively standard, number of prescriptions is a proxy for total exposure. We call this variable "amount," and categorize as (0), 1, 2, 3, 4+.
b. Broad v. narrow spectrum
    i. Narrow: penicillin, amoxicillin
    ii. Broad, to include intermediate and broader spectrum antibiotics: azithromycin, clindamycin, trimethoprim/sulfamethoxazole, cephalexin, cefazolin, 2nd and higher generation cepahalosporins, amoxicillin /clavulanic acid, others
c. Age at prescription (including specifying age at 1st prescription)
    i. 0-5 m, 6-11 m, 12-23 m
d. Combinations of amount, spectrum, and age at first prescription to address the main hypotheses that the risk of obesity-related outcomes will be most strongly related to a higher number of prescriptions of broad spectrum antibiotics prescribed in the first 6 months.

In addition, we will perform exploratory analyses that examine effects of

e. Broader v. narrower medication within class, using the example of amoxicillin/clavulanic acid v. amoxicillin
f. Prescription of any antibiotics v. none in the first 2 years of life. We do not consider this comparison to be of highest priority because it does not reflect comparative effectiveness and associations are more likely to be confounded than comparing type, timing, and amount.

Main outcomes

a. BMI (kg/m^2), continuous, expressed as z-score using CDC growth charts.[68] We will incorporate all available height/weight measures at each follow-up timepoint (4.0 to 5.9 y for primary outcome age, 9.0 to 11.0 y for secondary). For the distributed query analysis, we will choose the value closest to 5.0 or 10.0 years. In individual-level analyses, for which we can use mixed models, we will use all values in the age interval.

b. BMI (kg/m^2), categorical. We will use the 5- (and 10) -year estimates to categorize BMI as follows:
    1. <5th %ile, underweight
    2. 5th – 85th %ile, normal weight
    3. 85th- 95th %ile, overweight
    4. >95th %ile, obese
        1. >97th %ile or > 1.2 x 97th %ile, severely obese

    Our main approach will be to compare obese with normal weight. In an additional multinomial approach, the comparison will use normal weight as the referent group and the other categories as comparators.

While body mass index is not the most valid measure of body fat, it is highly related to patient-centered outcomes such as asthma symptoms, depression, and orthopedic maladies in childhood.[69, 70] It is also a determinant of cardiovascular risk

in childhood and school-age BMI predicts cardiovascular disease in adulthood.[71] Clinically measured weights are quite accurate, as are heights after age 3 years. Length in the 1st 2 years of life is overestimated by the paper and pencil method, and we will explore using a regression equation for correction.[72, 73] We avoid using length/height between 24 and 35 months, as recumbent length is measured in some and standing height in others; the difference is about 0.75 cm.[50] Age 5 years is an appropriate age for primary follow-up as most childhood obesity is incident by that age.[74]

Once length/height and weight measures are collected, it is important to exclude nonsensical values. The CDC algorithm to exclude biologically implausible values works well in repeated cross-sectional and longitudinal studies, as we have done in several studies of height and weight from electronic medical records.[47-50, 72] In these data we have also shown prediction of 5-year obesity from BMI (WHO growth charts) in the 1st 2 years of life is similar to prediction from weight-for-length (WHO or CDC), justifying the use of infant and toddler BMI in research studies.[75]

Main Covariates

1. Common Data Model 3.0 variables
    a. Diagnoses
        i. Conditions for which antibiotics for acute infection are prescribed, e.g., otitis media, pharyngitis, sinusitis, urinary tract infection
        ii. Conditions related to obesity for which primary care providers prescribe meds, with *a priori* focus on asthma defined by use of asthma-related medications, including primarily inhaled steroids and rescue beta agonists.
        iii. Conditions that are not related to obesity for which primary care providers prescribe meds, with *a priori* focus on G-E reflux
        iv. Conditions which are neither related to obesity nor an indication for medication, with *a priori* focus on bronchiolitis
    b. Prescription of medications that
        i. May cause obesity, with *a priori* focus on corticosteroids
        ii. Probably do not cause obesity, with *a priori* focus on reflux medications
    c. Other variables related to exposures and/or outcomes, including diagnoses and medications, identified through examination of pilot CDRN data (Section B.4.4.2)
2. Maternal variables (in a subset of node sites that have already performed maternal-child linkage)
    a. Maternal BMI before or early pregnancy. We will use primarily as continuous variable, but also use as a categorical: < 18.5 kg/m^2, underweight; 18.5-24.9, normal weight; 25-29.9, overweight; 30+, obese (including 30-34.9, stage I; 35-39.9, stage II, 40+, stage III)
    b. Gestational weight gain, defined in categories of inadequate, adequate, and excessive according to 2009 IOM guidelines, which are based both on maternal BMI and amount of weight gained.
    c. Gestational diabetes, defined by diagnosis code. We appreciate that various screening approaches provide differing estimates, but this study will not have the support to define GDM by laboratory values.
    d. Maternal use of antibiotics during pregnancy. As per definitions for the main exposure.
    e. Type of delivery, vaginal v. caesarean section, from the maternal medical record.
    f. Birth weight, from the maternal medical record
    g. Gestational age at birth, based on estimated delivery date from maternal medical record.

Although we prioritize data that are available in the current Common Data Model, characterizing potential availability of key covariates across the CDRNs, such as breastfeeding, may inform future studies. For breastfeeding, we propose to collect the following information:

a. Number of nodes that are able to capture breastfeeding data (and have been recording them in coded fields)
b. Type of breastfeeding data available (initiation, duration, exclusivity, etc.)
c. Number of children who have breastfeeding data available as a part of their medical records.

We will start our investigation with node sites that already know they have such data, and within them characterize the data as in *b.* and *c.* above. That process will inform how other sites could interrogate their electronic medical records for similar data, should they exist. We will request sites to carry out these interrogations, and will assess whether the sites

were able to do so. Among those who were, we will assess how successful they were in querying their data (i.e., as in *b.* and *c.* above).

The primary analysis will use aggregate-level data. Including only patients who meet the eligibility criteria for the study, we will use multiple imputation to address missing covariate and outcome information. An outcome of interest is BMI z-score at 5 and 10 years of age. It is a continuous variable; we will inspect the distribution of the variable and transform it as appropriate. The multiple imputation approach will generate m (e.g., 40) imputed datasets with complete data. Imputation will be conducted at each site with a distributed query program. The standard analysis procedure that we will apply to each imputed dataset is as follows: Suppose $\widehat{\theta_l}$ and $\widehat{W_l} \equiv \widehat{var}(\widehat{\theta_l})$ denote the point and variance estimates for the parameter of interest θ_0 from the lth imputed dataset using the standard analysis procedure, $l = 1,2,\dots,m$. Then the final point estimate $\hat{\theta} = \frac{1}{m}\sum_{l=1}^{m}\widehat{\theta_l}$, the average of the m complete-data estimates; and the final variance estimate $\widehat{W} = \frac{1}{m}\sum_{l=1}^{m}\widehat{W_l} + \left(1+\frac{1}{m}\right)\frac{1}{m-1}\sum_{l=1}^{m}\left(\widehat{\theta_l}-\hat{\theta}\right)^2$ which accounts for both within- and between-imputation variance.

We will fit identical models within each imputed dataset. For each comparison, we will first use a logistic regression model to estimate the propensity score (PS), which is defined as the probability of each exposure level, as outlined above. The PS model will be fit in a distributed model at each site, as the parameters associated with exposure level may well differ by site. We will then assess covariate overlap by comparing the PS distributions across each exposure group. If patients from different groups differ dramatically, we will consider trimming[76, 77] to exclude patients that are very unlikely to have received a different exposure. This typically happens when the probability of one exposure is near 1, so that there are few or no plausible non-exposed matches with a similar exposure probability.

We will run separate models for the 5-year and 10-year outcomes. For the continuous outcome of BMI score at each time point, we propose to use the distributed linear regression method[78, 79], to estimate the effects of antibiotics on weight adjusting for site and PS percentiles (e.g., deciles) as a categorical variable. Specifically, let $Y_{s,t}$ denote the 5-year BMI z-score for subject t at site s, $e_{s,t}$ denote the subject's PS, and $(q_{s,1}, q_{s,2}, \dots, q_{s,10})$ denote the site-specific PS deciles, then the mean model for $E\left[Y_{s,t}\right]$ includes site as a categorical variable, dummy variables to indicate which decile group $e_{s,t}$ falls in, and their interaction terms with site. Other terms including terms to reflect temporal trend will be added if deemed necessary.

Note that PS matching with replacement and PS stratification might be better adjustment methods using PSs, but PS regression allows us to use the distributed linear regression approach to conduct analysis using aggregate-level information only. Fitting a distributed linear regression model is a two-step process. At step 1, each site executes a distributed program locally and submits intermediate summarized statistical results to the lead team. The lead team combines the intermediate results, and computes the parameter estimates. At step 2, participating sites execute another distributed program and deliver the variance/covariance estimates of the parameter estimates to the lead team to compute the confidence intervals. We have previously employed the approach in another study.[80] Logistic and generalized logistic regression is a similar process, but it requires more iterations and the tool to automate this process is under development and may not be ready to use during the 2-year project period. Therefore, for the categorical outcome of obese, overweight, normal weight, and underweight status, we propose to use PS stratification. Only aggregate-level information on PS stratum, within each stratum the number of subjects within each exposure group and with each of the 4 weight status needs to be shared between each node site and the coordinating center for the final analysis using generalized logistic regression.

These aggregate analyses represent a key infrastructure-building activity as it will assess the feasibility and validity of using summary-level information to perform the same analysis that is conventionally done with pooled patient-level information. If proven feasible for PCORnet, the analytic method can further enhance the functionalities of the system by allowing sites to share less sensitive information while preserving the scientific rigor of the study.

As a secondary analysis **[IR-5]** using individual-level data, we will use the generalized mixed effects model approach to analyze repeated measures of continuous and categorical BMI outcomes, adjusting for site and PS percentiles (e.g., deciles) as a categorical variable.[81, 82] The number of outcome measures may vary among study subjects. Missing data will be handled using the multiple imputation approach the same as in the primary analysis.

As a sensitivity analysis, we will repeat the individual-level data analysis among the subset of individuals with linked maternal records and add additional covariates identified from the maternal records to the PS model. We also plan to

implement the PS calibration approach, which is a recently developed approach to combine PSs and regression calibration to allow better confounding adjustment using variables unmeasured in main study population but available for a subset of individuals.[83, 84] Evaluation of the secondary and sensitivity analyses results will provide information on the importance of the additional covariates identified from the matched maternal records and the validity and efficiency gain of the PS calibration approach.

A.4.5.2. Analysis plan for Aim 2--*To assess the comparative effects of different types, timing, and amount of antibiotics used during the first two years of life on the rates and patterns of childhood growth during the first 5 years of life.*

For Aim 2, the exposures and covariates remain the same from Aim 1. The main difference is in the outcome. Instead of examining BMI and prevalence of obesity at ages 5 and 10 years, in this Aim, we will examine growth trajectories from early infancy to age 5 years (and beyond should data availability allow). In addition to the criteria specified in Aim 1, for inclusion in this aim must also have at least two growth-evaluable encounters post-antibiotic exposure. Here the outcome is derived from all longitudinal measures of height and weight post-antibiotic exposure period.

Several ways exist to examine growth trajectories in early life. In fact, examining the pros and cons of different approaches was the topic of a recent symposium sponsored by the EU that Drs. Gillman and Kleinman helped to organize.[85, 86] Some approaches are driven only by length/height and weight data alone; others are linked to exposure or outcome. We propose to examine differences in the rate of growth in weight trajectory, height trajectory, and weight adjusted for height trajectory as alternate outcomes that will more accurately identify alterations in growth patterns.

We approach this question using a new analytic method we specifically designed to address this question based on a generalized, non-parametric growth trajectory. Height and weight measures will be modeled using Longitudinal Rate Regression.[87] The growth trajectories for height/weight will be estimated using a penalized spline equation. Antibiotic type, timing, and amount variables will be included in separate models at the mean and rate level with the rate level parameter being the parameter of interest. An additional model will be constructed using all three exposure characteristics to examine their joint effects on growth. Each model will be adjusted for additional confounders at both the mean and rate level. Confounders included at the rate level will be centered by the sample mean. The resulting estimate for the parameter of interest will be interpretable as the percent difference in the rate of change associated with each exposure characteristic based on a typical child from the sample. For example, the rate parameter for antibiotic type will estimate the percent difference in the rate of growth for a typical patient exposed to narrow antibiotics relative to the rate of growth estimated for a typical patient exposed to broad antibiotics. The estimated rate parameter of antibiotic exposure will be reported with an estimated 95% CI and a p-value for the test of no association. For added interpretation, the estimated height/weight trajectory will be used to calculate the growth between 2 years and 5 years of age for a patient in each exposure group. This value coupled with the rate parameter for each exposure characteristic will be used to estimate the additional height/weight gained by each exposure group relative to the reference group during this time period.

Sufficient density of longitudinal measurements is a concern for accurately characterizing any non-linear trajectory in growth. In particular, dense data are required for estimating the trajectory via a penalized spline model in this setting. However, we do not need penalized splines to model growth trajectories—that is just the most flexible and accurate method, provided the data are sufficiently dense. If the available data are not sufficiently dense, we can use a parametric non-linear function of time, such as regression splines.

Nonetheless, we anticipate sufficiently dense measurements to allow using penalized splines here. Weight and length/height are common measurements taken during outpatient visits, particularly at well-child visits. Typically, children are scheduled for well-child visits up to 9 times in the first 2 years of life and seen yearly thereafter for a total of 12 visits in the first 5 years. Because some children do not keep appointments and/or leave the delivery system, we do not anticipate full compliance with attending well-child visits and obtaining growth metrics. Nevertheless, some missing data will not invalidate our proposed trajectory analyses. For example, greater than 50% compliance on a majority of participants would be sufficient for effectively modeling non-linear trends in growth. This level of compliance is plausible, as in prior analyses within PEDSnet, we observed at least 6 visits with weight measures on a large majority of patients. Length/height measurements are less common, and underrepresented at sick visits, but they are also less rapidly varying and less sensitive to external influences, and can therefore be imputed or interpolated with higher confidence.

A.4.5.3 Analysis plan for Aim 3 – *To explore how the effects of different types, timing, and amount of antibiotics on childhood BMI, obesity risk and growth (Aims 1 and 2) vary according patient socio-demographic, clinical, and maternal characteristics*—as specified in Section B.1. This Aim addresses heterogeneity of treatment effects (HTE) **[HT-1, HT-2, HT-3, HT-4].** We propose to repeat the primary analyses within each of the subgroups defined for Aim 3 **[HT-2]**:

When there are 2 subgroups (e.g., males and females), we will examine the difference between the subgroup-specific effect estimates. The variance of the difference is the sum of the variances for the subgroup-specific effect estimates as the 2 effect estimates are independent. The standardized difference is expected to follow a normal distribution asymptotically. When there are more than 2 subgroups, the differences between the effect estimate from a reference subgroup and the other subgroup-specific effect estimates, after appropriate linear transformations, follows a multivariate normal distribution asymptotically. Then the sum of the differences squared follows a chi-square distribution with the degree of freedom being the number of subgroups minus 1.[88] A significant difference in treatment effect is detected if the corresponding p-value is ≤ 0.05 **[HT-3]**.The purpose of the HTE analysis is to explore any possible treatment effect heterogeneity, not formal hypothesis testing. Thus we will use the same 0.05 significance level for all contrasts **[HT-4]**.

A.5 Statistical Power for Aims 1-3 [RQ-4]

We first calculate power for Aim 1, to compare effects of different types, timing, and amount of antibiotic use in the first 2 years of life on BMI-z and obesity at age 5 and 10 years. We address several comparisons:

- Any v. no antibiotics in first 2 years.
- Number of antibiotic courses, 1, 2, 3, ≥4
- Broad v. narrow (Narrow is always narrow. Broad is ever broad.)
- Early v. late (0-5 m v. 6-23 m)
- A combination of all three characteristics of antibiotic use, specifically comparing the two extremes ([≥4, Broad, Early] v. [1, Narrow, late]) of the 16 categories formed from the combinations of #1, #2, and #3 above.

Our assumptions for sample proportions within each use category are from Bailey et al. (published[51] and personal communication). Except for comparison #5, we use estimates of BMI-z and %obese from Bailey, and calculate power for those differences. For comparison #5, cell-specific outcome estimates from Bailey were too unstable (a main reason to do this study), so we calculate least detectable differences for given power (80%, 90%) instead. Using the sample sizes from Table 3 (Section B.2.3), we estimate >99.9% power (**Table 5**) to detect the 5-year outcome differences at least as large as those in Bailey. For 10 years, power is substantially less but still >87% for BMI-z for the full sample and >66% for the sample with dispensing data.

Table 5: Power for Aim 1

Antibiotic Exposure in 1st 2 y	Mean BMI-z	Proportion Obese	Outcome Age (y)	Power for Total Sample		Power with Dispensing Data (~80%)		Power with Maternal-Child Linkage Data (~26%)	
				BMIz	*% Obese*	*BMIz*	*% Obese*	*BMIz*	*% Obese*
Any v. No Antibiotics	0.51	0.15	5	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%
	0.39	0.14	10	>99.9%	51.1%	99.8%	32.5%		
≥4 v. 1 Courses	0.55	0.16	5	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%
	0.45	0.14	10	87.9%	32.7%	66.4%	21.4%		
Broad v. Narrow	0.53	0.16	5	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%
	0.47	0.14	10	88.8%	36.5%	67.6%	23.3%		
Early v. Late	0.57	0.17	5	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%
	0.49	0.15	10	91.3%	69.3%	71.4%	47.2%		

For comparison #5, of [≥ 4, broad, early] v. [1 narrow, late] in the full sample, assuming BMI-z = 0.45 in the referent group (again from Bailey), we have 80% power to detect BMI-z difference of 0.02 between referent and exposed, and 90% power to detect 0.03. For proportion obese, assuming 0.14 in the referent group, the minimum detectable differences in prevalence are <0.01 for both 80% and 90% power. These are small effect sizes.

We do not calculate power for Aim 2, in which the outcome is growth trajectories. For Aim 3, which addresses heterogeneity of treatment effect, we use 2 examples of effect modifiers: sex, where we assume 50% male, 50% female, and race/ethnicity, where we assume 60% white, 25% Hispanic, and 15% black, similar to US children as a whole.[89] For

these comparisons, we evaluated 4+ v. 1 antibiotic course in the first 2 years of life. For pairwise comparisons (50% v. 50%, 15% v. 60%, 25% v. 60%), we have power to detect differences in differences in BMI-z of 0.03 to 0.05, and differences in differences of %obese of approximately 0.01 to 0.02.[90]

In sum, power for our Aims is ample based on observed data. In the actual analyses, sample sizes will be larger because we will use multiple imputation, albeit at the cost of larger standard errors.

B. Stakeholder Engagement in Planning and Study Design [PC-1, PC-4]

This section primarily addresses **Overall Goal III.)** Provide an opportunity for investigators, patients, and stakeholders from the PCORnet community to organize and collaborate in a multisite study and develop efficient, collaborative processes for doing so.

B.1 Identification of Key Stakeholder Groups. Since in essence, the proposed research embodies consumer preferences, gaps, and community values, patients and consumer stakeholders are engaged in all stages of proposal development. This includes formulating research questions and selecting outcomes that are of interest to the study population; aims, methods, monitoring study conduct and process; and designing and implementing dissemination plans to reach a number of audiences, including populations of interest.[91] These audiences are the backbone of PCORnet. They are also represented within other professional, governmental, policy, and community groups available to us outside of PCORnet, e.g., The Obesity Society, the American Academy of Pediatrics, the Centers for Disease Control and Prevention, the Centers for Medicare and Medicaid Services, the Administration's "Let's Move" campaign, and key community family health organizations. As part of our initial work to formulate this proposal we identified four key stakeholder groups critical to the success of our project. For the work of this project, we have created engagement teams comprising

1. *Patients and Caregivers*: parents and caregivers of young children, including those who were prescribed broad-spectrum antibiotics during infancy
2. *Health care Providers*: practicing primary care pediatric healthcare professionals, including physicians, physician assistants, and nurses
3. *Health care System or Organizational Leaders*: authorities on childhood obesity prevention/management and/or antibiotic use as well as organizational leadership within institutions that contribute to decisions regarding treatment, coverage, quality reporting, and/or clinical practice guidelines development and education
4. *Community and Advocacy Groups*: organizations outside of institutional healthcare systems that are authorities concerning community and public perspective on treatments in infancy and its implications for family community ecosystems

These four stakeholder group representatives are explicitly enlisted for the proposal overall. These include network participants of the Clinical Data Research Networks (CDRNs) and the Patient Powered Research Networks (PPRNs). Engagement is a key piece of the Secondary Aim as well (Section C).

B.2 Initial Engagement in Planning and Writing the Proposal. From drafting of the topic brief to PCORI in 2014, through the preparation of this proposal, we have involved stakeholders from the PCORnet CDRNs and PPRNs in developing and refining the research questions in this proposal. A systematic engagement process resulted in the plan and development of proposed activities. We integrated methods for incorporating multi-disciplinary stakeholder feedback throughout the planning and project period. From the earliest time of the topic brief development, and through the development of this proposal, at least one, and usually more, parents were involved. At the start of developing the full proposal, survey of the CDRN principal investigators elicited volunteers from the four stakeholder teams. The study team also engaged the PPRNs and invited nominations from their networks. Many of the networks nominated individuals who best fit the stakeholder descriptions and would be engaged during the proposal planning and writing period, be involved in study activities, and maintain a limited advisory role throughout the study period. We received eleven nominations. The proposal's investigative and stakeholder engagement teams selected eight people to represent all of the stakeholder groups. These stakeholders consist of parents, community pediatricians, health organization leaders, and community group representatives. These 8 individuals constitute the *Executive Antibiotic Stakeholder Advisory Group (EASAG)* (**Table 6**). We will include an additional parent/caregiver with lived experience and no professional role in the project to

be a member of the EASAG. This inclusion will result in 3 parent/caregiver members of the EASAG as compared with 2 from the other three stakeholder groups (provider, community/advocacy, health care systems leader), but given the primacy of the parent/caregiver experience we feel that this approach is appropriate. The other nominated individuals will be involved in other stakeholder activities.

Table 6. Members of the Executive Antibiotics Stakeholder Advisory Group (EASAG)

Category of Stakeholder	Organization	CDRN or PPRN	Name	Background
Patients and caregivers	Bank St. College of Education	NYC-CDRN	Jenny Ingber, PhD	• Parent of 3-year old who is participating in a microbiome study • PhD in science education
	NorthShore Univ Health System	CAPriCORN CDRN	Chelsea McKinney, PhD	• Research Manager, Center for Biomedical Research Informatics
Healthcare providers	NYU School of Medicine	NYC-CDRN	Mary Jo Messito, MD	• Pediatrician • Early childhood obesity disparities research
	Pennington Biomedical	REACHnet	Claire Hazlett	• Research pharmacist • Parent of 3 children
Community and advocacy organizations	Parent Project Muscular Dystrophy	Duchenne-Connect PPRN	Holly Peay, PhD CGC	• PI of PPRN whose population has obesity related to corticosteroid use
	LiveWell Colorado	PORTAL CDRN	Leslie Levine	• Runs organization leading statewide efforts in obesity prevention for Colorado
Systems leader	Rio Grande Valley Breastfeeding Coalition	GPC CDRN	Ivette Torres	• Childhood obesity – research and leading underserved community programs • Parent of 4 children
	Wake Forest School of Medicine	SCILHS	Joseph Skelton, MD	• Professor, Director of Brenner Families in Training Program • Pediatric obesity clinician and researcher

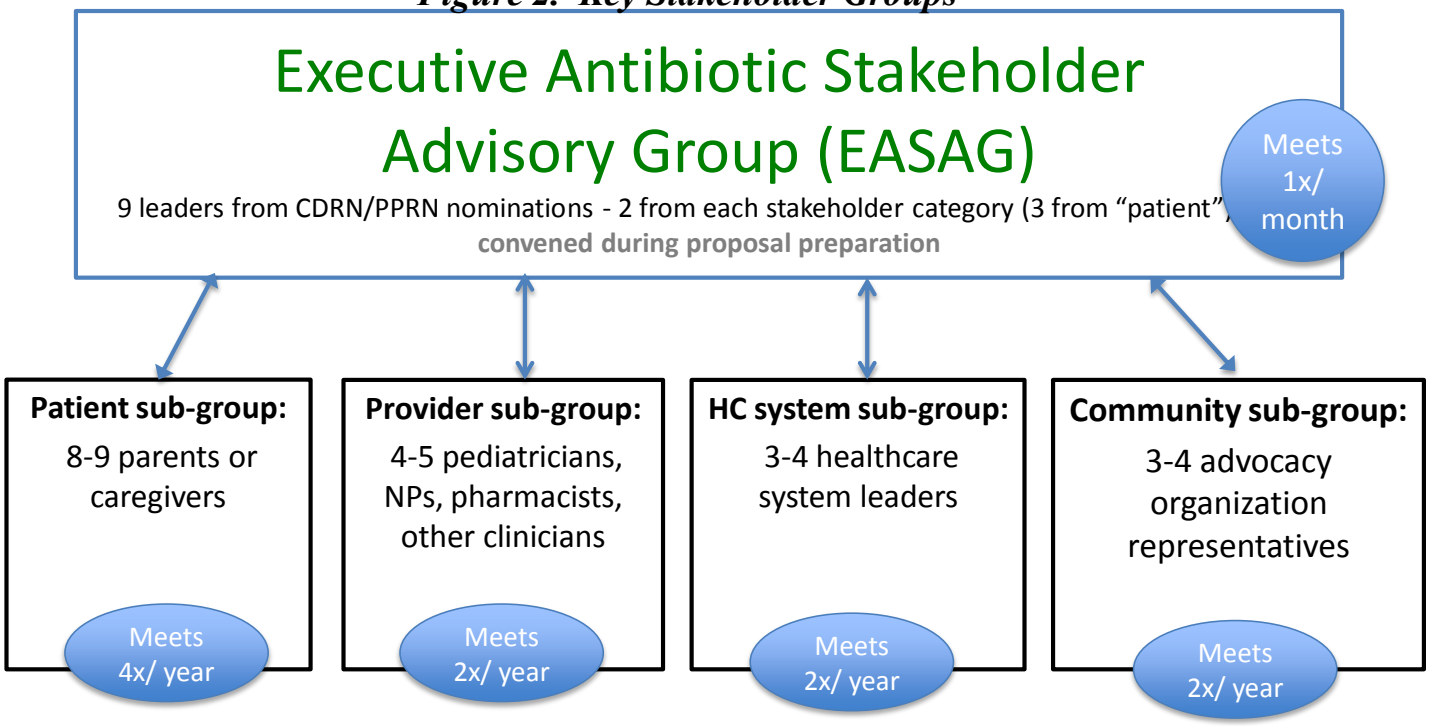
The EASAG took part in conference calls during the development of this proposal to discuss the basic plan for stakeholder engagement for the study (see below), as well as to provide feedback to the investigative team about the study aims, methods, and dissemination plans as they pertain to families and their systems of care. The initial meetings—as well as feedback received and ongoing communication by email—set the stage for continued rapport and communications to be maintained throughout the study period. Moving forward, the EASAG will be the core strategists in decisions and tactics for engaging participants from their respective stakeholder groups throughout the study. Letters of support from all EASAG members are included in the proposal. In addition to this input, Mr. Douglas Lunsford, one of the principal investigators of the proposed antibiotic study, is the parent participant in the PEDSnet Healthy Weight Network and has extensive experience in community engagement and advocacy.

The EASAG will be instrumental in working with the study team to ensure:

1. The study meets the needs of a truly representative population in their communities, including
    a. Diverse racial, ethnic, immigrant, and socioeconomic parent audiences;
    b. Parents and caregivers with various language and literacy skills;
    c. Diversity in family types, including caregivers other than biological parents (i.e. grandparents, foster parents);
    d. Coordination between the study team and the PCORnet Coordinating Center with the four Stakeholder Teams

2. Engagement with the well-established broad patient and provider networks as represented by the 10 participating CDRN communities and interested PPRNs, through community engagement in their geographic areas all over the United States.
3. Collaboration with community clinics and their patients through the varied representation of this stakeholder group.
4. Strategic development and dissemination of parent-friendly health communications materials about the ongoing research topics, throughout the project timeline, using appropriate literacy levels for professional and public audiences (see Dissemination section).

Figure 2. Key Stakeholder Groups



Sub-group members will be solidified after award

B.3 Stakeholder Engagement in Conducting the Study

The engagement structure for our proposed study is shown in **Figure 2.** The following section explains each group, its engagement, and its responsibilities while we conduct the proposed study.

B.3.1 Engagement of the Executive Antibiotic Stakeholder Advisory Group (EASAG). This group will function as a core study advisory group and the outcomes from group meetings will be highly valuable to the study team. In the proposal-writing period, the EASAG provided feedback regarding the study design and implementation. The EASAG will be very involved in study updates, findings, and dissemination to ensure relevance of study information and diversity of the population. It will have monthly meetings throughout the project facilitated by Sharon Terry, CEO of Genetic Alliance and citizen scientist with extensive expertise in patient engagement, community building, and PCORnet systems building (see Letter of Support). The EASAG will interact regularly with scientific team members including PIs Mr. Lunsford, and

Drs. Block and Forrest. Meetings will be conducted by telephone and webinar, to allow for easy nationwide participation. The EASAG will be responsible for designing and implementing the engagement plans for each of the four stakeholder teams. Each member of the EASAG will attend his or her respective stakeholder team meetings as a representative of the project and they will be responsible for assisting Ms. Terry and the scientific team in engaging each team. For example, the pediatric clinicians who are part of the EASAG will assist in the planning and engagement of the healthcare providers stakeholder team for the proposal.

B.3.2 Engagement of Patients and Caregivers. Participant/patient engagement is core to this PCORnet study. The CDRNs represent all of the major U.S. regions, with a reach of millions of individuals and families living in those areas. This stakeholder team, whose membership is already partly chosen through nominations from each CDRN and PPRNs, will be composed of four diverse members with a range of rich parental experiences that have informed their interest in this topic. The team will be engaged quarterly throughout the project, as well as use web-based resources such as message boards, community forums, and crowdsourcing to gather continuous feedback from the broader patient stakeholder group as the proposal develops, methods are developed and implemented, and preliminary results are made available. Parents/caregivers will also be critical in helping to identify methods of message dissemination and reach to vulnerable populations, including:

1. Racial and ethnic minorities (including Native Americans in isolated medical settings) and rural communities (also utilize partnership with HRSA regions).
2. Immigrants and caregivers with low levels of health literacy and/or low socioeconomic status (utilize social service providers serving these communities); and,
3. LGBTQ parent communities (utilize social service and adoption industries).

Quarterly meetings each year will be held as 90-minute webinars or phone calls for those who do not have access to the web. Updates will also be distributed by email and particular attention will be paid to stakeholders for whom email is not feasible or easily accessible.

Antibiotic patient and caregiver stakeholders will be paid $100 each for attendance at each quarterly meeting. They will not be incentivized for email/telephone correspondence or participation in any web-based communications. In the first quarter of the proposed study we will revisit the study aims and questions before beginning data collection. Subsequent meetings in the first year will serve to engage them in data collection and study refinement questions. In the second year of the grant, we will engage in understanding aggregate results using their input to help interpret results and solidify plans for dissemination.

B.3.3 Engagement of healthcare providers and health system/organizational leaders. This stakeholder group will be solidified through nominations from each CDRN and PPRN. The healthcare provider group, composed of four members, will contain broad cross-section of professionals involved in pediatric care and antibiotic prescribing and/or obesity including, but not limited to, a pediatrician, primary care provider, nurse, nurse practitioner, pharmacist, and hospitalist. An additional group will comprise two health system/organizational leaders, both individuals who work in national healthcare and are involved in decision-making regarding pediatric treatment guidance and leadership to provider and consumer audiences.

The combined group of providers and health system leaders will meet by phone for a 90-minute conference call twice per year during the study period. Between meetings, we will keep in regular contact with members of these two stakeholder groups, sending out updates every two months, but also contacting them with individual questions as-needed. In the first quarter of the proposed study, we will revisit the study aims and questions before beginning data collection. Subsequent meetings in the first year will serve to update them on data collection and refinement of study questions. In the second year of the grant, we will engage them around understanding aggregate results using their input to help interpret results and in dissemination.

B.3.4 Engagement of community and advocacy organizations. Participating CDRN leaders already have strong community relationships with trusted clinics, local condition advocacy organizations with diverse memberships, and local health departments and medical agencies. PPRNs also serve communities in these regions, and have wide-reaching constituencies from whom to draw, for both the primary and secondary aims. Local organizations are offered opportunities to learn and contribute to the trial before and after data completion and analysis. The study team will partner with participating CDRNs and PPRNs to offer bi-annual meetings for 3-4 key community stakeholders representing

advocacy groups and community organizations. Initial outreach to these stakeholders will include discussions about the study: its objectives and methods. Ideas and concerns from the community will be addressed and incorporated where feasible. This stakeholder group will also participate in dissemination of materials encouraging communication with, and recruitment of, individual patients and caregivers to the patient stakeholder group. This engagement will involve both community gatekeepers and members, representing grasstops and grassroots. The project will include strong participant-centric communities, and start envisioning methods for message dissemination once the study results are final and shared. Additionally, engagement of community and advocacy organizations ensures relationships, rapport and a sense of inclusion that will support sustainability for future participant- and community-centric research collaborations.

B.5 PCOR Engagement Principles

B.5.1 Theoretical Foundations Used for Engagement in the Proposed Study. We are using the PCORnet Methodology recommendations for engagement, including the Engagement Rubric,[91, 92] and especially regarding engaging the population of interest and other relevant stakeholders in ways that are appropriate and necessary in a given research context. Ensuring stakeholder comprehension of the study methods and results is critical to their role. In addition, we use the Community Based Participatory Research (CBPR) principles [93, 94] for patients/caregivers and community/advocacy groups. We will give careful attention to the architecture of the system we are building to be scalable, relevant and also durable beyond this proposal.

CBPR is distinct from community-based research in that although both are conducted in community settings, CBPR involves the stakeholders in all-important decisions about the research project. There is a particular emphasis on community assets rather than what communities lack when developing research questions. We will embrace the principles of CBPR by doing the following: recognize our patient partners and their families as a unit of identity, involve them in all phases of the research project, promote co-learning and empowerment among our patient and family partners (i.e. we have as much to learn from them as they do from us), use an iterative process in developing the project, create a shared dissemination strategy, and build a relationship with our patient partners and their families and the community/advocacy groups we engage that will extend beyond the grant to maintain the study cohort for future work.[93] This is consistent with all aspects of the Engagement Rubric as well as the fundamental principles of PCOR Engagement: reciprocal relationships; co-learning; partnership; and truth, transparency, and honesty.

B.5.2 PCOR Engagement Principles Applied to the Proposed Study.

1. *Reciprocal Relationships*: As described above, this study team values members of our stakeholder groups as critical research partners, and expect that they will continue to contribute in a meaningful way to study design and execution, as well as to assisting with dissemination of our results to a wider audience. While we do not expect any of our stakeholders to be expert scientific researchers, we do expect them to be "expert patients/professionals/leaders/advocates," helping us to understand their points of view, needs of the population being served, and experiences with the healthcare system, to better inform our research methodology.
2. *Co-Learning*: As noted above, we understand that we have a lot to learn from our stakeholders. The questions must be the ones that are critical to the stakeholders. Our EASAG and stakeholder teams will serve as experts on the pediatric experience from a number of viewpoints including the parent concerns, the provider interaction, the medication dispensed and attendant concerns, the institution, and the community. They will educate us on the best methods to engage families and clinicians, the outcomes and content important to them, and how best to communicate our findings to a wide variety of audiences. Meanwhile, to ensure that our study is scientifically rigorous, we will work with them to understand which of their many ideas can best be incorporated into the study design without compromising the quality of our findings. We have carefully chosen the different stakeholder groups based upon the expertise they bring to our proposed study. In all cases, these stakeholder groups provide the same value to the proposal as the scientific investigators. Without our stakeholders we could not hope to design a study that was meaningful and had any chance of being used to change guidelines, discourse, and provider-family decisions regarding antibiotic treatment in infancy.
3. *Partnership*: We will compensate our parent/patient stakeholders. We will also be sensitive to the needs of the various stakeholder groups and hold conference calls at times that work for their schedules. Though this often means evening calls for the parent team, this is planned to accommodate their needs. We will also hold one of the secondary aim focus groups in Spanish in an inner city community.

4. *Trust, Transparency, Honesty*: We plan to include all of our stakeholders in ongoing discussions not only about methodology, but also about how to interpret and disseminate our findings. We will create a safe and open environment for discussion where all stakeholders can honestly share their opinions and feel that their contributions are of value to the project. Meeting minutes will be distributed across all stakeholder groups so that they might understand the evolving perspectives and concerns of all of our partners involved in the planning, implementation, and dissemination of our proposed study. Decisions will be made with the patient principal investigator's input equal to the other principal investigators.

   We believe that all stakeholder contributors are critical research partners, and will look to their contributions as meaningful not only in study design, execution, and dissemination of our results to a wide, national audience, but also in examining how this study process will be refined over time to inform future network research efforts.

C. Secondary aim: Understanding How Parents' and Providers' Consider Antibiotics in Early Childhood

The aims of the overall study are to evaluate associations between taking antibiotics in early childhood and risk of excess weight gain and obesity later in childhood. If we find such associations, the application of this information to practice will require consideration of how decisions to prescribe (and take) antibiotics are made, and how the additional information from this study could be most helpful to parents and providers working together to make decisions about antibiotics. As summarized below, decades of prior research have identified a set of factors that might be considerations in these decisions, and some of the issues in parental-provider communication that can lead to overprescribing and consequent unnecessary risks of treatment. While the possibility of a link with future obesity is new, this potential risk must be integrated into the discussion of currently understood benefits and potential harms of antibiotic use for common infections in early childhood.

Shared decisions about antibiotic use. Over many decades, physicians largely underestimated the longer term individual and population-level consequences of antibiotic use, and overestimated the short term benefit on the patient in front of them—often not just on the course of illness but also meeting perceived or explicit parental desires. Now, there is widespread appreciation of the public health consequences of antibiotic overuse by both clinicians and the public. While antibiotic use rates for children in the U.S. increased dramatically in the 1980s with the highest rates among young children,[95] our analysis in the HMO Research Network showed cumulative decreases of 20%-25% in rates of antibiotic prescribing over a more recent 5-year period.[96] The CDC's National Ambulatory Care Survey showed a similar trend.[97] This decline was due, in part, to attention to the issues of resistance and antimicrobial overuse in the lay and professional press, and to public campaigns by CDC and others[98-101] to promote more judicious prescribing. In the absence of clear changes in infection rates, the variation in prescribing rates (in the US and between nations)[102] suggests that for common infections, antibiotic use is, at least somewhat, a preference-sensitive decision and appropriate for shared decision making.[103]

Parents bring to the encounter immediate concerns about their child's health, and a set of expectations and desires with regard to treatment.[104-108] These may vary by parents' educational level, cultural factors, and prior (and ubiquitous) experience with common infections themselves. Physicians also enter an episode of care with their own knowledge, judgments, and biases.[109, 110] In reality, this decision is not difficult in cases of serious bacterial infections such as meningitis, significant pneumonia, and bone and joint infections where antibiotics may be life saving. Rather, the decisions are most difficult when antibiotics are being considered for relatively minor infections, many of which resolve without treatment.

Specific elements of these decisions may promote overuse by clinicians. Clinical differentiation of bacterial disease from prolonged or severe viral symptoms is often difficult, resulting in some prescriptions written "just in case." And, parent demand for antibiotic treatment for viral upper respiratory illnesses is frequently perceived as pressure by clinicians.[107]

However, the vast majority of the increase in prescriptions for young children follows an increase in the reported diagnosis of otitis media (ear infections).[111] Otitis media (OM) is both the most common reason for antibiotic prescribing in early childhood, and brings into focus many of the issues that come into play in shared decisions. Treatment for OM likely speeds up recovery (by approximately one day) and slightly decreases the risk that the infection will involve the other ear or not resolve, which requires further treatment.[112] Estimates from some studies suggest that of 7-20 children who are treated with an antibiotic for an ear infection, 1 will benefit, depending on the outcome used.[112, 113] However, there is still controversy about impact on middle ear effusions and the wisdom of "watchful waiting," an option in recent guidelines.[114-116] Other risks of antibiotic use include side effects (diarrhea most common), allergic reactions, and development of antibiotic resistant bacteria in the individual (transiently for most children) and the population. This proposal may add the risk of future obesity into the assessment of benefits and risks considered by parents and providers. This potential risk is important because of the ubiquity of excess weight in the US population. Approximately 1/3 of children and adolescents age 6-19 years are overweight or obese, with the highest rates among non-whites.[117]

The Institute of Medicine report, "Communicating with Patients on Health Care Evidence," highlights the importance of providing patients and parents with the best available medical evidence for both the risks and benefits of treatment as part of supporting high-quality medical decision making.[118] However, effectively communicating risks and benefits is challenging due to both social factors[119] and challenges in understanding the mathematical concepts associated with assessing medical evidence. [120, 121] This challenge becomes even greater when comparing a current benefit to a potential distant risk such as obesity. When individuals make decisions, short-term gain, such as symptom relief from antibiotics, is often valued even if it comes with later risk[122] such as obesity, although it is unknown if this pattern holds for parents making decisions for their children. We do not know the content and delivery mode of information needed by parents and clinicians to effectively balance the future increased risk of obesity against the current real (albeit small) benefit of treatment, or how they might want such information presented. However, questions such as the following have been generated in the development of this proposal by the study investigators and parent and clinician stakeholders:

a. What is the expected benefit of the antibiotic being considered? And, is that benefit related to providing comfort (pain reduction) or shortening the length of illness for the patient, or is it tied to reducing the risk of long-term or added health outcomes?
b. How much added risk of obesity is there from one course of antibiotics? To what degree does the risk of overweight or obesity increase with each successive course of antibiotic treatment prescribed?
c. Does that risk vary by which antibiotic is prescribed?
d. How far in the future would weight be affected? To what degree is weight affected?
e. Is the effect permanent or transient?
f. [For future research:] Can the risk be mitigated by other actions (diet, physical activity) in the interim? If so, what level of dietary restrictions and physical activity are necessary for these factors to have a significant impact in deterring the potential risk of obesity accrued from antibiotic treatments?

The overall goal of this Secondary Aim is to understand how patients (parents and caregivers, in this case) and their providers assess information related to current and future benefits and risks, particularly for treatments with potential near-term benefits, but potential long-term risks. This includes exploring how information should be best presented to help parents understand the available data and its limitations.

Some of the issues we seek to understand include:

- How do parents assess the value and limitations of data related to current and future risks of antibiotic treatment? How does this vary by parent and family characteristics? How do both the magnitude and certainty of associations from these studies affect decisions?

- How does the future risk of obesity enter into this decision, and how would parents prefer this information presented? How does this vary by parent (and child) characteristics, including such factors as prior experience with antibiotic use (for themselves and their children), educational attainment, health literacy, parent overweight and obesity status, and elements of the provider/parent relationship?

- How do providers anticipate using data from studies such as this one in sharing decisions about antibiotic treatment?

Overview of Methods for Secondary Aim. Analysis of focus group discussions can provide awareness of the psychosocial dynamics and context, as well as knowledge, beliefs, and attitudes among participants. They can be particularly useful to understand differences among groups defined by race or ethnicity, economic status, or educational attainment. We will work with 4 selected CDRNs and PPRNs to conduct 8 focus groups, 2 in each of 4 locales, to gain insight into the communication channels and information that parents seek to help them make decisions regarding antibiotic use and future risks including about childhood obesity. The topics of discussion will be co-developed by the overall study EASAG, as well as researchers, as a semi-structured moderator's guide. To gather the perspective of prescribing clinicians, we will also conduct up to 20 semi-structured telephone interviews with pediatric physicians and nurse practitioners who commonly prescribe these agents to young children. The results of this qualitative analysis, results will be disseminated to stakeholders. We will then solicit suggestions on how to best incorporate these results in future antibiotic decision support materials.

To recruit participants for the focus groups, the study team will reach out to the participating PPRNs who initially expressed interest in participating, as well as to the 10 CDRNs and 2 new PPRNs to gauge their interest. We will determine which ones have the resources and capabilities to recruit and lead focus groups, including prior experience with recruitment, obtaining consent, and conduct of focus groups, including the technical aspects of tape recording to allow transcription, as well as ability to recruit from a general population rather than families of children with a particular health condition. Those PPRNs/CDRNs that have such capabilities and are interested will meet with a study team member to discuss the characteristics of the focus group that PPRN/CDRN would host. The study team, including stakeholders, will select PPRNs/CDRNs in a manner that balances the sociodemographic characteristics of participants (insurance type, child's age, geographic location, race/ethnicity) across the entire study, as well as experiences as parents of a children with chronic illness.

For the healthcare provider interviews, the study team will contact PPRNs and CDRNs to gauge their interest in participating and their ability to recruit 2-3 non-academic pediatric primary care providers. We will choose participating groups with the goal of achieving diversity of geography, practice type, years of experience, and race/ethnicity. We will continue recruiting and conducting interviews until we reach content saturation in key topic domains.

Focus Groups

Recruitment. We will conduct 2 focus groups in each of 4 different geographic locations (total of 8 groups). Each focus group will have 8-12 parent participants who will come together for one 2-hour session. English-speaking parents of children under age 7 will be recruited for participation. Only one parent per family will be eligible to participate. Moderators will be trained in appropriate discussion moderation and probing, and will have experience in translating medical information into appropriate lay language for low-health literacy audiences. Each CDRN or PPRN will recruit parents, including some with infants, using methods they deem appropriate in their setting. These may include electronic communications, placing posters and informational letters in waiting rooms of pediatric practices, outreach to childcare centers, or other family-centered locations in the community. We will aim to recruit a socio-demographically diverse group of participants, including parents with and without children with special health care needs. There will be a particular focus on recruiting families with varying levels of education and health literacy. A dedicated phone line and email for prospective participants to call for additional information and to sign up will be provided. An informed consent document, as well as a confirmation letter and directions, will be mailed to prospective participants. . The research assistant will either call, text, or email a reminder to the participant at least 3 days before the focus group. The consent will also be given to the parents for signature when they arrive for the focus group. We will provide a copy to participants and keep a copy for our records. Although our goal is conduct recruitment solely by posters and informational letters, if we do not have enough participants two weeks prior to a scheduled focus group, the research assistant may conduct in-person recruitment at community sites. We will provide a monetary incentive of $50 per participant, as well as a light meal.

Conducting the focus groups. Each group will be led by a professional (non-physician) moderator using a moderator's guide that will be co-developed with parent partners related to parents' perceptions of the risks and benefits of antibiotics. The guide will open with general questions aimed at understanding parents' prior experiences with antibiotics, as well as the perceived risks and benefits of their use. We will then inquire about parents' perceptions and concerns about childhood obesity, including the magnitude of the problem as well as contributing factors.

Following these general questions, participants will be presented with vignettes for discussion. Vignettes will provide a description of a typical infant or toddler with an ear infection. We will provide a general description of the well-established risks and benefits of the use of antibiotics and then explore the possibility of an increased risk of obesity with use of antibiotics. We will then ask parents to discuss the vignette and what information they would need to make a decision in each situation. We will include discussion probes designed to elicit the degree of information detail parents want, how they would like information provided to them and when, and the role they wish to have in decisions about antibiotic treatment. We will specifically explore how they interpret data from scientific studies on this topic, and how they consider the strength of the associations reported, the level of certainty of those results, and their relation to assessment of future risks.

In addition to the moderator, a research assistant who will not participate directly in the discussions will be present. The research assistant (RA) will help in managing the logistics of the discussions and take notes regarding key verbal and non-verbal reactions of participants to individual topics. All discussions will be audio-taped for later transcription and analysis. The groups will be semi-structured and are estimated to require 90 minutes. At the end of each group, we will ask participants to complete an optional brief survey asking general ethnographic questions such as their age, ages and number of children, racial/ethnic background, educational attainment, primary language spoken at home and a brief assessment of health literacy.

Interviews

Recruitment. We will conduct 20 in-depth interviews with healthcare providers involved in prescribing antibiotics to children (pediatricians, nurse practitioners or family physicians). Potential participants will be recruited by collaborating organizations (CDRNs and PPRNs). A dedicated phone line and email for prospective participants to call for additional information. A research associate will arrange a time for the telephone interview and then mail the participant a consent document and confirmation letter. The research assistant will either call, text, or email a reminder to the participant at least 3 days before the interview. We will provide a monetary incentive of $50 per participant.

Conducting the interviews. Interviews will be conducted by telephone from a central site by an interviewer with experience interviewing health care providers using an interview guide that will be co-developed with stakeholder partners related to provider's perceptions of the risks and benefits of antibiotics, as well as their approaches to working with families to make antibiotic choices. After obtaining verbal consent, the interview will open with general questions aimed at understanding providers' perceptions of the risks and benefits of antibiotic use. We will then inquire about their perceptions and concerns about childhood obesity, both the magnitude of the problem of obesity as well as earlier factors that contribute to it.

Following these general questions, participants will be presented with a vignette about of a typical infant or toddler with an ear infection. We will then ask the provider to discuss the vignette and how they would approach making the decision about antibiotics. If the risk of obesity is not highlighted by the provider, we will then ask how such a risk would alter their decision making approach. We will include discussion probes designed to elicit the degree of information they provide to parents and the role they want parents to have in decisions about antibiotic treatment.

All interviews will be audio-taped for later transcription and analysis. The interviews will be semi-structured and are estimated to require 30 minutes. At the end of each interview, we will ask participants to complete an optional brief survey asking general demographic questions such as length of time in practice, practice characteristics and personal demographic characteristics such as age and racial/ethnic background.

Data analysis. The audio recordings will be professionally transcribed and the transcription verified by the RA who attended the focus group or interviewer who conducted the interview. Data will then be qualitatively analyzed by a team that consists of both researchers and parent partners. We will perform a content analysis of the transcribed focus group discussions, incorporating the principles of the immersion-crystallization method.[76] This qualitative approach consists of repeated cycles of immersion into the collected data with emergence of an intuitive crystallization of the dominant themes.[76] Through this immersion process, the stakeholders will identify salient themes. These themes will be used as the basis of coding, by two members of the research team. We will attempt to identify important themes expressed by parents and the differences among the demographic groups represented. We will continue analysis until no new major themes emerge. Throughout this process we will also look for the explicit use of heuristics, decision-making shortcuts in reasoning, used by parents or providers.[126] Direct quotes from the transcribed focus groups and interviews will be used to

illustrate the findings. We will address recognized criteria for qualitative research: credibility, fittingness, auditability, and confirmability. We will establish credibility by including the moderator and research assistant in discussions to determine the concordance between derived themes and the live observer's experiences. In addition, having multiple investigators and parent representatives read the transcripts and listen to the groups will enhance credibility. Auditability will be established through using the audio tapes and notes taken during the discussions, as well as complete individual summary reports on each session produced by the RA. Through iterative discussions among the study team (including parents), these qualitative findings will be used to enrich our underlying conceptual model and inform future development of resources that promote shared decision-making that effectively weighs risks and benefits of the treatment, including parents' values regarding particular outcomes..

Dissemination plan. The conceptual model developed through qualitative analysis will be disseminated to both study participants and stakeholders. A printed brochure outlining the study findings and implications will be mailed along with a brief survey and stamped envelope to solicit feedback from participants and stakeholders. This feedback, along with the conceptual model, will be used in the future to develop tools for facilitating shared decision making about antibiotic treatment.

D. Project Milestones and Timeline

The study will take two years to address our scientific aims, appropriately engage stakeholders, and conduct an overall evaluation of the functionality of PCORnet and its research readiness. Start-up activities, including executing subcontracts and obtaining IRB approval will take approximately 4 months. We will hold an in-person study kick off meeting early in year 1, bringing members of the Core Scientific Team, Coordinating Center, and Executive Antibiotic Stakeholder Advisory Group together. Other key milestones over the course of the project include the development of a formal CC evaluation plan, and reports of interim and final study analyses. We will hold an in-person study closing meeting at the end of Year 2.

We will require up to one year after the study period has ended to meet remaining study milestones. The final evaluation and study reports will be submitted within 60 days after funding has ended. We will have all manuscripts accepted for publication and data sets available for request within one year following study completion. Formal plans for submission of all manuscripts and making datasets public availability will be provided with the final report. More information can be found in the milestones section of PCORI's online application system, in the table below**.**

Table 7. Timeline of Activities (each cell represents 2 months)

Milestones and Deliverables	**Year 1**						**Year 2**						**Post-study completion (1 year)**					
Study kick-off meeting held																		
CC submits formal evaluation plan																		
All subcontracts executed																		
Institutional Review Board (IRB) approval																		
CC evaluation: submits interim reports																		
Study: Interim progress reports																		
Cohort identification																		
Interim analyses																		
Final analyses																		
Study closing meeting: Final results																		
Focus group results																		
CC evaluation: Final Report																		
Study: Final report																		
All manuscripts accepted for publication																		
Data sets, analytic files, and codebooks																		

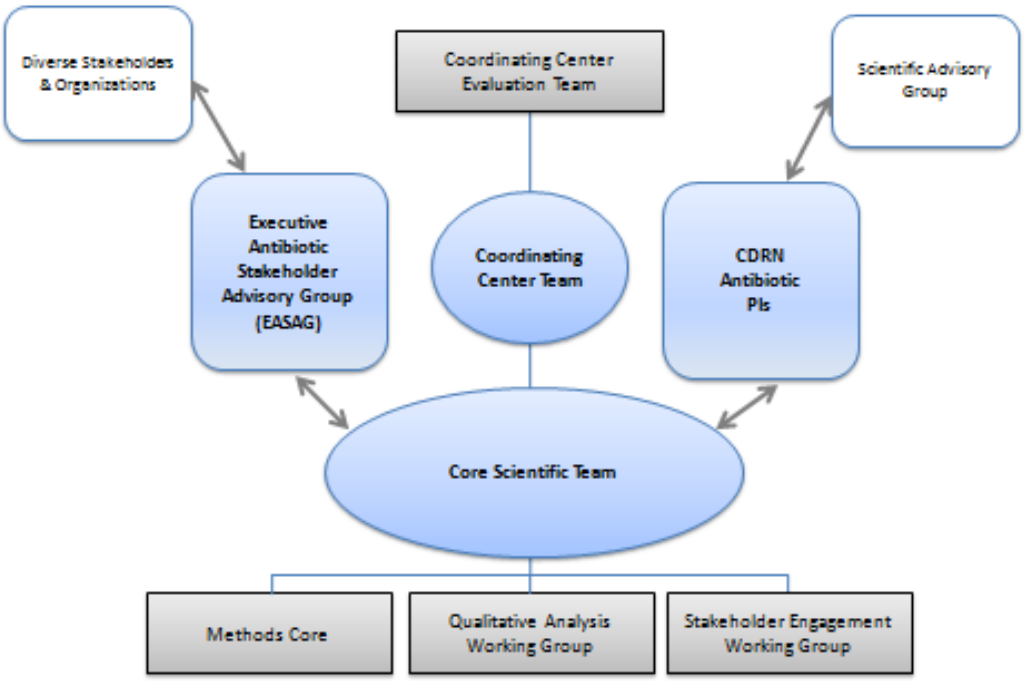
E. Research Team and Environment

We have developed a clear organizational structure that engages a core team of multidisciplinary researchers and stakeholders while also facilitating input broadly from CDRN and stakeholder partners (**Figure 3**). The **Core Scientific Team** is made up of the three Principal Investigators (Block, Forrest and Lunsford) whom PCORnet selected to lead this project, as well as the leader of the Methods Core (Sturtevant), a clinical informatician (Bailey), three biostatisticians (Li, Bryan, Kleinman) and expertise in antibiotic outcomes including obesity (Finkelstein, Trasande, Bailey). Other members of the Study Team have expertise in qualitative methodology and engagement (Lipstein, Terry), PCORnet infrastructure (Brown), and longitudinal and distributed data analyses (Brown, Li, Toh). The research team includes one parent (Lunsford), and it incorporates clinician, researcher, medical, epidemiologic, and statistical perspectives. The multidisciplinary team that we have assembled has the necessary clinical, operational, and methodological expertise (**Table 8**), and has laid important groundwork in the past three months to help ensure the successful execution of this study. Members of the study team have established highly productive collaborative relationships over the years through

numerous ongoing and completed multi-center projects. See Management Approach for additional detail and an Organizational Chart showing functional relationships among these and other team members.

The Core Scientific Team will work in close collaboration with the PCORnet CC led by Dr. Jeff Brown to rapidly establish working groups (**Figure 3**) that will focus on quantitative methods, qualitative analyses, and stakeholder engagement/ dissemination. The Core Scientific Team will be advised regularly by the Executive Antibiotic Stakeholder Advisory Group, and by the Science Advisory Group comprising CDRN Antibiotic PIs (**Table 9**), which includes researchers with experience in pediatric health services research, antibiotic trends, obesity, medical informatics, and multisite collaborative research. As we have described in Section C above, we will also engage a larger group of diverse

Figure 3. Overall relationships among study components



stakeholders**,** whom we will consult on a regular but less frequent basis. Similarly, while each CDRN will be represented by only one CDRN Antibiotic PI, we recognize that their constituent sites include additional researchers with an interest in engaging with the study. We will invite them to join the Scientific Advisory Group, to provide input to the design and conduct of the study.

Our Core Scientific Team will benefit from an ongoing close working relationship with the study team for the companion Obesity Observational Research Initiative proposal on long-term effects of bariatric surgery. To take advantage of potential efficiencies, we have created a virtual Methods Core that will work across both studies. This Methods Core will involve members of both study teams, including several (Drs. Block. Li, Toh) who are working on both studies, and will be the primary connection, along with study-specific analytic staff, with the CC team to develop and refine distributed queries and to develop infrastructure processes to facilitate future PCORnet observational research (see Management Approach for more detail).

The investigators' research institutions, in this project led by Harvard Pilgrim Health Care Institute (HPHCI**),** are also uniquely well-positioned to conduct the proposed study. HPHCI is a national leader in developing innovations in multisite research methods and governance. HPHCI leads the PCORnet Coordinating Center (Phase I), serves as the operations center for the FDA Sentinel program, and co-leads the Coordinating Center for the NIH-funded Health Care Systems Research Collaboratory. HPHCI maintains ongoing collaborative relationships with all the participating sites through other multi-center projects. HPHCI will support this project through its leadership in grants and contracts management and IRB review to insure timely execution of all of the study's administrative s tart-up activities. Further,

HPHCI will bring to bear a wealth of data and analytic resources, including expertise in programming and analysis of distributed research studies.

Table 8. Key personnel and expertise		*Roles in Study*
Jason Block, MD, PhD (PI) HPHCI	• Obesity health services and policy • Population medicine	• Overall leadership • Core Scientific Team
Chris Forrest, MD, PhD (PI) CHOP/PEDSnet CDRN	• Theory/measurement in childhood • Outcomes and effectiveness	• PCORnet leadership • Core Scientific Team
Doug Lunsford, BS (PI) PEDSnet CDRN	• Patient/parent/caregiver perspective • Educational innovation	• Engagement Leadership • Core Scientific Team
Charles Bailey, MD, PhD CHOP/PEDSnet CDRN	• Clinical informatics, incl. medications • Learning health care systems	• Core Scientific Team • Methods Core
Jeff Brown, PhD HPHCI	• Distributed health data networks • Sequential analytic methodologies	• Coordinating Center leadership
Matt Bryan, PhD	• Biostatistics • Growth trajectory analysis	• Core Scientific Team
Jon Finkelstein, MD, MPH HPHCI/PEDSnet	• Pediatric health services research • Antibiotic trends and determinants	• Qualitative Group leadership
Ken Kleinman, ScD HPHCI	• Missing data • Longitudinal analysis	• Core Scientific Team
Ellen Lipstein, MD, MPH Cincinnati Children's Hospital	• Pediatric health services research • Qualitative and quantitative methods	• Qualitative Group leadership
Sharon Terry, MA Genetic Alliance	• Transforming health systems for research • Stakeholder engagement	• Engagement leadership • Qualitative group
Darren Toh, ScD HPHCI	• Privacy-preserving analytic methods • CER and safety therapeutics	• Methods Core
Leo Trasande, MD, MPP NYU	• Clinician and epidemiologist • Antibiotics and obesity	• Core Scientific Team
Jessica Young, PhD HPHCI	• Simulation studies • Causal interference	• Core Scientific Team • Methods Core

Table 9. Scientific Advisory Board (CDRN Antibiotics PIs) and their Expertise			
CDRN	**Lead Site**	**Lead Site PI**	**Expertise**
ADVANCE	Oregon Community Health Information Network (OCHIN)	Janne Boone-Heinonen	Community influences on obesity and chronic disease
CAPriCORN	Northshore University Health System	Goutham Rao	HIT to improve delivery of preventative services
GPC	Univ. of Texas Health Science Center – San Antonio	Maria Rayas Alex Bokov	Development, nutrition, and early growth Informatics
Mid-South	Vanderbilt University	William Heerman	Health literacy and obesity
OneFlorida	University of Florida	David Janicke	Childhood obesity and dissemination of obesity interventions

NYC-CDRN	New York University	Melanie Jay	Behavior change in obesity
PEDSnet	Children's Hospital of Philadelphia	Ihuoma Eneli	Advocacy, prevention, medical weight management
PORTAL	Kaiser Permanente Colorado	Matthew Daley	Pediatric-focused outcomes in CER
REACHnet	Pennington Biomedical Research Center	Daniel Hsia	Childhood prevention of diabetes and obesity
SCILHS	Massachusetts General Hospital	Elsie Taveras	Early childhood obesity prevention

The study will include an evaluation component to provide an objective, ongoing assessment of the progress of the study and report on the readiness of PCORnet's infrastructure and the use of the DRN. The PCORnet CC will lead this effort and will report on the solutions implemented to resolve technical issues and evaluate the functionality of the network. The PCORnet CC's "evaluation team" will include an evaluation lead, project manager, and research assistant.

The evaluation will be conducted in three phases. The *introductory phase* will include refinement and initiation of the program evaluation plan and performance metrics (described in the following sections). The CDRN and PPRN PIs will be engaged during this period to review and refine the proposed evaluation plan and metrics. The PIs are key stakeholders as well as the leaders of the activities under evaluation; therefore, their contribution and buy-in is critical to the success of the process. The *formative evaluation phase* will consist of one-on-one qualitative interviews with members of the study team, CC personnel working on the individual studies, and lead site PIs of participating CDRNs; observations of calls and face-to-face meetings; analysis of documents provided by the CC and study team; and assessment of performance metrics. Finally, the *summative evaluation phase* will include an overall assessment of performance, including technical and operational achievements and challenges and suggested corrective actions. During each phase and for each priority area (listed below), a process evaluation will assess the development and dissemination of protocols and relevant information by the CC and the study team. An outcome evaluation will assess the effectiveness of the selected approaches, with particular attention to performance metrics.

The following sections describe the specific technical and operational priority areas of this evaluation, including the associated performance metrics by which each piece will be measured.

I.1 Evaluation of operational processes

The evaluation team will assess the CC's ability to operate as a multi-center distributed research network by assessing the degree to which the following are realized:

- The CC provides leadership, resources, and tools, and supports the fundamental requirement of coordination and productivity among obesity experts, programmers, CDRNs, PPRNs, and the PCORnet Coordinating Center Team (as assessed through one-on-one interviews)
- The milestones are completed in accordance with the proposed timeline
- The CC facilitates, manages logistics, and provides the necessary administrative support required by study participants (as assessed through documentation, such as Standard Operating Procedures, etc.)
- The principles of stewardship, creation of community value, and generation of products directly beneficial to the PCORnet research enterprise are modeled (as assessed through one-on-one interviews)
- The PCORnet Policies encourage collaboration and align with the study's aims and scope of work (as assessed through one-on-one interviews and observations of meetings)
- The participating institutions' governance policies align with the study's aims and scope of work (as assessed through one-on-one interviews and documentation)

In the report, the evaluation team will document accomplishments as well as conflicts and impediments to efficient and successful participation as assessed through the above key metrics, providing suggestions for specific corrective actions.

I.2 Evaluation of involvement of stakeholders in network oversight, use, and sustainability

The evaluation team will assess the stakeholder engagement strategy and describe the roles and activities of the various stakeholders throughout the project. In particular, they will focus on the roles and activities of the following groups:

1. *Patients and Caregivers*: These are parents and caregivers of young children who were prescribed and/or took broad-spectrum antibiotics during infancy.

2. *Healthcare Providers*: These stakeholders are community and other practicing pediatricians and pediatric healthcare professionals important to the proposed study (nurses, physician assistants) as well as those involved in the care and prescription of antibiotics to infants and young children.
3. *Healthcare System or Organizational Leaders*: These are authorities on childhood obesity prevention/management and/or antibiotic use and organizational leadership who make decisions regarding treatment, coverage, quality reporting, and/or clinical practice guidelines development and education.
4. *Community and Advocacy Groups*: These are organizations outside of the participating institutional healthcare systems that are authorities concerning community and public perspectives on treatments in early childhood and its implications for family community ecosystems.

The evaluation team will assess and summarize the degree to which the following key indicators are achieved, noting suggested corrective actions, as needed:

- The study meets the needs of a truly representative population in their communities, including:
  - Diverse racial, ethnic, immigrant, and socioeconomic parent audiences;
  - Parents and caregivers with various language and literacy skills;
  - Diversity in family types, including caregivers other than biological parents (i.e. grandparents, foster parents); and,
  - Coordination with the four stakeholder groups.
- Engagement with the well-established broad patient and provider networks as represented by the nine participating CDRN communities and interested PPRNs, through community engagement in their geographic areas all over the United States.
- Collaboration with community clinics and their patients through the varied representation of this stakeholder group.
- Strategic development and dissemination of parent-friendly health communications materials about the ongoing research topics, throughout the project timeline, using appropriate literacy levels for professional and public audiences

1.3 Evaluating contracting and IRB processes

The evaluation team will assess the efficiency of the contracting process between PCORI and the prime study site (HPHCI) and of the subcontracting process between the prime sites and subcontractors (i.e., participating CDRNs). The evaluation team will focus on measuring and reporting achievements and barriers related to:

- PCORI and HPHC's (prime institution) contracting process:
  - Proficiency and timeliness associated with executing the contract
  - Extent and timing of contract revision requests made by both parties
  - Time from contract initiation to contract execution
  - Degree to which PCORI and the CC provide effective and timely information to HPHC's Office of Sponsored Programs to encourage efficiency
- HPHC and the participating CDRNs' subcontracting process:
  - Proficiency and timeliness associated with executing each of the subcontracts
  - Extent and timing of contract revision requests made by all parties (documented for each CDRN)
  - Time from contract initiation to contract execution for each CDRN
  - Degree to which the study PM provides effective and timely information to HPHC's Office of Sponsored Programs to encourage efficiency
  - Whether and how a lead contracting CDRN site that is not serving in that lead in PCORnet Phase I affects the above measures

The evaluation team will also assess the process by which the CC develops, implements, and refines its IRB strategy. The evaluation team will work closely with the CC to determine whether the CC is able to successfully:

- Implement a centralized IRB approach which is appropriate for the study design, while appropriately addressing ethical and regulatory aspects of the research
- Obtain approval from the participating institutions' local IRBs to adopt the selected centralized approach and assess the timeliness of that process
- Assess the feasibility of the IRBShare approach based on the study protocol and quickly provide alternative solutions if IRBShare does not meet the needs of the study (including whether the research is "not human subjects research")
- Establish a streamlined, centralized IRB process to leave behind for future PCORnet studies

The evaluation team will report on the efficiency of the IRB process and will note the challenges and barriers related to use of a centralized review process.

I.4 Evaluating CDM 3.0 – Data standards, query functionality, security, and network infrastructure

The evaluation team will prepare a report on the readiness of participating sites' data sets as stored in Common Data Model Version 3.0 (CDM V3.0), data characterization activities, and data quality processes. The team will also report on the readiness of the data sets at the participating sites and list issues encountered at participating sites, solutions deployed, and time to resolution by assessing:

- Whether or not the CC maintains the PCORnet CDM V3.0 in accordance with the PCORnet CDM Guiding Principles
- The degree to which participating CDRNs have their data organized in accordance with PCORnet CDM v3.0
- The degree to which PCORnet CDM V3.0 provides the necessary data elements for the study
- The efficacy of query federation
- The quality and completeness of data and metadata based on queries from the CC that are generated for the project.

I.5. Evaluating the utility and efficiency of the DRN and the capacity of PCORnet CDRN node sites to conduct unmodified queries

The evaluation team will work to determine the functionality and efficiency of the PCORnet infrastructure, including any institutional barriers to participation. Additionally, the evaluation team will determine whether or not each participating CDRN is able to answer the queries identically and if sharing of confidential/proprietary information is avoided when possible and minimized at all other times.

The evaluation team will assess the CC's functions as they relate to the PCORnet DRN Query Tool process, including compiling and processing information provided by the Core Scientific Teams (query requestor), initiating the PCORnet Query, managing the querying process, and providing results to the Core Scientific Teams. In addition, any consultative support, with respect to issues of data quality, provision of meta-data, or preparation or review of queries, will be assessed.

The PCORnet DRN is built to serve as a secure infrastructure that will enable partners to (1) securely share information with one another and with external researchers, (2) allow distributed querying of local data resources to inform study design and conduct analyses while minimizing exchange of PHI, and (3) allow efficient reuse of research data sets by new investigators so they can perform analyses on these data. The legal, regulatory, privacy, and proprietary concerns of health systems with regard to data sharing often present insurmountable barriers to participation in multisite clinical research. The extent to which the DRN reduces the need to exchange PHI through the use of a distributed data model, thereby helping to overcome these barriers and expand the pool of clinical systems willing to participate in multicenter research, will be assessed.

The evaluation team will review each step of the query process with the participating study team members, CC programmer, and CDRNs to understand the successes and challenges encountered at each key phase of the process. The process is described below.

In collaboration with the Methods Core members, the study programmer[1] will develop functional specifications[2] and technical specifications[3] for each query that the team wants to run related first to the data characterization activities and then to the study aims. The CC programmer[4] will review the functional specifications with the study programmer and scientific team. Once the functional specification is completed, the study programmer will develop detailed technical specifications that will also be reviewed by the CC programmer. Once the technical specification is complete, the CC programmer will write the queries for distribution across the study participants.

For every distributed query required for the study, the following tasks will be completed by the CC programmer, with input from the Methods Core and scientific team programmer:

1) Develop the distributed code in a reusable format. Using the final technical specification, the CC programmer will write the distributed analytic code in SAS®. The goal is to develop reusable analytic code that includes a set of customizable parameters (i.e., macro variables) usable in a wide range of study settings,
2) Test the code locally to ensure it performs as expected using the PCORnet CDM v3.0,
3) Pilot the code at 1 CDRN (PEDSnet) and send results to the study team programmer with some iteration between teams and CC on the results/code,
4) Query all the participating DataMarts via PCORnet DRN Query Tool at all participating CDRNs and send results to the study team,
5) Review the results and program logs to ensure the query ran properly at all sites, and
6) Report on the readiness of the PCORnet infrastructure, listing all issues encountered at participating sites and solutions implemented to resolve technical issues.

In addition, the evaluation will include an assessment of the unique features of the antibiotic study including the quality (and specific limitations) of the data for the specific population under study (e.g. children), linkage to maternal data, and the prescription data, including but not limited to the following:

Assessing quality and limitation of the data

- Frequency distributions and descriptive statistics (means, variance, skewness, etc.) to assess the overall characteristics of the data, as well as differences across CDRN participating node sites
- Completeness of data, especially as it pertains to children. Data elements requiring special attention include Date of Birth (to calculate age at various events), Length/Height and Weight (to assess BMI and dosing)
- Validity of data, including face validity of anthropometric and antibiotic data given clinical covariates (e.g. age, overall clinical utilization, concurrent diagnoses), and external validity of data against existing widely-accepted benchmarks (e.g. NHANES survey for growth data, AAP recommended follow-up frequency for infants)
- Characterization of the cohort over time to assess loss-to-followup and identify covariates may explain why children leave or stay in the cohort over time

Assessing maternal/child linkage

- Proportion of child records link to their respective maternal records (linkage rate)
- Validation of linkage on a tractable subset from each link type (linked/non-linked)
- Frequency distributions and descriptive statistics of data elements, comparing the linked with non-linked records to identify those elements that are associated with linkage and non-linkage

---

[1] **Study programmer:** a SAS programmer who is a member of the scientific team and has familiarity with the PCORnet CDM and DataMart source data. This person works with the Coordinating Center programmer to define functional and technical requirements and specifications for the study programming effort.

[2] **Functional specifications:** a high-level description of the program purpose and scope (e.g., propensity score matching for cohort studies).

[3] **Technical specifications**: detailed description of the exact programming steps to be implemented, including use case and test cases. The technical specifications will be used by the CC programmer to develop the distributed programming code.

[4] **Coordinating Center programmer:** a SAS programmer who sits in the PCORnet Coordinating Center, who has experience with using the PCORnet Distributed Research Network (DRN) and understands the Common Data Model v3.0.

- Assessment of processes to establish linkage, identifying any problem areas or those that took more time than expected

Assessing prescriptions

- Proportion of prescription records with complete/incomplete data, for each data element available from the script
- Proportion of sites able to provide prescription data in the requested format
- Proportion of prescription records that can be associated with visit/diagnosis data at high confidence
- Proportion of sites that have dispensing data
- Proportion of patient records that link to dispensing data, and proportion of study period for which dispensing data are available
- Within linked subset, proportion of antibiotic prescription records without corresponding dispensing records, and vice versa, both overall and stratified by originating site, drug breadth, associated diagnos(es), and ordinality per treatment episode
- Proportion of prescriptions with errors (e.g., drug names that do not resolve to a national standard code)
- Proportion of children with two or more prescriptions for the *same* drug breadth with overlapping courses of therapy determined by the prescription date and the length of the course as noted on script
- Proportion of children with two or more prescriptions for the *different* drug breadth with overlapping courses of therapy determined by the prescription date and the length of the course as noted on script

I.6 Reports

The evaluation team will submit program evaluation plans and preliminary metrics within 60 days of the study kickoff meeting, and detailed reports of evaluation questions, metrics, instruments, and procedures every four months thereafter. The Final Evaluation Report will be submitted during month 24. Throughout the project, the evaluation team will respond promptly to any ad hoc requests required by either PCORI or the study team within the limits of the available budget.

I.7 Overall evaluation of PCORnet research readiness

I.7.1 Functionality of PCORnet

At the end of the project, the evaluation team will report on the functionality of PCORnet and its research readiness, highlighting the accumulated experience and infrastructure built to facilitate more efficient and collaborative PCORnet studies in the future. We envision that we will learn lessons in the following areas that help to build a "research commons":

1. Logistics: How to run a study like this. What did and did not work in governance, communication, personnel deployment, etc.?
2. Methodology and science: We will have developed and used a number of queries. For example, we will have learned a lot about harmonizing medication data. What processes and programs can we leave behind for future users?
3. Regulatory: What did we learn from IRB/human subjects protections, forging data use agreements and contracting that others should emulate?
4. Engagement: Which engagement strategies worked well and not so well? Are there lessons for dissemination, implementation, and decision making?
5. Distributed Research Network and CDM: did the DRN work efficiently? Was it useful to the study? What was the capacity of the PCORnet CDRN node sites to conduct unmodified queries?
6. Overall synopsis of lessons learned.

These performance measures will be used to provide descriptive guidance about progress toward a network with reusable infrastructure. The evaluation team will provide a descriptive analysis of the infrastructure, governance, and participation challenges experienced by all study participants and the opportunities for improvement of the network's governance procedures.

I.7.2 Defining Research Commons

Producing research as communities of engaged, geographically separate stakeholders, using a set of common resources has been termed "commons-based peer-production."[127] Well-known examples include open-source software development and Wikipedia. In a distributed learning health system, research communities include individuals from across the country, who use the governance, technical, social, regulatory, and scientific resources that PCORnet creates. No single person or institution "owns" these resources, rather, the Commons is a product of collaboration and an emergent phenomenon of the network. As part of PCORnet Phase II, the study team will work with the PCORnet Coordinating Center to share and disseminate products produced by this project through the PCORnet Commons. In particular, the CC PM will be responsible for curating the obesity study products to share.

I.8 Budget Implications

As described in the budget and associated justification, the lead evaluator and the evaluator project manager are budgeted to attend a kick-off meeting and an in-person presentation of the final results of the infrastructure testing activities.

I.9 CC Work Plan

The evaluation team will divide the evaluation into three phases: introductory phase, evaluation phase, and dissemination phase. The following describes the planned activities and associated timelines within each phase.

Activity	Timeline
Introductory Phase	
Attend kick-off meeting	2/25-2/26/2016
Engage study teams, participating CDRNs, and stakeholders to review and refine the proposed evaluation plan and metrics	3/3/2016
Submit formal evaluation plan and report template to PCORI	3/31/2016
Evaluation Phase	
Observe selected calls and meetings, review study documents, conduct key informant interviews, distribute follow-up surveys to the key informant interviews, meet with the CC data analytic team	Months 1-24
Conduct biweekly evaluation team meetings to discuss evaluation activities and prepare deliverables	Months 1-24
Submit interim evaluation report (6 months) to PCORI, based on previously approved evaluation plan	7/31/2016
Submit interim evaluation report (10 months) to PCORI	11/30/2016
Submit interim evaluation report (14 months) to PCORI	3/31/2017
Submit interim evaluation report (18 months) to PCORI	7/31/2017
Submit interim evaluation report (22 months) to PCORI	11/30/2017
Submit final evaluation report (24 months) to PCORI	1/31/2018
Dissemination Phase	
Share validated evaluation reports with the study teams and stakeholders	Following submission of evaluation

via iMeet Central	reports to PCORI (3/31/2016; 7/31/2016, 11/30/2016; 3/31/2017; 7/31/2017; 11/30/2017; 1/31/2018)

DISSEMINATION AND IMPLEMENTATION POTENTIAL

A critical role that our stakeholders will play in the proposed study is to help us plan, from the very start of the study, various dissemination and implementation strategies that meet PCORI's goals for rapid and meaningful dissemination of results to allow real time decision-making. This work will rely heavily on the four stakeholder teams because they are 'on the ground' in the places that need to be partners in this process. Many of the most important questions we have heard from our stakeholders during the development of the proposed study relate to dissemination. These included:

"How should parents and providers think about antibiotic use? What if there is very little effect from antibiotic use on obesity risk? How will we communicate the complexity of the findings and the process of the study that produced these findings? What about the influence of antibiotics from food sources, such as beef and chicken? Is the benefit of antibiotic use for less serious infections worth the risk of childhood obesity later? How can we connect patients, providers, and other stakeholders with each other to facilitate an exchange of information about any relationship between antibiotic use and childhood obesity?"

We believe these are important questions that have similarities with questions arising from any study to be conducted in PCORnet. Thus we propose that whatever we do implement is created in such a way as to leave 'durable' tools behind so that other studies can use the same materials and tools and simply customize them for the particular message and community. This is also no easy task but will be an important goal of this study.

It will be important for the stakeholder groups to carefully describe the aims to their stakeholder communities to provide the proper basis for understanding the results. This is described in the engagement section of the proposal. In addition, it will be critical for each of the four stakeholder groups to provide feedback that can improve the likelihood of success for this study. Stakeholder groups may disagree, requiring the EASAG to mediate and engage the relevant teams in discussion about how best to harmonize feedback.

Once results are available, the study team will determine whether it will apply for dissemination funding from PCORI. If so, in the final year of the proposal, time and attention could be devoted to creation and implementation of dissemination strategies within each stakeholder group. The EASAG could help us collate all of these strategies into one overarching approach that can be tailored for different segments of the population. This can be done with a special emphasis on reproducibility, durability and scale for other projects.

The secondary aim will be very important to determining the messaging since it will address many of the questions already asked by stakeholders. Messages developed could include a call to action, depending on the study results. Because content to be disseminated will not be ready until findings are final, the research and stakeholder team will remain fluid and open to messaging that makes sense for patient and caregiver empowerment.

We intend to utilize the EASAG as well as the rest of PCORnet's robust network for dissemination. The engagement stakeholders will play an important role in dissemination. We expect they will have opportunities to communicate to their own communities about study progress, present results in the field, write blogs and other communications, and contribute to publications.

Mindful that most parents of young children and their providers are overwhelmed by health messages, it would be critical to create succinct and useful messages for the parent-provider dyad. This could be in the form of a memorable slogan or mnemonic, that could serve as the foundation for a tool to facilitate decision-making, with support facts and information.

Within PCORnet, CDRNs and PPRNs have forged close relationships with clinicians and other healthcare providers, as well as health system organizations, who serve parents and families. We will draw on these relationships for dissemination and implementation. Once developed, communications toolkit, along with materials, videos, and digital communications (at appropriate literacy levels) for parents, can be shared through PCORnet and other professional networks (Grand Rounds, social media, and *Smartbriefs*) as well as through CDRN pediatric clinics and hospitals and from CDRN leadership to staff, patients, and communities. We will also plan to engage non-physician provider

communities, including but not limited to nurses and nurse practitioners, social workers and mental health workers, and physician's assistants. Further, dissemination can take place through all major medical associations serving obstetricians, pediatric providers, and family medicine providers, including but not limited to the American College of Obstetricians and Gynecologists, American Academy of Pediatrics and American Academy of Family Physicians.

Additional outreach methods would target consumers through non-medical communities serving children and families, including: faith-based communities; social services such as WIC and Healthy Start; daycares, preschools and public schools; prenatal settings; employee and insurance wellness programs, and other influential stakeholders for child caregivers such as "mommy bloggers" and other innovative programs such as Text4baby, Centering, and frequented web forums.

Potential barriers include the following:

- The high cost of moving messages in front of people, when they are constantly barraged other messages
    - A lack of systems to move such messages fluidly
    - A lack of incentives for various groups to collaborate on dissemination
- The complexity of the results and potential outcomes
    - Low science and medical literacy among the public and some of the other stakeholder groups
    - Difficulty understanding risk and communicating it, especially for future events
- Reticence of health systems for changing practices
- Challenges of shared decision making between providers and parents

We also realize the tasks we outline for dissemination and implementation are ambitious in the short time period designated for this study. However, we intend to start this process early and to continue it, in parallel with the enactment of the engagement plan, throughout the study period, not just at the end. We also expect that some of this work will extend beyond the study period, perhaps with the help of additional funding.

A. Ability to reproduce potentially important findings from this research in other data sets and populations. The study team will ensure that researchers examining other data sets and populations will be able to reproduce potentially important findings from this research. The SAS code from our data queries will be made publicly available – along with information about the dataset – so that comparable data sets can be assembled from patient populations that did not participate in this study. Likewise, focus group prompts will be made available. We will also share other pertinent study documentation such as the study protocol (initial and final versions) and data definitions. These materials will be posted on the study's website and their existence logged on the PCORnet Distributed Research Network, along with an email address for seeking more information about them. Manuscripts describing the analytic methods will be developed and published in a timely manner. In addition, we will develop and disseminate a white paper on the analytic process, detailing issues such as the challenges that arose, considered solutions, and rationale for decision-making.

B. Data availability following study completion. A complete, cleaned, de-identified electronic copy of the final data sets used for this study will be made available within 1 year after study completion. In addition to the quantitative (electronic medical record) aggregate and individual-level data, we will archive the de-identified focus group and interview transcripts, code libraries, and coded datasets. To protect the privacy of participating health systems, the de-identified electronic medical record data sets that are made available will not include information indicating which node sites are linked together into a clinical data research network. Likewise, in the case of focus group data, the affiliated PPRN or CDRN will not be recorded.

Data will be stored at Harvard Pilgrim Health Care Institute's (HPHCI) facilities for three years after they are made available and will be subject to routine data security standards. As a healthcare institution, HPHC is required to meet the standards defined by the Health Insurance Portability and Accountability Act of 1996 (HIPAA), and by subsequent Rules and Acts that further address privacy and security. HPHCI is an LLC of HPHC and adheres to the same standards.

Operating systems on HPHCI workstations and servers are kept current with patches and software updates, following industry best practices. Operating system updates are automated, and anti-virus software on workstations and servers is also updated automatically. A firewall prevents illegitimate contact with HPHC computers, protecting the systems from unauthorized Internet access or monitoring. Email attachments sent to HPHC addresses are scanned for malicious content and inappropriate content is removed.

Access to project files is restricted to approved study staff. HPHCI has well-defined password requirements to protect systems from unauthorized access and restricted access to study-specific folders. Accounts are ended at employment termination or job change. Access to HPHC and HPHCI information systems is granted based upon the user's job title/role and the standard of legitimate business need to know. HPHC conducts reviews of security level designations. Entry into the HPHCI facilities is restricted by swipe card access.

We will develop a brief application for data requests that will be housed on the HPHCI and PCORnet websites. If a researcher desires to access the dataset for additional research, the request must be (a) approved by a committee made up of the Antibiotic Study PIs, a project manager, and researchers involved with the study from the PCORnet Coordinating Center (n=1), and a co-investigator from participating CDRNs (n=1). This committee will assess proposed projects from a scientific perspective, its consistency with PCORnet goals, the degree of overlap between a proposed analysis and any previously approved work, and the potential for conflicts of interest on the part of the proposed investigators. A summary of any proposed analysis approved by the committee will be circulated to PIs of CDRNs and PPRNs which contribute data to the PCORnet Antibiotic Study, for discussion prior to final approval for data release. The data from a given CDRN or PPRN, or its node sites, will not be released without that network's (and node sites') written permission.

HPHCI will work with interested researchers to develop Data Use Agreements that are consistent with the data agreements used to assemble the dataset. Such data use agreements will specify the uses to which the data can be applied and the allowable duration of use (after which, data must be destroyed). Once approval is obtained from the

Antibiotic Data Use Committee and appropriate data use and IRB assurances are obtained, the de-identified data set will be transferred using a Secure File Transfer site.

Budget to cover costs of the data-sharing plan. Our budget for data sharing will be based on two items: the cost of ongoing data storage and an administrative fee for each request that covers the cost of proposal review, data extraction and data delivery. During the study period, we will determine reasonable costs that will be required for this work, and we will provide this information on the brief application. The personnel involved in this effort will include

Project Manager: The study project manager will track all data requests, coordinating with the data requestor and the review committee. If the project is determined to be of merit by the review committee, the Project Manager will coordinate with the institutions that had originally contributed data to obtain their written approval of releasing data. S/he will oversee the execution of any necessary regulatory agreements. The Project Manager will maintain a database to track progress over time, including presentations and publications; and, if applicable, confirm data destruction.

Programmer: The programmer will be responsible for compiling the dataset and documentation. S/he will also provide minimal technical assistance.

Application review committee**:** Each data request will be reviewed and deliberated by a committee of seven individuals, as above. Each individual will be compensated for review of all materials.

This study involves no therapeutic intervention and minimal risk of loss of confidentiality. All research conducted in the Department of Population Medicine at Harvard Pilgrim Health Care Institute (HPHCI), the prime site for this study, complies with the Department of Health and Human Services requirements for safeguarding the safety, rights and welfare of human subjects, regardless of the source of funding. We will strictly adhere to all guidelines for protecting participants' privacy and confidentiality.

1. Study Overview
The chief objectives are to assess the comparative effects of different types, timing, and amount of **antibiotic** use in the first 2 years of life with (Aim 1) body mass index (BMI) and obesity at ages 5 and 10 years and (Aim 2) growth trajectories to age 5 years. In Aim 3, we will examine effect modification (~heterogeneity of treatment effect) according to several a priori specified socio-demographic, clinical, and maternal variables. In our Secondary Aim, we will employ focus groups of parents and in-depth interviews of clinicians to explore how best to put the epidemiologic findings into everyday practice.

Specific Aim 1: To evaluate the comparative effects of different types, timing, and amount of antibiotics used during the first two years of life on body mass index and risk of overweight and obesity at ages 5 (primary outcome age) and 10 (secondary) years.

Specific Aim 2: To assess the comparative effects of different types, timing, and amount of antibiotics used during the first two years of life on the rates and patterns of childhood growth during the first 5 years of life.

Specific Aim 3: To explore how the effects of different types, timing, and amount of antibiotics on childhood BMI, obesity risk and growth (Aims 1 and 2) vary according to patient socio-demographic, clinical, and maternal characteristics, including socio-demographic, clinical, and maternal factors.

Secondary Aim: Through focus groups and in-depth interviews, to explore how parents and other caregivers and their providers assess information related to current and future benefits and risks, particularly for treatments such as antibiotics in early childhood with potential near-term benefits but potential moderate long-term risks. We will also explore how clinicians, health care organizations, and policy makers should best present study findings to help parents understand its strengths and limitations in the context of shared clinical decision-making.

2. **Description of Study Population and Recruitment**
2a. ***Inclusion and Exclusion Criteria***
Aims 1 to 3
Inclusion Criteria
To be eligible, patients must have entered care at the node site before 6 months of age, and adhere to the following criteria:

2. Data contained in PCORnet Common Data Model 3.0, including prescribed medications.
3. > 1 encounter with length and weight measured in each of the following age intervals: 0-5 m, 6-11 m, 12-23 m, and
4. > 1 encounter with height and weight measured in either or both of the following age intervals: 4.0 to 5.9 y ("age 5 years"), 9.0 to 10.9 y ("age 10 years"), or eligible to be followed to these ages for use in multiple imputation to account for missing data (Section B.4).

Exclusion criteria

c. Obesity with identified secondary cause (see below for ICD-10 codes and diagnoses)
d. Other clinical conditions that substantially alter growth (see below for ICD-10 codes and diagnoses)
e. Biologically implausible length/height or weight measurements (Section B.4.5)

For the **Secondary Aim,** we will work with 4 selected CDRNs and PPRNs to conduct 8 focus groups of parents and caregivers, 2 in each of 4 locales, to gain insight into the communication channels and information that parents seek to help them make decisions regarding antibiotic use and future risks including about childhood obesity. We also will 20 semi-structured interviews with clinical providers affiliated with these CDRNs/PPRNs. We have not yet chosen the specific CDRNs and PPRNs that will participate in this study; we will do so after the grant period begins (or before).

2b. Vulnerable Populations: Our study involves only children and will involve minorities, and, as such, the entire sample will come from a vulnerable population. We will request a waiver of informed consent to identify the electronic medical record cohort, which we will address vulnerable subjects. The focus groups that we conduct will be with parents and caregivers; the semi-structured interviews will include clinical providers. We do not expect to include vulnerable populations in the Secondary Aim. Individuals who are determined not to be able to provide consent will not be included in the study.

2c. Inclusion of women & minorities, and children: Aims 1 to 3 will include children, and we expect that we will have extensive representation across racial/ethnic groups and equal numbers of males and females. In addition, on a subset, we will obtain information on their mothers. There are no inclusion/exclusion criteria based on sex or race in either study cohort. We expect our study population to match the communities in which the CDRNs and PPRNs reside.

***2d. Stakeholders*:** For this study, we have recruited 8 stakeholders to be part of the Executive Antibiotic Stakeholder Advisory Group (EASAG), which will direct our engagement efforts. These stakeholders have provided feedback during the course of the proposal development and will remain involved in the study throughout the grant period. We are compensating these participants for their work. When the study begins, we will recruit a wider stakeholder group from the following 4 categories (these stakeholders are members of the research team, not research subjects):

(a) Parents/Caregivers. We will engage parents and caregivers of young children who were prescribed and/or took broad-spectrum antibiotics during infancy. This group of stakeholders will meet throughout the project to provide continuous feedback about the study aims and questions, methodology, data collection processes, study refinement, dissemination plans, and interpretation of the aggregate study results. These stakeholders will receive compensation for their work on this study.

(b) Healthcare providers and health system leaders. We will engage a broad cross-section of professionals directly involved in pediatric care and management of antibiotic use such as pediatricians, pharmacists, and hospitalists. This group will serve in helping refine the study aims and questions before beginning data collection. Later, we will gather input on interpreting results and in dissemination. These stakeholders will not receive compensation for their work on this study.

(c) Healthcare system leaders. We will engage people who have a say in decisions about childhood obesity prevention/treatment, antibiotics use, or obesity management in general. They could be leaders who decide coverage for certain treatments, determine quality reporting and/or create clinical practice guidelines around these topics, etc. This people will be integral during development of dissemination strategies for getting our findings into practice. They will also provide feedback to refine study aims and interpret results. These stakeholders will not receive compensation for their work on this study.

(d) Community advocacy groups. We will engage many groups whose mission is associated with obesity, childhood obesity, maternal and child health, antibiotic use, etc. These organizations may include condition advocacy groups, local health departments and medical agencies. Engagement activities will mirror those of the other stakeholder groups listed above. These stakeholders will not receive compensation for their work on this study.

3) IRB, recruitment, and informed consent procedures

3a. IRB procedures

Aims 1 to 3
Many sites may not consider the collection and use of de-identified aggregate data as activities that need to undergo IRB review. Currently, however, this determination must be made at the site level; this study involves 39 node sites across 10 CDRNs. As some sites will require IRB review, the CC will approach developing the process to use a centralized IRB of record with the assumption that IRB review will be required. The CC will also work towards a process in which, in the future, institutions could cede "not human subject research" determinations.

The study's **Secondary Aim**, which involves primary data collection in the form of focus groups of parents/caregivers and interviews of providers, will require IRB review. In addition to using a centralized IRB process for those node sites that require IRB approval for the primary Aims, the CC also will use this Secondary Aim to test out using a centralized IRB process.

We propose to use IRBShare as the mechanism to streamline IRB review. This process is outlined further in the management approach section. Briefly, IRBShare allows for streamlining IRB review of multisite studies in a single IRB review. This program has formal, written acknowledgement as an acceptable joint review model from the Office for Human Research Protections (OHRP) and Food and Drug Administration (FDA), and is supported by the Association for the Accreditation of Human Research Protection Programs (AAHRPP). IRBShare is based on a uniform national reliance agreement (i.e., institutions share regulatory responsibility) that allows IRBs to rely on the review of another IRB for all phases of review (initial study, annual/continuing reviews, and reviews of amendments) of multi-site studies at any risk level. This allows local IRBs to conduct the local context review and maintain study oversight between the times of reliance upon the lead IRB reviews.

Appropriate Data Use Agreement and Business Associate Agreement will be approved and signed by all sites contributing data to the analysis.

***3b. Recruitment*:** For **Aims 1 to 3, w**e will identify eligible patients from the Common Data Model at the 10 CDRNs participating in this study, comprising 39 node sites. These patients will be involved in data-only analyses. The estimated number of eligible patients is 1.6 million. We will request waivers of informed consent because this study presents minimal risks to the participants and it would be impractical to collect consent for such a large number of potential participants. We will take rigorous precautions to protect privacy, confidentiality, and security of personal health information (see below). In fact, we will be using either aggregate level data, without any individual level data, to conduct some analyses and de-identified patient level data to conduct other analyses. Patient informed consent would substantially increase the cost of this research and would introduce opportunity for selection bias associated with voluntary choice to disclose or withhold access to protected health information (PHI). Such bias would likely be of sufficient magnitude to reduce the generalizability of this study and its research value.

We believe that the proposed project clearly satisfies the criteria of 45 CFR 46:116 for waiver of written informed consent. Those criteria are:

- "The research involves no more than minimal risks to the subjects" – The only risk to participants from this procedure is violation of confidentiality, and we expect this risk to be minimal, considering the use of deidentified data.

- "The importance of the proposed research outweighs the minimal risk involved" – The importance of defining the risks of antibiotic use during early childhood is critical for public health. The risk will be minimal.

- "The research could not practicably be carried out without the waiver or alteration" – It would not be possible to identify potential participants without access to records data.

- "The proposed research will not adversely affect the rights of subjects" – Administrative data will be used solely for research purposes. Findings will be reported in aggregate only.

- "Obtaining written consent is not possible" – It would not be possible to obtain written consent from approximately 1.6 million patients and their parents/caregivers. In addition, attempting to contact this many patients would increase the risk of breach of confidentiality.

- "Identifiable information is necessary" – Use of consumer identifiers is necessary in order to link different data sources, but these linkages will be done at the datamarts and will not be accessible to the Coordinating Center or the study teams. Only the sites that are doing linkages could later reconnect random subject IDs to their true identifiers.

- "It is not appropriate to provide subjects with information after the study is completed" – Contacting 1.6 million children and their caregivers regarding use of records for research would increase the risk of breach of confidentiality.

For the **Secondary Aim**, we will recruit parents and caregivers of children directly, from the 4 CDRNs and PPRNs that will agree to participate in this study. We also will conduct semi-structured interviews with providers. During the recruitment phase, we will reach out to potential participants via existing CDRN stakeholder groups and pediatricians affiliated with the node sites. We will reach out to PPRN participants through the existing organizational structures of PPRNs. If necessary, we will target groups within institutions for participation, and we will obtain waivers of informed consent for this process. For recruitment of providers, we will work with the institutions to send emails directly to providers asking for participants in the semi-structured interviews, and we will use existing relationships already existing between CDRNs and provider stakeholders.

3c. Informed Consent

As mentioned previously in Section 3b, we will seek a waiver of informed consent for **Aims 1 to 3**. For the **Secondary Aim**, we will incorporate informed consent, either as written informed consent or a modified informed consent process. At the start of each focus group, we will reiterate to all participants that participation is voluntary, that they are free to not respond or to terminate involvement at any time, with no adverse consequences and participation in the study will have absolutely no bearing on participants' or their children's medical care (or their employment in the case of the providers). Participants will be told that they can decline to answer any questions and that they can stop participating at any time. We will request permission for audio-recording the focus groups and interviews. The focus groups and interviews involve no specific risk or discomfort beyond those of a standard clinical interview. The moderator will be a trained research interviewer sensitive to these issues.

4. Potential risks

The proposed study poses minimal risks to the participants. Potential risks are breach of privacy and loss of confidentiality. These events could, conceivably, lead to economic losses for subjects through medical underwriting practices and identify theft. Our procedures for protecting against such risks are described in the next section. For **Aims 1 to 3**, because we will be using aggregate or deidentified patient-level data, we anticipate that this risk will be minimal. All of the participating node sites are health care institutions, for which data are protected by HIPAA and other relevant federal and state laws. The data that we will use for these Aims will involve data collected through routine clinical practice and operations and not separately for research. Since the node sites will analyze their data behind their institutional firewalls, we expect risk of data use on participants in this study to be minimal.

For the **Secondary Aim**, there will be some risk of a confidentiality breach from the use of questionnaire forms, interview notes, audio-recordings, transcripts, and electronic data. There is a slight risk that research data files might be compromised, and obtained or viewed by unauthorized persons. Our procedures for protecting against such risks are described in the next section.

5. Protection of human subjects, potential benefits to participants, and important of knowledge to be gained

5a. Aims 1 to 3

The organizations proposing this study have systems, oversight, experienced personnel, and an organizational culture that supports the appropriate use, access and storage of confidential information. All persons collecting or handling data will

be trained in human subjects' procedures, confidentiality and privacy protection.

Data for all participants will be kept strictly confidential. For participants, a number of methods will be employed to maintain confidentiality of the participants. All source data files will be protected behind node site firewalls and will never leave their servers. We will query these data centrally using the distributed research network, but the data that will be returned to the Coordinating Center and then the study teams will be aggregated or deidentified patient-level data. Even with these extensive privacy-protecting analytic methods, we have additional safeguards in place to protect privacy. All investigators and project staff receive IRB and HIPAA training. Collected data will be used only for research purposes. Any published data be solely include aggregated information.

Physical safety processes are also in place. Access to the study site's buildings is restricted. All space accessible to the public is separated from research offices. Access to the computer systems is restricted. All research computers are located in locked departments and data files are password protected.

5b. Secondary Aim: This Aim will require the collection and use of identifiable patient data, at least for the purposes of recruitment, scheduling, and conduct of the focus groups and interviews. However, we will seek to deidentify data collected as soon as is feasible. For example, when the focus group and interview audiotapes are transcribed, we will remove all identifying information, such as names, that are discussed. We will keep records for the purposes of informed consent, which also will contain individually identifiable data. Yet, we will discard this information as soon as it is no longer necessary. In consultation with the relevant IRBs with jurisdiction over this Aim, we will explore using a modified consent process, in which we provide information materials to participants but do not require them to sign a consent form. All research files will be kept in locked file cabinets or a locked file room. Names and other identifiers will be kept in separate locked files. All audio-recordings of in-person interviews will be kept on secure computer servers that have firewalls and require password access. Once transcribed and checked, the audio-recordings will be destroyed. All computerized data will be kept on secured computers or networks at each site. These data will be accessible only to research staff, using confidential usernames and passwords.

Deidentified transcripts will be transferred between sites using password-protected files and secure encrypted web transfer procedures. All data will be used only for research purposes only; published data will not contain any individual identifiers.

5c. Potential benefits

It is uncertain whether individual participants will directly benefit from participation. Participants may learn something new about the impact of early childhood exposure to antibiotics and future risk of obesity. They may enjoy participating and may feel that doing so contributes to scientific knowledge in general. The focus group and interview participants will provide important information about how we should interpret and communicate results to the general public and clinicians. In so doing, these participants will have the opportunity to shape those messages and strategies.

5d. Importance of the knowledge to be gained

The products of this study will enhance scientific understanding of the relationship between antibiotic use and weight gain among young children and how to communicate with stakeholders about prescribing antibiotics in childhood and provide valuable insight into the impact of future growth following early childhood exposure to antibiotics.

APPENDIX

1.0 Glossary of Terms and Abbreviations

1.1 Glossary

Aggregate Data: Aggregated information across specified strata of individuals whose data is held by a Network Partner or a Network Data Affiliate. For example, counts of patients within a stratum that includes a particular age group, gender, and diagnosis.

Authorized Users: Individuals associated with and selected by a Network Partner who have been granted access to the PCORnet DRN Query Tool by the CC and are authorized to initiate queries through the PCORnet DRN Query Tool.

Breach: A use or disclosure of Protected Health Information in PCORnet Data or Network Data that is not permissible under HIPAA.

Clinical Data Research Network (CDRN): A clinical data research network that has executed a PCORnet Participation Agreement. 10 of the 13 PCORnet CDRNs will participate in this study.

Common data model (CDM): The standardized data model developed and maintained by the DSSNI Task Force that establishes the definition, content and format of data across networks that can be used for querying.

Coordinating Center programmer: a SAS programmer who sits in the PCORnet Coordinating Center, who has experience with using the PCORnet Distributed Research Network (DRN) and understands the Common Data Model v3.0.

DataMart: A specific data resource that can be uniquely defined and queried using the PCORnet DRN Query Tool. CDRNs havecreated their DataMart(s) through an Extract, Transform, and Load (ETL) of source data. PCORnet DataMarts are DataMarts that adhere to the CDM.

Data Standards, Security, and Network Infrastructure (DSSNI) Task Force: The PCORnet Task Force that develops and maintains the Common Data Model, including minimal data standards and technical specifications and models for querying, to support a functional distributed research network that facilities multi-site patient-centered research across Network Partners while allowing the Network Partner to maintain operational control over its data.

De-identified Data: As defined in 45 CFR Section 164.514(a) of the HIPAA Privacy Rule. Processes for de-identifying data are set forth in 45 CFR Section 164.514(b) of the HIPAA Privacy Rule.

Distributed Research Network (DRN): A research network in which there is no central data repository and in which network nodes are able to respond to network queries using agreed-upon systems (e.g., PCORnet DRN Query Tool).

Executive Antibiotic Stakeholder Advisory Group (EASAG): The Advisory Group comprised of 6 stakeholders. This group has provided feedback to the study team during the development of this proposal and will direct the 4 stakeholder groups once this study begins.

Functional specifications: a high-level description of the program purpose and scope (e.g., propensity score matching for cohort studies).

HIPAA: The Health Insurance Portability and Accountability Act of 1996, the Health Information Technology for Economic and Clinical Health Act, and all implementing regulations.

HIPAA Privacy Rule: The HIPAA Privacy Rule (45 CFR Part 160 and Subparts A and E of Part 164).

HIPAA Security Rule: The HIPAA Security Rule (45 CFR Part 160 and Subparts A and C of Part 164).

Query: These are statistical codes written centrally by the CC programmer that will be distributed to each of the node sites for execution. The queries will return aggregated or individual-level datasets to the CC for distribution to the study teams.

CDRN Lead Site PI: An investigator at the CDRN who will be site principal investigator at the lead site of that CDRN for this study. The 10 CDRN lead site PIs for this study will compose the initial Scientific Advisory Group.

Limited Data Set: A limited set of identifiable patient information as defined in the HIPAA Privacy Rule in 45 CFR Section 164.514 (e).

Methods Core: This group will work across both obesity demonstration projects to create efficiencies around working with the Coordinating Center, developing queries, and refining data and analyses. The group will include investigators working across both studies as well as members from each study team.

Minimum Necessary: As defined in the HIPAA Privacy Rule in 45 CFR Section 164.514(d).

Mini-Sentinel Tool: a reusable program, which was developed, tested, revised, and produced by Mini-Sentinel programming teams. These parameterized SAS programs are able run against the Mini-Sentinel common data model across all 18 data partners with disparate hardware and software configuration. Each program has a set of input parameters and output metrics that allow the stakeholder to define common concepts such as treatments, risk windows, study outcome, and inclusion and exclusion criteria. The programs address a range of common research questions, can be easily customized, and produce standardized output that enable rapid review.

Network Data: Data generated, collected, processed, maintained, held, or stored by either a Network Partner or a Network Data Affiliate in connection with their participation in PCORnet.

Network Data Affiliate/Node site: An individual site or organization that contributes data for use by a CDRN or PPRN.

Network Partners: The PCORnet Clinical Data Research Networks (CDRNs) and Patient-Powered Research Networks (PPRNs) who have executed a PCORnet Participation Agreement.

Patient-Centered: The question is important to patients, measures outcomes that are noticeable and meaningful to them, and produces results that help them weigh the value of healthcare options given their personal circumstances, conditions, and preferences.

Patient-Powered Research Network (PPRN): A patient-powered research network that has executed a PCORnet Participation Agreement.

PCORnet Coordinating Center (CC): The organization(s) with expertise in developing and maintaining an infrastructure to support research and research-related activities using the PCORnet DRN that are contracted by PCORI to serve as the PCORnet CC.

PCORnet Data: Data and/or analytics returned to an Authorized User who initiates a PCORnet Query.

PCORnet DRN Query Tool: The system used to operate the PCORnet distributed research network (DRN). The Query Tool includes a web portal and a software application (the DataMart Client) that are powered by PopMedNet.

PCORnet Network Partner Principal Investigator (PI): Lead scientist for a particular PCORnet Study.

PCORnet Participation Agreement: The legal agreement entered into by and between PCORI and each Network Partner that establishes the terms and conditions for each Network Partner's participation in PCORnet, including compliance with all PCORnet Policies.

PCORnet Policies: Policies adopted by the PCORnet Steering Committee regarding the operation and governance of PCORnet.

PCORnet Project Management Office (PMO): The PCORnet PMO, which is operated by the CC, assists the CDRNs and PPRNs with technical and logistical support and plays a critical role in fostering communication and coordination among them as well as disseminating best practices. It also supports program evaluation.

PCORnet Query: Queries of Network Data that use the PCORnet DRN Query Tool and Independent Queries.

PCORnet Steering Committee (SC): Includes representatives of the multiple CDRNs and PPRNs that have executed a PCORnet Participation Agreement and functions as an advisory group to PCORI leadership and as an oversight body for PCORnet.

PCORnet Study: A research activity that utilizes Network Data and is submitted to and approved by the PCORnet Executive Steering Committee or its designee as a PCORnet Study.

Protected Health Information (PHI): As defined in the HIPAA Privacy Rule at 45 CFR Section 160.103.

Requestor: An individual who is affiliated with a Network Partner and who develops and submits a request for a PCORnet Query and receives the results.

Study programmer: a SAS programmer who is a member of the scientific team and has familiarity with the PCORnet CDM and DataMart source data. This person works with the Coordinating Center programmer to define functional and technical requirements and specifications for the study programming effort.

Study Team: Also called the Core Scientific team. This will be the team that is responsible for all study specific activities, including defining the Aims, developing the functional and technical specifications, directing the work of the study, conducting final analyses, and writing manuscripts.

Technical specifications: a step-by-step description of the analytic code to be developed with detail and examples to ensure study team requirements will be achieved in programming.

1.2 Abbreviations

CC: Coordinating Center
CDM: Common Data Model
CDRN: Clinical Data Research Network
DUA: Data User Agreement
DRN: Distributed Research Network
DSSNI: Data Standards, Security, and Network Infrastructure
EASAG: Executive Antibiotic Stakeholder Advisory Group
EC: Executive Committee
HIPAA: Health Insurance Portability and Accountability Act
IRB: Institutional Review Board
PCOR: Patient-Centered Outcomes Research
PCORI: Patient-Centered Outcomes Research Institute
PHI: Protected Health Information
PI: Principal Investigator
PMO: Project Management Office
PPRN: Patient-Powered Research Network
RFA: Request for Funding
SC: Steering Committee
SQL: Structured Query Language

REFERENCES CITED

1. Cutler DM, Deaton AS, Lleras-Muney A. The Determinants of Mortality. National Bureau of Economic Research Working Paper Series 2006;No. 11963.
2. Deaton A. Policy Implications Of The Gradient Of Health And Wealth. Health Affairs 2002;21:13-30.
3. Jayachandran S, Lleras-Muney A. Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines. The Quarterly Journal of Economics 2009;124:349-97.
4. Jayachandran S, Lleras-Muney A, Smith KV. Modern Medicine and the 20th Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs. National Bureau of Economic Research Working Paper Series 2009;No. 15089.
5. McCaig LF, Besser RE, Hughes JM. Antimicrobial-drug prescrption in ambulatory care settings, United States, 1992-2000. Emerg Infect Dis [serial online] 2003 Apr [cited 19 Sep 2011]. Available from: URL: http://wwwnc.cdc.gov/eid/article/9/4/02-0268.htm.
6. McCaig LF, Besser RE, Hughes JM. Trends in Antimicrobial Prescribing Rates for Children and Adolescents. JAMA: The Journal of the American Medical Association 2002;287:3096-102.
7. Paul IM, Maselli JH, Hersh AL, Boushey HA, Nielson DW, Cabana MD. Antibiotic Prescribing During Pediatric Ambulatory Care Visits for Asthma. Pediatrics 2011;127:1014-21.
8. Hicks LA, Chien Y-W, Taylor TH, Haber M, Klugman KP. Outpatient Antibiotic Prescribing and Nonsusceptible Streptococcus pneumoniae in the United States, 1996–2003. Clinical Infectious Diseases 2011;53:631-9.
9. Finkelstein JA, Metlay JP, Davis RL, Rifas-Shiman SL, Dowell SF, Platt R. Antimicrobial Use in Defined Populations of Infants and Young Children. Arch Pediatr Adolesc Med 2000;154:395-400.
10. Finkelstein JA, Stille C, Nordin J, et al. Reduction in Antibiotic Use Among US Children, 1996–2000. Pediatrics 2003;112:620-7.
11. Grijalva CG, Nuorti JP, Griffin MR. Antibiotic Prescription Rates for Acute Respiratory Tract Infections in US Ambulatory Settings. JAMA: The Journal of the American Medical Association 2009;302:758-66.
12. Siegel RM. Acute Otitis Media Guidelines, Antibiotic Use, and Shared Medical Decision-Making. Pediatrics 2010;125:384-6.
13. McCormick DP, Chonmaitree T, Pittman C, et al. Nonsevere Acute Otitis Media: A Clinical Trial Comparing Outcomes of Watchful Waiting Versus Immediate Antibiotic Treatment. Pediatrics 2005;115:1455-65.
14. Coco A, Vernacchio L, Horst M, Anderson A. Management of Acute Otitis Media After Publication of the 2004 AAP and AAFP Clinical Practice Guideline. Pediatrics 2010;125:214-20.
15. Vaz LE, Kleinman KP, Raebel MA, et al. Recent trends in outpatient antibiotic use in children. Pediatrics 2014;133:375-85.
16. Jukes TH. Antibiotics in Animal Feeds and Animal Production. BioScience 1972;22:526-34.
17. Millar MR, Walsh TR, Linton CJ, et al. Carriage of antibiotic-resistant bacteria by healthy children. Journal of Antimicrobial Chemotherapy 2001;47:605-10.
18. Sjölund M, Wreiber K, Andersson DI, Blaser MJ, Engstrand L. Long-Term Persistence of Resistant Enterococcus Species after Antibiotics To Eradicate Helicobacter pylori. Annals of Internal Medicine 2003;139:483-7.
19. Dominguez-Bello MG, Costello EK, Contreras M, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proceedings of the National Academy of Sciences 2010;107:11971-5.
20. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the Human Infant Intestinal Microbiota. PLoS Biol 2007;5:e177.
21. Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. Nature reviews Microbiology 2011;9:279-90.
22. Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proceedings of the National Academy of Sciences 2011;108:4554-61.
23. Jakobsson HE, Jernberg C, Andersson AF, Sjölund-Karlsson M, Jansson JK, Engstrand L. Short-Term Antibiotic Treatment Has Differing Long-Term Impacts on the Human Throat and Gut Microbiome. PloS one 2010;5:e9836.
24. Hildebrandt MA, Hoffmann C, Sherrill–Mix SA, et al. High-Fat Diet Determines the Composition of the Murine Gut Microbiome Independently of Obesity. Gastroenterology 2009;137:1716-24.e2.

25. Muegge BD, Kuczynski J, Knights D, et al. Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans. Science 2011;332:970-4.
26. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. Science Translational Medicine 2009;1:6ra14.
27. Nieuwdorp M, Gilijamse PW, Pai N, Kaplan LM. Role of the microbiome in energy regulation and metabolism. Gastroenterology 2014;146:1525-33.
28. Power SE, O'Toole PW, Stanton C, Ross RP, Fitzgerald GF. Intestinal microbiota, diet and health. Br J Nutr 2014;111:387-402.
29. Weisberg SP, McCann D, Desai M, Rosenbaum M, Leibel RL, Ferrante AW, Jr. Obesity is associated with macrophage accumulation in adipose tissue. J Clin Invest 2003;112:1796-808.
30. Xu H, Barnes GT, Yang Q, et al. Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. J Clin Invest 2003;112:1821-30.
31. Union of Concerned Scientists Press Release: 70 Percent of All Antibiotics Given to Healthy Livestock. Cambridge MA; 8 January, 2001.
32. Jukes TH. Antibiotics in Animal Feeds and Animal Production. BioScience 1972;22:526-34.
33. Jukes TH, Stokstad ELR, Taylor RR, Cunha TJ, Edwards HM, Meadows GB. Growth-promoting effect of aureomycin on pigs. Arch Biochem 1950;26:324-5.
34. Stokstad ELR. Antibiotics in animal nutrition. Physiol Rev 1954;34:25-51.
35. Ajslev TA, Andersen CS, Gamborg M, Sorensen TIA, Jess T. Childhood overweight after establishment of the gut microbiota: the role of delivery mode, pre-pregnancy weight and early administration of antibiotics. Int J Obes 2011;35:522-9.
36. Trasande L, Blustein J, Liu M, Corwin E, Cox LM, Blaser MJ. Infant antibiotic exposures and early-life body mass. Int J Obes (Lond) 2013;37:16-23.
37. Azad MB, Bridgman SL, Becker AB, Kozyrskyj AL. Infant antibiotic exposure and the development of childhood overweight and central adiposity. Int J Obes 2014;38:1290-8.
38. Bailey L, Forrest CB, Zhang P, Richards TM, Livshits A, DeRusso PA. ASsociation of antibiotics in infancy with early childhood obesity. JAMA Pediatrics 2014;168:1063-9.
39. Mueller NT, Whyatt R, Hoepner L, et al. Prenatal exposure to antibiotics, cesarean section and risk of childhood obesity. Int J Obes 2014.
40. Huh SY, Rifas-Shiman SL, Zera CA, et al. Delivery by caesarean section and risk of obesity in preschool age children: a prospective cohort study. Archives of Disease in Childhood 2012.
41. Blustein J, Attina T, Liu M, et al. Association of caesarean delivery with child adiposity from age 6 weeks to 15 years. Int J Obes 2013;37:900-6.
42. Ajslev TA, Andersen CS, Gamborg M, Sorensen TI, Jess T. Childhood overweight after establishment of the gut microbiota: the role of delivery mode, pre-pregnancy weight and early administration of antibiotics. Int J Obes (Lond) 2011;35:522-9.
43. Barros FC, Matijasevich A, Hallal PC, et al. Cesarean section and risk of obesity in childhood, adolescence, and early adulthood: evidence from 3 Brazilian birth cohorts. Am J Clin Nutr 2012;95:465-70.
44. Goldani HA, Bettiol H, Barbieri MA, et al. Cesarean delivery is associated with an increased risk of obesity in adulthood in a Brazilian birth cohort study. Am J Clin Nutr 2011;93:1344-7.
45. http://www.pcornet.org/about-pcornet/.
46. http://mini-sentinel.org/data_activities/distributed_db_and_data/details.aspx?ID=131. DDaCDM.
47. Kim J, Peterson KE, Scanlon KS, et al. Trends in overweight from 1980 through 2001 among preschool-aged children enrolled in a health maintenance organization. Obesity (Silver Spring) 2006;14:1107-12.
48. Wen X, Kleinman K, Gillman MW, Rifas-Shiman SL, Taveras EM. Childhood body mass index trajectories: modeling, characterizing, pairwise correlations and socio-demographic predictors of trajectory characteristics. BMC Med Res Methodol 2012;12:38.
49. Wen X, Gillman MW, Rifas-Shiman SL, Sherry B, Kleinman K, Taveras EM. Decreasing prevalence of obesity among young children in Massachusetts from 2004 to 2008. Pediatrics 2012;129:823-31.
50. Taveras EM, Rifas-Shiman SL, Sherry B, et al. Crossing growth percentiles in infancy and risk of obesity in childhood. Arch Pediatr Adolesc Med 2011;165:993-8.

51. Bailey LC, Forrest CB, Zhang P, Richards TM, Livshits A, DeRusso PA. Association of antibiotics in infancy with early childhood obesity. JAMA Pediatr 2014;168:1063-9.
52. Brown JS, Kahn M, Toh S. Data Quality Assessment for Comparative Effectiveness Research in Distributed Data Networks. Medical care 2013;51:S22-S9.
53. Kahn M, G, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Medical care July 2012;50:S21-9.
54. Rubin DB. Inference and missing data. Biometrika 1976;63:581-92.
55. Rubin DB. Multiple imputation for nonresponse in surveys. New York, NY: John Wiley & Sons; 1987.
56. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. Stat Med 1991;10:585-98.
57. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc 1994;89:846-66.
58. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health 2006;60:578-86.
59. Toh S, Garcia Rodriguez LA, Hernan MA. Analyzing partially missing confounder information in comparative effectiveness and safety research of therapeutics. Pharmacoepidemiol Drug Saf 2012;21 Suppl 2:13-20.
60. Boyko EJ. Observational research--opportunities and limitations. J Diabetes Complications 2013;27:642-8.
61. Garabedian LF, Chu P, Toh S, Zaslavsky AM, Soumerai SB. Potential bias of instrumental variable analyses for observational comparative effectiveness research. Ann Intern Med 2014;161:131-8.
62. Benner JS, Glynn RJ, Mogun H, Neumann PJ, Weinstein MC, Avorn J. Long-term persistence in use of statin therapy in elderly patients. JAMA 2002;288:455-61.
63. Brookhart MA, Patrick AR, Schneeweiss S, et al. Physician follow-up and provider continuity are associated with long-term medication adherence: a study of the dynamics of statin use. Arch Intern Med 2007;167:847-52.
64. Karter AJ, Parker MM, Moffet HH, Ahmed AT, Schmittdiel JA, Selby JV. New prescription medication gaps: a comprehensive measure of adherence to new prescriptions. Health Serv Res 2009;44:1640-61.
65. Vrijens B, Vincze G, Kristanto P, Urquhart J, Burnier M. Adherence to prescribed antihypertensive drug treatments: longitudinal study of electronically compiled dosing histories. BMJ 2008;336:1114-7.
66. Fischer MA, Stedman MR, Lii J, et al. Primary medication non-adherence: analysis of 195,930 electronic prescriptions. J Gen Intern Med 2010;25:284-90.
67. Kuh D, Ben Shlomo Y, eds. A Life Course Approach to Chronic Disease Epidemiology. 2nd ed. Oxford: Oxford University Press; 2004.
68. National Center for Health Statistics. CDC Growth Charts, United States, 2000. http://www.cdc.gov/growthcharts/; Last updated: September 9, 2010, accessed May 25, 2015.
69. Daniels SR. Regulation of body mass and management of childhood overweight. Pediatr Blood Cancer 2005;44:589-94.
70. Field AE. Predictors and Consequences of Childhood Obesity. In: Hu F, ed. Obesity Epidemiology. New York: Oxford University Press; 2008:416-36.
71. Baker JL, Olsen LW, Sorensen TI. Childhood body-mass index and the risk of coronary heart disease in adulthood. N Engl J Med 2007;357:2329-37.
72. Rifas-Shiman SL, Rich-Edwards JW, Scanlon KS, Kleinman KP, Gillman MW. Misdiagnosis of overweight and underweight children younger than 2 years of age due to length measurement bias. MedGenMed 2005;7:56.
73. Smith N, Coleman KJ, Lawrence JM, et al. Body weight and height data in electronic medical records of children. Int J Pediatr Obes 2010;5:237-42.
74. Cunningham SA, Kramer MR, Narayan KM. Incidence of childhood obesity in the United States. N Engl J Med 2014;370:403-11.
75. Rifas-Shiman SL, Gillman MW, Oken E, Kleinman K, Taveras EM. Similarity of the CDC and WHO weight-for-length growth charts in predicting risk of obesity at age 5 years. Obesity (Silver Spring) 2012;20:1261-5.
76. Sturmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution--a simulation study. Am J Epidemiol 2010;172:843-54.
77. Li L, Kleinman K, Gillman MW. A comparison of confounding adjustment methods with an application to early life determinants of childhood obesity. J Dev Orig Health Dis 2014;5:435-47.

78. Karr AF, Lin X, Sanil AP, Reiter JP. Secure regression on distributed databases. J Comput Graph Stat 2005;14:263-79.
79. Karr AF, Fulp WJ, Vera F, Young SS, Lin X, Reiter JP. Secure, privacy-preserving analysis of distributed databases. Technometrics 2007;49:335-45.
80. Zhang, Li L, Abrams AM, K. K. Development and implementation of secure statistical analyses on distributed databases. http://effectivehealthcareahrqgov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=486 Accessed March 17, 2014 2010.
81. Breslow NE, Clayton DG. Approximate Inference in Generalized Linear Mixed Models. Journal of the American Statistical Association 1993;88:9-25.
82. Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. 2nd ed. New York: John Wiley & Sons; 2011.
83. Sturmer T, Glynn RJ, Rothman KJ, Avorn J, Schneeweiss S. Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. Med Care 2007;45:S158-65.
84. Sturmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Performance of propensity score calibration--a simulation study. Am J Epidemiol 2007;165:1110-8.
85. Regnault N, Gillman MW. Importance of characterizing growth trajectories. Ann Nutr Metab 2014;65:110-3.
86. Regnault N, Gillman MW, Kleinman K, Rifas-Shiman S, Botton J. Comparative study of four growth models applied to weight and height growth data in a cohort of US children from birth to 9 years. Ann Nutr Metab 2014;65:167-74.
87. Bryan M, Heagerty PJ. Direct regression models for longitudinal rates of change. Stat Med 2014;33:2115-36.
88. Casella G, Berger RL. Statistical Inference 2nd International ed: Cengage Learning; 2001.
89. Annie E. Casey Foundation, KidsCount Data Center, Child Population by Race. http://datacenterkidscountorg/data/tables/103-child-population-by-race?loc=1&loct=1#detailed/1/any/false/36,868,867,133,38/66,67,68,69,70,71,12,72/423,424:Last updated July 2014; Accessed May 25, 5.
90. VanderWeele TJ. Sample Size and Power Calculations for Additive Interactions. Epidemiologic Methods August 2012;1:159-88. http://www.degruyter.com/view/j/em.2012.1.issue-1/161-962X.1010/2161-962X.1010.xml.
91. PCORI (Patient-Centered Outcomes Research Institute) Methodology Committee. "The PCORI Methodology Report." Section 2: Standards Associated with Patient-Centeredness, PC-1.; 2013. pcori.org/research-we-support/research-methodology-standards.
92. PCORI Engagement Rubric. PCORI (Patient-Centered Outcomes Research Institute) website. http://www.pcori.org/sites/default/files/PCORI-Engagement-Rubricwith-Table.pdf. Published February 4, 2015. Updated February 11, 2015. Accessed May 18, 2015. 2015.
93. Israel BA, Schulz AJ, Parker EA, Becker AB. Community-based participatory research: policy recommendations for promoting a partnership approach in health research. Education for health (Abingdon, England) 2001;14:182-97.
94. Minkler M. Linking science and policy through community-based participatory research to study and address health disparities. Am J Public Health 2010;100 Suppl 1:S81-7.
95. McCaig LF, Hughes JM. Trends in antimicrobial drug prescribing among office-based physicians in the United States. JAMA 1995;273:214-9.
96. Finkelstein JA, Stille C, Nordin J, et al. Reduction in antibiotic use among US children, 1996-2000. Pediatrics 2003;112:620-7.
97. Office-related antibiotic prescribing for persons aged </= 14 years--United States, 1993-1994 to 2007-2008. MMWR Morb Mortal Wkly Rep 2011;60:1153-6.
98. Belongia EA, Knobloch MJ, Kieke BA, Davis JP, Janette C, Besser RE. Impact of statewide program to promote appropriate antimicrobial drug use. Emerg Infect Dis 2005;11:912-20.
99. Dowell SF, Marcy SM, Phillips WR, Gerber MA, Schwartz B. Principles of judicious use of antimicrobial agents for pediatric upper respiratory tract infections. Pediatrics 1998;101, No. Supplement 1:163-5.
100. Finkelstein JA, Huang SS, Kleinman K, et al. Impact of a 16-community trial to promote judicious antibiotic use in Massachusetts. Pediatrics 2008;121:e15-23.

101. Gonzales R, Steiner JF, Lum A, Barrett PH, Jr. Decreasing antibiotic use in ambulatory practice: impact of a multidimensional intervention on the treatment of uncomplicated acute bronchitis in adults. JAMA 1999;281:1512-9.
102. Goossens H, Ferech M, Vander Stichele R, Elseviers M. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. Lancet 2005;365:579-87.
103. Siegel RM. Acute otitis media guidelines, antibiotic use, and shared medical decision-making. Pediatrics 2010;125:384-6.
104. Finkelstein JA, Dutta-Linn M, Meyer R, Goldman R. Childhood infections, antibiotics, and resistance: what are parents saying now? Clin Pediatr (Phila) 2014;53:145-50.
105. Mangione-Smith R, McGlynn EA, Elliott MN, Krogstad P, Brook RH. The relationship between perceived parental expectations and pediatrician antimicrobial prescribing behavior. Pediatrics 1999;103:711-8.
106. Maor Y, Raz M, Rubinstein E, et al. Changing parents' opinions regarding antibiotic use in primary care. Eur J Pediatr 2011;170:359-64.
107. Stivers T, Mangione-Smith R, Elliott MN, McDonald L, Heritage J. Why do physicians think parents expect antibiotics? What parents report vs what physicians believe. J Fam Pract 2003;52:140-8.
108. Weissman J, Besser RE. Promoting appropriate antibiotic use for pediatric patients: a social ecological framework. Semin Pediatr Infect Dis 2004;15:41-51.
109. Finkelstein JA, Stille CJ, Rifas-Shiman SL, Goldmann D. Watchful waiting for acute otitis media: are parents and physicians ready? Pediatrics 2005;115:1466-73.
110. Kiang KM, Kieke BA, Como-Sabetti K, Lynfield R, Besser RE, Belongia EA. Clinician knowledge and beliefs after statewide program to promote appropriate antimicrobial drug use. Emerg Infect Dis 2005;11:904-11.
111. Zhou F, Shefer A, Kong Y, Nuorti JP. Trends in acute otitis media-related health care utilization by privately insured young children in the United States, 1997-2004. Pediatrics 2008;121:253-60.
112. Venekamp RP, Sanders S, Glasziou PP, Del Mar CB, Rovers MM. Antibiotics for acute otitis media in children. Cochrane Database Syst Rev 2013;1:CD000219.
113. Hoberman A, Paradise JL, Rockette HE, et al. Treatment of acute otitis media in children under 2 years of age. N Engl J Med 2011;364:105-15.
114. Coco A, Vernacchio L, Horst M, Anderson A. Management of acute otitis media after publication of the 2004 AAP and AAFP clinical practice guideline. Pediatrics 2010;125:214-20.
115. Lieberthal AS, Carroll AE, Chonmaitree T, et al. The diagnosis and management of acute otitis media. Pediatrics 2013;131:e964-99.
116. Pichichero ME. Antibiotics for acute otitis media: yes or no. JAMA 2015;313:294-5.
117. Ogden CL, Carroll MD, Flegal KM. Prevalence of obesity in the United States. JAMA 2014;312:189-90.
118. Alston C, Paget L, Halvorsen G, et al. Communicating with Patients on Health Care Evidence. Institute of Medicine Discussion Paper;September 2012.
119. Alaszewski A, Horlick-Jones T. How can doctors communicate information about risk more effectively? BMJ 2003;327:728-31.
120. Peters E, Hart PS, Fraenkel L. Informing patients: the influence of numeracy, framing, and format of side effect information on risk perceptions. Med Decis Making 2011;31:432-6.
121. Woloshin S, Schwartz LM. Communicating data about the benefits and harms of treatment: a randomized trial. Ann Intern Med 2011;155:87-96.
122. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science 1981;211:453-8.
123. Hillier TA, Pedula KL, Schmidt MM, Mullen JA, Charles MA, Pettitt DJ. Childhood obesity and metabolic imprinting: the ongoing effects of maternal hyperglycemia. Diabetes Care 2007;30:2287-92.
124. Angier H, Gold R, Crawford C, et al. Linkage methods for connecting children with parents in electronic health record and state public health insurance data. Matern Child Health J 2014;18:2025-33.
125. Heerman WJ, Bian A, Shintani A, Barkin SL. Interaction between maternal prepregnancy body mass index and gestational weight gain shapes infant growth. Acad Pediatr 2014;14:463-70.
126. Renjilian CB, Womer JW, Carroll KW, Kang TI, Feudtner C. Parental explicit heuristics in decision-making for children with life-threatening illnesses. Pediatrics 2013;131:e566-72.
127. Benkler Y, Nissenbaum H. Commons-based Peer Production and Virtue*. Journal of Political Philosophy 2006;14:394-419.

128. Little, R.J.A. Pattern-Mixture Models for Multivariate Incomplete Data. JASA 1993; 88: 125-134.
129. Carpenter & Kenward, Multiple Imputation and Its Application, New York: John Wiley & Sons, 2013