



# Data Sources for Community-Level Outcomes

AUGUST 2016

OAH TIER 1B  
EVALUATION  
TECHNICAL  
ASSISTANCE BRIEF

No. 2

## OVERVIEW

The OAH Tier 1B grant program is intended to infuse communities with evidence-based programming in multiple settings, provide teens with linkages and referrals to youth-friendly health services, and bolster community support and commitment to teen pregnancy prevention. The Tier 1B evaluations need data that can assess the impact of multiple diffuse strategies on an entire community or set of communities. This brief reviews general factors to keep in mind when selecting data sources for an evaluation of a Tier 1B grant project, and highlights the advantages and disadvantages of four data sources for outcome measures that might be appropriate for evaluation activities. Data sources for covariates, the community-level attributes (e.g., demographic characteristics) that are used to identify comparison communities and adjust for differences between treatment and comparison communities in analyses, are beyond the scope of this brief.

## KEY CONSIDERATIONS FOR SELECTING DATA SOURCES TO MEASURE OUTCOMES

Quasi-experimental (QED) evaluation designs require studying only those outcomes for which data for **both** the treatment and the comparison communities are available during the evaluation time period of interest. To study community-level outcomes, the OAH Tier 1B evaluations will use administrative data available at the community level to which the Tier 1B strategy is scaled. Because these data are often collected for non-research purposes, review the data source carefully to understand relevant features and to identify any limitations that will need to be accounted for or otherwise noted in the evaluation. There are four main features of the data source to review.

First, are data available at the appropriate community level? This is the geographic level at which all components of your strategy are available to your target populations, and at which a substantial proportion of eligible youth are affected by the components.<sup>1</sup> Clarify and understand the different populations represented by each administrative data source to ensure that the source measures the outcomes of interest for the target population (e.g., the age range of the teen birth data should match the age range of the teen birth outcome that is identified in your logic model). Any differences in measurement will need to be accounted for in the interpretation of findings.

- *Level and geographic units*
- *Ability to match/link datasets*
- *Timeliness*
- *Quality/completeness/reliability*

Next, are data sources linkable? If you plan to use multiple administrative datasets for your analysis, they should be comparable across treatment and comparison communities, and linked using a common geographic identifier (i.e., the geographically defined community such as city, county, or zip code).

Time lags in the availability of administrative data are common and vary by location and by data source. The data of interest for the evaluation will measure outcomes that occurred at some point after the implementation of the scale-up strategy components among the target cohort(s) of youth. For example, if eligible youth are those age 13-19,

<sup>1</sup> See Tier 1B Evaluation TA Brief No. 1: Defining Treatment Communities and Estimating Community Impacts.

## Data Sources for Community-Level Outcomes

community-level data on outcomes among youth age 13 - 22 could be used to study the impact of the scale-up strategy during the four-year time period of the evaluation (see Figure 1). Carefully plan the timing of your outcome measurement using information about when the necessary outcome data will be available. For most projects, administrative data will reflect the time period from mid-2016 through late 2019.

*It is possible to request vital statistics data by other age group combinations, but teen birth rate data are most commonly reported for the 15-19 age group.*

Lastly, because the evaluation results will help inform decisions about the effectiveness of the strategy, data need to be useful to stakeholders. Examine the relevance, quality, and credibility of the data sources selected for the evaluation, and discuss any data limitations with the agency or organization that created and/or owns the data. Many data issues can be addressed using statistical or analytic methods, and/or discussed in the evaluation report.

*Remember to check if provisional birth rate data are available from the vital statistics department in your state. These data include the full birth record, including mother's residence at time of birth, and may be accessible on a rolling basis as soon as three months after the birth event.*

**Figure 1. Multi-year implementation and timing of teen birth rate outcome measurement for hypothetical Tier 1B project targeting 13-19 year olds in County A**

Years					Outcome: County A birth rate among 15-19 year olds	
Year 1 2015/2016	Cell A Pilot/Planning Year				No effect on birth rate expected because community intervention has not yet started.	
	↓					
Year 2 2016/2017	Cell B 13-19 year olds 1 Yr Exposure	Age range exposed to one or more components of the full scale-up strategy: 13-19			Likely no effect on 2017 birth rate because not enough time has passed or saturation levels achieved for births to be affected significantly. Access provisional 2017 birth data by ~March 2018.	
	↓					
Year 3 2017/2018	Cell C 14-20 year olds 2 Yrs Exposure	Cell F New 13 - 19 year olds 1 Yr Exposure	Age range exposed: 13-20		More cohorts of youth are exposed to the strategy. May begin seeing effect of the strategy on teen birth rate due primarily to Cell C's exposure. Note that birth rate data will not include 13, 14, or 20 year olds also exposed to the strategy. Access provisional 2018 birth data by ~March 2019.	
	↓	↓				
Year 4 2018/2019	Cell D 15-21 year olds 3 Yrs Exposure	Cell G 14 - 20 year olds 2 Yrs Exposure	Cell I New 13-19 year olds 1 Yr Exposure	Age range exposed: 13 - 21	Saturation continues to increase. May see effect of the strategy on teen birth rate due primarily to Cell D and Cell G's exposure. Birth rate data will not include 13, 14, 20, or 21 year olds also exposed to the strategy. Access provisional 2019 birth data by ~March 2020.	
	↓	↓	↓			
Year 5 2019/2020	Cell E 16-22 year olds 4 yrs Exposure	Cell H 15-21 year olds 3 Yrs Exposure	Cell J 14-20 year olds 2 Yrs Exposure	Cell K New 13-19 year olds 1 Yr Exposure	Age range exposed: 13 - 22	Saturation continues to increase. Effect of the strategy on teen birth rate not possible to assess because 2020 birth data are only available in 2021, after the Tier 1B grants end.
<p>Notes: Years are grant years. Birth rate data are available by calendar year. In any given year, a mix of ages are exposed to the full scale-up strategy (EBPs, linkages and referrals to health services, community-wide activities). "Exposure" in the Tier 1B strategy means exposure to one or more components of the full scale-up strategy. Once a youth is exposed a component of the strategy (participates in EBP, gets a referral, is affected by community-wide programming, etc.) they are considered to be exposed. The strategy assumes that youth are exposed to strategy components multiple times over multiple years; youth could be reached by an EBP once, multiple times, or never, as well as exposed to other components of the strategy. The age at initial exposure is used to identify the age range of those exposed. Outcome data will include youth who are not reached by an EBP but are exposed to other components of the strategy.</p>						

### ADMINISTRATIVE DATA TO MEASURE COMMUNITY-LEVEL OAH TIER 1B GRANT PROGRAM OUTCOMES

The community outcomes that your Tier 1B strategy expects to impact should be outlined in your logic models, and might include community rates of STIs, births, pregnancies, sexual risk-taking behaviors, attitudes and skills, health care utilization, and academic performance among teens in the targeted communities. The best data sources for these evaluations are those that measure the expected outcomes of the scale-up strategy in both the treatment and the comparison communities. However, appropriate administrative data may not exist for certain short term or intermediate outcomes that the scale-up strategy expects to change in the target communities. If such data are unavailable, you will want to consider other outcomes that the strategy can reasonably be expected to affect, and for which data are available.

The remainder of this brief highlights key features, advantages, and disadvantages of five potential data sources for outcome measures that could be considered for use in your evaluations. Please note that this is not an exhaustive or prescriptive list; you may have identified other data sources that are specific or unique to your community that would be appropriate for evaluation.

#### Teen birth data

The ultimate health-related outcome measures of interest for Tier 1B program evaluations include teen pregnancies and teen births. In each state and Washington DC, live births, fetal deaths, and induced terminations of pregnancy are reported by physicians or other hospital personnel to the state department of health using established procedures and definitions. State registrars of vital statistics then compile, clean, and publish the data. Calculating pregnancy rates requires states to compile multiple datasets, so lags in data reporting can be lengthy.<sup>2</sup>

County-level birth data for counties with >100,000 population for 2014 and earlier are publicly available from CDC.<sup>3</sup> More recent data (including provisional statistics) and data at other geographic levels may be available from state or

county health departments. Many TPP scale-up strategies will be implemented and evaluated at lower level geographies (e.g., zip code) where only a small number of births occur. Typically, vital statistics counts of 1-9 are suppressed to protect data confidentiality and an Institutional Review Board process may be required to obtain these data.

#### Questions to ask when determining whether teen birth data are appropriate to use in the Tier 1B evaluation activities

- Are data available at the geographic level of interest in both the treatment and comparison locations?
- Given the lag time between the intervention and the outcome of a birth, when will data be available for the age groups affected by the intervention?
- Is data suppression to protect confidentiality anticipated?

#### Healthcare utilization data from community providers

Tier 1B grantees will be establishing and maintaining linkages and referrals to youth-friendly health care services, which might include federally-funded providers such as HRSA-funded Community Health Centers and OASH-funded Title X Family Planning Services, as well as (other) healthcare professionals who can provide high-quality, youth-friendly healthcare services for participants and their families (e.g., other community-based clinics, school-based health centers, and/or individual pediatric providers). The linkages and referrals should then lead to increased healthcare utilization among eligible youth in the target community. Healthcare access and utilization data relevant for a local evaluation can be requested from the service provider(s) (e.g., health center or clinic) in the treatment and comparison communities.

Federally-funded health centers and clinics that provide preventive and primary care to medically underserved populations are an important source of family planning and reproductive health services for youth. Annually, these health centers are required to report data on client characteristics, service providers, and provision of family planning and related preventive health services through the Uniform Data System.<sup>4</sup> If your community is partnering with a federally-funded health center to improve linkages

<sup>2</sup> Pregnancies are defined by the National Center for Health Statistics as the sum of the number of live births, reported induced terminations of pregnancies and reported fetal deaths of all gestations. Some states calculate pregnancy estimates using the estimated number of live births and induced terminations (i.e., excluding fetal deaths).

<sup>3</sup> <http://wonder.cdc.gov/natality.html>

<sup>4</sup> <http://bphc.hrsa.gov/uds/datacenter.aspx?q=d>

and referrals, the health center data on the number of teens accessing care for reproductive health purposes could be used for the evaluation (if data are also available for a comparison group).

### **Questions to ask when determining whether provider healthcare service data are appropriate to use in the Tier 1B evaluation activities**

- Which providers serve the target population, or are the target of the scale-up strategy's referral activities?
- Who are the comparable providers in the comparison community?
- Do clinics have aggregate data on the services of interest available for use by researchers? How many providers are represented in the data? Are similar measures captured across individual clinics/health centers/providers?

Title X grantees (state or local health departments, nonprofit organizations, community health centers (CHCs), and other federally qualified health centers (FQHCs)) are also required to report on clients and services. Specifically, Title X grantees are required to annually report data from all of their funded service sites on client characteristics, service providers, family planning, and related preventive health services.<sup>5</sup> If your community has a Title X-funded service site with which you are partnering for linkages and referrals, the Title X program utilization data reported in the Family Planning Annual Report (FPAR) could be used for the evaluation, if data are also available for a comparison community.

Other ambulatory practices that are used by the target population may be capturing data on health care utilization measures relevant to the Tier 1B program intervention for other reporting activities or for ongoing operational management and program improvement.<sup>6</sup>

### **Youth Risk Behavior Surveillance System (YRBSS) data**

The Youth Risk Behavior Surveillance System (YRBSS) monitors several sexual risk behavior outcome measures of

<sup>5</sup> <http://www.hhs.gov/opa/pdfs/fpar-reissued-oct13.pdf>

<sup>6</sup> E.g., the AHRQ-CMS Pediatric Quality Measures Program, which developed a measure of sexual activity status among adolescents (see <http://www.ahrq.gov/policymakers/chipra/factsheets/index.html> for details), or measures utilized by health insurers, such as the HEDIS measures on chlamydia screening rates.

interest for Tier 1B evaluations like teen sexual activity, sexual initiation, condom or other contraceptive use, and knowledge among middle and high school students in the United States. However, **the YRBSS might not be appropriate for your community context due to limited data availability and geographic scope.**

The YRBSS uses surveys of representative samples of students conducted in selected public middle and high schools every two years by CDC or state or local government agencies. The survey design allows CDC to calculate estimates of youth behaviors at the national and state level, as well as for over twenty large urban school districts. Because the design involves data collection from only a sample of schools in a given geographic area (state or school district), schools in the Tier 1B intervention and comparison communities may or may not be selected to participate in the survey in a given survey year. If some schools in your treatment communities conduct the YRBS independently, you will want to assess if it is conducted consistently over a period of time, the extent to which those schools align with the geographic boundaries of the treatment communities, are representative of the target community, and whether the same data are collected in comparison communities.

### **Questions to ask when determining whether YRBSS data are appropriate to use in the Tier 1B evaluation activities**

- Did the treatment communities and target age group (districts, high schools, or middle schools) participate in the YRBSS prior to the intervention and during the intervention evaluation period (i.e., in 2015 and 2017; ideally in 2010, 2013, 2015, 2017, and 2019)?
- Did the comparison communities and target age group also participate in the YRBSS prior to the intervention and during the intervention evaluation period (i.e., in 2015 and 2017; ideally in 2010, 2013, 2015, 2017, and 2019)?

### **Education data**

Some scale-up strategies might be expected to impact school-related outcomes in the aggregate, including rates of academic performance, graduation, or dropout.

## Data Sources for Community-Level Outcomes

For schools that receive public funds as their primary support, state education agencies are required to submit graduation rate data and other academic performance data to the US Department of Education (ED). The data are centralized in the ED Facts data management system.

Tables 1 – 4 summarize the features of each of these data sources and provide links to more information. If you have questions about any of these data sources or would like to discuss them with an evaluation TA provider, please contact your OAH project officer.

### Questions to ask when determining whether academic outcome data are appropriate to use in the Tier 1B evaluation activities

- Is the scale-up strategy expected to impact academic outcomes?
- Can we obtain the data more quickly by contacting school districts directly?
- Is the comparison community adjacent to the treatment community, with potential contamination for school-related outcomes?

**Table 1. Outcome Dimension – Births**

Data Source	Measures	Level/Units	Data Availability	Advantages	Disadvantages	Recommended Data Access
Vital statistics	Birth counts	<ul style="list-style-type: none"> <li>• County-level</li> <li>• Zip code-level may be available on request</li> </ul>	<ul style="list-style-type: none"> <li>• Provisional data: as soon as 3 months after the birth event</li> <li>• Final data: 9-12 months after close of calendar year</li> </ul>	<ul style="list-style-type: none"> <li>• Reliable.</li> <li>• High quality</li> <li>• Linked to maternal legal residence at time of birth</li> <li>• Historical trends</li> <li>• Available in comparison communities</li> </ul>	<ul style="list-style-type: none"> <li>• Defining rates</li> <li>• State regulations and/or small counts may limit data accessibility</li> <li>• Distal outcome</li> </ul>	State or county health department vital statistics department

**Table 2. Outcome Dimension – Health Care Utilization**

Data Source	Measures	Level/Units	Data Availability	Advantages	Disadvantages	Recommended Data Access
Federally-supported health centers <sup>7</sup>	<ul style="list-style-type: none"> <li>• Health center user characteristics<sup>8</sup></li> <li>• Number of patient visits for medical services</li> <li>• Number of patient visits for STI screening test, contraceptive management</li> </ul>	<ul style="list-style-type: none"> <li>• Clinic or health center</li> </ul>	<ul style="list-style-type: none"> <li>• Reported annually<sup>9</sup></li> <li>• As of July 2016, 2014 data publicly available</li> <li>• More timely data may be available directly from health center</li> </ul>	<ul style="list-style-type: none"> <li>• Reliable</li> <li>• Proximal outcomes</li> <li>• Historical trends.</li> <li>• State and national benchmarks</li> </ul>	<ul style="list-style-type: none"> <li>• Target clinics may not report</li> <li>• Defining rates</li> <li>• Up to 2 year lag.</li> <li>• Potential restrictions to accessing data</li> </ul>	Individual health centers
Title X program	<ul style="list-style-type: none"> <li>• Family planning user characteristics</li> <li>• STI screening rates</li> <li>• Family planning encounters</li> </ul>	Clinic	<ul style="list-style-type: none"> <li>• Reported annually</li> <li>• As of July 2016, 2014 data publicly available</li> <li>• More timely data may be available directly from grantee or clinic<sup>10</sup></li> </ul>	<ul style="list-style-type: none"> <li>• Reliable.</li> <li>• Proximal outcomes</li> <li>• Historical trends</li> </ul>	<ul style="list-style-type: none"> <li>• Target clinics may not report</li> <li>• Up to 2 year lag</li> <li>• Potential access restrictions</li> <li>• Comparison data</li> </ul>	Individual clinics or Title X grantee

<sup>7</sup> For locations of federally funded health centers, see <http://nachc.org/research-and-data/state-level-data-maps/>

<sup>8</sup> See <http://www.bphc.hrsa.gov/datareporting/reporting/2015udsmanual.pdf> for details on all measures.

<sup>9</sup> Historic data are available from HRSA: see <http://bphc.hrsa.gov/datareporting/index.html>

<sup>10</sup> Grantees funded by Title X are required to submit the FPAR. Grantee sub-recipients receive Title X funds via the grantee and follow grantee instructions for data collection and reporting. Currently, 78 grantees provide Title X services through approximately 4,000 clinics (<https://www.opa-fpclinicdb.com/>).



## Data Sources for Community-Level Outcomes

**Table 3. Outcome Dimension – Sexual Risk Behaviors**

Data Source	Measures	Level/Units	Data Availability	Advantages	Disadvantages	Recommended Data Access
Youth Risk Behavior Survey	Sexual behaviors related to unintended pregnancy and sexually transmitted infections, including HIV infection <sup>11</sup>	<ul style="list-style-type: none"> <li>State (weighted)</li> <li>Selected large school districts (weighted)</li> <li>Some individual schools (unweighted)<sup>12</sup></li> </ul>	<ul style="list-style-type: none"> <li>Several large districts for 1991-2015 available from CDC</li> <li>Final 2017 survey data available in 2018</li> <li>Final 2019 survey data available in 2020</li> </ul>	<ul style="list-style-type: none"> <li>Proximal outcomes</li> <li>Some schools may conduct YRBS independently</li> <li>Historical data may be available.</li> </ul>	<ul style="list-style-type: none"> <li>Target schools or districts may not participate</li> <li>Conducted biannually</li> <li>School-level data cannot be generalized to community</li> <li>Data may not be available in comparison communities.</li> </ul>	Individual school districts or CDC

**Table 4. Outcome Dimension – School Performance**

Data Source	Measures	Level/Units	Data Availability	Advantages	Disadvantages	Recommended Data Access
School-level performance data	<ul style="list-style-type: none"> <li>Test scores, proficiency levels.</li> <li>Cohort graduation rate</li> <li>Dropout rates</li> </ul>	<ul style="list-style-type: none"> <li>School</li> <li>Other geographies may be available on request from state or local education agencies</li> </ul>	<ul style="list-style-type: none"> <li>Reported annually</li> <li>Data from 2009-2010 through 2013-2014 are publicly available<sup>13</sup></li> <li>More recent data may be available directly from individual schools</li> </ul>	<ul style="list-style-type: none"> <li>Reliable</li> <li>Some data are available by age or grade</li> <li>Historical trends</li> <li>Can link data to other datasets</li> </ul>	<ul style="list-style-type: none"> <li>May not be relevant to the community-wide strategy</li> <li>Up to 2 year time lag in data availability<sup>14</sup></li> </ul>	US Department of Education EDData and/or individual schools or school districts.

11 For a list of all available measures, see the survey instrument and other documentation: <http://www.cdc.gov/healthyouth/data/yrbs/questionnaires.htm>; <http://www.cdc.gov/healthyouth/data/yrbs/data.htm>

12 The response rate for the survey varies by location. When the response rate is less than 60% the data are considered lower quality and are not weighted. Unweighted data represent only the students who completed the survey. With weighted data, it is possible to generalize; for example, “X% of students in location Y received HIV education.”

13 As of July 2016. <http://www2.ed.gov/about/inits/ed/edfacts/data-files/index.html>

14 Data are submitted between six and twelve months following the end of the school year.

### **Authors**

*This brief was written by Sara Donahue and Kim Francis from Abt Associates for the HHS Office of Adolescent Health under Contract No. HHSP233201500069I, Task Order No. HHSP23337003T*