

Review of Data Sources Relevant to Tracking Asthma in Environmental Public Health Tracking

Prepared by
Tulane University Center for Applied Environmental Public Health
Academic Partner of Excellence in Public Health Tracking

12/01/2008

Introduction

Environmental Public Health Tracking is defined as “the ongoing collection, integration, analysis, interpretation, and dissemination of data on human exposure to environmental hazards and health effects that may be related to those exposures”, (CDC website <http://www.cdc.gov/Features/PublicHealthTracking/> accessed 7/1/2008). The Centers for Disease Control and Prevention (CDC) is leading the effort to build a national Tracking Network that will integrate data from a variety of sources for the purpose of simultaneously monitoring environmental hazards and associated health. Information from the network will be available to a variety of users, including local, state, and federal public health professionals, policy makers, researchers, media, and the general public. The diversity of the target audience paired with the variety of potential data sources requires thorough documentation of the issues associated with using each data source, including data type and quality, availability and access issues, and uses and limitations, so that appropriate selection, analysis, and interpretation of data are possible. This paper examines these issues for asthma, one of the focal health outcomes of the Tracking Network.

Selection of the appropriate data to include in a tracking system for asthma and environmental hazards requires a complete understanding of the questions to be answered. There exist a number of methods of characterizing asthma diagnoses and events, as well as various indicators for asthma. While some research studies, particularly those aimed at determining etiology, may attempt to quantify asthma incidence, measures of prevalence are more readily available in routinely collected data, and are therefore more suited for use in tracking.

Various sources of asthma health outcome and indicator data are available for potential use in an environment/asthma tracking system. National surveys like the Behavioral Risk Factor Surveillance System (BRFSS) provide a national perspective on prevalence of asthma diagnosis as well as on prevalence of asthma exacerbation events. Data compiled at the state level are available for similar types of surveys, with geographic resolution at the county or smaller. State data on hospitalizations or emergency department visits for asthma and asthma related diagnoses provide a source of determining prevalence of acute exacerbations. An understanding of the differences between the data sources and the information they provide is essential to making meaningful inferences from the tracking system. National mortality data can also provide information regarding deaths attributable to asthma. School health data can provide information regarding asthma prevalence and covariates, although the Family Educational Rights and Privacy Act (FERPA) limits routine access to the data and therefore to its usefulness in Tracking. Administrative billing data, such as is available for Medicare and Medicaid patrons through the Centers for Medicare and Medicaid Services (CMS) is a potentially useful source of diagnostic and event prevalence data for sub-sets of the general population.

Issues related to data quality, access, benefits and limitations for each of these data sources are presented below.

Review of Data Sources

National Surveys

On a national level, information regarding asthma prevalence is provided by national surveys, namely the BRFSS, the National Health Interview Survey (NHIS), and the National Hospital Discharge Survey (NHDS). Attention must be paid to the wording of the asthma-related questions as they are presented within each of these surveys, as slight differences in wording from survey to survey, or within a survey over time, could change the meaning of

the question. This has the potential to result in data that lacks standardization from survey to survey or time period to time period. Understanding the source of the data will facilitate appropriate use of the data in the Tracking framework.

Behavioral Risk Factor Surveillance System

The BRFSS is a state-based survey conducted via a computer-assisted telephone interview (CATI) to respondents to provide responses for a representative sample of all households with telephones in each state. According to a BRFSS information page (<http://www.cdc.gov/brfss/about.htm>) on the CDC website, “data are collected monthly in all 50 states, the District of Columbia, Puerto Rico, the U.S. Virgin Islands, and Guam. More than 350,000 adults are interviewed each year, making the BRFSS the largest telephone health survey in the world.” BRFSS data are useful for identifying trends in *self-reported* asthma from year to year and by selected grouping factors (age, race, income) at the state and national levels. Additionally, data from the “The Selected Metropolitan/Micropolitan Area Risk Trends” (SMART) project provide these same data (minus grouping variables) at Metropolitan Statistical Area (MSA) or county level geography.

Data are generated through a telephone survey, and so generalizability of results is limited to similar households. Parameters of data quality, particularly response rates, are reported in the BRFSS Summary Data Quality Report, with 2007 being the most recent available, (<ftp://ftp.cdc.gov/pub/Data/Brfss/2007SummaryDataQualityReport.pdf>). According to the report, the response rate by state for 2007 ranged from 25.9% to 65.6% with a median of 50.58%. Contrasting this to the response rate of 72% in 1993 provides evidence that the participation rate for the BRFSS has declined over the past 10 to 15 years, (Fahimi, et. al., 2008). Fahimi and colleagues compared weighted estimates for selected information between the 2004 BRFSS and the 2004 NHIS and NHANES surveys, and found that certain health prevalence estimates were similar between the BRFSS and NHIS, and only one was similar between BRFSS and NHANES. In terms of asthma, the BRFSS estimates consistently exceeded NHIS estimates for prevalence of “ever having been told by a physician that the respondent had asthma”, both overall and for age, sex, and race sub-groups. This disparity in estimates may be due to the difference in the way the surveys are conducted. NHIS is conducted in person and has a significantly higher response rate (84% - 87%).

While the response rate indicates the frequency of non-response, it does not address any differences between respondents and non-respondents. This should be kept in mind when interpreting results of analyses using BRFSS data. If respondents are systematically different from non-respondents in relevant ways, the generalizability of the results is limited in ways that cannot be quantified from the available data.

Furthermore, BRFSS sampling methods were changed in 2004 to exclude telephone banks holding unlisted telephone numbers, (Fahimi, 2008). This further reduces the generalizability of the results to only those households with working telephones and publicly available telephone numbers.

State to state differences in response rates must also be considered when interpreting results of analyses using BRFSS data. These differences might be attributable to differences in the types of phone service used across states, differences in the underlying population, or to differences in the way the data are collected. The potential for these differences to affect comparability of the data from state to state should be considered.

Availability/Access – State level and SMART BRFSS data are available free of charge for public use through CDC’s BRFSS website. The website provides summary tables for user-

selected states and health-risk category, as well as complete data downloads with documentation found at http://www.cdc.gov/brfss/technical_infodata/surveydata.htm.

Benefits – BRFSS data are representative of the state population of households, and provide information at the state level and at a national level. These data are useful for establishing baseline prevalence, trend analysis, and comparing state to state and state to national prevalence estimates. Additionally, SMART data provide information at county or MSA levels, providing a smaller unit of analysis than just the state.

The survey collects demographic information, allowing for stratified and/or adjusted analysis. BRFSS data are useful for tracking trends in asthma prevalence over time. The data allow for tracking fluctuations in asthma prevalence, and can be incorporated into a health/environment tracking system to monitor concomitant fluctuations in asthma and relevant environmental parameters. This is particularly important for inclusion of BRFSS data in EPHT. BRFSS data are collected monthly, and are useful in situations where the time trend to be analyzed is month or greater.

The asthma-related questions in the 2007 BRFSS provide prevalence information for children and adults, as well as detailed asthma history, exacerbation, and healthcare use for adults. This information, in addition to just prevalence, is potentially useful in providing additional indicators for monitoring the impact of environmental conditions on respiratory health among individuals with asthma.

Limitations – BRFSS is a telephone survey, requiring respondents to have a working telephone in their homes. This limits the generalizability of results of analysis using BRFSS data to the population of households with phones. Some data indicate that income is inversely related to asthma exacerbation and healthcare use for asthma. To the extent that this is true for households that lack telephone coverage, the BRFSS might underestimate asthma prevalence and exacerbations by neglecting to capture the experience of this high-risk group.

Trends in asthma prevalence/exacerbations are available on a monthly basis. Analysis that relies on more frequent measures, such as might be the case in tracking asthma exacerbations and high ozone days, would not be supported by the temporal frequency of the BRFSS data.

CDC occasionally changes questions included in the survey, or changes the coding of responses. These changes have the potential to disrupt the ability to monitor trends over time, if the changes in wording or coding change the meaning of the information gathered.

Analytic Considerations – The BRFSS offers a relatively unique source of information about health and health risk behaviors for a largely representative sample of state and national populations. While the SMART data provides information at a smaller geographic area than the state, the geographic resolution of the general BRFSS survey results precludes regional subgroup analysis within the state.

As discussed above, the fact that the survey is conducted by telephone limits the ability to generalize results to the entire U.S. population. Households without a working telephone and a published telephone number are systematically excluded from the survey. Since this may be related to income, and income may be related to asthma exacerbation and healthcare use, this may represent the systematic exclusion of a group at high risk for asthma prevalence and exacerbation.

Data collection is coordinated at the state level rather than at the national level, meaning that differences may exist from state to state in the way data are collected. These differences should be considered when comparing BRFSS results from one state to another.

Applicability to EPHT – Asthma data generated by the BRFSS are useful for monitoring trends in asthma prevalence, history, exacerbation, and healthcare use over time and, to an extent, space. This data is useful for tracking in that it provides a routine and relatively standard source of asthma prevalence data that can be incorporated into a Tracking system and monitored in conjunction with relevant environmental data. Limitations in the data must be considered so that appropriate interpretations are made.

National Health Interview Survey (NHIS)

According to the NHIS descriptive website, (<http://www.cdc.gov/nchs/about/major/nhis/hisdesc.htm>), the National Health Interview Survey (NHIS) is a national, computer assisted personal interview (CAPI) conducted by the National Center for Health Statistics (NCHS) of the CDC. The survey is administered to participants selected through a multi-stage, probabilistic, stratified cluster sampling process, designed to provide a nationally representative sample. Survey modules gather information for households, randomly selected individuals in selected households, and randomly selected children in selected households. The survey provides information on the health of the U.S. civilian, non-institutionalized population. The primary purpose of the data collection is to provide information regarding national trends for a broad range of health topics, illness, and disability, including asthma.

Data Type/Quality – Asthma information from the NHIS indicates prevalence of an asthma diagnosis or complications from asthma, and is similar to the asthma information provided by the BRFSS. Questions related to asthma address whether the respondent has ever been told they have asthma, whether they still have asthma, whether they have experienced an asthma attack in the previous 12 months, and whether they have used emergency or urgent care services because of asthma in the preceding 12 months. These questions are asked in both the adult and child modules. An additional question in the adult module asks for the respondent to indicate for how long they have had asthma.

As mentioned above, the response rate for the in-person NHIS is much higher than for the telephone-conducted BRFSS. Depending on the module, (family, individual, child), response rates for the NHIS range from 84% to 87%. Nevertheless, certain issues related to the way respondents are selected have the potential to influence the generalizability of the results. “Because of technical or logistical problems, several segments of the population are not included in the sample or in the estimates from the survey. Examples of persons excluded are patients in long-term care facilities; persons on active duty with the Armed Forces (though their dependents are included); persons incarcerated in the prison system; and U.S. nationals living in foreign countries”, (National Health Interview Survey Description web-page: <http://www.cdc.gov/nchs/about/major/nhis/hisdesc.htm>, accessed 11/08/08). These exclusions must be considered when interpreting findings from an asthma/environment tracking system, particularly regarding the applicability of the results to the general population.

Availability/Access – National-level NHIS survey data are publicly available for download free of charge at the NCHS website (http://www.cdc.gov/nchs/about/major/nhis/quest_data_related_1997_forward.htm). Certain data-use restrictions apply, including the use of the aggregate data for “health statistical reporting and analysis”, and restrictions on efforts to identify individual participants. State identifiers are not included in the public release data, but can be

requested through the NCHS Research Data Center. Requesting data in a format other than downloading from the internet is subject to fees.

Benefits – One of the primary benefits of the NHIS data is that the survey collects detailed information regarding socioeconomic and demographic characteristics of respondents. These variables can be used to perform stratified analysis, resulting in detailed profiles of health statistics for specific age, sex, race, and socioeconomic sub-groups in the U.S. Additionally, the data are sampled to provide a representative sample of non-institutionalized, non-military residents of the U.S. The response rate for the NHIS varies from 84% to 87%, depending on the module. This is much better than the response rate of ~51% for the BRFSS, and may indicate reduced response bias for the NHIS when compared with the BRFSS.

Limitations – Although the sample is drawn from each state and the District of Columbia, the sample is too small to provide state level data with acceptable precision for all but large states. To achieve sufficient power to estimate results for a single state, data results for the state must be pooled over time. Because of this, the geographic resolution of one year of data is limited to the entire U.S. and is therefore of limited usefulness to Tracking for most states. Use of the data at geographic levels of precision more refined than the national level requires pooling of the data over time.

Additionally, the data represent annual survey results. The data are not available at a temporal precision of less duration than year. While this does not preclude the data from usefulness, it limits the data to providing annual trends of asthma diagnostic and episodic prevalence. Tracking projects requiring more frequent asthma-related estimates, such as monitoring asthma exacerbations and fluctuations in routinely collected air quality parameters would not be supported by NHIS asthma estimates.

Applicability to EPHT – The usefulness of NHIS data to Tracking is limited due to the fact that this is a national survey providing data at a national geographic resolution and annual temporal resolution. Using NHIS asthma data for Tracking within a single state, or by state for a group of states, would require pooling the data over time for each state. Estimates at geographic levels lower than the state are not supported. Temporal precision shorter than a year is not available.

National Hospital Discharge Survey (NHDS)

The National Hospital Discharge Survey (NHDS) is one of the health-care provider-based surveys in the suite known as the National Health Care Surveys. The NHDS utilizes a multi-stage probability sampling design to collect information from randomly selected persons as they are discharged from hospital. The sampling strategy is designed to provide a national probability sample of inpatient stays in non-federal, short-term hospitals, and is not intended to provide information at geographic levels more precise than national. The unit of analysis is the discharge event rather than the person, so that an individual who experienced more than one hospital stay and discharge during the calendar year could be selected more than once for inclusion. For this reason, the survey is not useful for providing prevalence estimates. Federal, military, VA, and institutional hospitals (e.g. prisons) are excluded from eligibility.

Type/Quality – The NHDS collects information about short-term hospitalizations at the time of hospital discharge. The unit of analysis is the discharge, meaning that an individual person can be counted more than once during a calendar year. Information gathering for the NHDS involves two methods; the first involves record abstraction for sampled discharge events and the other involves the purchase of electronic medical record data from which

systematic samples are taken. Information collected relevant to the hospital stay includes admission and discharge dates, patient demographic information, payer source, type of admission, patient's status on discharge, diagnostic codes and procedural codes. Diagnostic codes are used to identify the frequency of events that list a particular health outcome, such as asthma, as either a principal or secondary reason for the hospitalization.

The data are compiled in annual installments, with shorter duration time periods of data unavailable. Hospital response rates have improved over the past two decades from 79.5% in 1988 to over 90% in 1997 and 92% in 2006.

Availability/Access – According to the NHDS website, (<http://www.cdc.gov/nchs/about/major/hdasd/nhdsdes.htm>), “Data from the NHDS are available annually and in publications, on public-use data tapes, data diskettes, CD-ROMs and downloadable files from the FTP server. Information pertaining to date of birth and patient’s residential ZIP code are considered protected health information, and are not available in the public use dataset. These variables can be used for research pending approval by the NCHS Research Data Center.

Benefits – The NHDS has continuously collected national information regarding hospital discharges since 1965 and was redesigned in 1988 to make use of new technologies and analytic techniques. The sampling frame is periodically updated to ensure that newly eligible hospitals are given an opportunity for selection.

The NHDS is a potential source of information related to identifying trends in asthma hospitalizations from year to year, with the potential to control for a number of covariates. Covariate data collected by the NHDS is vast and allows for stratified analysis of asthma hospitalizations by age, race, sex, marital status, payer type, and discharge status. Asthma is diagnosed by a physician and identified in the record through the use of ICD-9CM codes.

Limitations – Data are reported annually for the NHDS, precluding analysis for shorter duration time intervals, as may be relevant for an asthma/environment Tracking system. Additionally, the geographic resolution is at the national level. While information regarding respondents’ residential ZIP codes is collected, it is not routinely available in the public use dataset. While it is accessible through the NCHS Research Data Center, the sample for the NHDS is drawn to be nationally representative, meaning that the data lacks power for more precise geographic level analyses.

Analytic Considerations – The primary consideration for using NHDS data in Tracking lies in the fact that the primary unit of measure is the hospital discharge and not the individual person. This fact precludes use of this data for estimation of asthma prevalence as an individual can be counted more than once per year. The data is useful, however, to monitor annual trends in inpatient hospital use for asthma and asthma-related illnesses.

Applicability to EPHT – Because the geographic resolution of the NHDS is the nation, and because the temporal resolution of the data is annual, applicability of the NHDS for use in Tracking is limited. Linking asthma health outcome data with environmental parameters requires data points of shorter duration, with daily intervals appropriate in some cases.

The National Hospital Ambulatory Medical Care Survey (NHAMCS) and the National Ambulatory Medical Care Survey (NAMCS) are two components of the National Health Care Surveys, and are similar to the NHDS in data collection, variables, benefits, limitations, and applicability to Tracking.

State Hospitalization/Emergency Department Data

States collect information related to hospitalizations and emergency department (ED) visits for a number of purposes, including monitoring health and healthcare utilization trends and to provide information for resource allocation, policy planning, and evaluation. These datasets are typically compiled by state agencies such as the health department, although some states use the services of private entities. Because hospital discharge/ED data efforts are conducted at the state level, “differences among the states abound...with regard to the specific data elements collected, how they are defined, data completeness, voluntary vs. mandatory data submission, and policies regarding data release,” (Schoenman, et. al., 2007). Nevertheless, much of the data collected is consistent from state to state and amenable to multi-state analysis. Additionally, these data generally provide more detailed information than data available through surveys.

Type/Quality – The manner in which the data are compiled varies from state to state. While some states rely on voluntary reporting, others mandate that hospital discharge data be submitted to the system. Generally, the completeness of hospital discharge data is very good, due to the fact that most or all of the information collected through hospital discharge surveillance is required for billing payers for service. Because of this, the data, at least with respect to those items necessary for billing, tends to be complete.

In general, these hospital discharge/ED data collection systems compile information regarding patient information including demographics (age or age group, race, sex); patient residential geographic information; admission and discharge dates; diagnostic and procedure codes; provider information; facility information; payer source; patient discharge status; and some measure of monetary expenditure for the stay/visit. A convenience sample of four hospital discharge data systems was assessed for this review and each contained measures related to these variable groups. There were, however, noticeable differences in the specific information collected, as well as in the precision of the information. For instance, one of the states collects the patient’s residential ZIP code, and provides this data in the public use data file. The remaining four systems reviewed either did not provide ZIP code, or provided only the first 3 digits in the public use files. Most systems provided the patient’s county.

Hospitalizations associated with a diagnosis of asthma can be identified by the diagnostic codes (ICD-9CM) included for all of the hospital discharge data reviewed. In addition to primary diagnosis code, each of the sources reviewed accommodates several secondary diagnosis codes, as well as procedural codes that identify billable procedures performed during the hospital stay.

Availability/Access – The availability and access constraints of state level hospital discharge public use data varies by state. For two states reviewed, hospital discharge data was downloadable from the states’ websites. Both files contained record-level data, stripped of variables that could be used to identify an individual. Dates of admission and discharge were limited to quarter and year for one state, and limited to year only for the other. Both states provided 3-digit ZIP codes for patients’ residence.

Accessibility to one of the remaining two states included in this review required that a request be submitted to the health department, and a successful application included payment of a nominal fee. According to the data documentation, patient’s ZIP code and county are listed for each discharge record, and the month and year of discharge are included, along with length of stay.

The final dataset assessed required that individuals outside of the health department contact the state's hospital association for information about acquiring the data. The state health department was provided with the data by the hospital association from which they calculated summary statistics (counts and rates) by county, which were posted for public consumption.

From this relatively small sample, it is evident that access to states' hospital discharge data varies widely by state, and that publicly available and easily accessible data understandably omits precise geographic and temporal information in an effort to preclude identification of individuals.

Benefits – Schoenman and colleagues (2007) conducted a review of state based hospital discharge datasets and suggested various benefits and limitations. Among the benefits they identified were cost, accuracy and completeness, representativeness and inclusiveness, and range of data elements and potential for data linkage. State based hospital discharge datasets are often very thorough, both in terms of the scope of data collected, and the population covered by the system. Paired with the relative low cost of compiling and acquiring this data compared with primary data collection, the ratio of benefit to cost is high.

Limitations – While many state hospital discharge datasets collect information that is representative of the population of hospitalizations, some states exclude certain types of hospitals and or providers from contributing to the dataset, for example military or VA hospitals. While a within-state analysis would be appropriate with reference to the limitations to generalizability, between-state comparison of states that identify sources of discharge data differently, without adjusting for these differences, threatens the validity of the analysis.

Furthermore, problems may be introduced by the lack of consistency in the specific variables collected by different state hospital discharge systems. While this may present a problem with respect to very specific information, the systems we reviewed all had the same primary components. Another source of this data that may mitigate this issue is the Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample (NIS). According to the Agency for Healthcare Research Quality, "The NIS is the largest all-payer inpatient care database that is publicly available in the United States, containing data from 5 to 8 million hospital stays from about 1,000 hospitals sampled to approximate a 20-percent stratified sample of U.S. community hospitals." The NIS currently has 38 state participants, and draws from each state's hospital discharge data collection efforts.

Additional problems potentially are introduced due to differences in the way illnesses are diagnosed or coded from state to state, or even from area to area within a state. Previous reports have indicated the possibility that asthma diagnoses differ by region and other characteristics such as race, (Mannino, et. al.). While these issues are largely intangible, they should nevertheless be considered as a possible alternative explanation for findings.

Differences in the precision with which data are recorded, particularly temporal information regarding the hospitalization, and patient geographic information, introduce a potential problem in linking the data with other state hospital data and with environmental data. For an asthma and environment tracking system, the health outcome data will be linked to the environmental data by geographic and temporal variables. Weekly or even daily fluctuations in air quality parameters are relevant to a potential impact on asthma hospitalizations. Environmental data will require temporal and spatial aggregation to

successfully link it with asthma hospitalization data. While possible, this practice dilutes the information and limits the meaning of the linkage.

Applicability to EPHT – Asthma related hospital discharge/ED data are applicable for use in EPHT and represent one of the most complete routinely collected health outcome datasets with specific information regarding asthma exacerbations. While certain issues may need to be addressed in using this data for tracking, state based hospital discharge data represents one of the best available routine resources for asthma health outcomes.

Access to more detailed information may require application to state health departments and development of data share agreements, etc.

Centers for Medicare and Medicaid Services (CMS) Data

The Centers for Medicare and Medicaid Services (CMS) makes available an array of data products created from billing data. While the data exist for the purpose of billing, they are valuable resources for health research, containing a host of clinical information related to Medicare and Medicaid beneficiaries' health care utilization.

CMS provides two tiers of data potentially relevant for Tracking. Research Identifiable Files (RIF) and Limited Data Sets (LDS) are standard analytic files (SAF) that contain utilization information from inpatient, outpatient, physician's office, skilled nursing facility, home health, hospice, and durable medical equipment utilization, in addition to a denominator file that contains beneficiary information. The difference between the RIF and the LDS is the level of precision provided for certain identifying variables. In the LDS, service dates are replaced by quarter and year, beneficiary's age is aggregated to 5 year intervals, and the geographic precision is the county as contrasted with ZIP code in the RIF. The RIF files contain person-record level data for providers and beneficiaries, allowing identity reconstruction. The files relevant to tracking asthma in EPHT are described briefly below. This information was obtained from the Research Data Assistance Center (RESDAC) at the University of Minnesota,

(http://www.resdac.umn.edu/Medicare/file_descriptions.asp#carrier, accessed 8/19/2008):

Inpatient SAF – contains claims data submitted by hospitals; includes information regarding diagnosis (ICD-9), procedural codes, dates of service, and demographic information for the beneficiary.

Outpatient SAF – contains claims data submitted by institutional outpatient providers; includes information regarding diagnosis, procedural codes, dates of service, and demographic information for the beneficiary.

Skilled Nursing Facility SAF – contains claims data submitted by SNF providers. Some of the information contained in this file includes diagnosis and procedure (ICD-9 diagnosis and ICD-9 procedure code), dates of service, reimbursement amount, SNF provider number, and beneficiary demographic information.

Home Health Agency SAF – contains final action claims data submitted by HHA providers. Some of the information contained in this file includes the number of visits, type of visit, diagnosis (ICD-9 diagnosis), the dates of visits, and beneficiary demographic information.

Carrier SAF – contains claims data submitted by non-institutional providers, including physicians, physicians assistants, nurse practitioners, and others; includes information regarding diagnosis and procedure codes, dates of service,

and beneficiary demographic information.

Denominator File - contains demographic and enrollment information about each beneficiary enrolled in Medicare during a calendar year. Some of the information contained in this file includes the beneficiary unique identifier, state and county codes, ZIP code, date of birth, date of death, sex, race, and age.

Like the Medicare data, Medicaid claims data are contained in a series of files, called Medicaid Analytic Extract (MAX) files. Billing data obtained from Medicaid represent the health experience of a subset of the population, specifically, persons of low-income. The majority (58%) of Medicaid beneficiaries are younger than 21 years of age (Wenzlow, et al., 2007).

Medicaid MAX files are arranged similarly to Medicare SAF files, with separate data components contained in separate files. Files relevant to asthma tracking are described below. Descriptions of these files were obtained from the RESDAC website (http://www.resdac.umn.edu/Medicare/file_descriptions.asp#carrier, accessed 8/18/2008):

Inpatient MAX File - contains complete stay records for enrollees who used inpatient services; data includes diagnosis codes, procedure codes, and discharge status.

Drug MAX File – contains paid drug claims and related information; data includes National Drug Codes (NDC codes), which can be used to identify the product and dosage information.

Other Therapy MAX File – contains claim records for non-institutional Medicaid services, including physician services, lab/X-ray, clinic services; data includes diagnosis codes, procedure codes, and date of service.

In addition to these three clinically relevant files, Medicaid MAX files include a Person Summary file that contains relevant information regarding the beneficiary, including demographic data, date of birth, and geographic information down to ZIP code.

Type/Quality – Data available through CMS is healthcare utilization data compiled for the purpose of billing for services. As such, research is a secondary use of the data. This could result in underreporting or missing information for research relevant fields if these fields are not required for processing payments.

CMS data are record level data for health services claims, utilizing standard coding schemes for recording diagnoses and procedures (ICD-9). The data are useful for identifying asthma diagnoses and estimating prevalence (among the specific population represented) as well as complications resulting in healthcare use, including outpatient, inpatient, and health care providers' offices.

Medicare claims data is a source health data for U.S. citizens age 65 years and older, who are enrolled in the program. A large majority of U.S. citizens in this age group are enrolled in Medicare. Individuals with certain disabilities and End Stage Renal Disease are also eligible for Medicare, although those enrolled under these criteria are identifiable in the dataset.

Medicaid availability is based on income and other criteria defining eligibility groups. Medicaid is administered by each state and eligibility varies by state, with criteria related to age, disability, income, and citizenship.

Availability/Access –

Access to CMS data requires completion of a rigorous request process, which includes submitting a study protocol, completing a data use agreement, Institutional Review Board (IRB) approval, evidence of funding, completion of a CMS data request form, a CMS disclaimer user agreement, a CMS privacy board review, and a letter of support from the funding agency project officer. While these criteria for access are reasonable given the nature of the data, the process restricts rapid incorporation of the data into a tracking system. Furthermore, because the data is current to between one and two years prior, near real-time analysis is not possible. However, this limitation is true for almost all sources of health data, including the health surveys.

Data use is approved on a project-specific basis. If the data is intended to be used for multiple purposes, each of these purposes must be vetted and approved by CMS. In addition to the administrative approval process, access to CMS data is fee-based. Depending upon the type of data and number of states for which data is obtained, it can be relatively expensive.

Medicaid programs are administered separately by individual states. While this review addresses Medicaid claims data made available by CMS, individual states' data may be accessible through the state programs. Access and availability through this mechanism will vary by state.

Benefits – Medicare claims data is particularly useful due to the completeness with which the target population is enrolled. A large majority of U.S. citizens age 65 years or older are enrolled in Medicare, and utilizing their health care billing data for asthma tracking can provide a relatively complete picture for this age group.

Both Medicare and Medicaid data provide a vast array of covariate information, allowing for detailed analysis, and enhancing the ability to identify high risk groups. Furthermore, although the data represent subsets of the general population, the sample sizes are large, permitting valid analyses for sex and race subgroups. In one study, researchers found that "sample sizes for Asians, blacks, and Hispanics are generally adequate to permit meaningful comparisons with whites." (Escarce, 2003)

In addition to providing the ability to identify primary and secondary diagnoses and related information for the target population of seniors, Medicare data allows for the tracking of individuals over time, providing some measure of the importance of severity or progression of disease. While following individuals in time is not a requisite of an asthma tracking system, it adds value to the dataset.

The RIF files contain detailed geographic (ZIP code) and temporal (date) information, contributing to more precise linkage ability for tracking asthma in conjunction with environmental conditions. While the LDS files do not provide this level of precision, beneficiaries' residential county and quarter/year of service are provided.

Limitations

Although CMS data provide a large and robust picture of the populations represented, they do represent a special subset of the general population. Results of tracking the coexistence of environmental conditions and asthma in this population are not applicable to the general

public. Nevertheless, the elderly and low-income children may be at increased risk of adverse outcomes related to the impact of the environment on asthma. Additionally, certain racial/ethnic groups within these subsets may be underrepresented. Escarce and McGuire found the potential for bias in rates calculated for Asians and Native Americans, (Escarce, 2003).

The population eligible for Medicare is stable and easily defined relative to Medicaid eligible/enrolled individuals. Medicaid eligibility criteria result in a dynamic population and dynamic eligibility for individuals, making it difficult to determine a person's eligibility and enrollment status at a given point in time. Depending on the state's eligibility criteria, a person's eligibility/enrollment may fluctuate in a given calendar year. This has the potential to present a challenge to data management efforts, as well as to defining inclusion/exclusion criteria for tracking or research. Defining inclusion as ever having been eligible or enrolled in a give time period could help to mitigate these concerns.

Finally, access to CMS data requires a rigorous application process and the payment of significant fees. Additionally, the lag time for availability of Medicare data is in excess of a year, and for Medicaid data is longer, two to three years.

Applicability to EPHT

The data available from CMS are well suited for tracking asthma and environmental conditions, with the ability to provide relatively complete coverage of events within the defined populations, demographic and geographic (ZIP code) precision beneficiary information, temporally precise information for services, and the ability to determine the denominator. Use of standard diagnostic and procedural codes contributes to the efficient determination of health outcomes of interest. Applicability of the data to asthma tracking is diminished, however, by the cost and time associated with acquiring the data.

Mortality Data

United States mortality data is collected and compiled by the Vital Statistics Cooperative Program of the Centers for Disease Control and Prevention's National Center for Health Statistics (NCHS) for each of the 50 states and the District of Columbia. Data are acquired from death certificates filed with state vital statistics registration offices. Data contained within the Mortality file represents essentially every death that occurs in the U.S. in a given year. In excess of 99% of U.S. deaths are reported and subsequently included in this dataset, (Kung, et. al.).

Prior to 1989, state and county of residence and date of death were included in the record for each decedent in the file, and between 1989 and 2004, precise geographic information was included only for counties with a population of at least 100,000. As a result of stricter standards regarding confidentiality, beginning in 2005, no geographic information is provided, and only month, year, and day of the week (but not date) are given for temporal identification in the public use data file.

Other aspects of the file have remained functionally unchanged, and include information regarding age group, race, sex, ethnicity, and marital status of the decedent, and underlying cause of death as identified by ICD-10 codes. These variables are potentially useful for performing stratified analysis of cause of death, and identifying demographic subsets of the population at increased risk of death from asthma. Deaths from asthma may provide insight regarding high-risk groups, and serve as sentinel events in terms of the general population. However, the absence of geographic or temporal precision severely limits the data's usefulness in EPHT.

Type/Quality – Mortality files represent a virtual census of deaths occurring to U.S. residents inside the country. As stated previously, it is estimated that more than 99% of all deaths occurring in the country are documented on death certificates, and these are all captured in this data file.

Potential issues related to compatibility of the data from state to state or region to region are introduced by the possibility that coding or diagnostic practices differ by geography. To the extent that physicians utilize incompatible nosological conventions, the comparability of mortality data from one area to another is jeopardized. Additionally, investigations in other countries have suggested that asthma is over-diagnosed as the underlying cause of death on death certificates, (Wright, et. al., 1994; Sidenius et. al., 2001). An older U.S. study concluded that asthma as cause of death on death certificates had low sensitivity and high specificity, with the authors concluding that asthma mortality rates were underestimated, (Hunt, et. al., 1993). Accuracy of current U.S. asthma mortality estimates from death certificates may have been improved by the widespread adoption of standard coding practices, and by state vital statistics registrants' active clarification of nebulous or incomplete information, (Hoyert and Lima, 2005). Nevertheless, differences in coding practices should not be ruled out as a possible explanation for findings of regional differences in asthma mortality rates obtained from death certificate data.

Availability/Access – Annual public-use micro-data death files can be downloaded from the NCHS website or obtained on CD-ROM or DVD. It is possible to request custom files in which geographic and/or temporal precision are included. According to the NCHS Data Release and Access Policy, approval for use of these detailed data is limited to situations involving "critical issues related to public health surveillance," (NCHS Data Release and Access Policy for Micro-data and Compressed Vital Statistics Files, http://www.cdc.gov/nchs/about/major/dvs/NCHS_DataRelease.htm, accessed 10/27/2008). To be considered for approval to use micro-data files containing geographic and temporal detail, a request must be made to the NCHS and a state representative who is authorized to approve use of the state's detailed mortality data. Upon approval, the researcher(s) must submit a data use agreement ensuring that the data will be used only for the purposes stated.

Benefits - As stated, national public use mortality files represent a relatively complete source of asthma mortality, capturing virtually all deaths occurring in the U.S. The data captured by the death certificates provides a rich source of covariate information with which to conduct stratified analyses.

NCHS has developed standard forms and procedures to foster consistent recording of the data across the country, and they work with local jurisdictions to encourage the adoption of these standards.

Limitations

Death from asthma is a relatively rare event, with 0.3 deaths/100,000 children and 1.4 deaths/100,000 adults in 2003, (Akinbami, 2006). While the mortality files are essentially complete, these numbers would not seem to provide the power necessary for detailed demographically- and geographically-stratified analysis.

As discussed previously, data files from 2005 forward do not contain specific geographic or temporal data related to the decedent and date of death. Depending on the type of study being conducted, this may or may not hinder analysis. For the purpose of monitoring concomitant fluctuations in asthma mortality and air quality or other environmental

parameters, however, more detailed geographic and temporal information is necessary for meaningful linkage of the datasets.

Finally, there appears to be a considerable lag between the end of the reporting year and the availability of mortality files. Currently, the most recent public use mortality micro-data file available is from 2005. More timely data may be needed to achieve the objectives of Tracking.

Applicability to EPHT

Mortality data is of limited usefulness for asthma tracking due to a lack of geographic and temporal detail and due to the fact that mortality attributable to asthma is a relatively rare event (0.3/100,000 among children and 1.4/100,000 among adults). To achieve power to conduct detailed, stratified, linked analyses, multiple years of mortality data would need to be pooled. Furthermore, small numbers and the absence of detailed geographic information for decedents preclude geographic comparisons and small-area analyses.

School Health Data

Information from school health records are a potentially valuable means of determining prevalence of asthma and other health conditions in school-aged children. Several states conduct some form of school-based asthma surveillance that can be used for monitoring prevalence over time. Information from this type of surveillance is potentially useful for tracking prevalence and other asthma indicators at the state or county level, and may even be useful for linking with environmental hazard data. However, school-based asthma surveillance efforts lack the standardization necessary for routine monitoring of asthma across state, regional, or national areas.

Type/Quality - The type and quality of data generated by school health surveillance efforts varies from state to state, and is dependent on the methods employed to capture and record the information. Typically, school-based asthma surveillance provides an estimate of asthma prevalence, but it can also provide information regarding asthma complications, supplemented by documentation of limitations on physical activity and absences. The quality of the estimates of asthma prevalence generated and its subsequent comparability with data from other states depends on how the data are gathered, the questions that are asked and how they are asked, and who qualifies as a respondent.

Data may be generated for each school in the state or for individual school districts, public schools, private schools, or both. Asthma prevalence among students of a public school will probably look much different than the prevalence observed in private schools due to social and economic factors. The population targeted by the school-based asthma surveillance system should be considered when interpreting results and generalizing findings of asthma surveillance.

Some states conduct routine annual surveillance using standardized reporting forms and have strict standards defining the information that qualifies as confirming an asthma diagnosis, while others are less rigorous. Some states use periodic surveys or school health record abstraction to determine a child's diagnosis of asthma. The quality of the data generated will vary somewhat depending on the method of data collection.

Data quality and comparability can also be impacted by the source of information. Survey respondents range from a school nurse or administrator, to a parent, to the student. Varying levels of reliability of responses from these participants would be expected, influenced by the amount and quality of information available. In one study that looked at agreement between students and their parents relative to indicating an asthma diagnosis,

the student report was “determined to be a good indicator of probable current asthma,” (Magzamen, 2005).

Another consideration regarding data quality involves the definition of an asthma diagnosis, and what constitutes the diagnosis. Some states may require a written diagnosis from the child’s physician to allow the child to receive medication on school grounds; others may require the parent’s self-report. Differences in the way the diagnosis is identified or confirmed can affect data quality. It is essential to be aware of potential “case-definition” differences when data from multiple school-based asthma surveillance systems are being combined for analysis.

Availability/Access – Record level data from school-based surveillance efforts is not likely to be available outside of the state’s Department of Education or Department of Public Health. In response to the Family Educational Rights and Privacy Act (FERPA), the availability of school health data is limited. If available at all outside of the state health or education departments, data will likely consist of count or rate data, geographically and demographically aggregated.

Benefits – School-based asthma surveillance records provide estimates of asthma prevalence using data from a large segment of the general population of school-aged children. These children, particularly elementary and junior high students, represent the sub-group of the general population at greatest risk for an asthma diagnosis and/or complications.

Additionally, school-based data contains potentially useful covariates that can be used for stratifying analysis, and also for supplementing the diagnostic information, for example, information on absenteeism. In many cases, school-based surveillance provides a routine mechanism for collecting state and local-level prevalence data that may be otherwise unavailable.

Limitations – Many of the limitations of school health surveillance are related to the comparability of the data with similar data collected in other locations. Lack of a standard list of pertinent variables or method for collecting the data introduces difficulty in comparing data collected in different states, or even within a state by different school districts. Prevalence of asthma determined by a physician’s diagnosis probably differs from prevalence determined by self-report. Because of this, the data are not comparable. This has direct implications for EPHT from the national level regarding pooling of state based datasets to increase power for statistical analysis, although the data remain useful for small, defined areas.

Additionally, there is no standard schedule by which school health data are collected. Some states conduct annual surveillance, others use a periodic routine every two or more years, and still others conduct sporadic surveys.

Finally, the diversity of school districts and systems within a state or smaller level geography means that, unless every system is represented, results of surveillance will probably not represent the general population of school children. Furthermore, certain sub-groups of school-aged children, such as home-schooled or those in specialty schools will almost certainly be excluded. The magnitude of the impact of these exclusions is probably minimal in most cases, but should be considered. Differences between public schools and private schools are reflected in the diversity in their students. To the extent that these differences and disparities are also related to having been diagnosed with asthma, the experience of one type of school is not indicative of the experience of the other type. The

inclusion criteria for contributing to the surveillance system should be considered when contemplating the use of school-health data for EPHT.

Applicability to EPHT – Depending on the Tracking application, asthma information obtained from school health records or through school based surveillance is potentially useful for tracking asthma with environmental hazards. Care must be taken when combining this data from disparate sources to ensure that the data are comparable. School health data may be more appropriate for small area projects rather than for pooling data across large geographic regions represented by diverse school systems and school health programs.

Conclusion

There are a variety of sources of asthma data that are relevant and applicable for use in tracking asthma health measures and environmental conditions. When considering a source for inclusion in a tracking system, a number of issues must be considered in determining its appropriateness.

Tracking asthma and environmental measures over a large geographic area will require the compilation of data from diverse data sources. The source population represented and the manner in which the data is collected from the population must be clearly identified so that the tracking results and subsequent public health recommendations can be aimed at the appropriate audience. Ensuring the comparability of the data, or adjusting the data to make it comparable, is an essential step in achieving valid results.

Linkage of asthma and environmental databases requires common geographic and temporal information. The precision of these measures available in the asthma data source must contribute to the successful linkage of the datasets, and to meaningful analysis. The protection of confidentiality related to health outcome data is imperative. Finding the most appropriate dataset for small area linkage analysis while maintaining both individuals' right to privacy and valid and consequential investigation is a challenge to the tracking effort.

As the Tracking Network becomes functional, the practical usefulness, feasibility, and issues involved with using these data sources for asthma tracking will be elucidated through field testing.

References

AHRQ Agency for Healthcare Research Quality. Healthcare Cost and Utilization Project - HCUP A Federal-State-Industry Partnership in Health Data – Introduction to the HUCP Nationwide Inpatient Sample (NIS), 2006.

Akinbami L. Asthma Prevalence, Health Care Use and Mortality: United States, 2003-05. NCHS website, accessed 8/2008, Nov. 2006. (<http://www.cdc.gov/nchs/products/pubs/pubd/hestats/ashtma03-05/asthma03-05.htm>)

Bloom Y, Figgs LW, Baker EA, Dugbatey K, Stanwyck CA, Brownson RC. Data uses, benefits, and barriers for the behavioral risk factor surveillance system: a qualitative study of users. [Journal Article. Research Support, U.S. Gov't, P.H.S.] Journal of Public Health Management & Practice. 6(1):78-86, 2000 Jan.

Escarce JJ, McGuire TG. Methods for Using Medicare Data to Compare Procedure Rates among Asians, Blacks, Hispanics, Native Americans, and Whites Health Serv Res. 2003 October; 38(5): 1303–1318.

Fahimi M, Link M, Schwartz DA, Levy P, Mokdad A. Tracking chronic disease and risk behavior prevalence as survey participation declines: statistics from the Behavioral Risk Factor Surveillance System and other national surveys. Prev Chronic Dis 2008;5(3). http://www.cdc.gov/pcd/issues/2008/jul/07_0097.htm. Accessed [12/08/2008].

Hoyert DL, Lima AR. Querying of death certificates in the United States. Public Health Reports. 120(3):288-93, 2005 May-Jun.

Hunt LW Jr, Silverstein MD, Reed CE, O'Connell EJ, O'Fallon WM, Yunginger JW. Accuracy of the death certificate in a population-based study of asthmatic patients. JAMA 1993, 269(15):1947-52.

Krishnan V, Diette GB, Rand CS, Bilderback AL, Merriman B, Hansel NN, Krishnan, JA. Mortality in patients hospitalized for asthma exacerbations in the United States. Am J Resp and Critical Care Med, 2006; 174: 633-638.

Kung HC, Hoyert DL, Xu JQ, Murphy SL. Deaths: Final data for 2005. National vital statistics reports; vol 56 no 10. Hyattsville, MD: National Center for Health Statistics. 2008.

Magzamen S, Mortimer KM, Davis A, Tager IB. School-based asthma surveillance: a comparison of student and parental report. Pediatr Allergy Immunol, 2005; 16(8):669-78.

Mannino DM, Homa DM, Akinbami LJ, Moorman JE, Gwynn C, Redd SC. Surveillance for Asthma – United States 1980-1999. MMWR, 2002; 51(SS01): 1–13.

Nelson DE, Powell-Griner E, Town M, Kovar MG. A comparison of national estimates from the National Health Interview Survey and the Behavioral Risk Factor Surveillance System. AJPH 2003; 93(8): 1335-1341.

Schoenman JA, Sutton JP, Elixhauser A, Love D. Understanding and enhancing the value of hospital discharge data. Med Care Res Rev 2007 64: 449-468.

Sidenius KE. Munch EP. Madsen FF. Lange P. Viskum K. Soes-Petersen U. [Reliability of recorded deaths from asthma in Denmark during a 1-year period 1994-1995]. [Danish] [English Abstract. Journal Article. Research Support, Non-U.S. Gov't] Ugeskrift for Laeger. 163(44):6131-3, 2001 Oct 29.

Wright SC. Evans AE. Sinnamon DG. MacMahon J. Asthma mortality and death certification in Northern Ireland. [Journal Article] Thorax 1994, 49(2):141-3.

Wenzlow AT, Finkelstein D, Le Cook B, Shepperson K, Yip C, Baugh D. The Medicaid Analytic eXtract Chartbook. 2007. Located on the web and accessed 12/08/08 from: http://www.cms.hhs.gov/MedicaidDataSourcesGenInfo/Downloads/MAX_Chartbook_2007.pdf.