



# **Assessment of Selected Ohio Department of Health Databases for Collection of Race and Ethnicity Data**

**March 2003**

**The Ohio Department Of Health**

[www.odh.state.oh.us](http://www.odh.state.oh.us)

*To protect and improve the health of all Ohioans*

**Bob Taft  
Governor**

**J. Nick Baird, M.D.  
Director of Health**

# **Assessment of Selected Ohio Department of Health Databases for Collection of Race and Ethnicity Data**

Mary Lynn  
Researcher

Health Care Data and Statistics Unit  
William Ramsini, Ph.D., Chief

Center For Vital and Health Statistics  
Jim Pearsol, Assistant Director

**Ohio Department Of Health**

**March 2003**

## Table of Contents

Executive Summary .....	1
Introduction .....	5
Methodology.....	6
Overall Summary.....	7
Summary of Race and Ethnicity Results .....	11
Conclusions and Discussion.....	26
Appendix.....	30

## List of Figures

Figure 1. Purpose for Data Collection.....	7
Figure 2. Primary Population of Data Collection.....	7
Figure 3. Data Collection at the Individual or Aggregate Level .....	8
Figure 4. Method of Data Collection.....	8
Figure 5. Frequency of Data Collection .....	9
Figure 6. Frequency of Data Compilation .....	9
Figure 7. Number of Years for Data Collection .....	10
Figure 8. Race and Ethnicity Collected as Data Items.....	10
Figure 9. Multi-Race Data Collection.....	10

<b>Figure 10. Programs Mandated to Switch to Multi-Race Data Collection.....</b>	<b>13</b>
<b>Figure 11. Programs Voluntarily Switching to Multi-Race Data Collection .....</b>	<b>13</b>
<b>Figure 12. Racial/Ethnic Electronic Data Category Storage and Collection.....</b>	<b>14</b>
<b>Figure 13. Data Collection Mandate .....</b>	<b>15</b>
<b>Figure 14. Restrictions on Data Distribution .....</b>	<b>15</b>
<b>Figure 15. Private/Confidential Data Release in Aggregate Form .....</b>	<b>16</b>
<b>Figure 16. Geographic Units Recorded.....</b>	<b>16</b>
<b>Figure 17. Quality of Racial/Ethnic Data Collected .....</b>	<b>18</b>
<b>Figure 18. Decline in Percentage of Missing Race/Ethnicity Data .....</b>	<b>19</b>
<b>Figure 19. Edit Checks Performed on Dataset .....</b>	<b>20</b>
<b>Figure 20. Reports Produced that Include Information by Race/Ethnicity .....</b>	<b>20</b>
<b>Figure 21. Frequency of Reports that Include Information by Race/Ethnicity .....</b>	<b>21</b>
<b>Figure 22. Format of Reports that Include Information by Race/Ethnicity .....</b>	<b>22</b>
<b>Figure 23. Focus Area of Reports that Include Information by Race/Ethnicity.....</b>	<b>22</b>
<b>Figure 24. Need Identified for Collection of Additional Racial/Ethnic Groups .....</b>	<b>23</b>
<b>Figure 25. Collection of New Racial/Ethnic Groups with Existent Data Tools .....</b>	<b>23</b>
<b>Figure 26. Challenges with the Collection of Race/Ethnicity Data .....</b>	<b>24</b>

**Figure 27. Recommendations for the Collection of Race/Ethnicity Data ..... 24**

**Figure 28. Recommendations to Increase Data Analysis and Dissemination ..... 25**

### **List of Tables**

**Table 1. Racial Data Collection by Category ..... 11**

**Table 2. Ethnic and Other Group Data Collection by Category ..... 12**

**Table 3. Percent of Data Observations that Represent Racial/Ethnic Minorities..... 14**

**Table 4. Data Categories by Race/Ethnicity ..... 17**

**Table 5. Percent of Missing Race/Ethnicity Data ..... 19**

**Table 6. Reports Produced without Race/Ethnicity Data ..... 21**

# ***Executive Summary***

## **Key Findings**

The majority of databases collected at the Ohio Department of Health (ODH) focus on service programs and disease surveillance. A typical database includes information collected at the individual level, with data obtained from facilities such as laboratories and clinics. Data collection was most frequently conducted on a daily basis while the most frequent type of formal reporting was annual. Nearly half of the databases have been collected for five years or less.

Most databases are collected to fulfill either federal or state requirements (i.e. birth and death records, abortion reporting). The remaining data collections are typically made as program decisions to determine program effectiveness and verify information submitted by grantees.

More than two-thirds of the data surveyed at the Ohio Department of Health had data restrictions. Some datasets could not be released due to federal agreements (i.e. Minimum Data Set and Home Health - OASIS). Other data sets required a data user agreement, while still others would only permit access to part of the dataset. Larger population-based surveys were typically available to the public.

The majority of databases surveyed include the collection of race and ethnicity data elements. Data was collected for the race/ethnicity categories of white, black, American Indian, Asian, Hawaiian/Pacific Islander and Hispanic by nearly all of the databases and for the category unknown race by two-thirds of the databases. Some additional variances were noted including one-third of databases that had a category called "other race" which was a fill-in-the blank type category. Also represented were checkboxes for "other race" and checkboxes for multi-race. Of the 40 databases surveyed, several collected Hispanic as a race and one collected Mediterranean and Middle Eastern as a race. A category for Somalian as an ethnicity was included in some databases, while two had a separate category for Amish.

Most respondents indicated the correct racial and ethnic groups were currently being collected. Of those who indicated additional racial/ethnic groups should be collected the groups were Hispanic, Somalian and Amish.

Overall, race/ethnicity data quality was considered good, very good or excellent by two-thirds of those surveyed. The majority of databases had low percentages of missing data (0-10 percent). More than 80 percent indicated that some forms of edit checks were performed on the data.

Some challenges noted in collecting race/ethnicity data included a lack of understanding and training on the part of the person completing the data form, locating the appropriate population and language barriers. In some cases, ensuring that race/ethnicity were self-reported was a priority.

Multi-race data was collected in nearly half of the databases. Most multi-race data collection was mandated by federal grantors. Of those that did not collect multi-race data, 25 percent indicated they would be required to do so by federal mandate, while 40 percent of those not mandated or already collecting multi-race data would be voluntarily switching to that format.

The collection of multi-race data, while important, is limited in Ohio. Nearly all of the databases that collected multi-race data needed to re-assign the data to other race categories for the purpose of data analysis. Typically, the number of people in Ohio who identify their race as “multiple race” was such a small number that meaningful analysis could not be conducted.

Although slightly more than half of the databases surveyed were utilized to produce reports that included information on health disparities, the focus of the reports varied from health disparities information as a minor part of a much larger report, to full reports devoted to health disparities. Typical reasons for not producing reports on health disparities included data restrictions, race/ethnicity not being a factor for data items reported, not enough data and lack of resources.

### **Recommendations**

The use of standardized definitions for race and ethnicity enhances the value of the data and also increases the opportunity for merging and linking of various databases. It is encouraging that the majority of databases include racial and ethnic data and plan to switch to new federal standards for data collections. To increase uniformity, ODH programs should consider using the same questions with regard to race and ethnicity as recommended by the ODH Data Standards Committee. The use of “other” category is of concern and should be addressed.

Because socioeconomic factors play a major role in health disparities, uniform collection of these variables should also be considered to better understand underlying causes for racial and ethnic disparities. Presently only 60 percent of databases collect socioeconomic data and programs that do not collect this information should be encouraged to do so. The Public Health Data and Research Committee should monitor this process and offer expertise and assistance to programs.

Respondents to this survey rated the quality of racial and ethnic data as fair or poor for 35 percent of the databases. Primary reasons for this were

missing data and lack of verification. As mentioned before, 70 percent of the databases had missing data in the 0-10 percent range, and for 12.5 percent of the databases the missing data range was greater than 30 percent. More efforts are needed to rectify this problem. Selected audits should also be performed to determine the quality of key racial and ethnic databases such as vital records.

In recent years, the response rate for data collection efforts has gone down, especially for minority populations. Extra efforts and resources are needed to increase the response rate and to determine the effect of this low response rate on key findings.

ODH should also encourage the collection of variables that would allow for geocoding of the data to determine variations across the state. Only 62.5 percent of databases capture the address and 12.5 percent include Census tract information, which is needed for geocoding.

No reports on health disparities are produced from about half of the databases because of data restrictions and lack of sufficient resources or data. Some databases at the department may not have sufficient observations or have a small sample size for meaningful analysis for some minority groups. One way of dealing with this issue is to combine data for several years. Another approach involves increasing the sample size for minority populations by over-sampling them or conducting data collection efforts targeted at small racial/ethnic groups. Recent efforts by the center to increase the sample size for minority populations in the Behavioral Risk Factors Surveillance System (BRFSS) and the Ohio Family Health Survey (OFHS) are positive steps and should continue. Each division should also consider devoting more resources for data analysis.

ODH programs should make maximum use of their existing data by including health disparities in their reports. Because our research resources are limited, ODH should make its data readily available to those researchers who have a particular interest in minority health. The recent ODH publication *A Guide to Selected ODH Databases*, includes information on availability of racial and ethnic data and will be useful to identify promising data sources for research.

The establishment of the Minority Health Data Unit at the Center for Vital and Health Statistics is a positive development and should continue to be given the needed resources to perform its functions. In the last two years, the unit contributed funds to increase the sample size for minority populations in the BRFSS and OFHS, produced several reports on minority health and provided funds to external researchers to study minority health issues.

In conclusion, the databases collected at the Ohio Department of Health are evolving to support the effort of reducing and eliminating health disparities in

Ohio. Many programs have improved data quality, added the collection of multiple race data and are reporting findings in a useful manner. While program-specific data provides very useful information about health disparities for target populations in Ohio, continuing the effort to include and expand population-based surveys such as BRFSS or OFHS is necessary to provide population-based estimates on topics such as health insurance, health status and access to health care for Ohioans. In addition, continued effort to produce more longitudinal data will assist us in determining the effectiveness of our efforts. Such future endeavors will be imperative to understanding and reducing health disparities in Ohio.

## ***I. Introduction***

Racial and ethnic health disparities in the United States have become the focus of considerable public attention during the past decade. The limited data and research has made the task of understanding the problems, and developing appropriate solutions and interventions particularly difficult. In recent years, a number of both public and private sector efforts have been made to better understand the factors that contribute to health disparities between racial and ethnic groups. While economic and environmental factors undeniably contribute to racial and ethnic differences in morbidity and mortality, other factors such as insurance, access to care, health care utilization, bias in health care practices and cultural practices all may contribute to the gap in health status.

Addressing health disparities is a priority for the department and efforts are underway to increase the Ohio Department of Health's activities in this area. To measure disparities and document the progress in reducing disparities requires complete and accurate data. This report discusses the availability and limitations of racial and ethnic data at the Ohio Department of Health (ODH) and offers some recommendations for improvement.

As part of the effort to reduce health disparities, ODH recently created a research unit focused on minority health. One of the tasks of the minority health research unit was to closely examine selected data collected by ODH, particularly in program areas and disease surveillance. The goal of the project was to gain a better understanding of the collection of racial and ethnic information throughout the department that may be utilized to help us better understand health disparities. Basic information collected as part of this effort included the purpose and population of the database, the frequency of data collection and data-content categories. The collection of detailed information included racial and ethnic data categories, quality of racial and ethnic data, reports on health disparities produced from the dataset and information regarding data gaps. Comments were solicited on issues related to the collection of race and ethnicity data and recommendations for improving the collection of racial and ethnic health data.

*The Assessment of Selected Ohio Department of Health Databases for Collection of Race and Ethnicity Data* provides an analysis of current issues and practices related to the collection of race and ethnicity data collected at the Ohio Department of Health.

## **II. Methodology**

Interviews for the *Assessment of Selected Ohio Department of Health Databases for Collection of Race and Ethnicity Data* were conducted during the summer and fall of 2002. The interview was approved by the Ohio Department of Health Strategic Management group, Leadership group and Ohio Public Health Data/Research Policy Advisory Committee, and pilot tested with selected ODH programs.

Each ODH division chief was contacted, provided an explanation of the study, a copy of the survey tool and asked to provide key contact personnel for databases. Administrative, financial and paper and pencil databases were excluded from the study. In an effort to reduce the burden on other ODH staff and develop a consistent understanding of the survey questions, a decision was made to conduct the survey as a personal interview. Every effort was made to comprehensively examine each dataset in selected areas. In total, 46 datasets from the Ohio Department of Health were included in this survey. Some of the data sources examined are designed for public use while others are primarily utilized by ODH programs providing health services to Ohioans.

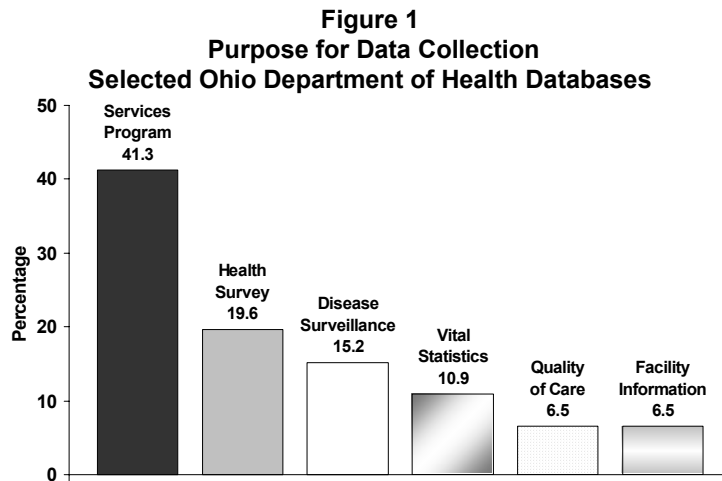
The survey design included two sections; the first brief section provided very general information about the database, while the second section provided more detailed information regarding the collection of race and ethnicity data. If databases did not include information on race and ethnicity, only section one of the survey was completed. Following the interview, responses were categorized by the researcher for the project and forwarded to the participant for corrections, comments and review prior to finalizing each survey form. The data gathered was self-reported and unverified. A copy of the survey instrument is included in the Appendix of this report,

It is important to note that information specific to each database is diminished when categorizing some data areas (i.e. purpose and population of the data collection). A more comprehensive understanding of issues related to specific databases can be found in our document titled; *A Guide to Selected Ohio Department of Health Databases*. The guide is intended for public use and is posted on the ODH Web site at [www.odh.state.oh.us](http://www.odh.state.oh.us).

### III. Overall Summary

#### Purpose of Data Collection

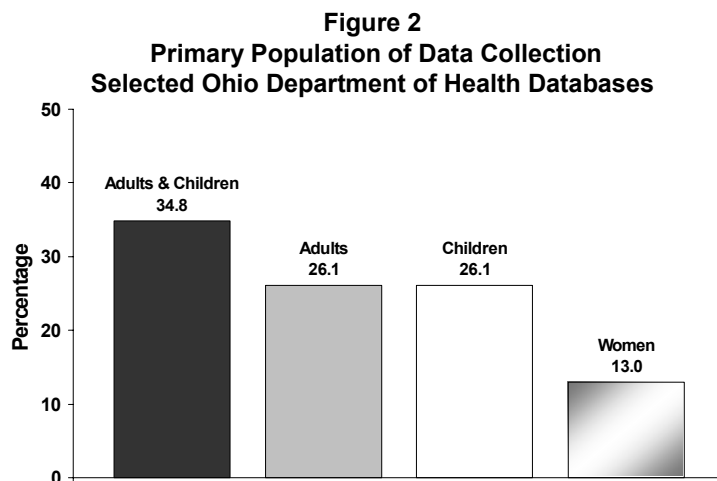
Participants in the study were asked to describe the purpose for the data collection. Detailed responses were recorded and collapsed into the six categories described in figure 1. The majority of data was collected in the



Family and Community Health Services area as part of the assessment of public health programs. The services program category was followed by larger population-based health surveys, disease surveillance, vital statistics and quality of care. Facility information included data collected at the aggregate level (i.e. hospital level).

#### Covered Populations

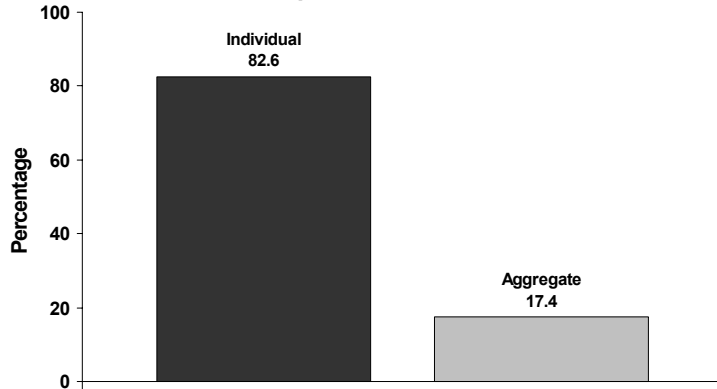
The majority of databases surveyed included both adults and children (34.8 percent). Just over one out of four databases included primarily adults only or primarily children only. Thirteen percent of databases had a primary population of women (see figure 2).



**Aggregation Level**

More than 80 percent of the databases surveyed collected data at the individual level. The majority of the individual level databases are accounted for by services programs, disease surveillance, vital records and large population surveys (see figure 3).

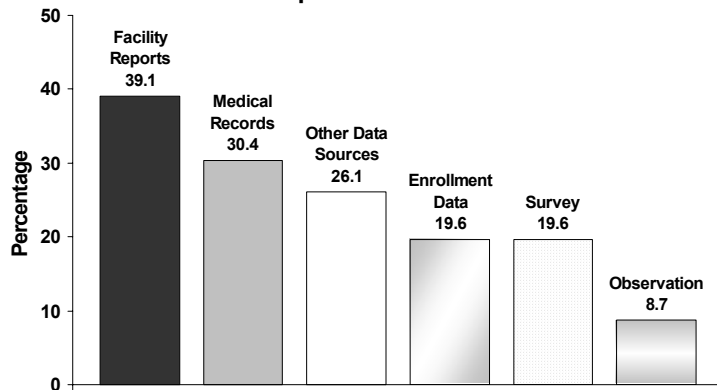
**Figure 3**  
**Data Collection at the Individual or Aggregate Level**  
**Selected Ohio Department of Health Databases**



**Data Source**

A number of programs collect data from more than one data source. For example, some information in a dataset may be collected through enrollment forms, while other data is obtained from laboratory reports. Programs were asked to identify all sources of data collection. The main source of data was facility reports which could come from a number of sources, including hospitals and home health agencies. Medical records were the second-most

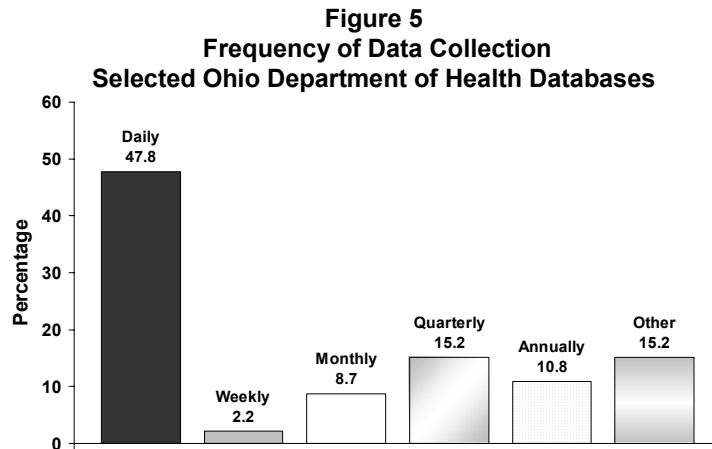
**Figure 4**  
**Method of Data Collection**  
**Selected Ohio Department of Health Databases**



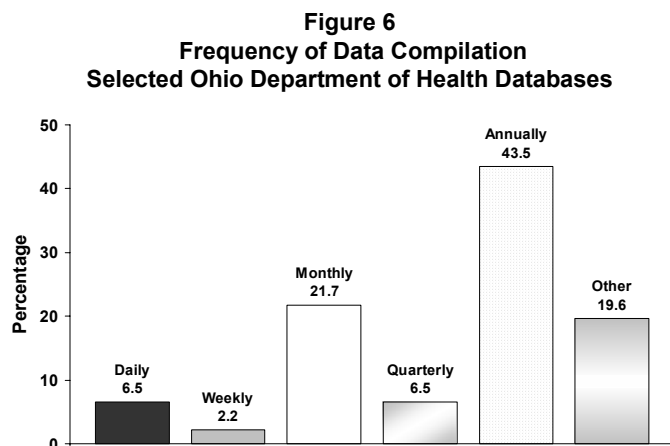
likely source of data, followed by other data sources (primarily customized ODH forms). About one out of five programs obtained data through enrollment or survey information. Observation was the least likely way to gather data (see figure 4).

## **Frequency of Data Collection and Compilation**

Nearly half of the participants surveyed indicated data was collected on a daily basis. This is particularly true for data collected in the prevention area. Larger survey data was frequently completed annually or in some cases every other year or every five years (see figure 5).



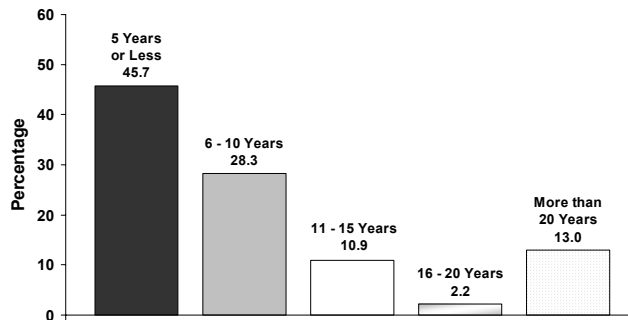
Data was most likely to be compiled annually (43.5 percent) followed by other time frames (19.6 percent) which were typically accounted for by those surveys that collect data intermittently (i.e. every two or every five years). Monthly reporting was third-most likely (21.7 percent) followed by quarterly (6.5 percent), daily (6.5 percent) and weekly (2.2 percent). A number of participants also indicated data could be compiled more frequently upon request (see figure 6).



## **Duration of Data Collection**

Nearly half of the databases surveyed have been collecting data for less than five years. Just over a quarter of the databases have collected data for six to ten years. Thirteen percent of databases have collected data for more than 20 years. Five of the six databases that have collected data for more than 20 years are vital statistics data (see figure 7).\*

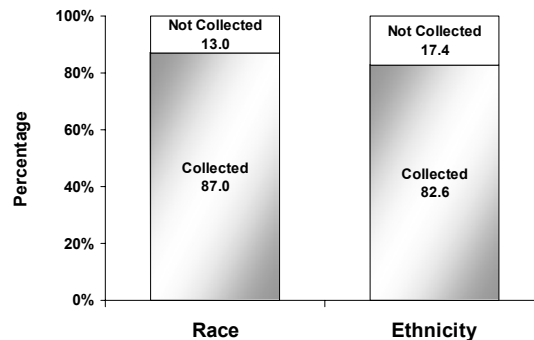
**Figure 7**  
**Number of Years for Data Collection**  
**Selected Ohio Department of Health Databases**



## **Prevalence of Collection of Racial and Ethnic Data**

The majority of databases collect race and ethnicity as data elements. The most frequently cited reason for not collecting race and ethnicity indicated that those data items were not perceived as pertinent to the information being collected (see figure 8).

**Figure 8**  
**Race and Ethnicity Collected as Data Items**  
**Selected Ohio Department of Health Databases**



---

\*Data that has been collected intermittently was coded for the number of years collected (i.e. oral health survey has been collected three separate times with five years between surveys – this would be coded as three years of data). Only electronically available data was included for this analysis. Some programs have paper reports or data available prior to implementing electronic databases.

#### **IV. Summary of Race and Ethnicity Results**

*The summary of race and ethnicity results includes only those databases that collect information by race and ethnicity. Of the 46 databases surveyed, 40 collected race and ethnicity information.*

##### **Data Collection Format**

All of the programs surveyed collected data for the race category black or African American. Data was collected in all databases surveyed for the race categories white, American Indian and Asian with the exception of the Pregnancy Risk Assessment Monitoring System (PRAMS), which extracts data from the birth certificate and categorizes in two categories, black and non-black. Nearly two-thirds of the databases surveyed have a category of unknown for race. Almost one-third provide a fill-in-the-blank category for a write-in race. One out of five programs had a checkbox for “other” race, while several databases included a checkbox for multi-race (see table 1).\*

**Table 1  
Racial Data Collection, by Category  
Selected Ohio Department of Health Databases**

<b>Category</b>	<b>Percentage</b>
<b>Black</b>	<b>100.0</b>
<b>White</b>	<b>97.5</b>
<b>American Indian</b>	<b>97.5</b>
<b>Asian</b>	<b>97.5</b>
<b>Hawaiian/Pacific Islander</b>	<b>92.5</b>
<b>Unknown</b>	<b>65.0</b>
<b>Other Fill-in-the Blank</b>	<b>30.0</b>
<b>Other Checkbox</b>	<b>20.0</b>
<b>Multi-Race Checkbox</b>	<b>7.5</b>

Hispanic ethnicity was collected in 35 of the 40 databases surveyed. Of the five not reporting Hispanic ethnicity as a category, three collected Hispanic under the category of race. The category other ethnicity included five databases that collected further breakout of Hispanic ethnicity and one database that had a checkbox for other ethnicity.

---

\*Several databases collected race as a fill-in-the-blank type response only. For the purposes of categorizing, fill-in-the-blank responses were coded positive for the race categories listed in the table.

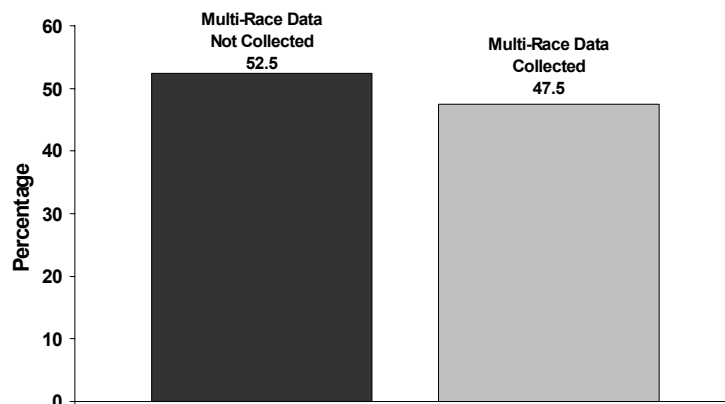
**Table 2**  
**Ethnic and Other Group Data Collection, by Category**  
**Selected Ohio Department of Health Databases**

Category	Percentage
Hispanic Ethnicity	87.5
Other Ethnicity	15.0
Other Groups	22.5

A category called “other groups” was created to capture a number of databases that included some specialized groups. In this category six databases had a category for Somalian collected as an ethnicity, two databases had a category for Amish while one database had a category selection of Mediterranean and Middle Eastern as a race (see table 2).

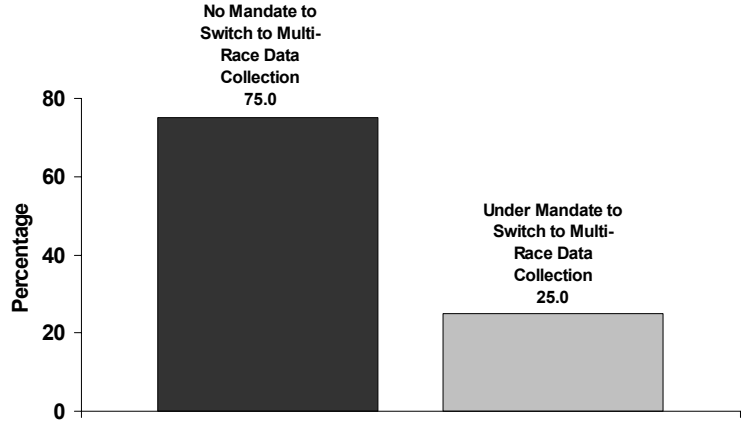
Multi-race data was collected by 19 of the 40 databases surveyed. The collection of multi-race data was handled in different ways. Databases that included checkboxes for race often included an instruction of “check all races that apply”, others allowed a fill-in-the-blank response for all applicable races while some included a separate check box for multiple race (see figure 9).

**Figure 9**  
**Multi-Race Data Collection**  
**Selected Ohio Department of Health Databases**

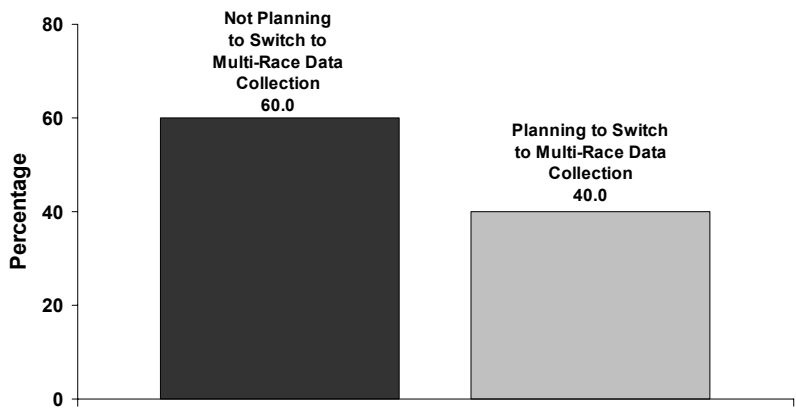


Of the databases currently not collecting multi-race data, only 25 percent indicated they would be required by federal mandate to collect multi-race data (see figure 10). Of those who would be required to switch to multi-race data collection, four out of five would be making the switch within the year (see figure 10).

**Figure 10**  
**Programs Indicating they will be Mandated to Switch to Multi-Race Data Collection**  
**Selected Ohio Department of Health Databases**



**Figure 11**  
**Programs Indicating they will Voluntarily Switch to Multi-Race Data Collection**  
**Selected Ohio Department of Health Databases**

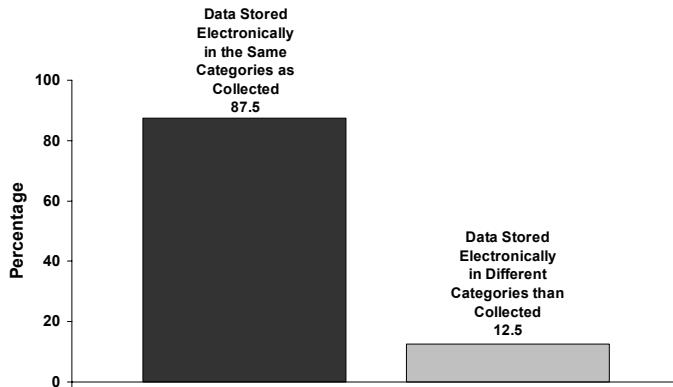


Those surveyed who indicated they would not be mandated to switch to multi-race data collection were further asked whether they would be voluntarily switching to multi-race data collection. Six programs (40 percent) indicated

they would voluntarily be switching to multi-race data collection. Two of the nine programs that indicated they would not be collecting multi-race data were unable to collect multi-race data due to the method of their data collection. These programs relied on alternative data sources to provide race and ethnicity information (see figure 11).

**Data Storage Format**

**Figure 12  
Racial and Ethnic Electronic Data Category  
Storage and Data Collection  
Selected Ohio Department of Health Databases**



Thirty-five out of forty databases stored race and ethnicity data electronically in the same format as collected. Of the five databases that stored electronic data in different categories than collected, three databases collected fill-in-the-blank data for race on the original data form, but utilized a complex series of rules to collapse the data into different categories for electronic

data storage. A fourth program collected multiple-race data on original form but stored the data as “other” or unknown in the electronic database, while the fifth database collapsed race data that was not black or African American into a category called “other” (see figure 12).

**Percentage of Racial and Ethnic Data**

**Table 3  
Percent of Data Observations that Represent  
Racial/Ethnic Minorities  
Ohio Department of Health Selected Databases**

Percent of Minority Data Observations, by Category	Number of Databases	Percent of Total ODH Selected Databases
0-10%	4	10.8
11-20%	13	35.1
21-30%	7	18.9
31-40%	5	13.5
41-50%	5	13.5
Greater than 50%	3	8.1

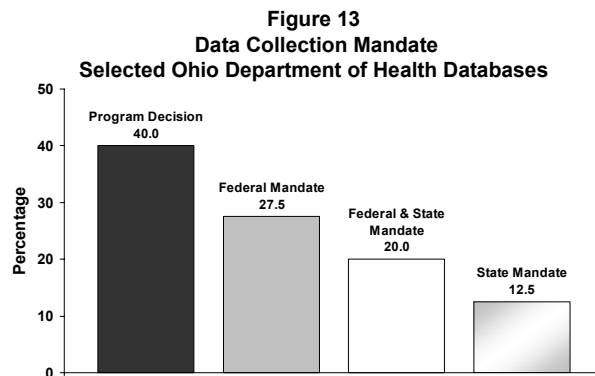
More than half of the databases surveyed (54 percent) reported the percentage of minority observations in their dataset was 21 percent or higher. Just over one out of five databases had minority observations of 41 percent or more of their total data observations (see table 3)\*

---

\*Three programs surveyed were unable to provide the percentage of minority observations in their database.

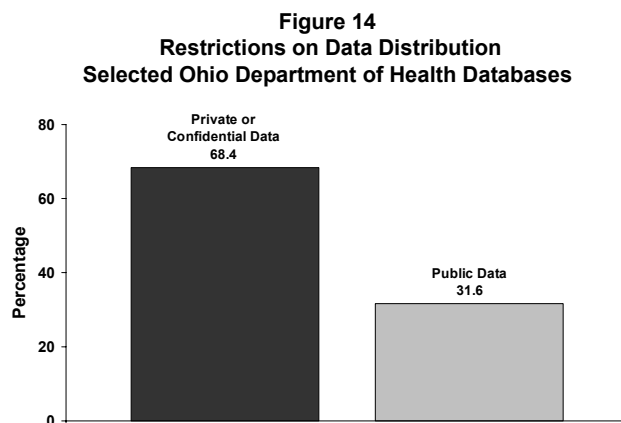
## **Data Collection Mandate**

The majority of data (60 percent) was collected as mandated federally, by the State of Ohio or by both. As an example, the majority of disease surveillance data is mandated. Other data was collected as part of programs decisions to assist the department in achieving our Healthy Ohioans' goal. Examples of this type of data would include the Ohio Family Health Survey and Oral Health Survey. It is important to note that while a number of data collections may not be mandated, data is necessary to apply for federal grants (see figure 13).



## **Data Sharing Restrictions**

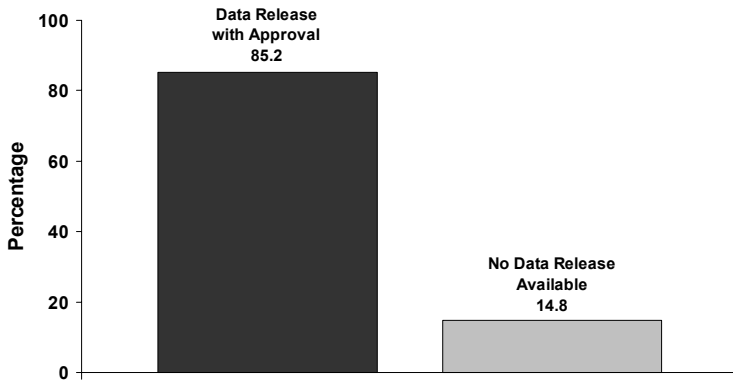
More than two-thirds of the databases surveyed had restricted distribution. In some cases, the data set was stripped of identifying variables such as name and address. In other cases, a data user agreement was required to release the data. Still other datasets could not be released, only summary information was available. A number of programs indicated some uncertainty about data release. Typically, they had not been requested to release the electronic data set, and had released only summary information. If the dataset were to be requested, it is possible some form of approval would be necessary (see figure 14).\*



---

Birth and fetal death data are not included in this graph; both datasets are partially public, and partially private or confidential.

**Figure 15**  
**Private/Confidential Data Release in Aggregate Form**  
**Selected Ohio Department of Health Databases**

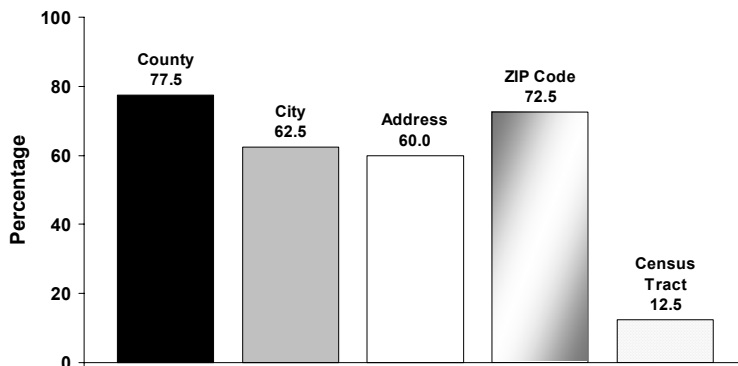


Of those datasets considered private or confidential, all but four (or 15 percent), indicated they could release the data, but with some form of restriction on variables released or proper authorization required for data release (see figure 15).

**Geographic Units**

More than 77 percent of databases collected race and ethnicity by county and 73 percent by ZIP code. City (63 percent) and address (60 percent) were also commonly collected data fields by race and ethnicity. Census tract data by race and ethnicity was collected in just five (or 13 percent) of databases surveyed. Although a number of datasets collected race and ethnicity data by smaller geographic levels (i.e. city or ZIP code) it is possible that the smaller geographic unit may be too small to conduct meaningful data analysis (see figure 16).

**Figure 16**  
**Geographic Units Recorded**  
**Selected Ohio Department of Health Databases**



**Data Collection Areas**

**Table 4  
Data Categories by Race/Ethnicity  
Ohio Department of Health Selected Databases**

<b>Data Category</b>	<b>Number of Databases that Collect Data</b>	<b>Percentage of Databases that Collect Data</b>
Demographic	40	100.0
Health Care Utilization	25	62.5
Socioeconomic	24	60.0
Other Data*	23	57.5
Behavior Risk Factors	22	55.0
Morbidity	21	52.5
Health Insurance	20	50.0
Mortality	19	47.5
Health Status	13	32.5
Prevention	12	30.0
Access to Care	12	30.0
Payment	11	27.5
Health Care Cost	11	27.5
Immunization	10	25.0
Health Behavior/Promotion	9	22.5
Cancer	8	20.0
Quality of Life	7	17.5
Genetics	4	10.0
Quality of Care	4	10.0

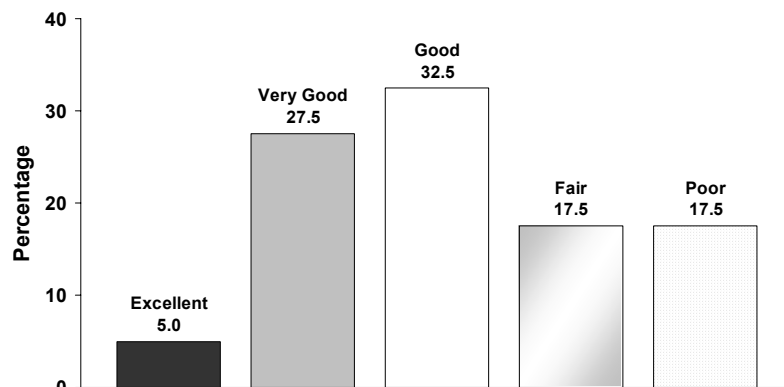
\*The "other data" category includes specific areas (i.e. domestic violence, type of treatment received) that do not fit into the broader categories collected.

Rather than attempt to collect individual data items, we choose to categorize data into a number of topic areas. All of the datasets included some demographic information such as gender or age. More than half of the databases surveyed included information on health care utilization, socioeconomic data and other types of data (items that did not closely fit in one of the main categories), behavior risk factors, morbidity and health insurance. Information on mortality, health status, prevention, access to care, payment information, health care cost and immunization was collected in 25 to 50 percent of the databases surveyed. Categories collected in less than 25 percent of the databases surveyed included behavior risk factors, cancer data, quality of life, genetics and quality of care (see table 4).

**Data Quality**

Those surveyed were asked to rate the overall quality of the race and ethnicity data for their dataset, taking into account the amount of missing data and how the information is obtained (i.e. self reported versus observation). Nearly one-third (32.5 percent) of the databases surveyed had data quality rated as either very good or excellent.

**Figure 17**  
**Quality of Racial/Ethnic Data Collected**  
**Selected Ohio Department of Health Databases**



A good rating was the most common for racial and ethnic data with nearly one-third (32.5 percent) of databases surveyed in this category. The remaining third (35 percent) of databases were rated as either fair or poor for the quality of racial and ethnic data. Typical reasons for the lower ratings included a large percentage of missing data in the race and ethnicity category, observation method utilized to obtain race and ethnicity information and inability to verify data (see figure 17).

Of the databases surveyed, 70 percent had between 0-10 percent missing race and ethnicity data. The remaining databases had various percentages of missing race and ethnicity information. Typically, databases that had a higher percentage

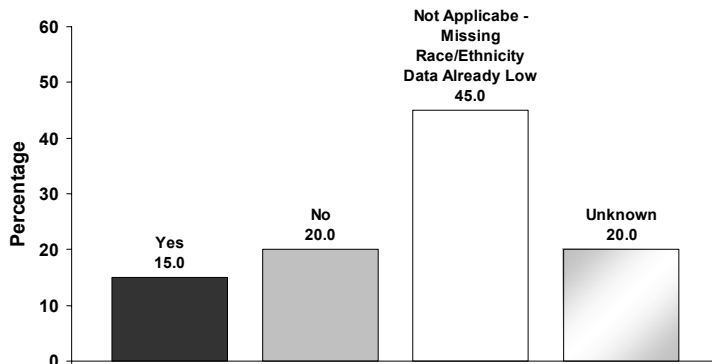
**Table 5**  
**Percent of Missing Race/Ethnicity Data**  
**Selected Ohio Department of Health Databases**

Percent of Missing Minority Data, by Category	Number of Databases	Percent of Total ODH Selected Databases
0-10%	28	70.0
11-20%	1	2.5
31-40%	2	5.0
41-50%	2	5.0
Greater than 50%	1	2.5
Unknown	6	15.0

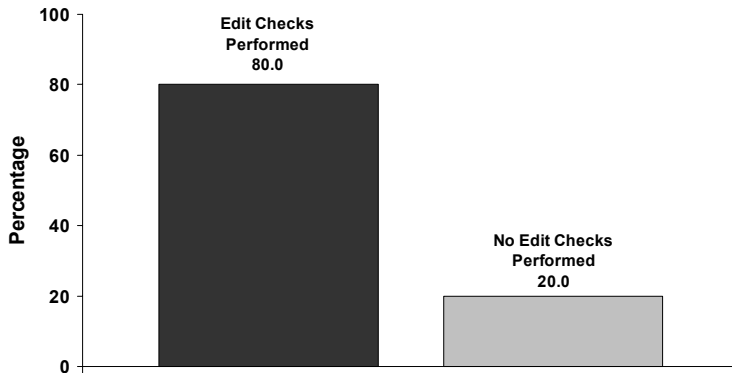
of missing race and ethnicity data did not consider race and ethnicity a factor in the type of data that was collected. The majority of databases in the unknown category were collected in a manor that it was not possible to verify data accuracy (see table 5).

When asked whether the percentage of missing race and ethnicity data had declined over the years, 45 percent of those surveyed indicated that the percentage of missing race and ethnicity data had remained consistently low (missing less than two percent). Typically this was the case when race and ethnicity were considered required fields. Twenty percent of respondents indicated that there had been no change in the percentage of missing data, while 15 percent reported a decline in the percentage of missing race and ethnicity data (see figure 18).

**Figure 18**  
**Has the Percentage of Missing Race/Ethnicity Data Declined?**  
**Selected Ohio Department of Health Databases**



**Figure 19**  
**Edit Checks Performed on Dataset**  
**Selected Ohio Department of Health Databases**

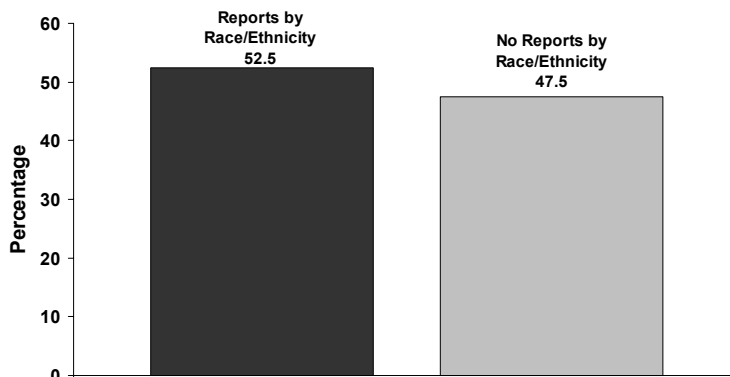


Four out of five of the databases surveyed indicated that some edit checks were performed on the dataset in an effort to increase the accuracy of the data. In the cases where edit checks were not performed, the manner of data reporting made it very difficult to perform edit checks (see figure 19).

**Data Analysis**

Respondents to the survey were asked whether they produced analysis/reports that included health care information by race and ethnicity. More than half of the databases (52.5 percent) produced reports that included health care information broken out by race and ethnicity. Several programs had reports dedicated to looking at health care issues by race and ethnicity, while other programs reported limited items by race and ethnicity as part of a broader report (see figure 20).

**Figure 20**  
**Percent of Databases from which Reports are Produced that**  
**Include Information by Race and Ethnicity**  
**Selected Ohio Department of Health Databases**



Survey respondents who indicated they did not produce any health information by race and ethnicity were asked why reports with that type of information were not produced. The majority of reasons were too varied to be categorized but included reasons such as federal government restrictions, race and ethnicity

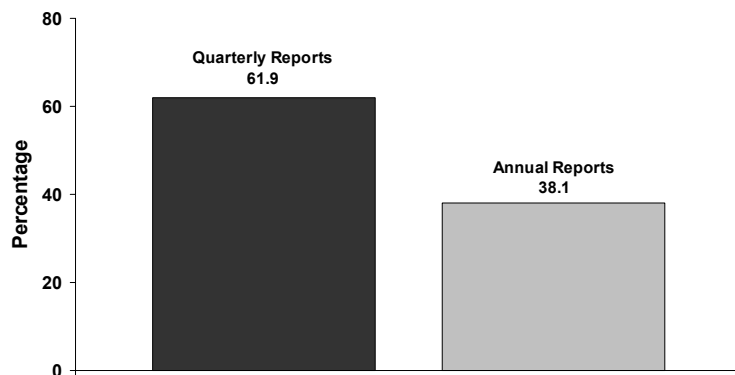
**Table 6**  
**Reports Produced Without Race/Ethnicity Data**  
**Selected Ohio Department of Health Databases**

Reasons for Not Reporting Data by Race and Ethnicity	Number of Databases (Total=19)	Percent of Databases
Miscellaneous Reasons	12	63.2
Not Enough Time	3	15.8
Not Enough Staff	3	15.8
Not Enough Data	3	15.8
Not a Priority	1	5.3
Poor Quality Data	1	5.3

believed not to be a factor with the type of data collected and information being in the development stage or transition to name a few. Other reasons included not having enough time or staff to complete the analysis, and not having enough data to provide meaningful data analysis (see table 6).

The majority of reports that included information by race and ethnicity were produced on a quarterly basis (62 percent), the remaining reports (38 percent) were produced on an annual basis (see figure 21).

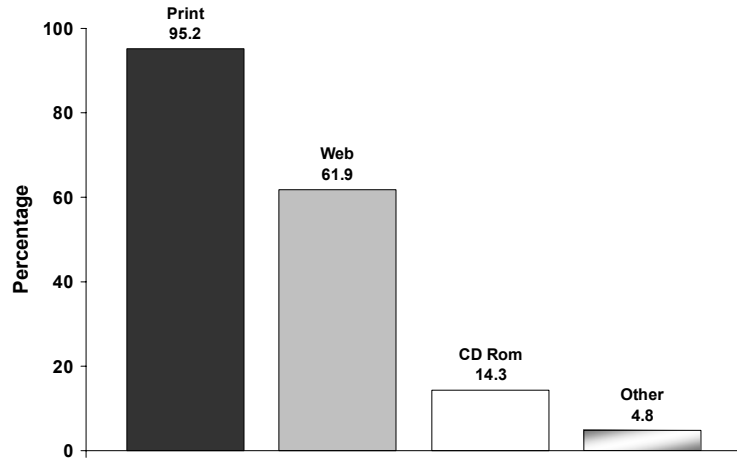
**Figure 21**  
**Frequency of Reports Produced that Include Information**  
**by Race and Ethnicity**  
**Selected Ohio Department of Health Databases**



## **Data Reports Format**

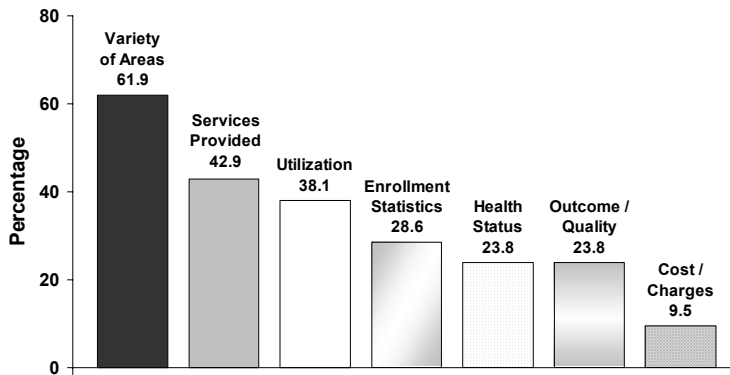
Respondents were asked to indicate all of the formats for the reports they produced. Reports that included information by race and ethnicity were most likely to be printed reports (95 percent). The second most likely format of reports was web based (62 percent) followed by CD-ROM (14 percent) and other types of reporting methods (see figure 22).

**Figure 22**  
**Format of Reports that Include Information by Race and Ethnicity**  
**Selected Ohio Department of Health Databases**



Reports produced from the datasets surveyed most often included a variety of health information (62 percent). When categorizing data into specific types of reports, health services provided was most frequently a report focus (43 percent), followed by health care utilization (38 percent), enrollment statistics (29 percent), health status and outcome/quality (24 percent) and cost/charges (10 percent). It is important to note that more than one focus area could be selected for each database surveyed (see figure 23).

**Figure 23**  
**Focus Areas of Reports that Include Information by Race and Ethnicity**  
**Selected Ohio Department of Health Databases**

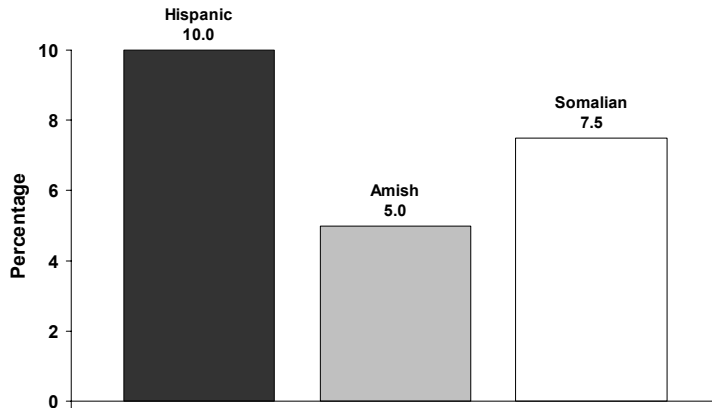


---

\*The "variety of areas" category includes reports that cover topics in several areas.

## Need for Additional Data

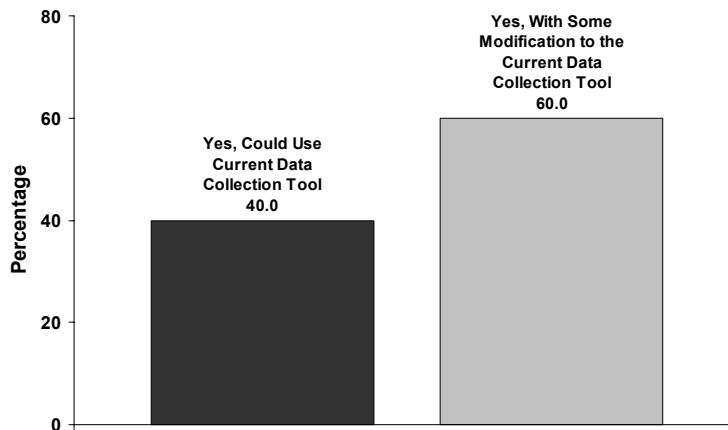
**Figure 24**  
**Need Identified for Collection of Additional**  
**Racial and Ethnic Groups**  
**Selected Ohio Department of Health Databases**



The majority of those surveyed (82.5 percent) indicated that there was not a need to collect data on additional racial and ethnic groups. Of those who indicated additional racial and ethnic groups should be added to the dataset, Hispanic was most likely to be identified (10 percent), followed by Somalian (7.5 percent) and Amish (5 percent) as a specific group (see figure 24).

The respondents surveyed indicated it would not be difficult to accommodate the collection of additional race or ethnic groups. The majority (60 percent) indicated that the collection of new race and ethnic groups would require some modification to the current data set, while 40 percent indicated that they could utilize their existent data tool (see figure 25).

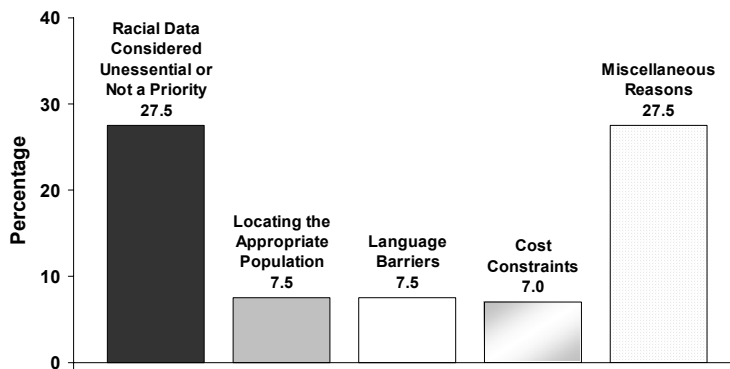
**Figure 25**  
**Collection of New Racial and Ethnic Groups**  
**with Existent Data Collection Tools**  
**Selected Ohio Department of Health Databases**



## Data Collection Challenges

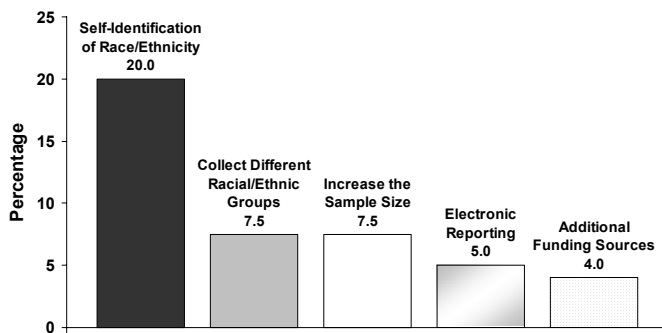
More than a quarter of those surveyed who indicated they faced challenges with the collection of race and ethnicity data stated one of the problems was a lack of understanding of the importance of the information on the part of the person completing the information. As an example, in some facilities admissions clerks complete the race and ethnicity information, while in other facilities physicians complete the information. Often times other areas of the form will be considered priority items. It is difficult to account for the quality of

**Figure 26**  
**Challenges Identified with the Collection of**  
**Race and Ethnicity Data**  
**Selected Ohio Department of Health Databases**



the information when different people with varying levels of expertise complete the data forms. In some cases it was deemed difficult to locate young black and Hispanic males. Language barriers were also a constraint in sampling the appropriate population. Cost was also noted as a factor. For example, additional funds would be necessary to reach populations difficult to locate such as migrant farm workers (figure 26).

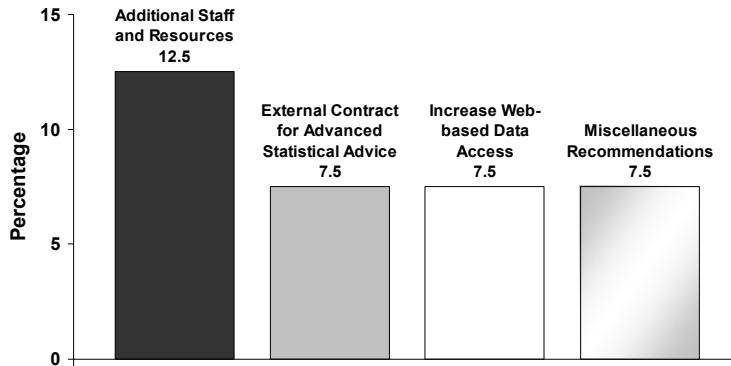
**Figure 27**  
**Recommendations for the Collection of**  
**Race and Ethnicity Data**  
**Selected Ohio Department of Health Databases**



Of those making recommendations for improving the collection of race and ethnicity data, the primary recommendation was to ensure race and ethnicity items were self-reported. Other recommendations included the collection of different racial and ethnic groups, increasing the same size, electronic reporting as a quality

control to reduce the amount of missing information and ensuring additional funding for the collection of race and ethnicity (see figure 27).

**Figure 28**  
**Recommendations to Increase Data Analysis and Dissemination**  
**of Race and Ethnicity Data**  
**Selected Ohio Department of Health Databases**



To increase data analysis and dissemination, additional staff and resources were the most frequently reported recommendations. Additional recommendations included an Ohio Department of Health contract for advanced statistical advice that could be utilized by different programs on an as-needed-basis and additional Web-based development for data access (see figure 28).

## ***V. Conclusions and Discussion***

### **Key Findings**

The majority of databases collected at the Ohio Department of Health focus on service programs and disease surveillance. A typical database includes information collected at the individual level, with data obtained from facilities such as laboratories and clinics. Data collection was most frequently conducted on a daily basis while the most frequent type of formal reporting was annual. Nearly half of the databases have been collected for five years or less.

Most databases are collected to fulfill either federal or state requirements (i.e. birth and death records, abortion reporting). The remaining data collections are typically made as program decisions to determine program effectiveness and verify information submitted by grantees.

More than two-thirds of the data surveyed at the Ohio Department of Health had data restrictions. Some datasets could not be released due to federal agreements (i.e. Minimum Data Set and OASIS). Other data sets required a data user agreement, while still others would only permit access to part of the dataset. Larger population-based surveys were typically available to the public.

The majority of databases surveyed include the collection of race and ethnicity data elements. Data was collected for the race/ethnicity categories of white, black, American Indian, Asian, Hawaiian/Pacific Islander and Hispanic by nearly all of the databases and for the category unknown race by two-thirds of the databases. Some additional variances were noted including one-third of databases that had a category called "other race" which was a fill-in-the-blank type category. Also represented were checkboxes for "other race" and checkboxes for multi-race. Of the 40 databases surveyed, several collected Hispanic as a race and one collected Mediterranean and Middle Eastern as a race. A category for Somalian as an ethnicity was included in some databases, while two had a separate category for Amish.

Most respondents indicated the correct racial and ethnic groups were currently being collected. Of those who indicated that additional racial/ethnic groups should be collected the groups were Hispanic, Somalian and Amish.

Overall, race/ethnicity data quality was considered good, very good or excellent by two-thirds of those surveyed. The majority of databases had low percentages of missing data (0-10 percent). More than 80 percent indicated some form of edit checks were performed on the data.

Some challenges noted in collecting race/ethnicity data included a lack of understanding and training on the part of the person completing the data form and locating the appropriate population and language barriers. In some cases, ensuring that race/ethnicity were self-reported was a priority.

Multi-race data was collected in nearly half of the databases. Most multi-race data collection was mandated by federal grantors. Of those that did not collect multi-race data, 25 percent indicated they would be required to do so by federal mandate, while 40 percent of those not mandated or already collecting multi-race data would be voluntarily switching to that format.

The collection of multi-race data, while important, is limited in Ohio. Nearly all of the databases that collected multi-race data needed to re-assign the data to other race categories for the purpose of data analysis. Typically, the number of people in Ohio that identify their race as “multiple race” was such a small number that meaningful analysis could not be conducted.

Although slightly more than half of the databases surveyed were utilized to produce reports that included information on health disparities, the focus of the reports varied from health disparities information as a minor part of a much larger report, to full reports devoted to health disparities. Typical reasons for not producing reports on health disparities included data restrictions, race/ethnicity not being a factor for data items reported, not enough data and lack of resources.

### **Recommendations**

The use of standardized definitions for race and ethnicity enhances the value of the data and also increases the opportunity for merging and linking of various databases. It is encouraging that the majority of databases include racial and ethnic data and plan to switch to new federal standards for data collections. To increase uniformity, ODH programs should consider using the same questions with regard to race and ethnicity as recommended by the ODH Data Standards committee. The use of “other” category is of concern and should be addressed.

Because socioeconomic factors play a major role in health disparities, uniform collection of these variables should also be considered to better understand underlying causes for racial and ethnic disparities. Presently only 60 percent of databases collect socioeconomic data and programs that do not collect this information should be encouraged to do so. The Public Health Data and Research committee should monitor this process and offer expertise and assistance to programs.

Respondents to this survey rated the quality of racial and ethnic data as fair or poor for 35 percent of the databases. Primary reasons for this were missing

data and lack of verification. As mentioned before, 70 percent of the databases had missing data in the 0-10 percent range, and for 12.5 percent of the databases the missing data range was greater than 30 percent. More effort is needed to rectify this problem. Selected audits should also be performed to determine the quality of key racial and ethnic databases such as vital records.

In recent years, the response rate for data collection efforts has gone down, especially for minority populations. Extra effort and resources are needed to increase the response rate and to determine the effects of this low response rate on key findings.

ODH should also encourage the collection of variables that would allow for geocoding of the data to determine variations across the state. Only 62.5 percent of databases capture the address and 12.5 percent include Census tract information, which is needed for geocoding.

No reports on health disparities are produced from about half of the databases because of data restrictions and lack of sufficient resources or data. Some databases at the department may not have sufficient observations or have a small sample size for meaningful analysis for some minority groups. One way of dealing with this issue is to combine data for several years. Another approach involves increasing the sample size for minority populations by over-sampling or conducting data collection efforts targeted at small racial/ethnic groups. Recent efforts by the center to increase the sample size for minority populations in the Behavioral Risk Factors Surveillance System (BRFSS) and the Ohio Family Health Survey (OFHS) are positive steps and should continue. Each division should also consider devoting more resources for data analysis.

ODH programs should make maximum use of their existing data by including health disparities in their reports. Because our research resources are limited, ODH should make its data readily available to those researchers who have a particular interest in minority health. The recent ODH publication *A Guide to Selected ODH Databases*, includes information on availability of racial and ethnic data and will be useful to identify promising data sources for research.

The establishment of the Minority Health Data Unit at the Center for Vital and Health Statistics is a positive development and should continue to be given the needed resources to perform its functions. In the last two years, the unit has contributed funds to increase the sample size for minority populations in the BRFSS and OFHS, produced several reports on minority health and provided funds to external researchers to study minority health issues.

In conclusion, the databases collected at the Ohio Department of Health are evolving to support the effort of reducing and eliminating health disparities in

Ohio. Many programs have improved data quality, added the collection of multiple race data and are reporting findings in a useful manner. While program-specific data provides very useful information about health disparities for target populations in Ohio, continuing the effort to include and expand population-based surveys such as BRFSS or OFHS is necessary to provide population-based estimates on items such as health insurance, health status and access to health care for Ohioans. In addition, continued effort to produce more longitudinal data will assist us in determining the effectiveness of our efforts. Such future endeavors will be imperative to understanding and reducing health disparities in Ohio.

## Appendix

**Ohio Department of Health  
Center for Public Health Data and Statistics  
Racial/Ethnic Health Data Survey**

**Data set name:**

**Division name:**

**Bureau:**

**Program:**

**Contact person:**

Telephone number:

e-mail address:

**1. Purpose for the data collection?**

**2. What population is covered by the database?**

**3. Is the data collected at the individual level or in aggregate form?**

Yes

No

Individual

Aggregate (i.e. hospital,  
county, grantee)

***Please specify***

**4. How is the data collected? *Please circle all that apply***

Survey (*indicate type, i.e. mail, telephone*)

Enrollment data

Patient Medical Records

Facility Reports (*i.e. hospital or nursing home records*)

Observation (*i.e. report by surveyor, clinician observation*)

Other (*please specify*)

<p><b>5. How frequently is the data collected? <i>Please circle</i></b></p> <p>Daily  Weekly  Monthly  Annual  Other (<i>please specify</i>)</p>		
<p><b>6. How frequently is the data compiled? <i>Please circle</i></b></p> <p>Daily  Weekly  Monthly  Annual  Other (<i>please specify</i>)</p>		
<p><b>7. What is the most recent period for which data is available?</b></p>		
<p><b>8. How many years have the data been collected and a dataset produced?</b></p>		
<p><b>9. Do you collect Race/Hispanic Origin as data items?</b></p> <p>If the response is yes, continue with section two of the survey  <i>If the response is no, answer only questions 10 and 11 to complete the survey</i></p>		
<p>Yes <span style="float: right;">No</span></p>		
Race		
Hispanic Origin		
<p><b>10. Reason for Not collecting Race/Hispanic Origin as data items? <i>Please circle applicable reasons</i></b></p> <p>Not applicable  Cost  Other (<i>please specify</i>) -</p>		
<p><b>11. Would you like to collect Race/Hispanic Origin data?</b></p> <p style="text-align: center;">Yes <span style="float: right;">No</span></p>		

## Section Two

*This section should only be completed for those databases that contain race/ethnicity data.*

### A. General Information

#### 1. Do you collect racial/ethnicity data for the following categories?

	Yes	No
White		
Black		
American Indian or Alaska Native		
Asian		
Hawaiian/Pacific Islander		
Unknown		
Other Race (please specify)		
Hispanic Ethnicity		
Other Ethnicity (please specify)		

#### 2. Do you allow for multiple race selections in your data collection?

Yes

No

#### 3. Will you be required to comply with a Federal Mandate for the collection of multiple race?

Yes

No

**If yes,  
What is the expected date for compliance?**

**If no,  
Will you be voluntarily changing your categories for race/ethnicity to comply with Federal Mandate standards for the collection of multiple races?**

<b>4. Is the data stored electronically in the same race/ethnicity categories that it is collected?</b>		
Yes		No
<b>If no, how do racial/ethnic electronic categories differ from the racial/ethnic categories collected?</b>		
<b>5. Please estimate the percentage of data observations that represent racial/ethnic minorities?</b>		
<i>Please circle appropriate response</i>		
0-10%		
11-20%		
21-30%		
31-40%		
41-50%		
Greater than 50%		
<b>6. Is the data collection mandated by any of the following entities?</b>		
Yes		No
Federal		
State		
Program Requirement		
Other <i>(please specify)</i>		
<b>7. Restrictions on data use: Is the data considered?</b>		
Yes		No
Public		
Private or Confidential		
If private or confidential can the data be released in any form? (i.e. aggregate, stripped of identifiers, approved by a review committee)		
<b>8. Geographic units recorded:</b>		
Yes		No
State		
County		
City/Town		
Zip Code		
Address		
Census Tract		

**9. Indicate whether or not the dataset includes data by Race/Ethnicity in the following areas:**

	Yes	No
Demographic data		
Socioeconomic data		
Behavior risk factors		
Health care utilization		
Health care cost		
Payment		
Mortality		
Morbidity		
Prevention		
Immunization		
Cancer data		
Quality of life		
Genetics		
Health Insurance		
Access to care		
Health status		
Quality of care		
Health behavior/Promotion		
Other ( <i>please specify</i> )		

**B. Data Quality**

**1. How would you describe the quality of the racial/ethnic data you collect?**

- Poor
- Fair
- Good
- Very Good
- Excellent

**2. What is the percentage of missing race/ethnicity data?**

- 0-10%
- 11-20%
- 21-30%
- 31-40%
- 41-50%
- Greater than 50%

**3. Is the percentage of missing race/ethnicity data declining?**

- Yes
- No
- Not applicable

If yes, what improved your missing data percentage?

**4. Do you perform edit checks on your dataset?**

- Yes
- No

**C. Reports**

**1. Have you produced any reports from the dataset that include information on health disparities?**

- Yes
- No

If yes, what are the titles?

If no, have you produced other reports from the dataset?

- Yes
- No

What are the titles?

If you have not produced any reports on health disparities, please circle all applicable reasons for not producing such reports

- Not a priority
- Not required
- Not enough time
- Not enough staff
- Poor quality data
- Not enough data
- Other (*please specify*)

**2. How often are the reports on health disparities produced?**

- Monthly
- Annually
- Other (*please specify*)

**3. Have you made the reports on health disparities available to the public in the following format?**

- Print
- Web
- Diskette
- CD ROM
- Other (*please specify*)

**4. Which areas are the focuses of the health disparities reports?**

Circle all that apply

- Enrollment statistics
- Services provided
- Health status
- Utilization
- Cost/charges
- Outcome/quality
- Other (*please specify*)

**D. Gaps, Issues and Recommendations**

**1. Do you think there is a need to collect additional racial/ethnic data? In what areas?**

**2. Could the collection of racial/ethnic data be accomplished utilizing existent tools or would new tools/surveys be necessary?**

**3. What kinds of challenges do you have attempting to collect racial/ethnic data?**

**4. Do you have any recommendations for the collection of racial/ethnic data?**

**5. Do you have any recommendations to increase data analysis/disseminations for racial/ethnic groups?**