

Collection of Data on Race, Ethnicity, Language, and Nativity by US Public Health Surveillance and Monitoring Systems: Gaps and Opportunities

Public Health Reports
2018, Vol. 133(1) 45-54
© 2017, Association of Schools and
Programs of Public Health
All rights reserved.
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0033354917745503
journals.sagepub.com/home/phr



Alfonso Rodriguez-Lainz, PhD, DVM, MPVM¹, Mariana McDonald, DrPH, MA²,
Maureen Fonseca-Ford, MPH¹, Ana Penman-Aguilar, PhD, MPH³,
Stephen H. Waterman, MD, MPH⁴, Benedict I. Truman, MD, MPH⁵,
Martin S. Cetron, MD¹, and Chesley L. Richards, MD, MPH⁶

Abstract

Objective: Despite increasing diversity in the US population, substantial gaps in collecting data on race, ethnicity, primary language, and nativity indicators persist in public health surveillance and monitoring systems. In addition, few systems provide questionnaires in foreign languages for inclusion of non-English speakers. We assessed (1) the extent of data collected on race, ethnicity, primary language, and nativity indicators (ie, place of birth, immigration status, and years in the United States) and (2) the use of data-collection instruments in non-English languages among Centers for Disease Control and Prevention (CDC)–supported public health surveillance and monitoring systems in the United States.

Methods: We identified CDC-supported surveillance and health monitoring systems in place from 2010 through 2013 by searching CDC websites and other federal websites. For each system, we assessed its website, documentation, and publications for evidence of the variables of interest and use of data-collection instruments in non-English languages. We requested missing information from CDC program officials, as needed.

Results: Of 125 data systems, 100 (80%) collected data on race and ethnicity, 2 more collected data on ethnicity but not race, 26 (21%) collected data on racial/ethnic subcategories, 40 (32%) collected data on place of birth, 21 (17%) collected data on years in the United States, 14 (11%) collected data on immigration status, 13 (10%) collected data on primary language, and 29 (23%) used non-English data-collection instruments. Population-based surveys and disease registries more often collected data on detailed variables than did case-based, administrative, and multiple-source systems.

Conclusions: More complete and accurate data on race, ethnicity, primary language, and nativity can improve the quality, representativeness, and usefulness of public health surveillance and monitoring systems to plan and evaluate targeted public health interventions to eliminate health disparities.

Keywords

population surveillance, health disparities measurement, ethnic groups, nativity, immigrant, language barriers

¹ Division of Global Migration and Quarantine, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA

² Office of the Director, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA

³ Office of Minority Health & Health Equity, Centers for Disease Control and Prevention, Atlanta, GA, USA

⁴ Division of Vector-Borne Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA

⁵ Office of the Director, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA, USA

⁶ Office of Public Health Scientific Services, Centers for Disease Control and Prevention, Atlanta, GA, USA

Corresponding Author:

Alfonso Rodriguez-Lainz, PhD, DVM, MPVM, Division of Global Migration and Quarantine, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, 3851 Rosecrans St, MS-P575, San Diego, CA 92110, USA.

Email: jq13@cdc.gov

Since 1965, the US population has become increasingly diverse in culture, race, ethnicity, and language use. Immigration from numerous countries is a major cause of these profound demographic changes.¹ During 1965 to 2015, an estimated 59 million immigrants came to the United States, accounting for more than half of the nation's population growth. During that same period, the proportion of foreign-born people increased from 5% to 14% of the total US population.¹ In 2014, 21% of the US population spoke a language other than English at home.²

For decades, multiple federal agencies, advisory groups, and researchers have recommended that US national health data systems increase their efforts to accurately assess the racial, ethnic, and linguistic diversity of the US population as an essential step in eliminating health disparities.³⁻¹⁴ Recommendations have included collection of more detailed data on race and ethnicity in addition to the current minimum federally required Office of Management and Budget (OMB) standard categories.¹⁰ Proficiency in English and language spoken at home have been proposed for documenting primary language. In addition, indicators of nativity (eg, place of birth, immigration status, and years in the United States) have been increasingly recognized as crucial.^{3-9,11-13,15,16} In 2016, the Centers for Disease Control and Prevention's (CDC's) Advisory Committee to the Director also recommended collection of more detailed data on race, ethnicity, language preference, and immigrant status "in all research studies, all health care settings, and all public health data sets [emphasis added], as an essential component to achieving health equity."¹⁷ Because a substantial proportion (9%) of the US population has limited-English proficiency, translation of questionnaires and access to interpreters have also been recommended to improve the representativeness and quality of data collected.^{5,18}

US public health surveillance and health monitoring systems are critical for quantifying changes in population health, identifying and responding to emerging health challenges and health disparities, and evaluating the effectiveness of public health programs.¹⁹ In addition to data systems designed for surveillance of notifiable diseases, health monitoring systems include population-based surveys, vital records, disease registries, and hospital discharge data systems.²⁰ CDC coordinates, operates, and supports activities of national health data systems through standardization, analysis, and dissemination and by providing funding and other resources.²¹ The amount and detail of data collected by those systems may depend on multiple factors, including purpose, stakeholder information needs, data sources, resources, and local, state, and federal policies. A principal question is whether national data systems have adapted their procedures to properly capture data on changing US demographic characteristics.

The primary objective of this study was to assess the extent to which CDC surveillance and health monitoring systems collected indicators of race, ethnicity, primary language, and nativity. A secondary objective was to assess these systems' use of data-collection instruments in languages other than English.

Methods

We identified CDC-supported public health surveillance and health monitoring data systems that were active anytime from 2010 through 2013 by searching online federal government health data.²²⁻²⁵ We included only those data systems that collected and reported data periodically or on an ongoing basis (eg, nationally notifiable disease systems, annual health surveys, disease registries, vital records, and hospital discharge data systems).²⁶ We excluded data systems monitoring only environmental conditions (eg, air, water, animal vectors, or animal health), as well as CDC-supported global health monitoring systems that were implemented exclusively overseas.

Variables of interest were race, ethnicity, primary language, and nativity indicators. For each data system, we searched the website, technical documentation, data-collection instruments (case report forms, questionnaires, and internet-based data-entry applications), and publications to assess if the demographic variables of interest were collected and with what level of detail, as well as if data-collection instruments were available in languages other than English. If needed, we contacted a CDC data system representative to obtain information about the system.

We used the following criteria to categorize the detail of information collected by data systems:

- **Race and ethnicity:** We defined data on race and ethnicity as basic if the system used some version of the 1997 OMB federally required minimum data standards 5 racial categories (American Indian or Alaska Native, Asian, black or African American, Native Hawaiian or other Pacific Islander, and white) and 2 ethnicity categories (Hispanic or Latino and not Hispanic or Latino).²⁷ We defined data on race and ethnicity as detailed if the system collected more detailed data on race (eg, Vietnamese) or Hispanic/Latino ethnicity (eg, Puerto Rican).
- **Place of birth:** We defined data on place of birth as basic if the system collected only dichotomous data on variables (eg, US-born and foreign-born) and detailed if data on the country of birth were collected.
- **Immigration status:** We defined data on immigration status as a yes if the system collected data about citizenship, refugee status, or other legal immigration category. We defined immigration status as a no if such data were not collected. Data on unauthorized immigration status are not collected by CDC-supported health monitoring systems.
- **Primary language:** We defined data on primary language as a yes if the system collected data on preferred language, language spoken at home, English-speaking ability, language of the interview, or the need for an interpreter to interview a participant.
- **Years in the United States:** We defined data on years in the United States as a yes if the system collected questions about the year of arrival or the number of years living in the country.

Table 1. Examples of Centers for Disease Control and Prevention (CDC)–supported surveillance and health monitoring systems, by category, United States, 2010-2013^a

Surveillance and Health Monitoring Systems	URL
Case-based	
Adult Blood Lead Epidemiology and Surveillance	http://www.cdc.gov/niosh/topics/ABLES/ables.html
GeoSentinel	http://www.istm.org/geosentinel
National HIV Surveillance System	https://www.cdc.gov/hiv/statistics/surveillance/systems/index.html
National Tuberculosis Surveillance System	https://www.healthypeople.gov/2020/data-source/national-tb-surveillance-system
National Respiratory and Enteric Virus Surveillance System	http://www.cdc.gov/surveillance/nrevss/
Sexually Transmitted Diseases Surveillance Network	https://www.cdc.gov/std/ssun/default.htm
Viral Hepatitis Surveillance Program	http://www.cdc.gov/hepatitis/Statistics/index.htm
Population survey	
Behavioral Risk Factor Surveillance System	https://www.cdc.gov/brfss/index.html
National Adult Tobacco Survey	http://www.cdc.gov/tobacco/data_statistics/surveys/nats/index.htm
National Agricultural Workers Survey	http://www.doleta.gov/agworker/naws.cfm
National Health and Nutrition Examination Survey	https://www.cdc.gov/nchs/nhanes/index.htm
National Health Interview Survey	http://www.cdc.gov/nchs/nhis/nhis_2014_data_release.htm
National Immunization Surveys (NIS-Children, NIS-Teen, and NIS-Adult)	https://www.cdc.gov/vaccines/imz-managers/nis/about.html
Registry	
National Amyotrophic Lateral Sclerosis Registry	https://wwwn.cdc.gov/als/
National Occupational Respiratory Mortality System	https://webappa.cdc.gov/ords/norms.html
National Program of Cancer Registries	http://www.cdc.gov/cancer/npcr/
National Spina Bifida Patient Registry	https://www.cdc.gov/ncbddd/spinabifida/NSBPRregistry.html
Administrative	
National Assisted Reproductive Technology Surveillance System	http://www.cdc.gov/art/NASS.htm
National Electronic Injury Surveillance System—Occupational Supplement	https://wwwn.cdc.gov/wisards/workkrisqs/about.aspx
National Healthcare Safety Network	http://www.cdc.gov/nhsn/index.html
National Hospital Care Survey	http://www.cdc.gov/nchs/nhcs.htm
Multiple sources	
Asthma Surveillance	http://www.cdc.gov/asthma/asthmaadata.htm
Chronic Kidney Disease Surveillance System	http://www.cdc.gov/ckd/surveillance
National Diabetes Surveillance System	https://www.cdc.gov/diabetes/data/index.html
National Violent Death Reporting System	http://www.cdc.gov/ViolencePrevention/NVDRS/index.html

^aCase-based refers to individual case reports or aggregated counts of reported cases. Population survey refers to periodic collection from a population probability sample. Registry refers to a structured system to track all cases of a disease or condition, births, or deaths for a defined population (eg, cancer registries, vital records). Administrative refers to data collected mainly for administrative purposes (eg, hospital discharge data). Multiple sources refers to a system that compiles variables from various data sources (eg, asthma surveillance using vital statistics, hospital discharges, and surveys).

For analysis, we grouped the data systems as follows: (1) case-based (individual case reports or aggregated counts of reported cases [eg, National Notifiable Diseases Surveillance System])^{28,29}; (2) population survey (periodic collection from a population probability sample [eg, National Health Interview Survey])³⁰; (3) registry (structured system to track all cases of a disease or condition, births, or deaths among a defined population [eg, cancer registries, vital records]); (4) administrative (data collected mainly for administrative purposes [eg, hospital discharge data]); and (5) multiple sources (system that compiles data on variables from various sources [eg, asthma surveillance using vital statistics, hospital discharges, and surveys]) (Table 1).²⁰

Results

We identified 125 CDC-sponsored surveillance and health monitoring systems. Diseases and conditions being

monitored included, but were not limited to, infectious diseases, chronic conditions, injuries, mortality, and birth defects. We classified systems as case-based (n = 54), population survey (n = 22), registry (n = 16), administrative (n = 16), and multiple sources (n = 17) (Table 2).

Of the 125 data systems reviewed, 100 (80%) collected data on race and ethnicity; 2 of the systems that collected data on ethnicity did not collect data on race. Twenty-six (21%) systems collected detailed data on race, and 27 (22%) systems collected detailed data on ethnicity. Thirteen (10%) systems collected data on primary language. For nativity indicators, 40 (32%) systems collected data on place of birth, of which 31 (25%) systems collected detailed information. Fourteen (11%) systems collected data on immigration status (most frequently citizenship), and 21 (17%) systems collected data on years in the United States. Fourteen or fewer ($\leq 11\%$) systems collected data on the 5 questions about the respondent's parents. The wording of questions used to collect data on variables of

Table 2. Data on race/ethnicity, primary language, and nativity indicators collected by 125 Centers for Disease Control and Prevention (CDC) surveillance and health monitoring systems, by type of system, United States, 2010-2013^a

Variable	Type of Data System ^a					
	Total (n = 125) No. (%)	Case-Based (n = 54) No. (%)	Population Survey (n = 22) No. (%)	Registry (n = 16) No. (%)	Administrative (n = 16) No. (%)	Multiple Sources (n = 17) No. (%)
Race						
Yes	100 (80)	37 (69)	21 (95)	14 (88)	13 (81)	15 (88)
Basic ^b	74 (59)	32 (59)	10 (45)	9 (56)	10 (63)	13 (76)
Detailed ^c	26 (21)	5 (9)	11 (50)	5 (31)	3 (19)	2 (12)
No	25 (20)	17 (31)	1 (5)	2 (13)	3 (19)	2 (12)
Ethnicity						
Yes	102 (82)	39 (72)	21 (95)	14 (88)	13 (81)	15 (88)
Basic ^b	75 (60)	37 (69)	7 (32)	8 (50)	10 (63)	13 (76)
Detailed ^c	27 (22)	2 (4)	14 (64)	6 (38)	3 (19)	2 (12)
No	23 (18)	15 (28)	1 (5)	2 (13)	3 (19)	2 (12)
Primary language						
Yes	13 (10)	1 (2)	10 (45)	2 (13)	0 (0)	0 (0)
No	112 (90)	53 (98)	12 (55)	14 (88)	16 (100)	17 (100)
Place of birth						
Yes	40 (32)	17 (31)	10 (45)	10 (63)	3 (19)	0 (0)
Basic ^b	9 (7)	0 (0)	4 (18)	3 (19)	2 (13)	0 (0)
Detailed ^c	31 (25)	17 (31)	6 (27)	7 (44)	1 (6)	0 (0)
No	85 (68)	37 (69)	12 (55)	6 (38)	13 (81)	17 (100)
Immigration status						
Yes	14 (11)	9 (17)	5 (23)	0 (0)	0 (0)	0 (0)
No	111 (89)	45 (83)	17 (77)	16 (100)	16 (100)	17 (100)
Years in the United States						
Yes	21 (17)	11 (20)	10 (45)	0 (0)	0 (0)	0 (0)
No	104 (83)	43 (80)	12 (55)	16 (100)	16 (100)	17 (100)
Parental race or ethnicity						
Yes	6 (5)	2 (4)	0 (0)	2 (13)	0 (0)	2 (12)
No	119 (95)	52 (96)	22 (100)	14 (88)	16 (100)	15 (88)
Parental language						
Yes	4 (3)	1 (2)	3 (14)	0 (0)	0 (0)	0 (0)
No	121 (97)	53 (98)	19 (86)	16 (100)	16 (100)	17 (100)
Parental country of birth						
Yes	14 (11)	7 (13)	6 (27)	1 (6)	0 (0)	0 (0)
Basic ^d	6 (5)	1 (2)	5 (23)	0 (0)	0 (0)	0 (0)
Detailed ^e	8 (6)	6 (11)	1 (5)	1 (6)	0 (0)	0 (0)
No	111 (89)	47 (87)	16 (73)	15 (94)	16 (100)	17 (100)
Parental immigration status						
Yes	1 (1)	0 (0)	1 (5)	0 (0)	0 (0)	0 (0)
No	124 (99)	54 (100)	21 (95)	16 (100)	16 (100)	17 (100)
Parental years in the United States						
Yes	9 (7)	4 (7)	4 (18)	1 (6)	0 (0)	0 (0)
No	116 (93)	50 (93)	18 (82)	15 (94)	16 (100)	17 (100)

^aCase-based refers to individual case reports or aggregated counts of reported cases. Population survey refers to periodic collection from a population probability sample. Registry refers to a structured system to track all cases of a disease or condition, births, or deaths for a defined population (eg, cancer registries, vital records). Administrative refers to data collected mainly for administrative purposes (eg, hospital discharge data). Multiple sources refers to a system that compiles variables from various data sources (eg, asthma surveillance using vital statistics, hospital discharges, and surveys).

^bCollects minimum federally required 1997 White House Office of Management and Budget data on race or ethnicity categories as follows: race (American Indian or Alaska Native, Asian, black or African American, Native Hawaiian or other Pacific Islander, white); ethnicity (Hispanic or Latino and not Hispanic or Latino).

^cCollects more detailed data on race (eg, Filipino or Vietnamese) or ethnicity (eg, Mexican or Puerto Rican) group information.

^dUS-born or foreign-born.

^eCollects data on country of birth.

interest varied by system. Twenty-nine (23%) data systems had a data-collection instrument translated into non-English languages, all of which were population surveys or multiple-

source systems using survey data. Most of these systems (27/29) provided questionnaires in Spanish, and 2 provided questions in Chinese.

Data on variables and their degree of detail varied by type of data system. For example, of the 22 population surveys, 15 collected data on race or ethnicity, 11 collected data on detailed race categories, and 14 collected data on detailed ethnic categories. Of the 54 case-based systems, 37 (69%) collected data on race (only 5 of which collected detailed data) and 39 (72%) collected data on ethnicity (only 2 of which collected detailed data). Data collection on primary language ranged from 10 of 22 population surveys to none of the administrative or multiple-source systems. Data collection on place of birth varied from 10 of 16 registries to none of the multiple-source systems. No registry, administrative, or multiple-source system collected data on immigration status or years of residence, whereas 5 of 22 population surveys collected data on immigration status and 10 of 22 population surveys collected data on years in the United States.

Data collection also varied within data systems. For example, in the National Notifiable Diseases Surveillance System, only basic data on race and ethnicity were collected on the various data forms used for all of the approximately 100 notifiable conditions monitored by the system. Data on additional variables were collected only for a few conditions.

Discussion

Many CDC-sponsored surveillance and health monitoring systems have not fully adapted their procedures for completely and accurately capturing data on the increasing diversity of the United States by gathering data on race/ethnicity, primary language, and nativity. Gaps in data collection were greatest for primary language and nativity indicators and varied by type of data system. We also observed a lack of standardization in questions used to obtain data on variables of interest, limiting comparability across data systems. In addition, most systems do not translate their data-collection instruments, which could prevent non-English speakers from participating, particularly for self-administered questionnaires or when collecting data directly from people. Similar gaps in US national data systems have been reported, but published assessments focused only on collection of data on race and ethnicity or health conditions or included only a few national data sources.^{5,9,10,16,31-36}

Our findings may be partly explained by the diversity of systems included in the assessment in terms of main objectives, data needs, personnel and financial resources, who collects the data, and whether data are collected directly from people or extracted from existing records and subsequently reported to CDC. Current OMB standards for data collection relate only to race/ethnicity and are not required for all data systems.⁵ Lack of resources and the need to limit the burden to reporters and maintain comparability with historic data might be important barriers to changing the system. For some surveillance systems, state and local partners determine which data are collected and what to report to CDC.²⁰ Other surveillance systems may limit monitoring to distinct indicators, such as over-the-counter prescription sales, and

detailed data on demographic characteristics may be less relevant and/or not available.²⁰ Case-based systems frequently focus on disease and risk factors and clinical course of illness, with limited demographic data gathered by health care providers and laboratory personnel. In contrast, data for population surveys include detailed data on demographic characteristics, usually collected by professional interviewers, to assess differences in health among population groups.²⁰

Race and Ethnicity

Our findings revealed that approximately 20% of data systems did not collect *any* data on race or ethnicity. Among those that did, most collected only basic information. Currently, federal data systems are not required to collect data on race and ethnicity. However, extensive evidence indicates that omitting detailed data on race and ethnicity, or aggregating diverse communities into simplified racial/ethnic categories, limits public health practitioners' ability to identify health disparities among certain populations.^{4,7,8,13,16,37,38} For example, smoking prevalence among Hispanic people overall (13.5%) is substantially lower than among non-Hispanic white people (23.8%). However, smoking prevalence varies among certain Hispanic-origin subgroups, with a rate of 21.6% for Puerto Ricans. Such a disparity affecting Puerto Ricans would likely remain hidden and unaddressed if data were collected only on Hispanic ethnicity and no subgroups.¹³

Primary Language

Few data systems collected data on primary language. Data on primary language are important because people who are not proficient in English have multiple health disparities, including poorer health status, less access to health care, lower quality of health care, and less usage of preventive services, as compared with English-proficient people.^{7,39-42} Data on primary language are also useful for guiding decisions about which languages should receive high priority for use in questionnaires or for translation services delivered to the public.

Nativity Indicators

The importance of collecting data on nativity indicators has long been recognized.³ Foreign-born people are an increasing proportion of the US population.¹ International migration is an important factor in the global spread of emerging infectious diseases and can have profound effects on disease epidemiology at both national and local levels,⁴³ yet most data systems in our assessment did not collect data on any nativity indicators.

Nativity indicators are important for identifying health disparities.^{5,12-14,36,44} Although foreign-born people may have more favorable indicators for selected conditions (eg, breast and lung cancer) than does the US-born population,^{13,16} they have substantial disparities in other conditions (eg, infectious diseases, access to health care) compared with US-born

populations of the same race and ethnicity and after adjusting for socioeconomic factors.^{16,34} Among foreign-born people, health insurance and vaccination coverage typically are lower for non-US citizens, newer immigrants, and those born in Latin America.⁴⁵ Longer length of US residence, in contrast, is associated with increased risk for such health conditions as diabetes, obesity, and substance abuse.^{16,35}

One-quarter of all US children live with at least 1 foreign-born parent.² Children of immigrant parents, even for US-born children, have risk factors, health outcomes, and barriers to health access that differ from those of children with US-born parents.^{16,42} Collection of data on parents' primary language and nativity indicators is, therefore, essential for monitoring US children's health.

Data-Collection Instruments in Foreign Languages

Our finding that most data systems exclusively use data-collection instruments in English is of concern only for systems that collect data directly from people, particularly if surveys are self-administered. In other settings, translation services can be made available as needed. Otherwise, respondents with limited English proficiency may be underrepresented or excluded from participating in data collection. Even if they do participate, there may be inaccuracies in the data. Consequently, the representativeness, validity, reliability, and completeness of such data are likely to be reduced.⁴⁶⁻⁴⁸ This observation is particularly relevant for those geographic regions, health conditions, and racial/ethnic groups with high proportions of people with limited English proficiency.⁴⁹

Limitations

Although extensive, the data systems included in this assessment do not represent all CDC-supported data systems for population health monitoring but should rather be considered a convenience sample. We did not provide a full list of the data systems because many systems were inactive in 2017, had made changes in the data that were collected, or did not have a website describing the system. Thus, our findings applied only to the data systems included in this study and to the data they collected during the study period. Also, we based the assessment primarily on public-use documentation available; as such, it may not reflect all data pertaining to each system. Finally, we based the grouping of data systems into various categories based on our interpretation of available documentation; thus, some data systems may have been misclassified.

Lessons Learned

The findings from this assessment, along with evidence from previous reports (including those from federal agencies and national advisory groups^{3-5,7,9-14,18}), suggest the following strategies for data systems to consider for improving the quality, standardization, and representativeness of data on the health of the US population:

Collection of detailed data on race, ethnicity, primary language, and nativity: Collection of these indicators using US Census Bureau–validated questions is suggested for comparability and to provide appropriate population denominators (Figure).^{3,50}

Translation of data-collection instruments into non-English languages: For data systems that collect data directly from people, languages for translation can be prioritized based on the most prevalent languages spoken at home by members of the non-English-speaking target population. Such information is available from US Census Bureau data and sometimes from collaborating organizations. Validating the cultural appropriateness of translated documents is necessary for ensuring equivalence of meaning across languages. In addition, using trained bilingual interviewers can ensure higher-quality data collection.^{5,48,51-53}

The relevance and degree of implementation of recommended strategies by data systems may depend on their purpose and available resources, among other factors. Data systems could prioritize new variables based on their public health needs and resources. Integration of strategies may be easier for new data systems in the planning stages. Changes in a data system, including modification and translation of data-collection instruments and databases, may impose logistic or resource challenges and may increase the burden of data collection on data providers and the public. However, system changes can provide gains in the quantity, quality, and completeness of the resulting data. Data systems with greater resources (eg, national surveys and enhanced population-based surveillance systems) may have more capacity than those based on passive reporting from health care providers to implement the recommended strategies.⁵⁴ Multiple-source systems may choose to extract data on the recommended variables if they are available in one of the data sources they use. Emerging technologies may also minimize burdens in collecting new data and in translating data-collection instruments.⁵³

The feasibility of these strategies is demonstrated by multiple national (eg, National Health Interview Survey, Tuberculosis Surveillance System) and state (eg, California Health Interview Survey) data systems that have implemented them for decades.⁵⁵ More recently, the National Notifiable Diseases Surveillance System and ArboNET (for Zika case reporting) added country of birth to their report forms.²⁸ The CDC Listeria Initiative also added detailed subcategories on race, ethnicity, country of birth, and primary language and made its questionnaire available in Spanish.⁵⁶

Adopting these strategies is also crucial from an ethics perspective, to prevent the potential exclusion of people from federal data-collection activities and to allow the identification of populations with health disparities that may otherwise remain invisible and underserved. To complement those strategies, innovative and efficient data-collection and analysis approaches, such as periodic targeted surveys, data modeling, and linking across data sets, have been recommended.⁵ Finally, enhanced data collection and analysis require appropriate safeguards to protect the privacy and confidentiality of

<p>1. Is this person of Hispanic, Latino, or Spanish origin?</p> <p><input type="checkbox"/> No, not of Hispanic, Latino, or Spanish origin</p> <p><input type="checkbox"/> Yes, Mexican, Mexican American, Chicano</p> <p><input type="checkbox"/> Yes, Puerto Rican</p> <p><input type="checkbox"/> Yes, Cuban</p> <p><input type="checkbox"/> Yes, another Hispanic, Latino, or Spanish origin <i>Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.</i> _____</p>	<p>These categories roll up to the Hispanic or Latino category of the OMB minimum standard</p>
<p>2. What is this person's race? Mark (X) on 1 or more boxes.</p> <p><input type="checkbox"/> White</p> <p><input type="checkbox"/> Black or African American</p> <p><input type="checkbox"/> American Indian or Alaska Native <i>Print name of enrolled or principal tribe.</i> _____</p> <p><input type="checkbox"/> Asian Indian</p> <p><input type="checkbox"/> Chinese</p> <p><input type="checkbox"/> Filipino</p> <p><input type="checkbox"/> Other Asian <i>Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.</i> _____</p> <p><input type="checkbox"/> Japanese</p> <p><input type="checkbox"/> Korean</p> <p><input type="checkbox"/> Vietnamese</p> <p><input type="checkbox"/> Native Hawaiian</p> <p><input type="checkbox"/> Guamanian or Chamorro</p> <p><input type="checkbox"/> Samoan</p> <p><input type="checkbox"/> Other Pacific Islander <i>Print race, for example, Fijian, Tongan, and so on.</i> _____</p> <p><input type="checkbox"/> Some other race – Print race _____</p>	<p>These categories are part of the current OMB minimum standard</p> <p>These categories roll up to the Asian category of the OMB minimum standard</p> <p>These categories roll up to the Native Hawaiian or other Pacific Islander category of the OMB minimum standard</p>
<p>3. Language:</p> <p>a. Does this person speak a language other than English at home?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p>b. What is this language? _____ <i>For example: Korean, Italian, Spanish, Vietnamese</i></p> <p>c. How well does this person speak English?</p> <p><input type="checkbox"/> Very well</p> <p><input type="checkbox"/> Well</p> <p><input type="checkbox"/> Not well</p> <p><input type="checkbox"/> Not at all</p>	
<p>4. Nativity indicators:</p> <p>a. Where was this person born?</p> <p><input type="checkbox"/> In the United States</p> <p><input type="checkbox"/> Outside the United States (<i>Print name of foreign country, or Puerto Rico, Guam, etc.</i>) _____</p> <p>b. In what country was your (a) mother, (b) father born? (List countries.) (a) _____ (b) _____</p> <p>5. When did this person come to live in the United States? <i>If this person came to live in the United States more than once, print most recent year.</i> Year _____</p>	
<p>6. Is this person a citizen of the United States?</p> <p><input type="checkbox"/> Yes, born in the United States</p> <p><input type="checkbox"/> Yes, born in Puerto Rico, Guam, the US Virgin Islands, or Northern Marianas</p> <p><input type="checkbox"/> Yes, born abroad of US-citizen parent or parents</p> <p><input type="checkbox"/> Yes, US citizen by naturalization</p> <p><input type="checkbox"/> No, not a US citizen</p>	

Figure. Examples of questions used in surveys conducted by the US Census Bureau to collect data on race, ethnicity, primary language, and nativity indicators of the US population.⁵⁰ Abbreviation: OMB, Office of Management and Budget.

respondents and to prevent stigmatization of racial/ethnic minority populations.⁷ CDC has strict privacy and security policies and procedures for collecting, storing, and releasing personally identifiable data by surveillance and health monitoring programs, in compliance with federal regulations.⁵⁷

Conclusion

The US population is becoming increasingly diverse in race, ethnicity, language, and nativity. To protect and improve the health of all US populations,²⁰ surveillance and health monitoring systems may need to adapt to changing US demographic characteristics and capture complete and accurate data on the nation's diversity.

Gaps in data collection identified in this article can be filled by using feasible strategies based on strong scientific and ethical justification. The recommended strategies can enhance the scientific quality of information needed to support public health practice and make the nation better equipped to respond to emerging health challenges and eliminate health disparities.

Authors' Note

The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

1. Pew Research Center. Modern immigration wave brings 59 million to U.S., driving population growth and change through 2065: views of immigration's impact on U.S. society mixed. 2015. http://www.pewhispanic.org/files/2015/09/2015-09-28_modern-immigration-wave_REPORT.pdf. Accessed September 19, 2016.
2. Zong J, Batalova J. *Frequently Requested Statistics on Immigrants and Immigration in the United States*. Washington, DC: Migration Policy Institute; 2015. <http://www.migrationpolicy.org/article/frequently-requested-statistics-immigrants-and-immigration-united-states-4>. Accessed September 19, 2016.
3. National Committee on Vital and Health Statistics. *Migration, Vital, and Health Statistics: A Report of the United States National Committee on Vital and Health Statistics*. Washington, DC: US Department of Health, Education, and Welfare, Public Health Service, Health Services and Mental Health Administration; 1968. http://www.cdc.gov/nchs/data/series/sr_04/sr_04_009.pdf. Accessed September 19, 2016.
4. National Research Council, Panel on DHHS Collection of Race and Ethnicity Data; Ver Ploeg M, Perrin E, eds. *Eliminating Health Disparities: Measurement and Data Needs*. Washington, DC: National Academies Press; 2004.
5. US Department of Health and Human Services, National Committee on Vital and Health Statistics. *Eliminating Health Disparities: Strengthening Data on Race, Ethnicity, and Primary Language*. Rockville, MD: HHS; 2005. <http://www.cdc.gov/nchs/data/misc/elihealthdisp.pdf>. Accessed September 19, 2016.
6. Blendon RJ, Buhr T, Cassidy EF, et al. Disparities in health: perspectives of a multi-ethnic, multi-racial America. *Health Aff (Millwood)*. 2007;26(5):1437-1447.
7. Institute of Medicine. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. Washington, DC: National Academies Press; 2009.
8. Islam NS, Khan S, Kwon S, Jang D, Ro M, Trinh-Shevrin C. Methodological issues in the collection, analysis, and reporting of granular data in Asian American populations: historical challenges and potential solutions. *J Health Care Poor Underserved*. 2010;21(4):1354-1381.
9. Beltran VM, Harrison KM, Hall HI, Dean HD. Collection of social determinant of health measures in U.S. national surveillance systems for HIV, viral hepatitis, STDs, and TB. *Public Health Rep*. 2011;126(suppl 3):41-53.
10. Dorsey R, Graham G. New HHS data standards for race, ethnicity, sex, primary language, and disability status. *JAMA*. 2011;306(21):2378-2379.
11. US Department of Health and Human Services, Administration for Children and Families. Survey data elements to unpack diversity of Hispanic populations. OPRE Report 2014-30. 2014. http://www.acf.hhs.gov/sites/default/files/opre/brief_survey_data_to_unpack_hispanic_final_03_27_2014.pdf. Accessed September 19, 2016.
12. Penman-Aguilar A, Talih M, Huang D, Moonesinghe R, Bouye K, Beckles G. Measurement of health disparities, health inequities, and social determinants of health to support the advancement of health equity. *J Public Health Manag Pract*. 2016;22(suppl 1):S33-S42.
13. Dominguez K, Penman-Aguilar A, Chang MH, et al. Vital signs: leading causes of death, prevalence of diseases and risk factors, and use of health services among Hispanics in the United States—2009-2013. *MMWR Morb Mortal Wkly Rep*. 2015;64(17):469-478.
14. Dean HD, Roberts GW, Bouye KE, Green Y, McDonald M. Sustaining a focus on health equity at the Centers for Disease Control and Prevention through organizational structures and functions. *J Public Health Manag Pract*. 2016;22(suppl 1):S60-S67.
15. Koch-Weser S, Grigg-Saito D, Liang S, et al. Health status of Cambodians and Vietnamese—selected communities, United States, 2001-2002. *MMWR Morb Mortal Wkly Rep*. 2004;53(33):760-765.
16. Singh GK, Rodriguez-Lainz A, Kogan MD. Immigrant health inequalities in the United States: use of eight major national data systems. *Scientific World J*. 2013;2013:512313.
17. Richardson LD. Integrating health equity into practice and policy. *J Public Health Manag Pract*. 2016;22(suppl 1):S107-S109.

18. US Department of Health and Human Services/Office of Minority Health. *National Standards for Culturally and Linguistically Appropriate Services in Health and Health Care: A Blueprint for Advancing and Sustaining CLAS Policy and Practice*. Rockville, MD: HHS; 2013. <https://www.thinkculturalhealth.hhs.gov/pdfs/EnhancedCLASStandardsBlueprint.pdf>. Accessed September 19, 2016.
19. Thacker SB, Qualters JR, Lee LM. Public health surveillance in the United States: evolution and challenges. *MMWR Morb Mortal Wkly Rep*. 2012;61(3):3-9.
20. Lee LM, Teutsch SM, Thacker SB, St. Louis ME, eds. *Principles and Practice of Public Health Surveillance*. New York, NY: Oxford University Press; 2010.
21. Richards CL, Iademarco MF, Anderson TC. A new strategy for public health surveillance at CDC: improving national surveillance activities and outcomes. *Public Health Rep*. 2014;129(6):472-476.
22. US Department of Health and Human Services. Healthdata.gov. 2016. <http://healthdata.gov/dataset/search>. Accessed December 22, 2016.
23. US Department of Health and Human Services. Guide to HHS surveys and data resources. 2012. <http://aspe.hhs.gov/sp/surveys/index.cfm>. Accessed September 19, 2016.
24. National Center for Health Statistics. Health indicators warehouse. 2016. <http://www.healthindicators.gov/Resources/DataSources>. Accessed December 22, 2016.
25. US Department of Health and Human Services. Healthy people 2020: data sources. 2016. <https://www.healthypeople.gov/2020/data-search/Data-Sources>. Accessed September 19, 2016.
26. Hall HI, Correa A, Yoon PW, Braden CR. Lexicon, definitions, and conceptual framework for public health surveillance. *MMWR Morb Mortal Wkly Rep*. 2012;61(03):10-14.
27. Office of Management and Budget. Revisions to the standards for the classification of federal data on race and ethnicity. *Fed Regist*. 1997;62(210):58782.
28. Centers for Disease Control and Prevention. National Notifiable Diseases Surveillance System. <https://www.cdc.gov/nndss>. Accessed September 19, 2016.
29. Centers for Disease Control and Prevention. *Manual for the Surveillance of Vaccine-Preventable Diseases*. 6th ed. Atlanta, GA: US Department of Health and Human Services, CDC; 2013. <https://www.cdc.gov/vaccines/pubs/surv-manual/index.html>. Accessed September 19, 2016.
30. National Center for Health Statistics. 2014 National Health Interview Survey (NHIS): public use data release. 2015. http://www.cdc.gov/nchs/nhis/nhis_2014_data_release.htm. Accessed September 19, 2016.
31. Centers for Disease Control and Prevention. CDC health disparities and inequalities report—United States, 2011. *MMWR Morb Mortal Wkly Rep*. 2011;60:1-116.
32. Bilheimer LT, Klein RJ. Data and measurement issues in the analysis of health disparities. *Health Serv Res*. 2010;45(5 pt 2):1489-1507.
33. Waksberg J, Levine D, Marker D. *Assessment of Major Federal Data Sets for Analyses of Hispanic and Asian or Pacific Islander Subgroups and Native Americans. Task 3 Report: Extending the Utility of Federal Databases*. Rockville, MD: Westat; 2000. https://archive.org/stream/assessmentofmajor00west_0#page/n0/mode/2up. Accessed December 22, 2016.
34. Johnson PJ, Blewett LA, Davern M. Disparities in public use data availability for race, ethnic, and immigrant groups: national surveys for healthcare disparities research. *Med Care*. 2010;48(12):1122-1127.
35. Oza-Frank R, Cunningham SA. Highlighting resources to study cardiovascular disease and diabetes in US immigrants. *ISRN Public Health*. 2012;2012:198983.
36. Steege AL, Baron SL, Marsh SM, Menéndez CC, Myers JR. Examining occupational health and safety disparities using national data: a cause for continuing concern. *Am J Ind Med*. 2014;57(5):527-538.
37. Avila RM, Bramlett MD. Language and immigrant status effects on disparities in Hispanic children's health status and access to health care. *Matern Child Health J*. 2013;17(3):415-423.
38. Holland AT, Palaniappan LP. Problems with the collection and interpretation of Asian-American health data: omission, aggregation, and extrapolation. *Ann Epidemiol*. 2012;22(6):397-405.
39. Lee S, Nguyen HA, Jawad M, Kurata J. Linguistic minorities in a health survey. *Public Opin Q*. 2008;72(3):470-486.
40. Kimbro RT, Gorman BK, Schachter A. Acculturation and self-rated health among Latino and Asian immigrants to the United States. *Soc Prob*. 2012;59(3):341-363.
41. Dorsey R, Graham G, Glied S, Meyers D, Clancy C, Koh H. Implementing health reform: improved data collection and the monitoring of health disparities. *Annu Rev Public Health*. 2014;35:123-138.
42. Flores G, Tomany-Korman SC. The language spoken at home and disparities in medical and dental health, access to care, and use of services in US children [published erratum appears in *Pediatrics*. 2009;124(4):1265]. *Pediatrics*. 2008;121(6):e1703-e1714.
43. Gushulak BD, Weekers J, MacPherson DW. Migrants and emerging public health issues in a globalized world: threats, risks and challenges, an evidence-based framework. *Emerg Health Threats J*. 2009;2:e10.
44. Centers for Disease Control and Prevention. CDC health disparities and inequalities report—United States, 2013. *MMWR Morb Mortal Wkly Rep*. 2013;62(suppl 3):1-189.
45. Lu PJ, Rodriguez-Lainz A, O'Halloran A, Greby S, Williams WW. Adult vaccination disparities among foreign-born populations in the US, 2012. *Am J Prev Med*. 2014;47(6):722-733.
46. Pearson WS, Garvin WS, Ford ES, Balluz LS. Analysis of five-year trends in self-reported language preference and issues of item non-response among Hispanic persons in a large cross-sectional health survey: implications for the measurement of an ethnic minority population. *Popul Health Metr*. 2010;8:7.
47. Pastor PN, Reuben CA, Duran CR. Reported child health status, Hispanic ethnicity, and language of interview: United States, 2011-2012. *Natl Health Stat Rep*. 2015;82:1-11.
48. Hunt S, Bhopal R. Self-report in clinical and epidemiological studies with non-English speakers: the challenge of language and culture. *J Epidemiol Community Health*. 2004;58(7):618-622.

49. Landrine H, Corral I. Advancing research on racial-ethnic health disparities: improving measurement equivalence in studies with diverse samples. *Front Public Health*. 2014;2:282.
50. US Census Bureau. Foreign-born, surveys/programs. <https://www.census.gov/topics/population/foreign-born/surveys-programs.html>. Accessed November 5, 2017.
51. Fujishiro K, Gong F, Baron S, et al. Translating questionnaire items for a multi-lingual worker population: the iterative process of translation and cognitive interviews with English-, Spanish-, and Chinese-speaking workers. *Am J Ind Med*. 2010;53(2):194-203.
52. Pan Y, de la Puente M. *Census Bureau Guideline for the Translation of Data Collection Instruments and Supporting Materials: Documentation on How the Guideline Was Developed. Survey Methodology 2005-06*. Washington, DC: US Census Bureau; 2005. <https://www.census.gov/srd/papers/pdf/rsm2005-06.pdf>. Accessed September 19, 2016.
53. Li RM, McCardle P, Clark RL, Kinsella K, Berch D, eds. *Diverse Voices: The Inclusion of Language-Minority Populations in National Studies: Challenges and Opportunities*. Bethesda, MD: National Institute on Aging and National Institute of Child Health and Human Development; 2001. https://www.nichd.nih.gov/publications/pubs/documents/Diverse_Voices.pdf. Accessed December 22, 2016.
54. Liu SJ, Iqbal K, Shallow S, et al. Characterization of chronic hepatitis B cases among foreign-born persons in six population-based surveillance sites, United States 2001-2010. *J Immigr Minor Health*. 2015;17(1):7-12.
55. Edwards S, Fraser S, King H. *California Health Interview Survey (CHIS) 2011-2012 Methodology Series: Report 2—Data Collection Methods*. Los Angeles, CA: UCLA Center for Health Policy Research; 2014. http://healthpolicy.ucla.edu/chis/design/Documents/chis2011-2012-method-2_2014-02-21.pdf. Accessed September 19, 2016.
56. Centers for Disease Control and Prevention. CDC listeria initiative case report form, version 2.0. 2016. <https://www.cdc.gov/listeria/pdf/listeria-case-report-form-omb-0920-0004.pdf>. Accessed September 19, 2016.
57. Centers for Disease Control and Prevention. CDC/ATSDR policy on releasing and sharing data. Updated 2005. <https://www.cdc.gov/maso/policy/releasingdata.pdf>. Accessed April 14, 2007.