WORKING PAPER #25001

February 28, 2025

HHS Data Council Workgroup Report on Data Gaps During the COVID-19 Pandemic and New Data Sources Implemented During the Public Health Emergency

Disclaimer: The views expressed in this article are those of the authors, and no official endorsement by the U.S. Department of Health and Human Services, Administration for Children and Families, Administration for Community Living, Agency for Healthcare Research and Quality, Administration for Strategic Preparedness and Response, Centers for Disease Control and Prevention, Centers for Medicare & Medicaid Services, Health Resources and Services Administration, Indian Health Service, Office of the Assistant Secretary for Planning and Evaluation, Substance Abuse and Mental Health Services Administration, or Federal Emergency Management Agency is intended or should be inferred. This report is based on interviews conducted from October 2021 to April 2022. Writing was completed in March 2023, with only minor edits applied thereafter, during final editing and clearance. Description of the "current" status of data sources and other items reflects the status when interviews were conducted. The authors did not attempt to systematically update the report after March 2023.

Table of Contents

HHS Data Council Workgroup Report on Data Gaps During the COVID-19 Pandemic and New Data Sources Implemented During the Public Health Emergency

I. Executive Summary

Many U.S. Department of Health & Human Services (HHS) Operating Divisions (OpDivs) and Staff Divisions (StaffDivs) were asked by HHS leadership, Congress, and state and local governments for data and analyses to assist in planning and implementing an effective response to the COVID-19 pandemic. Throughout HHS, data gaps hindered the responses to such requests.

To help address data gaps, the HHS Data Council constituted the Data Gaps Work Group. This workgroup was asked to identify data gaps encountered in responding to public health emergencies (PHEs) like the COVID-19 pandemic and develop recommendations for addressing them. From October 2021 to April 2022, members of the Work Group conducted 82 semi-structured interviews with people involved in the HHS response to COVID-19. Based on these interviews, the Work Group identified data gaps related to the supply side of the healthcare system, specifically data on health and human services providers. The Work Group considered gaps in data gathered from individual providers (such as doctors, nurses, and social workers) and facilities (such as hospitals, nursing homes, and childcare providers).

This report reflects the findings of the HHS Data Council's Data Gaps Work Group. It addresses (1) data requests from senior leadership, (2) data resources used in the pandemic response, (3) data gaps experienced during the pandemic, and (4) recommendations for addressing data gaps.

A. Data Requests

The first section of this report describes data requests pertaining to the pandemic received by U.S. federal government staff from federal government leadership, including the executive branch and Congress, as well as state and local governments (Table 1). Such data requests were related to four distinct areas of the response: hospital capacity and readiness, nursing homes, other healthcare providers, and human services providers. Because hospitals treated the most severe COVID-19 cases and COVID-19 outbreaks affected many nursing homes, these two types of providers received the most attention. The requests informed a variety of policy objectives, including resource allocation and situational awareness.

At the beginning of the COVID-19 response, HHS's priority was to gather data to track hospital capacity and readiness. Hospitals were critical to response efforts because they treated the sickest COVID-19 patients. Timely data were necessary to understand the availability of hospital resources to treat COVID-19 patients and inform efforts to aid states. Employees from various divisions within HHS, including the Agency for Healthcare Research and Quality (AHRQ), Office of the Assistant Secretary for Planning and Evaluation (ASPE), Administration for Strategic Preparedness and Response (ASPR), Centers for Disease Control and Prevention (CDC), and Centers for Medicare and Medicaid Services (CMS), as well as partner agencies, such as the Federal Emergency Management Agency (FEMA), were asked to gather and report hospital data. Hospital data informed many government efforts, including the allocation of resources to states and individual hospitals. Hospital data were also used to inform community guidance. COVID-19 hospitalization data were also a key input into various planning tools developed by the U.S. federal government and its partners.

The second priority of HHS's provider data collection was nursing homes. These providers care for older patients who are particularly vulnerable to COVID-19. At the beginning of the pandemic, many outbreaks of infections were concentrated in nursing homes. Timely data from these facilities were used for situational awareness and to inform resource allocations and infection control initiatives.

The COVID-19 response also required data about healthcare providers other than hospitals and nursing homes. During the pandemic, many providers faced interruptions in operations, which had consequences for patient care as well as provider finances and employment. HHS needed data to track these problems and implement responses, such as the Provider Relief Fund and telemedicine policies. Furthermore, contact tracing and vaccination efforts required participation from a wide network of providers. HHS operating and staff divisions such as ASPE, CDC, CMS, Health Resources and Services Administration (HRSA), Indian Health Service (IHS), and Substance Abuse and Mental Health Services Administration (SAMHSA) played important roles in gathering and reporting the necessary data.

Lastly, the response also required data on human services providers, which fall under the purview of the Administration for Children and Families (ACF) and Administration for Community Living (ACL). Like healthcare providers, many human services providers interrupted services during the pandemic. One area of particular focus was the availability of childcare, which was critical for employees to return to work and for the economy to rebound. Data from these providers was requested to inform distribution of the American Rescue Plan and the Coronavirus Aid, Relief, and Economics Security Act. Data were also requested to inform waivers and recommendations and to target technical assistance.

B. Data Resources Used

The second section of this report summarizes the HHS data resources used in the pandemic response (Table 2). We describe data resources for each of the four provider types previously discussed, as well as cross-cutting data on healthcare providers. The Unified Hospital Data Surveillance System (UHDSS) and the National Healthcare Safety Network's (NHSN) long-term care facilities COVID-19 module were introduced to provide high-frequency provider data during the pandemic and were among the most used resources. For other healthcare providers, the National

Plan and Provider Enumeration System (NPPES) and the Medicare Provider Enrollment, Chain, and Ownership System (PECOS) were extensively used to track provider location. Agencies also used administrative data to respond to requests for information from policymakers. HRSA collected a variety of data on Federally Qualified Health Centers and their look-alikes . Human services provider data were sparser but included important data about childcare and care for older Americans.

Efforts to collect daily hospital data started in an ad hoc manner. Within a few months, however, two main data collection methods were implemented, one through NHSN and the other eventually becoming the UHDSS. Other databases used in responding to the pandemic included the Healthcare Cost and Utilization Project (HCUP), which collects discharge-level data from hospital billing records; the National Syndromic Surveillance Program (NSSP), which includes encounter-level electronic health records (EHR) from emergency departments (EDs), urgent and ambulatory care centers, inpatient healthcare settings, and laboratories; the Coronavirus Disease 2019 (COVID-19)-Associated Hospitalization Surveillance Network (COVID-NET), which provides detailed data on hospitalizations from COVID-19; and National Center for Health Statistics (NCHS) surveys. NCHS conducts two nationally representative surveys of hospitals: the National Hospital Ambulatory Care Survey, which covers hospital EDs and outpatient facilities, and the National Hospital Care Survey, which primarily uses EHRs to track data on inpatient, ED, and outpatient hospital facilities. In addition, several other hospital datasets were used to inform the pandemic response, including the American Hospital Association Annual Survey Database, the Healthcare Cost Report Information System, and Definitive Healthcare. These datasets include hospital locations and various characteristics.

Nursing home data resources used included the National Healthcare Safety Network (NHSN) long-term care facilities COVID-19 module, a surveillance system for tracking healthcare-associated emerging infections and antibiotic resistance. Other resources used included the National Plan and Provider Enumeration System (NPPES) and the Medicare Provider Enrollment, Chain, and Ownership System (PECOS), which contains the national provider identifier (NPI) for all HIPAA-covered providers and providers who bill Medicare for their services; HRSA Program Data, which provides detailed information on the provider workforce, the Nurse Corps, and the National Service Corps; National Center for Health Statistics (NCHS) provider surveys, including the National Ambulatory Medical Care Survey (NAMCS), and the National Post-acute and Long-term Care Study (NPALS); SAMHSA data on behavioral health providers, including buprenorphine waivers, the National Substance Use and Mental Health Services Survey (N-SUMHSS), the Treatment Episode Data Set (TEDS), and SAMHSA treatment locators; and IHS data on providers that serve American Indians and Alaska Natives.

Human services provider data used included the National Survey of Early Care and Education (NSECE), which is the only nationally representative source of data on all childcare providers; Head Start data from the Head Start Enterprise System and the

Payment Management System; and ACL data on the services provided via the Older Americans Act and Titles III, VI, and VII.

C. Data Gaps

Despite the vast array of data employed by HHS, numerous data gaps hindered the response to COVID-19. These gaps applied to a wide spectrum of providers; thus, we organized them thematically (Table 3). We identified three types of gaps: (1) data collection, (2) data quality, and (3) data accessibility. Some data collection gaps were addressed during the pandemic; after an initial period of insufficient data on hospitals and nursing homes, data on these providers was improved over time. Other data collection gaps, such as data on most human services providers, were not addressed. Data quality issues had three dimensions: timeliness, completeness, and accuracy. Timeliness was particularly important in addressing policy priorities quickly, and it affected many HHS surveys. Completeness limited the usefulness of many data resources because they were not representative. Accuracy was affected by several factors, including differences in data definitions across sources and manual data entry. Data gaps related to accessibility stemmed from coordination difficulties across HHS divisions and with external partners, such as FEMA. Each HHS division has unique responsibilities, authorities, and rules governing its operations, making it difficult to share data.

D. Recommendations

In the fourth section, we provide a set of recommendations to address our identified data gaps (Table 4). Our recommendations span four areas: (1) key data elements that should be collected, (2) data management and coordination, (3) development of PHE-specific data plans, and (4) data collection technology and tools. We first identify key data elements that are necessary to mount an effective response during a PHE, including high-frequency data on hospitals and nursing homes and yearly data on provider location. Second, we provide recommendations for improving data management and coordination, with the aim of improving the quality and accessibility of data. Third, we identify elements that should be incorporated into effective PHE plans. We acknowledge that it is not possible to collect all necessary data in advance of a PHE, but an effective response requires a detailed PHE plan to quickly implement additional data gathering and distribution capabilities. We conclude with recommendations for data collection technology and tools, which can help reduce data collection costs for HHS and the burden of reporting on providers.

Our recommendations include the following:

A. Maintain and develop capabilities for gathering key data elements.
    1. Maintain existing critical data collection capabilities for hospitals and nursing homes.
    2. Gather yearly data on provider location.

    3. Gather yearly data on childcare.
    4. Gather demographic data to track differences in health outcomes across populations.
    5. Improve identification of telehealth visits.

B. Improve data management and coordination across HHS
    1. Increase communication and coordination across HHS OpDivs and with essential partners like FEMA.
    2. Develop a data collection plan that identifies which OpDivs will be responsible for collecting new data elements and recommends funding the identified OpDivs for this purpose.
    3. Establish public data sources of record for key variables.
    4. Validate data sources of record.
    5. Agree on universal data definitions.
    6. Reduce redundancies in data collection.

C. Develop a data plan to respond to additional needs during PHEs
    1. Develop a coordinated plan for communicating with providers.
    2. Develop and maintain a platform for secure data sharing across Divisions.
    3. Streamline the processes for securely sharing data within the federal government for the purposes of responding to PHEs.
    4. Improve the expedited review process during PHEs.
    5. Share data with stakeholders, including states, Tribes, territories, providers, and the public.
    6. Develop capabilities for rapid and preliminary data analyses.
    7. Include human services (especially childcare) and behavioral health in the PHE-specific data plan.

D. Modernize data collection technologies and tools
    1. Automate reporting processes through investments in technology.
    2. Provide resources to smaller providers that do not have the ability to automate reporting.
    3. Develop nationally representative sentinel networks.

4. Provide support to smaller OpDivs so they can use and modify data collection systems developed by larger OpDivs.

Table 1. Data Requests From Senior Leadership

	Request	**Data elements**	**Policy importance**
Hospitals	List of all U.S. hospital facilities	Number of hospitals, locations, and contact information	• Situational awareness regarding overall pandemic and hotspots • Resource allocation to states and individual hospitals • Community guidance from the CDC and local and state governments • Modeling initiatives such as ASPR's long-term scenarios for COVID-19
	Daily data on hospital utilization	COVID-19 patients, COVID-19 deaths, non-COVID-19 patients, and non-emergency elective surgeries	
	Daily data on hospital resources	Staffed beds by hospital department, supply of personal protective equipment (PPE) and ventilators, supply of COVID-19 therapeutics, and staffing shortages	
	Data on infection control measures	Staff vaccination rates, staff health, and triage protocols for protecting staff and patients	
Nursing homes	A list of nursing homes	Number of nursing homes, locations, and contact information	• Situational awareness regarding distribution of COVID-19 in nursing homes • Information on facilities facing shortages and prioritization of requests for supplies and assistance • Infection control initiatives (e.g., provide training on evidence-based practices to mitigate the spread of COVID-19)
	Weekly data on COVID-19 burden	COVID-19 case counts with information on type of test (PCR/antigen), person characteristics, and case severity, and deaths	
	Weekly data on resources	Count of available staff, staff shortages, PPE and supply shortages, access to supplies, length of time necessary to get COVID-19 test results	
	Data on vaccination rates	Vaccination level (i.e., 1 dose, 2 doses, or booster) for staff and patients	

Table 1. Data Requests From Senior Leadership (continued)

	Request	Data elements	Policy importance
Other Healthcare Providers	Data on Federally Qualified Health Centers and look-alikes	Number of providers and facilities, locations, contact information, specialty, and credentials and licenses	• Distribution of provider relief funds • Regulations to address interruptions in provision of healthcare (e.g., CMS telehealth waivers and policies for practicing across states) • Logistical efforts (e.g., identifying providers that could administer vaccines) • Provider-specific initiatives such as targeted training and technical assistance for Federally Qualified Health Centers
	Data on resource availability	Staff vacancies, PPE availability and shortages, supply of COVID-19 vaccines and therapeutics (distribution, utilization, and amount on hand)	
	Services delivered	Utilization (including vaccinations performed), demographics served, operational status, telehealth (capabilities, barriers, and visit counts)	
	Data on provider health and finances	Vaccination status, measures of provider stress and mental health, COVID-19 cases, deaths, and finances	
Human Services Providers	Data on services delivered	Type of services, location of facilities, operational status, utilization, and demographics served	• Situational awareness regarding operational status of human services providers • Distribution of funds from COVID-19 relief packages, such as American Rescue Plan • Inform ACF waivers and recommendations • Targeting of technical assistance
	Data on inputs required for operations	Counts of employees, employee vaccinations and mental health, and PPE availability	
	Data on provider finances	Revenues and spending of pandemic relief funds (American Rescue Plan)	
	Barriers to operation	Use of ACF waivers, social distancing, and difficulty communicating with families	

Table 2. Data Requests Used During the COVID-19 PHE Response

Panel A. Hospitals

Data Resource/ Owner	Description	Use
Unified Hospital Data Surveillance System (UHDSS) / CDC	Daily reports of capacity, utilization (including COVID-19 cases), supplies, and workforce shortages. Covers all U.S. hospitals and available within 1-2 days of collection.	Situational awareness, resource allocation, COVID-19 forecasts/scenario planning, and community guidance
Healthcare Cost and Utilization Project (HCUP) / AHRQ	Detailed discharge-level data from hospital billing records from EDs, inpatient stays, and ambulatory care. Inpatient data available on a quarterly level, while other data available annually. Covers most states.	Situational awareness, resource allocation, historical data
National Syndromic Surveillance Program (NSSP) / CDC	Encounter-level EHRs from EDs, urgent and ambulatory care centers, inpatient healthcare settings, and laboratories. Covers 71% of EDs, with coverage rate varying by state.	Detailed studies of COVID-19
COVID-19 Associated Hospitalization Surveillance Network (COVID-NET) / CDC	Detailed data on COVID-19 hospitalizations, such as patient demographics, admission dates, health conditions, and outcomes. Data from 250+ hospitals, in 99 counties in 14 states.	Detailed studies of COVID-19
National Hospital Ambulatory Medical Care Survey (NHAMCS) / CDC	Nationally representative data from hospital EDs and outpatient facilities. Annual data, with preliminary estimates available via COVID-19 dashboard.	Retrospective research
The National Hospital Care Survey (NHCS) / CDC	Nationally representative EHRs on inpatient, ED, and outpatient hospital facilities. Annual data, with preliminary estimates available via COVID-19 dashboard.	Retrospective research
Other Data	The American Hospital Association Annual Survey Database, the Healthcare Cost Report Information System, and Definitive Healthcare contain information on hospital locations and other characteristics.	Resource allocation, various other

Table 2. Data Resources Used During the COVID-19 PHE Response (continued)

Panel B. Nursing Homes

Data Resource/ Owner	Description	Use
National Healthcare Safety Network (NHSN) long-term care facilities COVID-19 module / CDC	Weekly reports of capacity, utilization (including COVID-19 cases), supplies, infection control measures, vaccinations, workforce. Questions about staff and PPE shortages, and vaccination rates in nursing homes.	Resource allocation, situational awareness, COVID-19 forecasts/scenarios, PPE demand modeling

Panel C. Other Healthcare Providers

Data Resource/ Owner	Description	Use
National Plan and Provider Enumeration System (NPPES) / CMS	National provider identifier (NPI), provider specialty, and location for all HIPAA-covered providers and providers who bill Medicare for their services.	Identification of providers and their locations for workforce analyses and resource allocation (e.g., Provider Relief Fund)
Medicare Provider Enrollment, Chain, and Ownership System (PECOS) / CMS	Billing and enrollment information for Medicare-enrolled providers.	
HRSA Program Data / HRSA	Detailed information on the provider workforce, the Nurse Corps, and the National Service Corps.	Track provider workforce and inform mobilization efforts
Health center grant application data / HRSA	Location of service delivery sites for health centers.	Resource allocation (such as vaccines), training, and health center assistance
Health Center COVID-19 Survey / HRSA	COVID-19 testing, PPE, virtual visits, vaccinations, therapeutics, antivirals, and vaccines received. Gathered weekly or every 2 weeks. Offered to all health centers and required for participants in Health Center COVID-19 Vaccine Program.	
Health Center Program Uniform Data System (UDS) / HRSA	Collects data on patient characteristics, health center personnel, quality of care, and costs. Annual data for Health Center awardees and look-alikes.	

Table 2. Data Resources Used During the COVID-19 PHE Response (continued)

Panel C. Other Healthcare Providers (continued)

Data Resource/ Owner	Description	Use
The National Ambulatory Medical Care Survey (NAMCS) / CDC	Provider characteristics and visit data from electronic records, including patient demographics, conditions, and services rendered. Annual, nationally representative sample of office-based physicians and community health centers.	Retrospective research
The National Post-Acute and Long-Term Care Study (NPALS) / CDC	Annual, nationally representative data on multiple types of providers, including adult day service centers, nursing homes, and hospice.	Resource allocation (e.g., PPE and testing supplies)
Buprenorphine Waivers / SAMHSA	Provider location for each type of buprenorphine waiver.	Track buprenorphine waivers, and treatment accessibility
National Substance Use and Mental Health Services Survey (N-SUMHSS) / SAMHSA	Annual survey of public and private substance use disorder and mental health treatment facilities. Collects data on facility characteristics.	Provider location for treatment locator database
The Treatment Episode Data Set (TEDS) / SAMHSA	Annual demographic and drug history data on individuals undergoing substance abuse treatment.	Provider location for treatment locator database
Inventory management system / IHS	Data on supplies and locations of IHS facilities.	Resource allocation (e.g., PPE, testing, therapeutics, vaccines)
Enterprise Human Capital Management System (EHCM) / IHS	Human resources management system for IHS.	Identify provider locations and vacancies
Resource and Patient Management System (RPMS) / IHS	Clinical and administrative information for healthcare facilities affiliated with IHS. Includes patient registration information and EHR.	Identify telehealth visits

Table 2. Data Resources Used During the COVID-19 PHE Response (continued)

Panel D. Cross-cutting data on healthcare providers

Data Resource/ Owner	Description	Use
Tiberius / CDC	Distribution of COVID-19 vaccines and therapeutics to providers. Included manufacturer shipments, deliveries to providers, and product utilization.	Distribution of vaccines and therapeutics
Supply Chain Control Tower / ASPR	Data on production and distribution of products from major manufacturers of supplies.	Supplies distribution for hospitals and nursing homes
Medicare and Medicaid Claims Data / CMS	Medical claims including national provider identifiers. Available for analysis at ASPE within 1-2 months of date of service.	Preliminary insight into services obtained by Medicare fee-for-service beneficiaries

Table 2. Data Resources Used During the COVID-19 PHE Response (continued)

Panel E. Cross-cutting data on healthcare providers

Data Resource/ Owner	Description	Use
National Survey of Early Care and Education (NSECE) / ACF	Nationally representative source of data on all childcare providers collected every 5 years. Data on the type of care provided, characteristics of staff and populations served, and affordability of care. COVID-19 follow-up survey collected data on the operational status of facilities, barriers to normal operations, and concerns about the safety of children, families, and staff.	Retrospective research
Head Start Enterprise System / ACF	Location of Head Start facilities, number of children served, and funding allocated for the year.	Tracking of operational status and spending of relief funds
Payment Management System / ACF	Detailed financials, including spending and grants received by Head Start facilities.	Spending of relief funds
Administration for Community Living (ACL) data / ACL	Yearly operational data on community-based centers serving older adults and persons with disabilities. Includes measures of utilization (such as the number of rides provided or meals served), funds allocated to provide services, demographics of populations served, and full- and part-time staff.	Resource allocation and distribution
Ad hoc data collection / ACF and ACL	Manually collected data on operational status of ACF grantees.	Recommend service delivery methods (i.e., in-person vs. remote), resource and funding allocation

Table 3. Data Gaps and Obstacles Encountered During the COVID-19 PHE Response

Panel A. Data Collection	
Not collected before PHE	• A definitive list of hospitals and their locations • High frequency hospital and nursing home data • COVID-specific questions in various HHS databases
Still not collected	• Data on human services providers (excluding Head Start) • Health, finances, and licensure status of most providers • Demographics on populations served • Severity of COVID-19 cases in hospitals
Panel B. Data Quality	
Timeliness	• Time-consuming approval process for new data elements • Low-frequency data collection • Long data processing times • Delayed focus on childcare and behavioral health
Completeness	• Voluntary participation in data collection • Lack of compliance with mandated reporting • Lack of regular updates • Information not collected on some types of providers
Accuracy	• Inconsistent definitions for key data elements • Errors due to manual data entry • Inaccurate facility level information submitted by owners of multiple sites • Difficulties gathering inventory data • Misreporting because of provider incentives
Panel C. Data Accessibility	
Accessibility obstacles	• Difficult to access data across OpDivs and OpDiv subdivisions • Limited visibility regarding what other OpDivs are working on leads to duplication and coordination difficulties • Restrictive data use agreements • Providers initially reluctant to share data

Table 4. Recommendations to Address Data Gaps Encountered During COVID-19 PHE Response

	Recommendation	Potential short-run Data Council initiative?
Key Data Elements	• Maintain critical data collection capabilities for hospitals (daily reporting to Unified Hospital Data Surveillance System) and nursing homes (weekly reporting to NHSN).	No; funding needed
	• Gather yearly data on provider location.	No; funding needed
	• Gather yearly data on childcare.	No; funding needed
	• Gather demographic data to track differences in health outcomes across populations.	No; funding needed
	• Improve identification of telehealth visits.	Yes
Data Management & Coordination	• Increase communication and coordination across HHS OpDivs and with essential partners like FEMA.	Yes
	• Develop data collection plan that identifies which OpDivs will be responsible for collecting new data elements and funds identified OpDivs for this purpose.	No; funding needed
	• Establish public data sources of record for key variables.	Yes
	• Validate data sources of record.	No; appropriate for OpDivs, GAO, or researchers
	• Agree on universal data definitions.	Yes
	• Reduce redundances in data collection.	Yes; Data Council may help identify redundancies
PHE-Specific Data Plan	• Develop a coordinated plan for communicating with providers.	Yes
	• Develop and maintain a platform for secure data sharing across divisions.	No; funding needed
	• Streamline processes for securely sharing data within the federal government for the purposed of responding to PHEs.	Yes
	• Improve the expedited review process during PHEs.	No; Requires departmental action.
	• Share data with stakeholders including states, Tribes, providers, and the public.	No; OpDiv action required
	• Develop capabilities for rapid and preliminary data analyses.	No; OpDiv action required
	• Include human services (especially childcare) and behavioral health in the PHE-specific data plan.	Yes

Table 4. Recommendations to Address Data Gaps Encountered During COVID-19 PHE Response (continued)

	Recommendation	Potential short-run Data Council initiative?
Data Collection Technology & Tools	• Automate reporting processes through investments in technology	No; funding needed
	• Provide resources to smaller providers that do not have the ability to automate reporting.	No; funding needed
	• Develop nationally representative sentinel networks	No; funding needed
	• Provide support to smaller OpDivs so they can use data collection systems developed by larger OpDivs	No; requires funding and new administrative infrastructure

II. Introduction

Many HHS OpDivs and StaffDivs were asked by HHS leadership, Congress, and state and local governments for data and analyses to assist in planning and implementing an effective response to the COVID-19 pandemic. Throughout HHS, data gaps hindered the responses to such requests. To help alleviate this problem, the HHS Data Council constituted a Departmental Work Group to identify data gaps encountered in responding to PHEs and provide recommendations for addressing them. The Work Group interviewed people directly involved in the HHS response to COVID-19. Based on these interviews, it identified data gaps related to the supply side of the healthcare system, specifically data on health and human services providers.

This report reflects the findings of the HHS Data Council's Data Gaps Work Group. From October 2021 to April 2022, members of the Work Group conducted 82 semi-structured interviews with people involved in the HHS response to COVID-19. The Work Group then synthesized the findings to understand how provider data were used in the pandemic response and where there were gaps in the data that hindered the response. The Work Group then developed a series of recommendations for addressing the data gaps so that the federal government might be better positioned to respond to future PHEs.

The first section of this report briefly discusses the study objectives and the methodology employed for the interviews. The second section summarizes data requests received from leadership within HHS and the broader federal government and discusses the policy importance of these requests. The third section provides an overview of the key data resources used in the response, and the fourth discusses the data gaps that hindered the response. The report concludes with the Work Group's recommendations for addressing the data gaps identified in its investigation.

III. Objectives

The objective of the Data Gaps Work Group was to develop a department-wide perspective about data on health and human services providers in the context of a PHE. The Work Group focused on data about health and human services providers to match the project's scope with its members' expertise. The Work Group defined provider data as information gathered from individual providers (such as doctors, nurses, and social workers) and facilities (such as hospitals, nursing homes, and childcare providers). Many other types of data were used during the pandemic response but were not within the scope of this project. Examples of data that were not in scope include data about manufacturers of supplies, COVID-19 case counts, and vaccine effectiveness and side effects.

The Work Group further tailored the project to focus on provider data that was needed for the HHS response to COVID-19. These policy needs typically prioritized high-frequency, up-to-date data to respond efficiently to the emergency. The data needs for PHEs can differ from those for public health or social science research, which is typically conducted over a longer period of time. Study participants were identified based on their role in the pandemic response, and the questions for the semi-structured interviews also emphasized data needs related to the pandemic response.

IV. Methods

The Work Group developed a project plan to maximize insight into provider data in the context of the HHS COVID-19 response. The Work Group first identified HHS OpDivs and StaffDivs that used provider data and were most engaged in the response. (Notably, OpDivs such as the National Institutes of Health (NIH) were excluded because they primarily work with clinical data rather than provider data.) Members of the Work Group then contacted staff in leadership roles at the OpDivs and StaffDivs to identify people who personally contributed to the response and could speak to the data needs and data gaps. An interview request was sent to the identified staff, and up to two reminder emails were sent to staff who did not respond. Staff were given the opportunity to respond to a written questionnaire if they were not available for an interview. People who participated in interviews were also asked if they had any other suggestions about who the Work Group should reach out to. Multiple people were contacted from each division, but the number of people who participated in interviews was at the discretion of that division.

Overall, the Work Group sent interview requests to 104 people, and 82 participated in the study. Three people participated exclusively via written response, while the rest participated in one-on-one or group interviews. The participants spanned 11 HHS OpDivs and StaffDivs and two organizations that partnered extensively with HHS in the response, FEMA and the Johns Hopkins University Applied Physics Laboratory (JHU APL). Because of their centrality to the response and focus on provider data, the divisions with the greatest number of participating staff were ASPR, CDC, and HRSA. Thus, the study provides a more in-depth accounting of data resources and gaps faced by those organizations. The people who participated in the study were distributed as follows:

- ACF 4
- ACL 5
- AHRQ 5
- ASPE 3
- ASPR 10
- CDC 17
- CMS 7
- FEMA 2
- HRSA 16
- IHS 7
- JHU APL 2
- HHS Office of the Secretary (OS) 3
- SAMHSA 1

Members of the Data Gaps Work Group developed a set of questions to serve as the basis of the semi-structured interview. The questions were pilot-tested for clarity before the interviews began. Some incremental changes were made after the first interview; thereafter, interviewees were asked the same set of questions. The Work Group also developed a set of probes to provide additional clarifications for the interviewees and elicit more in-depth responses. The same questions were included in the written questionnaire, but probes were listed as examples. The complete set of questions is shown in Exhibit 1.

Members of the Data Gaps Work Group included:

Joel Cohen (AHRQ, chair)
Giacomo Meille (AHRQ, interviewer)
Jessica Monnet (AHRQ, interviewer)
Shariece Evans (ACF)
Pamela Owens (AHRQ)
Jenny Schnaier (AHRQ)
Herbert Wong (AHRQ)
Krycia Cowling (ASPE)
Pranav Samineni (ASPE)
Scott Smith (ASPE)
Leremy Colf (ASPR)
Jennifer Layden (CDC)
Alexander Strashny (CDC)
Megan Walters (CDC)
Carla Hodge (CMS)
Meagan Khau (CMS)

Nancy De Lew (ASPE)
Violanda Grigorescu (ASPE)
Susan Queen (ASPE)
Rachel Zuckerman (ASPE)
Daniel Duplantier (HRSA)
Susan Monarez (HRSA)
Peter Bosse (HRSA)
Kirk Greenway (IHS)

Exhibit 1. Interview Questionnaire

1) **Please think about requests for data or information on providers, that you received from senior leadership, the Department, Congress, or others, during the COVID-19 pandemic.**
   a) What types of healthcare and human services providers did the request(s) ask about?

   b) What data were required?

   Interviewer Probes:

   a) *Providers may be organizations (i.e., hospitals, nursing homes, etc.) or people (e.g., physicians, nurses, PAs, and social workers).*
   b) *Types of data requested may include counts of providers (e.g., physicians, nurses, etc.) and provider outcomes (e.g., utilization, revenues, capacity, financing, and volumes).*

2) **a) What was the objective of the data request?**
   b) What decisions (e.g., prospective, retrospective, funding, operations, etc.) relied on the data requested?

3) **What data or data estimation methods did you use to respond to the request?**
   Interviewer Probes:

   a) *Who owns the data you used?*

4) **What were the most important limitations of the data that you used to respond to the request?**
   Interviewer Probes:

   a) *Were data elements missing?*
   b) *Were the data timely?*
   c) *Was a census required, or was a sample sufficient?*

5) **If we removed barriers to accessing data, what existing data sources (private or public) would you have wanted to use? Please explain the benefits of these data sources and the barriers to accessing them.**
   Interviewer Probes:

   a) *Would the preferred data source have addressed the problems identified in the previous question?*

6) **What should be done to address these information gaps?**
   Interviewer Probes:

   a) *Given your response, it seems as though one of the options below may have been an ideal solution.*
      a. *Amend existing data collections*
      b. *Procure from private sector sources*
      c. *Develop a new data collection*
      d. *Collaborate and coordinate current data sources to see if integrating data sources could address information gaps*
      e. *Use statistical techniques to estimate the data*
   b) *Could you rank the suggested solutions?*
   c) *How feasible are the suggested solutions?*

7) **Do you know anyone else who answered similar requests from senior leadership?**

All interviews were conducted by the same two members of the Work Group, who alternated between asking questions and taking notes. Interviews were scheduled for 30 minutes to 1 hour, with most interviews lasting 1 hour. Probes and follow-up questions were asked at the interviewer's discretion. At the start of the interview, participants were given additional information about the study's objectives and how the information would be used. To encourage participation and honest assessments of areas for potential improvements in the HHS response, interviewees were informed that their names would not be used in the report and that the report would focus on identifying broad areas for improvement rather than specific problems.

Once interviews were completed, the report was drafted by the Work Group. In limited cases, assessments of data gaps conflicted with each other. The Work Group used its best judgment to resolve such conflicts. Though participants in the study were asked to suggest recommendations, not all suggestions were adopted by the Work Group. Members of the Work Group used their best judgment to prioritize the data gaps and identify appropriate recommendations for addressing the gaps. Participants were given the opportunity to read over the report to ensure that their responses were not misinterpreted and that the report accurately represented their experiences.

V. Data Requests from Senior Leadership

During the COVID-19 pandemic, leadership from the U.S. federal government, including the executive branch and Congress, as well as state and local governments, requested data from numerous parts of HHS to aid response efforts. In general, data requests were related to four distinct areas of the response: hospital capacity and readiness, nursing homes, other healthcare providers, and human services providers.

Hospital Capacity and Readiness: At the beginning of the COVID-19 response, HHS's priority was to gather data to track hospital capacity and readiness. Hospitals were critical to response efforts because they treated the sickest COVID-19 patients. Timely data were necessary to understand the availability of hospital resources to treat COVID-19 patients and inform efforts to aid states. Employees from various divisions within HHS, including AHRQ, ASPE, ASPR, CDC, and CMS, as well as partner agencies, such as FEMA, were asked to gather and report hospital data.

Questions and data required:

Requests for hospital data covered four key areas:

1. A definitive list of all U.S. hospital facilities that included key information such as address and contacts.
2. Daily data that tracked hospital utilization by COVID-19 and non-COVID-19 patients.
3. Daily data that tracked hospital resources, including staff, equipment, and supplies.
4. Data on infection control measures such as staff vaccination, triage protocols, and staff health.

Details about the requested data elements are listed below.

- Facility information:
    - Number of hospitals
    - Locations
    - Contact information
- Utilization:
    - COVID-19 patients, total and by age group
    - COVID-19 deaths
    - Non-COVID-19 patients (with breakout for flu)
    - Non-emergency elective surgeries
- Resources:

    - Staffed beds by hospital area (emergency department, inpatient, intensive care units)
    - Supply of PPE and ventilators
    - Supply of COVID-19 therapeutics (such as Remdesivir and life support drugs for ventilators)
    - Staffing shortages
- Infection control measures:
    - Staff vaccination rates
    - Staff health
    - Triage protocols for protecting staff and patients

Policy importance: Hospital data contributed to situational awareness regarding the overall status of the pandemic, COVID-19 hotspots, and geographies where hospitals were strained. For situational awareness, COVID-19 hospitalization data had several advantages compared to COVID-19 testing data. First, hospital data were not affected by testing rates, which varied widely over the course of the pandemic and across states. Second, as the pandemic progressed, a lower percentage of cases resulted in hospitalization, and rapid home tests became more widely used, reducing the completeness and usefulness of positive case reports. Hospital data not only informed the Federal Government but was also shared with states, local governments, public health departments, and hospitals. HHS released a weekly "Governor's Report" for each state government and updated "Community Profile Reports" daily for the public.

Hospital data informed many government efforts, including the allocation of resources to states and individual hospitals. During the COVID-19 pandemic, many states and hospitals requested help from the HHS because hospital systems were overburdened. Daily hospital data informed state decisions about when to request resources from HHS and helped HHS prioritize requests for resources. Teams from ASPR and FEMA used hospital data to complement qualitative information relayed by local sources with quantitative data that could be compared across geographies and hospitals. The data informed a variety of decisions regarding the allocation of federal government resources, including deployments of National Disaster Medical Assistance Teams (a set of 50 teams of healthcare professionals which ASPR routinely deployed over the pandemic), decisions about when and whether to release resources (such as ventilators) from the Strategic National Stockpile, allocation of resources from the Strategic National Stockpile, and allocation of COVID-19 therapeutics with emergency use authorization. Data were also used to coordinate the movement of ventilators across hospitals, and data on hospital supply shortages were shared with PPE and supply manufacturers to target distribution.

Hospital data were also used to inform community guidance. For example, the usefulness of COVID-19 case counts declined in 2022 due to a lower percentage of cases resulting in hospitalization and increased use of rapid at-home tests. Thus, for a period of time, the CDC based community guidance, such as masking recommendations, on COVID-19 hospitalization rates. Local and state governments also used hospital data to inform policies such as travel restrictions, school closures, and masking requirements.

COVID-19 hospitalization data were also a key input into a variety of planning tools developed by the U.S. federal government and its partners. During the pandemic, several modeling initiatives were developed, including ASPR's long-term scenarios for COVID-19 rates, and forecasts by the CDC Center for Forecasting and Outbreak Analytics and the JHU APL. Each of these initiatives utilized data on COVID-19 hospitalizations as one of the key inputs for their models. Such initiatives helped the U.S. federal government anticipate future needs and plan its response.

Nursing Homes: The second priority of HHS's provider data collection was nursing homes. These providers care for older patients who are particularly vulnerable to COVID-19. At the beginning of the pandemic, many outbreaks of infections were concentrated in nursing homes. Timely data from these facilities were used for situational awareness and informed resource allocations and infection control initiatives. Employees across HHS were asked to gather and report nursing home data.

The HHS response required four types of data about nursing homes:

1. A list of nursing homes with locations and contact information.
2. Weekly data about the COVID-19 burden by nursing home.
3. Information about the resources that nursing homes had at their disposal, including staff and supplies.
4. Information about vaccination rates.

Details about the required data elements are listed below.

- Facility information:
    - Number of nursing homes
    - Locations
    - Contact information
- COVID-19 burden:
    - Case counts, with information on:
        - Identification method: Rapid antigen test or PCR
        - Person characteristics:
            - Nursing home staff or patients
            - Race/ethnicity
            - Vaccination status
        - Case severity

    - Deaths
- Resources:
    - Count of available staff
    - Shortages of staff
    - Shortages of PPE and supplies
    - Access to supplies, including PCR testing kits
    - Length of time necessary to get COVID-19 test results
- Vaccination rates, with information on:
    - Staff and patients
    - Vaccination level: 1 dose, 2 doses, or booster

Policy Importance: Data from nursing homes contributed to creating and maintaining situational awareness within the government and the public. HHS used these data to track the distribution of COVID-19 cases across nursing homes, monitor large outbreaks, and track the ability of facilities to respond to the PHE via employees, supplies, vaccines, and infection control measures. CMS also shared the data with the public via its website: data.cms.gov.

Nursing home data also played an essential role in resource allocation decisions. HHS used the data to understand which facilities faced shortages and to prioritize requests for supplies and assistance. Every week, the CDC team responsible for the National Healthcare Safety Network (https://www.cdc.gov/nhsn/index.html) sent a list of facilities experiencing supply shortages to ASPR, which was responsible for allocation decisions. ASPR used the data to inform a variety of allocation decisions, such as providing ventilators to facilities with shortages and sending HHS strike teams to facilities with worker shortages or large outbreaks of COVID-19.

Nursing home data also helped inform infection control initiatives. AHRQ was allocated $250 million and tasked with providing training on evidence-based practices to mitigate the spread of COVID-19. To register nursing home staff, it needed the contact information and locations of the facilities. In a parallel effort, CMS's Center for Clinical Standards and Quality (https://www.cms.gov/About-CMS/Agency-Information/CMSLeadership/Office_CCSQ) relied on the nursing home data to understand how COVID-19 spreads, develop infection control measures, and target technical assistance to nursing homes. CMS identified facilities that were experiencing surges and worked with them to implement infection control measures, such as visitation policies. It also used information on testing and disease burden to identify the best testing strategies to limit the spread of COVID-19 in nursing home facilities. Finally, CMS used the data to examine the importance of vaccinating nursing home staff to reduce COVID-19 cases. Based on the evidence about vaccine effectiveness, CMS reached out to nursing homes with low vaccination rates and eventually issued requirements for vaccinating staff.

Other Healthcare Providers: The COVID-19 response also required data about healthcare providers other than hospitals and nursing homes. During the pandemic, many providers faced interruptions in operations, which had consequences for patient care as well as provider finances and employment. HHS needed data to track these problems and implement responses, such as the Provider Relief Fund and telemedicine policies. Furthermore, contact tracing and vaccination efforts required participation from a wide network of providers. HHS divisions such as ASPE, CDC, CMS, HRSA, IHS, and SAMHSA played important roles in gathering and reporting the necessary data.

The HHS response required data about Federally Qualified Health Centers and look-alike community health centers, behavioral health providers, dialysis centers, and critical sectors of the healthcare workforce, such as nurses. The data covered four categories:

1. Location and contact information by type of provider.
2. Information about operational status and telehealth capabilities.
3. Information on resource availability
4. Information on provider health and finances.

Details about the requested data elements are listed below.

- Location and contact information by type of provider:
    - Number of providers/facilities
    - Locations
    - Contact information
    - Specialty
    - Credentials and licenses
- Service delivery:
    - Utilization, including vaccinations performed
    - Demographics served
    - Operational status (closed/open in-person/open via telehealth)
    - Telehealth
        - Capabilities
        - Barriers
        - Counts of visits
- Resources:
    - Staff vacancies
    - PPE availability and shortages
    - Supply of COVID-19 vaccines and therapeutics
        - Distribution
        - Utilization
        - Amount on hand
- Provider health and finances:

    - Vaccination status
    - Measures of provider stress and mental health
    - COVID-19 cases
    - Deaths
    - Finances

Policy importance: Data on other healthcare providers informed a variety of initiatives during the COVID-19 response. These initiatives included the distribution of money from the Provider Relief Fund (PRF), regulations to address interruptions in healthcare, logistical efforts, and initiatives specific to health centers and behavioral health providers.

HRSA was tasked with distributing the PRF, which totaled over $100 billion. The Agency identified providers who qualified for the funds, targeted providers who needed funds the most, determined the amount of the payments (based on estimates of lost revenues), and made the transactions. These steps required information on provider licensing and credentialing, location, and finances. The PRF ensured that providers did not go out of business because of disruptions in healthcare services. Timely data were required to distribute the funds quickly. Similar efforts were taken to distribute other funds from the American Rescue Plan.

HHS also modified regulations to address interruptions in the provision of healthcare. Provider data informed CMS telehealth waivers and policies for practicing across states. CMS also worked with SAMHSA and ASPE to adjust prescribing rules for behavioral health providers. Treatment of opioid use disorder was especially affected by the pandemic because of the strict rules that govern the distribution of medication-assisted treatment. Data on treatment facilities and prescribing were requested to inform changes in allowable take-home doses for medications used to treat opioid use disorder and regulations about whether providers need to be on-site to prescribe medication.

HHS also relied on provider data to inform several logistical efforts that were central to the COVID-19 response. ASPE helped states identify providers that could administer vaccines, and CDC and ASPR coordinated the distribution of vaccines and therapeutics. These efforts were conducted in partnership with HRSA's Bureau of Primary Health Care, which worked to distribute supplies (including PPE and vaccines) to health centers, and IHS, which distributed resources to Tribes and facilities that served American Indian and Alaska Native populations. IHS distributed a wide variety of resources, including supplies and manpower, which came from volunteers and Federal employees from the Commission Corps, Department of Defense (DOD), and CDC. HHS also coordinated with FEMA to set up pop-up vaccination sites, repurposing parking lots and other structures to increase vaccination rates in areas with high demand or few providers.

Other initiatives targeted specific providers. HRSA used health center facility information to provide targeted training and technical assistance to health centers.

SAMHSA-led efforts focused on opioid use disorder, which was exacerbated by disruptions in healthcare and strains on mental health during the pandemic. It created publicly accessible websites with the location of treatment centers for behavioral health conditions, including opioid use disorder (findtreatment.org and findtreatment.samhsa.org). Similarly, it used data on buprenorphine prescribing to understand the supply of these providers and consider changes to treatment policy.

Human Services Providers: The COVID-19 pandemic also impacted human services providers. These providers deliver a wide variety of services that are monitored and funded by the ACF, including childcare, foster care, child welfare, services for unaccompanied migratory children, and refugee resettlement. In addition, ACL grantees, primarily providers for older Americans in community living arrangements (such as food services, caretakers, and transportation), were also impacted by the pandemic. Leadership requested data to track interruptions in the supply of services and develop policies to address the interruptions. One area of particular focus was the availability of childcare, which was critical for employees to return to work and for the economy to rebound.

Information requests about human services providers fell into four categories.

1. Information about services delivered, such as the operational status of facilities and demographics of populations served.
2. Information about inputs required for operations, such as employees and supplies.
3. Information about finances, including pandemic relief funds.
4. Information about other barriers to operations, including take-up of ACF waivers.

The specific data elements requested are outlined below.

- Services delivered:
    - Type of services
    - Location of facilities
    - Operational status
        - Open/closed
        - Virtual/hybrid/in-person
        - Reductions in hours
    - Utilization
    - Demographics served
- Inputs for operations:
- Employees
    - Total employment
    - Employee health
        - Vaccination status of providers

        - Mental health
    - PPE availability
- Finances:
    - Revenues
    - Spending of pandemic relief funds (American Rescue Plan)
- Other barriers to operations:
    - Use of ACF waivers
    - Social distancing
    - Difficulty communicating with families

Policy importance: Human services provider data were requested to improve situational awareness and policy responses to the COVID-19 pandemic. Many providers shut down for extended periods of time during the pandemic. These shutdowns occurred because human services, such as childcare and eldercare, tend to be delivered in congregate settings and cannot be delivered remotely. For this reason, it was important to track interruptions in the services provided and barriers to operation, including PPE availability, finances, and employee health. The delivery of human services had a direct impact on children and families and an indirect impact on the labor force participation of parents and caretakers.

Data were also requested to inform policy responses to the pandemic. The U.S. federal government enacted legislation that provided funds to human services providers, and ACF was responsible for the targeted distribution of substantial portions of those funds. Billions of dollars were distributed from the American Rescue Plan and the Coronavirus Aid, Relief, and Economics Security Act (commonly referred to as the CARES Act). Data on provider finances and service interruptions were necessary to target the funds appropriately. ACF also responded to the pandemic through waivers, guidance, and technical assistance. ACF waivers and flexibilities performed a variety of functions during the pandemic, such as allowing for electronic reporting, simplifications in application processes, extensions in enrollment timelines, and relaxations in the way that some funds could be used. It also issued recommendations, such as what kinds of services can be delivered remotely. Finally, many grantees turned to the ACF for technical assistance during the pandemic because of interruptions in services and changes in regulations. ACF provided technical assistance on a variety of topics, including program eligibility, school closings, and connection with public health departments. Data on operational barriers and waiver take-up were requested to help ACF craft more useful waivers and recommendations and target technical assistance.

VI. Data Resources

Many provider data resources were used in the HHS pandemic response. In this section, we provide a brief overview of these data resources, with more in-depth descriptions of the data sources that were most used by interviewees involved in the response. The data covered hospitals, nursing homes, other healthcare providers, and human services providers. Most data were created by HHS, but some were purchased from private vendors. Some important data sources, such as the Unified Hospital Data Surveillance System (UHDSS) and the National Healthcare Safety Network (NHSN) COVID-19 module, were created specifically to aid the pandemic response. Others, such as the National Plan and Provider Enumeration System, predate the pandemic.

Hospital Data

1. Unified Hospital Data Surveillance System (UHDSS). UHDSS gathered daily information from all hospitals in the U.S. on overall utilization, treatment of COVID-19 cases, supplies, and staffing. The dataset was created during the pandemic and underwent numerous improvements over time, increasing the response rates of hospitals and quality checks for the data. Because the data were high frequency, timely, and covered all hospitals, they were used in numerous areas of the response and by multiple HHS operating divisions (OpDivs). The data informed situational awareness, resource allocation, COVID-19 forecasts and scenario planning, and policy decisions.

Efforts to collect daily hospital data started in an ad hoc manner, and multiple official memos were issued by leaders of the pandemic response asking hospitals to submit such data. Within a few months, two main data collection methods were implemented, one through NHSN and the other maintained by HHS Office of the Chief Information Officer (OCIO). To reduce confusion, minimize the burden of reporting, and improve data quality, in August 2020, CMS mandated daily reporting by hospitals, and HHS decided that all reporting should be done through the HHS/OCIO database. This database became the UHDSS.

The UHDSS collects data on utilization, COVID-19 counts, supplies, and workforce at the hospital facility level. In total, data are collected for approximately 130 variables. Hospitals are required to report daily for most variables, but for a minority of variables, hospitals report weekly or reporting is optional. The exact variables included and whether reporting was mandatory for each variable changed over time as a function of the policy needs at different moments during the pandemic. For example, information about ventilator and PPE supplies was required during the beginning of the pandemic but not by January 2022, when most shortages affecting these supplies were resolved.

HHS/OCIO initially developed the UHDSS with input from several HHS OpDivs and coordination by the HHS Data Strategy Execution Workgroup. It was eventually transitioned to the CDC, but the database remains a departmental effort, as the authority to mandate reporting comes from CMS, and the data is used extensively by ASPR. Hospitals submit data through TeleTracking software, and some hospitals have automatic feeds for reporting data elements. Some hospitals report directly to HHS, while others report to state governments, and then the relevant information is passed on to HHS. At HHS, the data is stored in HHS Protect, a system that uses Palantir Foundry[1] and updates, integrates, and distributes the data securely. The data is available to HHS within 1-2 days of the reference date. Historical facility-level data on hospital capacity, aggregated to the weekly level, is made available to the public at healthdata.gov.

For many variables, systematic quality checking of the data began in December 2020 and was incrementally improved thereafter. When we conducted interviews at the end of 2021 and the beginning of 2022, the quality-checking process was managed by a team that met daily. Outliers and suspicious data were flagged by algorithms, which suggested replacement values, primarily based on the 7-day average for the facility. The quality-checking team reviewed the suggested changes and decided which data to replace and whether to follow up with states or hospitals to check potentially problematic data. Relatively few replacements were made; out of approximately 130 elements for 6,000 hospitals, approximately 100 changes were made per day. Agencies and organizations that use the data, such as ASPR, CMS, and JHU APL, routinely perform additional data cleaning.

2. Healthcare Cost and Utilization Project (HCUP). HCUP collects discharge-level data from hospital billing records. It is a voluntary program in which states can provide data to AHRQ, which processes the data and makes it available to the public. Data collection began over 20 years ago, and as of 2021, it included ED data from 40 states and inpatient data from 48 states and Washington, D.C. HCUP creates several data products, including state-level ED and inpatient databases and databases that are nationally representative. Data include a wealth of information for each discharge, including clinical condition codes, procedure codes, and patient demographics. The exact variables that are publicly released are at the discretion of states and vary. To protect patient privacy, some variables used in the development of public databases, such as exact admission dates, are only available internally at AHRQ.

In the past, HCUP collected data from states in yearly increments. To improve the timeliness of data, HCUP began collecting quarterly inpatient files for many states during the COVID-19 pandemic. These data were used to understand hospital utilization and capacity. Key data on hospital utilization by broad condition categories (including COVID-19) and demographic groups at the state-by-month level was made publicly available on the HCUP-US website.

3. National Syndromic Surveillance Program (NSSP). The NSSP includes encounter-level EHRs from EDs, urgent and ambulatory care centers, inpatient healthcare settings, and laboratories. It is gathered automatically through the

---

[1] Foundry is Palantir's central platform for data-driven decision making and situational intelligence. It serves as the operating system that will enable effective end-to-end asset and risk management. See https://www.palantir.com/foundry-explained-get-demo/#:~:text=Foundry%20is%20our%20central%20platform,end%20asset%20and%20risk%20management.

Biosense Platform, with many data available to analysts at the CDC within 24 hours of a patient visit. The database contains records from 71% of U.S. EDs. Coverage varies widely by state, with low participation from facilities in some states like California, Iowa, Maryland, and Minnesota.

4. Coronavirus Disease 2019 (COVID-19)-Associated Hospitalization Surveillance Network (COVID-NET). COVID-NET provides detailed data on hospitalizations from COVID-19. It includes patient demographics and information about the admission, such as admission dates, health conditions, and outcomes. It collects data from over 250 hospitals, which are located in 99 counties in 14 states, and cover 10% of the U.S. population. The data are collected by trained surveillance officers using a standardized case reporting form.

5. National Center for Health Statistics (NCHS) surveys. NCHS conducts two nationally representative surveys of hospitals: the National Hospital Ambulatory Care Survey, which covers hospital EDs and outpatient facilities, and the National Hospital Care Survey, which primarily uses EHRs to track data on inpatient, ED, and outpatient hospital facilities. These surveys added questions specific to COVID-19 during the pandemic, such as whether staff contracted COVID-19 and whether hospitals had to set up emergency spaces like tents and pop-up facilities. To aid the pandemic response, NCHS released preliminary estimates from its surveys, which are typically released 1-2 years after data are collected. These data are publicly accessible.

6. Other data. Several other hospital datasets were also used to inform the pandemic response. Examples include the American Hospital Association Annual Survey Database, the Healthcare Cost Report Information System, and Definitive Healthcare. These datasets include hospital locations and various characteristics.

Nursing Home Data

1. National Healthcare Safety Network (NHSN) long-term care facilities COVID-19 module. The CDC's NHSN is a surveillance system for tracking healthcare-associated emerging infections and antibiotic resistance. During the pandemic, a COVID-19 module was added for long-term care facilities, including nursing homes. In May 2020, CMS mandated that all certified nursing homes submit data for the COVID-19 module on a weekly basis; and in September 2020, the mandate was updated with financial penalties for nonreporting. During the pandemic, the number of nursing homes reporting data to NHSN increased from approximately 3,000 to over 15,000.

The COVID-19 module consists of four parts. The Resident Impact and Facility Capacity Pathway asks about the number of beds, patients, COVID-19 patient cases, patient deaths, patient vaccinations, and facility supplies. The Staff and Personnel Impact part asks about COVID-19 cases among staff and the vaccination status of staff. The final two parts of the module ask about the use of therapeutics and report individual test results for COVID-19 with basic demographic data. The exact data elements required have changed over time based on policy needs.

Nursing homes report to NHSN via a web-based application. Within each facility, staff are provided with credentials to log in and are required to report by the end of the week. NHSN has over 90% compliance with its weekly reporting requirement. The CDC receives the data and performs quality checks before sharing data with other HHS agencies. The CMS Center for Clinical Standards & Quality conducts additional data validation checks and then reports data publicly through its website. Each week, data for the prior week is made available.

During the COVID-19 response, NHSN data informed resource allocation, situational awareness, COVID-19 forecasts, PPE demand modeling, and scenario planning.

Other Data on Healthcare Providers

1. The National Plan and Provider Enumeration System (NPPES) and the Medicare Provider Enrollment, Chain, and Ownership System (PECOS).

CMS created the NPPES in response to the Administrative Simplification provisions of the Health Insurance Portability and Accountability Act (HIPAA), which required all HIPAA-covered entities (individual providers and organizations) to have a unique identifier. NPPES contains the national provider identifier (NPI) for all HIPAA-covered providers and providers who bill Medicare for their services. It also includes provider specialty and location information. CMS releases a full NPPES downloadable file each month and maintains a searchable NPPES registry online.

All providers and suppliers that bill Medicare or provide services to Medicare beneficiaries are required to register with CMS through PECOS. PECOS contains the billing and enrollment information for each provider. Data from PECOS are used to construct the Medicare Fee-For-Service Public Provider Enrollment dataset, which is publicly available.

CMS requires providers to update NPPES and PECOS within 30 days of a change in address. Providers are periodically required to revalidate information for PECOS (many types of providers are required to revalidate every 5 years). If providers do not revalidate information, then their records may be deactivated. Despite these requirements, studies have found that the two databases are not regularly updated. For example, a 2013 Office of Inspector General (OIG) study found that 48% of NPPES records contained inaccurate information, 58% of PECOS records contained inaccurate information, and data were inconsistent between NPPES and PECOS for 97% of records (Levinson, 2013).

During the COVID-19 response, initiatives within HHS (such as the Provider Relief Fund) used the NPPES and PECOS to identify providers and their locations.

2. HRSA Program Data. HRSA Program Data provides detailed information on the provider workforce, the Nurse Corps, and the National Service Corps. To collect this data, states work with licensing boards and send surveys to providers. The resulting dataset is proprietary and cannot be shared outside of HRSA. Past surveys have tracked the number of hours per week providers spend at a given facility and the number of hours per week providers spend treating Medicaid patients. During the COVID-19 response, this data (along with data from the NPPES) was used to identify the provider workforce, state resources, and inform mobilization efforts.

3. Health Center Data. The Bureau of Primary Health Care at HRSA uses three datasets to gather and report data on health centers.

a. Health center grant application data. HRSA health center grants are awarded to improve healthcare access in underserved communities. During the COVID-19 response, HHS relied on health center grant application data to identify the number of service delivery sites and the location of these sites for the allocation of resources such as vaccines, training, and health center assistance.
b. Health Center COVID-19 Survey. HRSA's Health Center COVID-19 Survey began during the COVID-19 pandemic and was designed to capture COVID-specific information in a timely and flexible manner (that allows questions to be quickly added or removed). The survey is required of all participants in the Health Center COVID-19 Vaccine Program and is offered to any health center look-alike. The first part of the survey asks about COVID-19 testing, PPE, virtual visits, vaccinations, therapeutics, and antivirals, with some information broken down by race and ethnicity. A second part of the survey asks specifically about vaccines received through the Health Center COVID-19 Vaccine Program and is only answered by participants in the program. Initially, the survey gathered weekly data; as of August 2022, data were gathered every 2 weeks.
c. Health Center Program Uniform Data System (UDS). Health Center (awardee and look-alike) facilities are required to report a core set of performance data defined in the UDS. These data include measures representing patient characteristics (such as number served and socio-demographics), health center personnel (such as FTE counts and visits attributed to medical, dental, mental health, substance use, vision, other professional, and enabling service providers), metrics on the quality of care delivered (such as childhood immunizations, hypertension control, depression screening, cancer screenings), and costs (such as patient-related revenues, funding sources). UDS data are collected annually, and data from the previous calendar year's performance are typically released in August. During the COVID-19 response, UDS data were used to inform health center resource allocation.

4. National Center for Health Statistics (NCHS) provider surveys. In addition to the NCHS hospital surveys (previously discussed), NCHS conducts nationally representative surveys of other types of providers. Data for these surveys are typically released 1-2 years after collection. NCHS added questions specific to COVID-19 to these surveys in 2020 and released preliminary results on its website.

a. National Ambulatory Medical Care Survey (NAMCS). The NAMCS is a nationally representative sample of non-federally employed office-based physicians and community health centers conducted by the National Center for Health Statistics. This survey was first conducted in 1973 and has been conducted annually since 1989. During the NAMCS induction interview, surveyors collect information about physician and practice characteristics. Information on visits for a 1-week period is gathered through electronic records. Visit information includes patient demographics, conditions, and services rendered by the provider. During the pandemic, questions specific to COVID-19 were added to the survey. These questions asked about PPE shortages, the capability of COVID-19 testing in the office, the need to turn away patients with COVID-19, providers that tested positive, and the use of telemedicine.

b. National Post-acute and Long-term Care Study (NPALS). The NPALS gathers nationally representative data on multiple types of providers. The study (previously named the National Study of Long-Term Care Providers) has been conducted every two years since 2012. Primary data from adult day services centers and residential care communities are gathered via a survey. Administrative data is gathered from home health, nursing home, hospice, inpatient rehabilitation, and long-term care hospitals. Together, these providers comprise the post-acute and long-term care industries. Questions were added to the 2020 survey to capture the effects of COVID. Examples include the number of COVID-19 cases, hospitalizations, deaths among services users and staff, PPE shortages, shortage of testing supplies, need to turn away patients with COVID-19, use of telemedicine, and infection control strategies.

5. Data on behavioral health providers. SAMHSA collects several datasets on behavioral health providers, with a focus on providers who treat patients with substance use disorders.

a. Buprenorphine Waivers. Buprenorphine waivers allow approved providers to administer, dispense, and prescribe buprenorphine to treat opioid use disorders. The waivers require registration with SAMHSA, which maintains a database of providers, which includes their location and type of waiver (the maximum number of patients a provider can prescribe to varies by waiver type). Waiver types and the type of providers who can apply have changed over time. Most recently, the Substance Use Disorder Prevention that Promotes Opioid Recovery and Treatment for Patients and Communities Act of 2018 expanded the types of providers allowed to prescribe buprenorphine. During the COVID-19 response, data on buprenorphine waivers were used to track changes in waivers over time, accessibility to treatment, and to better understand the impact of COVID-19 on behavioral health.

b. National Substance Use and Mental Health Services Survey (N-SUMHSS), the Treatment Episode Data Set (TEDS), and SAMHSA treatment locators.
c. The N-SUMHSS is administered to all public and private substance use disorder and mental health treatment facilities and collects data on facility characteristics. In 2021, to reduce reporting burden and improve data quality, the National Survey of Substance Abuse Treatment Services and the National Mental Health Services Survey were combined into the N-SUMHSS.
d. The TEDS collects demographic and drug history data on individuals undergoing substance abuse treatment. This dataset includes the admission (TEDS-A) and discharge (TEDS-D) records for individuals 12 years and older, demographic data, and substance abuse characteristics (e.g., age at first use). TEDS-A was first collected in 1992, and TEDS-D was first collected in 2000.
e. Both the N-SUMHSS and the TEDS are updated annually. Information from the datasets feeds into treatment locator databases, which SAMHSA maintains. The treatment locator databases are searchable online, allowing people to find the nearest providers that treat substance use disorders and mental health disorders.
f. During the COVID-19 response, the above resources were used to understand the location of behavioral health providers, evaluate the impact of COVID-19 in specific geographic regions, and better understand the impact of COVID-19 on behavioral health.

6. IHS data. The IHS maintains several databases that were used to inform the pandemic response for providers that serve American Indians and Alaska Natives.

a. Inventory management system. The IHS inventory management system collects data on supplies and locations of facilities. The data are used by the National Supply Service Center, which coordinates and manages the acquisition and distribution of pharmaceuticals and medical supplies. The Emergency Management Point of Contact for each IHS area provides additional information about the supplies needed. During the COVID-19 response, IHS staff relied on these systems to inform resource allocation of PPE, lab products, therapeutics, vaccines, and other supplies.

b. Enterprise Human Capital Management System (EHCM). EHCM is the human resources management system for IHS. It provides benefits management and personnel administration for all civilian employees. IHS employees manually add and remove members of the Commissioned Corps to track the location of all providers at IHS facilities. During the COVID-19 response, IHS staff used this system to identify provider locations and vacancies.

c. Resource and Patient Management System (RPMS). RPMS is a decentralized management system for clinical and administrative information at healthcare facilities affiliated with IHS. RPMS gathers patient registration information and EHRs, supports administrative functions such as appointment scheduling and billing, and provides development and communication tools for internal users. RPMS is used by Federal facilities that IHS operates directly, as well as tribal and urban facilities. IHS has access to RPMS EHRs for the facilities that it operates directly, while tribal and urban facilities can opt to share data. During the COVID-19 response, IHS staff used RPMS data to identify telehealth visits.

Cross-Cutting Data on Healthcare Providers

1. Logistical data. HHS used a variety of logistical data to aid resource distribution efforts. One of the most important systems was used for tracking the distribution of COVID-19 vaccines and therapeutics. HHS tracked distribution through Tiberius, which became operational in September 2020 and was built on Palantir's Foundry platform. The system tracked manufacturer shipments, deliveries to providers, and product utilization. However, it did not track the demographics of populations who received vaccines or therapeutics.

A second set of logistical data was maintained by the ASPR Supply Chain Control Tower. This group gathered data on the production and distribution of products from major manufacturers of supplies, such as PPE. The data complemented reports by hospitals and nursing homes about supplies on hand and supply shortages.

2. Claims data. HHS also used claims data from Medicare and Medicaid for various analyses of providers. These claims include national provider identifiers and can be used to identify providers of specific types of medical services, such as vaccinations, and track the volume of services over time. Claims can take time to be submitted and approved. ASPE determined that the majority of Medicare Fee-For-Service claims are submitted within 3-6 weeks of the date of service. Therefore, the agency used unadjudicated claims data to provide preliminary insight into services obtained by the population covered by Medicare Fee-For-Service within 1-2 months.

Human Services Provider Data

1. National Survey of Early Care and Education (NSECE). The NSECE is the only nationally representative source of data on all childcare providers. It was fielded in 2012 and 2019 and is scheduled to be conducted again in 2024. Soon after the pandemic began, staff at the ACF added a follow-up survey that re-interviewed the 2019 panel of childcare providers to gather data on the effects of the pandemic. This was possible because NORC at the University of Chicago conducted the 2019 survey and was already contracted to gather the data for the 2024 survey. Wave one of the follow-up surveys was conducted from October 2020 to February 2021, and Wave two was conducted from October 2021 to February 2022. NORC released data from the follow-up survey in July 2023.

The survey consists of four components: a household survey of parents of children under 13, a home-based provider survey, a center-based provider survey, and a workforce provider survey of staff from center-based providers. It measures a variety of important variables, such as the type of care provided, characteristics of staff and populations served, and the affordability of care. The follow-up survey asked questions to understand the effects of COVID-19 on the industry. It covered information about the operational status of facilities and services provided, barriers to normal operations, and concerns about the safety of children, families, and staff.

2. Head Start data. Head Start programs provide services to children younger than 5 years old from low-income families. In contrast to most other providers who receive funds through block grants distributed through states or other organizations, the ACF works directly with Head Start providers. For this reason, Head Start was one of the few providers for which the ACF could gather representative, timely data.

The data came from two sources: the Head Start Enterprise System and the Payment Management System. The Head Start Enterprise System includes the location of facilities, the number of children served, and funding allocated for the year, whereas the Payment Management System tracks detailed financials, including spending and grants received. During the COVID-19 pandemic, these systems allowed the agency to track the spending of relief funds, such as those from the American Rescue Plan. In December 2020, the ACF received permission to add a data element to the Head Start Enterprise System, which tracked the operational status of facilities (open or closed, in-person, remote, or hybrid.) By the 15th of each month, grantees reported the operational status of each Head Start facility over the prior month.

3. Administration for Community Living (ACL) data. The ACL provides services to adults aged 60 or over and people with a disability who live in community settings. The agency collects a variety of data on the services provided via Titles III, VI, and VII of the Older Americans Act and the Rehabilitation Act. Data includes the number of units served (such as the number of rides provided or meals served), funds allocated to provide services, demographics of populations served, and full- and part-time staff. It is collected yearly at the state or tribal level.

4. Ad hoc human services data. For most types of human services providers, the ACF and ACL do not collect timely, high-frequency data. As a result, staff in these OpDivs commonly addressed data gaps through ad hoc data collections. Examples included data calls and surveys to estimate key metrics, such as the percentage of providers that could provide remote electronic services. Staff also gathered qualitative data via formal and informal interviews with organizations that represent groups of grantees (such as industry groups) and large grantees.

References:

Levinson, D. (2013). *Improvements needed to ensure provider enumeration and Medicare enrollment data are accurate, complete, and consistent* (Report No. OEI 07-09-00440). Office of Inspector General.

VII. DATA GAPS

In response to the COVID-19 pandemic, HHS used many data resources that tracked health and human services providers. However, even in the presence of these resources, our interviews revealed a variety of data gaps and limitations that hindered the response. These gaps applied to a wide spectrum of providers; thus, we organized them thematically. We identified three types of gaps: 1) data collection, 2) data quality, and 3) data accessibility.

Data Collection: In some cases, key data were not collected prior to the pandemic. In such cases, HHS either set up systems to collect the data or executed its pandemic response without key data.

Data that HHS began collecting during the COVID-19 pandemic. To better respond to the pandemic, HHS began collecting several datasets, including:

- A definitive list of U.S. hospitals. A definitive list of hospitals was necessary to track hospital strain comprehensively and distribute resources equitably and efficiently. Multiple lists of hospitals, including lists maintained by CMS and the AHA Annual Survey Database, are available, but they were deemed incomplete. Thus, in the beginning months of the pandemic, HHS devoted substantial efforts to developing a complete list of U.S. hospitals.
- Daily hospital data and weekly nursing home data. Prior to the COVID-19 pandemic, timely, high-frequency data on hospitals and nursing homes was not collected. Developing these critical data collection systems took time. For example, hospitals were not initially required to submit data to the Unified Hospital Data Surveillance System. After CMS mandated reporting, the response rate increased, and the data were considered reliable starting in August 2020. Over 2020 and 2021, systems for checking data quality were implemented, resulting in additional improvements.

- COVID-specific questions in various HHS databases. COVID-specific questions were added to a variety of HHS databases, including the Head Start Enterprise System, the Health Center Program Uniform Data System, and the National Ambulatory Medical Care Survey. Obtaining approvals to add such questions took several months, delaying data collection and processing efforts.

These data were ultimately very useful to the response, but setting up the data collection infrastructure during the pandemic posed challenges. First, putting the infrastructure in place to collect the data took several months, which meant that much of the data were not available until the end of 2020. Second, there was limited time to develop and test data collection infrastructure, which resulted in lower data quality at the start of data collection efforts.

Data still not collected: During interviews, HHS staff identified other federal data that were missing and not collected during the pandemic, including:

- Data on most human services providers. With the exception of Head Start, ACF collects limited data on human services providers. For example, many service providers closed or operated remotely in 2020 and 2021. Thus, the operational status of human services facilities was an important data point for policymakers. However, ACF did not track the operational status of most human services providers, which primarily receive funds via block grants to states and other grantees.
- Health, finances, and licensure status of most providers. HHS collects limited data on these subjects. Provider health was important for tracking safety during the beginning of the pandemic and workforce issues, such as burnout, during the later stages of the pandemic. Provider finances and licensure information were important for distributing emergency funds, such as the Provider Relief Fund. However, HHS only has visibility into finances for a few types of providers (such as hospitals), and licensure information is collected by states and not typically shared with HHS.
- Demographics of populations served. Most Federal provider databases do not track the demographics of the populations that the providers serve. For example, daily hospital data collected by the Unified Hospital Data Surveillance System and weekly nursing home data collected by NHSN are collected at the facility level and do not contain counts of patients by race or ethnicity. Vaccination data were similarly limited. For most data collection systems that were set up during the pandemic, the original focus was on the stress of the healthcare system and the total COVID-19 burden; therefore, they were not set up to track patient demographics. Many interviewees

reported that the lack of demographic data limited efforts to track differences in health outcomes across populations during the COVID-19 response.

- Severity of COVID-19 cases in hospitals. HHS did not have a mechanism to classify the severity of COVID-19 cases in hospitals. This issue became particularly salient in the middle and later stages of the pandemic, as COVID-19 variants with lower severity became dominant and the mortality rate decreased.

Data Quality: Quality limited the usefulness of data resources during the COVID-19 pandemic. We focus on three dimensions of quality, which affect a broad range of data resources: timeliness, completeness, and accuracy.

Timeliness: Timely data were critical to COVID-19 response efforts. Policy objectives such as distribution of relief funds, allocation of resources from ASPR and FEMA, and situational awareness required up-to-date data. During our interviews, the following timeliness data gaps were identified:

- Timely approvals for new data elements. Though HHS instituted an expedited COVID-19 review procedure, staff throughout HHS reported that approvals to amend data collections regularly required 3-6 months. Many agencies were affected, including ACF, CDC, and HRSA. For example, in early 2022, new COVID-19 therapeutics received emergency authorization from the FDA. These therapeutics were initially distributed by the federal government, and HHS added questions about therapeutic use to the Unified Hospital Data Surveillance System to help target distribution. However, obtaining approval to add these elements to reporting requirements took several months. Similar delays also affected other data collection efforts, such as the National Survey of Early Care and Education and data on Head Start providers.
- Low-frequency data collection. Data collection can be complicated and burdensome. For this reason, many data resources collect data once per year or less frequently. This issue affects many in-depth surveys, which place a high burden on the responder. In some cases, data is only collected once—when a provider registers for a particular system. This issue is particularly problematic for the National Plan and Provider Enumeration System and the Provider Enrollment Chain and Ownership System, which collect data on the specialty and location of providers. Though providers are required to report changes in location within 30 days, many providers do not follow these requirements, causing inaccuracies in the data (Levinson, 2013). Out-of-date location and contact information had important policy consequences, such as delaying distribution of the Provider Relief Fund.

- Data processing time. Many data resources require long processing times and release data 1-2 years after collection. Examples include hospital data from the Healthcare Cost and Utilization Project, Medicare claims data, EHR data collected by the National Ambulatory Medical Care Survey, and most HHS surveys. For some provider types, HHS exclusively collects data through surveys that require long processing times. Two examples are:
    - Childcare. The National Survey of Early Care and Education is the only nationally representative dataset on childcare in the United States. It was conducted in 2012 and 2019, and a COVID-19 module was fielded to the 2019 sample of providers. Wave one of data collection for the COVID-19 module was conducted from October 2020 through February 2021, and the data are expected to be released at the beginning of 2023.
    - Long-term care facilities other than nursing homes. The National Post-acute and Long-term Care Study (NPALS) is the only data source on long-term care facilities other than nursing homes. This study has been conducted every 2 years since 2012. While COVID-specific questions were added in 2020, HHS does not track these providers with high frequency, and it takes 1-2 years to release the NPALS. Long-term care providers, such as assisted living and residential care communities, disproportionately serve older adults, who have the highest risks of severe infection and death from diseases such as COVID-19.
- Delayed focus on childcare and behavioral health. During the pandemic, HHS prioritized data about hospitals and nursing homes because these providers treated many COVID-19 cases. As a result, there was a delayed focus on childcare and behavioral health, impacting the timeliness of related data collection efforts.

Completeness: Many data resources are missing data about a substantial fraction of providers. Multiple problems can reduce the completeness of data resources, including non-response to voluntary data collection efforts and the design and implementation of data collections, which may systematically exclude certain types of providers.

Problems associated with incomplete data:

- Incomplete data hindered the HHS response to the COVID-19 pandemic. Data from incomplete sources is not representative of the population of providers unless missingness is random. In practice, missingness is never completely random because it is affected by the design of data collections, their implementation, and provider characteristics that affect the probability of response. Because they are not representative, it is difficult to use incomplete data resources to inform policy objectives comprehensively and equitably.

- Incomplete data resources are less extensively used, even if they contain valuable information. For example, data from the National Syndromic Surveillance Program are not used in many settings because facilities in certain large states, like California and Colorado, have low participation rates. COVIDNet, which contains detailed data on COVID-19 cases for 14 states, is not used in a wide range of settings for the same reason.
- Imputation of missing data requires assumptions about the statistical properties of the missingness and is burdensome. In some cases, it is possible and appropriate to statistically model and impute missing data. However, decisions about whether to impute such data must balance costs and benefits. Imputing data requires a team of people with expertise in statistical methods, and even the most rigorous estimates will not be as reliable as observed data. With these caveats, it may be worthwhile to impute data in cases when the agency that collects the data has information that can help impute missing data more precisely but cannot be publicly released for reasons such as privacy protection. Data imputation can also be useful when the data is widely used, and multiple downstream users will need to impute data if it is not imputed by the agency that releases it. For example, multiple groups, including staff at ASPR and the JHU APL, reported that they developed algorithms to impute missing data from the Unified Hospital Data Surveillance System.
- Overlapping data collection efforts can exacerbate missing data problems. When multiple agencies collect data from the same providers, response rates may drop if participation is voluntary. For example, CDC recently began collecting data from substance abuse treatment centers, which can elect to contribute data to the Treatment Episode Data Set maintained by SAMHSA. As a result, participation in the Treatment Episode Data Set declined.

Reasons for incomplete data:

- Voluntary participation in data collection. Many data collection efforts do not mandate participation and are affected by voluntary participation. Examples include the Treatment Episode Data Set and the Health Center COVID-19 Survey. Additionally, at the start of the COVID-19 pandemic, participation in the daily hospital and weekly nursing home data collections was voluntary. Several months after the pandemic began, CMS mandated reporting for hospitals and nursing homes, increasing the reliability of these data resources.

- Lack of compliance with mandated reporting. After reporting was mandated for the Unified Hospital Data Surveillance System, the proportion of hospitals that did not report ranged from 1%-5% each day.
- Lack of regular updates. The National Plan and Provider Enumeration System (NPPES) does not include providers without a National Provider Identifier (NPI). Examples of providers without NPIs include behavioral health and social services providers, nurse practitioners, and physician assistants. Moreover, many providers do not regularly update their information in the NPPES and the Provider Enrollment Chain and Ownership System (PECOS), so it can be challenging to identify the facilities in which providers work. These limitations make it difficult to use the NPPES and PECOS to track the location of providers. This data gap substantially affected efforts to allocate the Provider Relief Fund.
- Information not collected. Data on behavioral health providers and behavioral health facilities are especially lacking. This affected efforts such as the allocation of PPE to behavioral healthcare providers and distributions of the Provider Relief Fund. Some of the limitations related to behavioral health providers and behavioral health facilities include:
    - Behavioral health providers are not included in NPPES, and there is no national credentialing for many specialties.
    - Many patients receive care at Institutions for Mental Disease (IMDs), psychiatric hospitals, or other residential treatment facilities that have more than 16 beds. Medicaid does not pay for treatment at IMDs; therefore, data on these providers is sparse. Data about IMDs varies by state, with some states not collecting any information. The American Hospital Association (AHA) is one of the only data sources that tracks inpatient psychiatric beds.

Accuracy: Inaccuracies also reduced the usefulness of provider data during the COVID-19 response. HHS staff identified a wide variety of accuracy problems, including varying definitions of key data elements and data entry errors:

- Standardized data definitions. Definitions of key data elements can vary by state and within the Federal Government. For example, COVID-19 hospitalizations reported by Arizona and the Unified Hospital Data Surveillance System (UHDSS) often conflict due to a difference in the definition. Similarly, NHSN’s nursing home modules and the UHDSS sometimes use slightly different definitions for similar concepts. Such differences in definitions reduce the comparability of data sources.

- Manual data entry. In many instances, providers manually submit data, introducing a variety of errors. Such errors even affect some high-frequency data collections, like NHSN. While some nursing homes use a software vendor to query facility records and automatically upload data, others (usually smaller nursing homes) manually enter data. Examples of manual data entry errors include:
    - Transcription errors.
    - A respondent may not update a question answer, especially if it is binary.
    - Staff turnover may change how questions are interpreted.
- Corporate data entry. To reduce response burden, corporations that own many facilities may respond to NHSN nursing home modules on behalf of their facilities. At times, the corporation does not have complete information about their facilities, leading to erroneous data submissions.
- Inventory data. Inventory data were critical for distributing supplies, especially given localized COVID-19 hotspots, newly introduced therapeutics and vaccines, and supply chain disruptions. Despite the importance of these data, they were difficult to gather. Several data collection efforts were affected, including the following:
    - Hospital inventory data. Hospitals were asked to provide data on supplies via the Unified Hospital Data Surveillance System (UHDSS). Hospitals reported that it would be difficult to provide the exact number of supplies in their inventories; thus, UHDSS questions typically asked whether hospitals were experiencing or anticipated a shortage. Because these answers did not provide precision about the amount of supplies, HHS modeled inventories using distribution data from manufacturers and burn rates that estimated the speed at which supplies were used. However, comparing such data across hospitals remained difficult because hospitals employed different conservation strategies to reduce burn rates and had varying contingency plans for shortages.
    - IHS inventory data. During the COVID-19 pandemic, IHS worked with federal partners, including other HHS agencies, DoD, and FEMA, to distribute supplies to tribal and urban healthcare facilities. To aid these efforts, IHS was asked to provide inventory data on various supplies at the facility level. At the beginning of the COVID-19 pandemic, IHS had

an antiquated inventory management system, which made it difficult to share data, identify inventory levels, and identify the location of all facilities. Currently, IHS is upgrading the inventory management system.

- Misreporting because of provider incentives. Provider incentives can bias responses to data collection efforts. Such problems arise because a certain response is desirable for regulatory purposes or interactions with the U.S. federal government. For example, ASPR used the Unified Hospital Data Surveillance System to help make allocation decisions for hospitals. Thus, there was an incentive for hospitals to report that they experienced staffing shortages or high utilization relative to bed counts.

Data Accessibility: In many cases, data were collected, but data sharing constraints hindered COVID-19 response efforts. The following limitations were identified by HHS staff:

- Constraints related to sharing data within HHS. Data sharing within the department is complicated because HHS has 11 Operating Divisions (OpDivs) with unique responsibilities, authorities, and rules governing their operations. Because of these differences, HHS OpDivs often act as distinct units rather than as parts of a single organization. Staff from one OpDiv often do not know what data is collected in another OpDiv, leading to coordination difficulties, such as duplicative data collection efforts. Even when staff are aware of data in another OpDiv, the data may not be shareable because of restrictions related to privacy or terms of data use agreements. If data can be shared, there are often hurdles, such as project review procedures and monetary costs to share the data. These data-sharing constraints can hinder even the best-intentioned response efforts.
- Constraints related to sharing data between HHS and external entities. In some cases, providers were reluctant to share data because they were concerned about proprietary information or punitive consequences. For example, nursing homes worried that providing data about infection control measures or COVID-19 training could result in punishments for noncompliance, and hospitals worried that their supplies might be reallocated. Similarly, manufacturers of supplies such as PPE were concerned about sharing confidential data on the production and distribution of supplies. Generally, such hurdles were overcome via reporting mandates or data use agreements. In such cases, HHS must weigh the cost and benefits of restrictive data use agreements, which enable it to collect additional data but limit how the data can be shared within the department and with the public.

VIII. RECOMMENDATIONS

Our study revealed data gaps that were significant for policy purposes and affected the COVID-19 response. Based on these gaps and the suggestions from the interviews we conducted, we developed a set of recommendations described below. Our recommendations span four areas: key data elements that should be collected, data management and coordination, development of PHE-specific data plans, and data collection technology and tools.

Key Data Elements: To successfully respond to future PHEs, the federal government should maintain and develop capabilities for gathering key data elements. These data elements, known as essential elements of information in the emergency management literature, capture information that is critical to coordinating an effective operational response to a PHE. The requests for data, creation of data assets, and data gaps encountered during the COVID-19 pandemic highlight the importance of the data collection capabilities that follow. While the causes of a PHE, and thus the specific data required to mount an effective response, can vary, there is a core set of data that is necessary and widely applicable. Developing and maintaining the capabilities to collect and use these data will ensure that they are available at the start of a PHE and can be immediately utilized to mount an effective response. We recommend taking the steps below to address these needs.

1. Maintain critical data collection capabilities for hospitals and nursing homes. During the COVID-19 pandemic, HHS developed regulatory authorities and data infrastructure to collect daily hospital data through the Unified Hospital Data Surveillance System and weekly nursing home data through the National Healthcare Safety Network. (The previous Data Resources section describes these data capabilities in more detail.) Setting up these surveillance systems was costly and time-consuming, but nearly all HHS staff interviewed agreed about the importance and eventual success of the data collection efforts. Therefore, we recommend maintaining and improving these systems so that they can be utilized to respond to future PHEs.

As the U.S. federal government mounted its pandemic response, it immediately realized that it needed timely data on hospitals and nursing homes. These data are necessary because hospitals treat the highest severity patients, and nursing homes are a congregate environment for older Americans who are in poor health. For the vast majority of PHEs, these providers serve critical functions. Staff across HHS emphasized the importance of this data for the COVID-19 pandemic and other regularly occurring disasters, such as hurricanes and wildfires.

To maintain and improve these data collections, HHS needs to be thoughtful when adjusting the specific data elements collected. The Department must be able to reduce the number of data elements collected during periods in which the country is not experiencing a PHE and quickly amend data collection efforts when necessary. Improving the flexibility of these data assets will reduce the burden on hospitals while maintaining critical data collection capabilities.

In our interviews, there was near-universal consensus that HHS must thoughtfully reduce the number of data elements collected during periods when the U.S. is no longer in a PHE. When we finished conducting interviews in early 2022, the UDHSS required hospitals to submit daily data on approximately 130 elements, some of which were specific to COVID-19. This set of data elements was developed and adjusted throughout the pandemic based on the perceived importance of specific data elements at different points in time. Going forward, HHS should identify a reduced set of data elements that are widely applicable to PHEs and need to be regularly reported. To do this effectively, it must partner with experts and stakeholders. For example, it should develop a more in-depth understanding of which metrics effectively measure hospital stress. Identifying stressed hospitals can be difficult because hospitals contain costs by adjusting staffing to meet utilization. Thus, for many hospitals, the ratio of utilization to staffed beds is regularly above 80%.

The ability to quickly amend these data collections is also critical. If the number of data elements that are regularly collected is reduced, then HHS must maintain the capability to add data elements during PHEs. Flexible data collection is also important because the data elements of interest may vary based on the type of PHE. For example, data on ventilators were very important during COVID-19 because it was a respiratory disease, but it may not be required for an emergency such as a hurricane or a non-respiratory disease. Similarly, data needs change as PHEs progress. Throughout the COVID-19 PHE, the needs of hospitals changed over time, shifting from ventilators and PPE to therapeutics (as they were developed), and staff (as the U.S. faced labor shortages in the healthcare sector).

2. Gather yearly data on provider location. As discussed in the data gaps section, yearly data on provider location are not currently available; however, such data were often requested during the pandemic for initiatives ranging from the Provider Relief Fund to contact-tracing efforts. The national provider identifier (NPI) system is the most comprehensive current system that collects data on different types of providers, but it has a number of significant drawbacks. One of the primary problems is that providers often do not comply with the requirement to update their location if it changes. Thus, the data is not only out of date, it does not provide a census of provider locations at any point in time. Additionally, providers may have multiple NPIs, and some types of providers, such as nurses or behavioral health providers, are not required to register for an NPI at all. Additionally, NPIs are not consistent with other commonly used identifiers, including DEA numbers, buprenorphine waiver IDs, and CMS Certification Numbers. These drawbacks make it very difficult to use the current system to analyze provider location.

Numerous solutions are possible for gathering yearly data on healthcare provider locations. Several interviewees suggested that amending the NPI system would be the easiest to implement rather than introducing a new identifier. If that approach is taken, the first step should be to enforce requirements to update provider location, possibly by requiring providers to resubmit this information once per year. Subsequent steps could be taken to reduce duplicate NPIs, require NPIs for all healthcare providers, and either use NPIs throughout HHS or create crosswalks to other identification systems.

Various solutions are also possible for human services providers, whose locations are largely not tracked. One suggestion was to require ACF grantees to submit and regularly update a list of facility locations. However, interviewees noted that the vast majority of funds distributed to grantees are used to provide services. Thus, the allocation of additional funds specifically tied to data modernization and reporting, or requirements that a certain percentage of funds are used for data modernization, may be necessary to collect location data for these providers.

3. Gather yearly data on childcare. Childcare is a policy priority both during PHEs and during other times. It is also an important part of the HHS portfolio of providers, with the ACF allocating $19 billion in 2022 (30% of its total budget) to childcare and Head Start. Despite the importance of the childcare industry, the federal government does not collect representative yearly data on providers. As described earlier, The National Survey of Early Care and Education (NSECE) was conducted in 2012 and 2019, and there are plans to conduct it in 2024. A COVID-19 module was fielded to the 2019 sample, but the infrastructure to quickly process the data does not currently exist; thus, the data was not released until July 2023. Gathering yearly data on childcare and building the infrastructure to quickly release it should be a priority for HHS. Possible solutions include conducting the NCESE annually or partnering with other agencies like the Census or BLS to field a streamlined annual survey.

4. Gather demographic data to track differences in health outcomes across populations. Data on the race and ethnicity of patients served was not collected in most provider databases. Interviewees commonly reported this as a data gap that limited their ability to track differences in health outcomes and the federal government's resource allocation across populations. While race and ethnicity data are collected in individual surveys, such as the National Health Interview Survey and the Medical Expenditure Panel Survey, individual-level surveys usually take 1-2 years to be released. Race and ethnicity are also tracked in data such as the National Syndromic Surveillance Program, but this data is not available in all states. The CDC's mortality records are one of the few datasets that collects this information for all states and quickly releases it.

In a PHE response, it is essential that HHS is able to track differences in health outcomes across populations, without waiting for 1-2 years to observe outcomes other than death. Providers should be asked to submit data on the demographics of the patients they serve, such as race and ethnicity. Provider-level surveys are processed much more quickly than individual-level surveys, and these surveys can be used to track whether PHE responses are equitable and responsively serve vulnerable populations.

5. Improve identification of telehealth visits. The COVID-19 pandemic highlighted the importance of telehealth, which was used to treat many patients when in-person visits were limited or suspended. Despite the importance of telehealth, many current data collection projects do not have the ability to differentiate between telehealth and in-person visits. Investments should be made to measure utilization and access to this service, the importance of which is growing quickly as technological improvements increase its availability.

Data Management and Coordination: Improvements in data management and coordination across HHS are necessary to address many of the data gaps identified by this study. Many of the recommendations in this section have been previously identified but have been difficult to implement. Implementing these changes will require input and support from the OpDivs that constitute HHS via mechanisms such as the HHS Data Council, the HHS Evidence and Evaluation Council, the HHS Data Governance Board, and the Chief Information Officer, as well as coordination and direction from the Office of the Secretary. Due to the unique responsibilities, authorities, and rules governing their operations, HHS OpDivs often operate in a siloed manner. While HHS made substantial progress in coordinating and collaborating across OpDivs during the pandemic, implementing the suggested changes will be difficult and require further progress in coordination.

1. Increase communication and coordination across HHS OpDivs and with essential partners like FEMA. As discussed above, improving coordination and collaboration among HHS OpDivs is essential to efficiently achieve the department's goals and improve its response to PHEs. In HHS, data are collected by specific OpDivs or subgroups of OpDivs. HHS staff often do not know what data is collected by OpDivs outside their own division or have access to it. In cases where data is shareable across OpDivs, staff must apply and sometimes pay a fee to use data gathered by another OpDiv. Sometimes the easiest way to access data from another OpDiv is via a trusted contractor who works with both OpDivs. The barriers to sharing data across HHS increase the importance of coordinating across the department and with other government agencies, such as FEMA, which are key partners during PHE responses.

Numerous changes could be made to improve communication and coordination at HHS. We propose three specific recommendations:

a. Use interagency working groups to coordinate strategic operations. A number of interagency work groups, such as the HHS Data Council and the Data Strategy Execution Workgroup (DSEW), exist to facilitate communication across HHS. Such interagency groups must meet regularly and be empowered to coordinate strategic operations. Our interviews suggest that DSEW, which was responsible for coordinating the HHS response, was very successful during the COVID-19 pandemic. This interagency group was primarily composed of staff from ASPR, CDC, CMS, and the Office of the Secretary. During the first few months of the pandemic, it also included staff from FEMA. Among other things, DSEW was responsible for managing and sharing the data resources created during the COVID-19 pandemic, such as the Unified Hospital Dataset Surveillance System. Staff reported many difficulties working across the OpDivs that were most involved in the response, especially during the beginning months of the pandemic. DSEW was praised for increasing communication between the ASPR, CDC, and CMS. As of May 2022, DSEW continued to operate with widespread support from these OpDivs. Going forward, this group and other similar groups should continue to be empowered, and attention should be given to integrating staff from other parts of HHS, like HRSA, which was essential to the response but was not as integrated in DSEW.

b. Develop or improve a searchable database of HHS databases. Given the large size of HHS and the siloed relationship of OpDivs across the department, it is important to maintain a searchable database that lists the data elements collected in various datasets. The importance of such a resource was evidenced in the first months of the pandemic when key staff kept and developed multiple Excel databases to summarize the data resources available for the response. Multiple attempts have been made to develop this kind of database, with the latest effort being the website "healthdata.gov." Detailed descriptions for several of the key databases developed during the pandemic, like the Unified Hospital Dataset Surveillance System, are available on this website. However, descriptions of other databases that pre-date the pandemic are often overly general, incomplete, or do not have working links. Therefore, regular updating and additional information about key databases are necessary to make this resource useful for discovering data assets collected by HHS.

c. Coordinate messaging from leadership to emphasize the importance of sharing data. To reduce the barriers to collaborating across HHS, leadership needs to develop clear messaging that emphasizes the importance of sharing data. During interviews, staff noted that the willingness to share data varied by OpDiv and over time. Differences in messaging across the leadership of OpDivs and presidential administrations explained some of this variation, emphasizing the importance of leadership for sharing data. Interviewees also noted that OpDivs have incentives not to share data, such as maintaining control over data resources and increasing the importance of their OpDiv. Without clear guidance from leadership, these incentives act as barriers to sharing data. Legal reasons sometimes prevent data from being shared across agencies, but guidance could be developed to recommend that OpDivs share data whenever it is legally possible. Alternatively, if data sharing is not possible, OpDivs should have a stand-ready capacity to provide aggregated statistics or information from their data sources to the requesting agencies during PHEs.

2. Develop a data collection plan that identifies which OpDivs will be responsible for collecting new data elements and recommends funding the identified OpDivs for this purpose. Gathering the key data elements identified in the first recommendation section will require new data collection efforts. HHS must convene the relevant OpDivs to identify which ones are best positioned to collect the new data elements, assign responsibility to them, fund them, and track their progress. Unless a data collection plan is developed, new data resources will not systematically address the identified gaps, resulting in data gaps that remain unresolved and others that are addressed in duplicative fashions.

3. Establish public data sources of record for key variables. Establishing data sources of record is essential for crafting swift policy responses. At times, multiple data sources measure similar metrics, and a preferred source must be identified and agreed to. Data sources of record must also be made publicly available. If not, people and organizations will gather alternate data. Having a single publicly available data source of record is essential so that stakeholders work with the same information and are able to coordinate their responses. For example, in the early months of the pandemic, the CDC's COVID-19 case counts and hospitalizations were not widely available. As a result, competing data sources were developed by *The New York Times* and JHU APL. Operational meetings were at times stymied by discussions about differences between the data sources, slowing response efforts.

4. Validate data sources of record. Establishing a data source of record does not mean that other data sources should be abandoned. On the contrary, efforts to validate data sources of record are important to establish their credibility and identify potential improvements.

For example, data on COVID-19 hospitalizations gathered by the National Syndromic Surveillance Program (NSSP) and Healthcare Cost and Utilization Project (HCUP) should be used to validate and improve the data collected by the Unified Hospital Data Surveillance System. The Unified Hospital Data Surveillance System has important advantages, such as collecting daily data from all hospitals in all states and releasing the data nearly immediately. The NSSP and HUCP data come from EHRs and billing records, which may be less prone to data entry errors and useful for validation.

As a second example, HHS should solicit input on how best to measure staffing shortages and other measures of hospital stress using surveys. The Unified Hospital Data Surveillance System required that hospitals submit responses about measures of hospital stress, but it was set up rapidly during the pandemic and was the first survey to attempt to measure these outcomes. It may be possible to improve the metrics that are collected.

5. Agree on universal data definitions. During the pandemic, different jurisdictions and organizations collected similar data elements using varying definitions. These differences made it difficult to compare data across sources. The U.S. federal government should lead the discussion to standardize data definitions, with input from HHS OpDivs, other federal departments, state governments and divisions of public health, Tribes, and providers. One important example is how to gather data on the number of available hospital beds: one approach is to subtract occupied beds from total beds, while another is to measure available beds directly. During the pandemic, the approach varied by state, making it difficult to compare some states to each other and prioritize the distribution of scarce resources.

6. Reduce redundancies in data collection. Redundancies in government data collections can occur for several reasons. In some cases, a single provider is required or asked to respond to the same question via different surveys or reporting mechanisms. (This can happen if different agencies collect the data for different purposes, and the data are not easy to share.) More often, the provider is asked to respond to several different questions on different surveys. Another possibility is that a provider must submit daily or weekly responses to a question, but the response rarely changes.

In some cases, data sources should be consolidated to reduce such redundancies. This is important because redundancies increase the burden of data submission for providers. Data submission is costly to providers, requiring them to devote resources to data collection rather than patient care or client service. In addition, data quality may be reduced because of redundancies: providers that face a high reporting burden may not respond to optional surveys or may provide low-quality responses to required surveys. Interviewees noted that providers, including nursing homes, substance abuse treatment centers, and Federally Qualified Health Centers reported burdens from submitting similar information to multiple government data collection systems.

PHE-Specific Data Plan: During a PHE, the U.S. federal government needs additional data to facilitate its response. For example, during COVID-19, senior leadership needed timely data to inform resource allocation, situational awareness, and infection control policies. However, it would be costly to gather data to address all the anticipated needs during a PHE and impossible to predict all the data required. Thus, a PHE-specific data plan is required to expand data collections quickly and facilitate access to data. The data plan should include protocols and platforms that are ready to be implemented immediately on the first day of a PHE. We recommend taking the following steps:

1. Develop a coordinated plan for communicating with providers. During the COVID-19 pandemic, multiple authorities within the federal government and HHS sent providers memos outlining procedures for submitting data. Interviewees stated that the procedures outlined in these memos occasionally conflicted, resulting in confusion among providers regarding how data should be submitted and concern about the consequences of not submitting data. As a result, some providers reported data using multiple pipelines, resulting in duplicative records in the Unified Hospital Data Collection System. A coordinated communication plan would prevent unnecessary confusion among providers, reduce data reporting costs among providers, and improve data quality by reducing duplicative records.

2. Develop and maintain a platform for secure data sharing across divisions. Coordinating an effective response during a PHE requires increased data sharing across OpDivs and StaffDivs. For this reason, HHS should develop and maintain a platform to securely share data across the department.

During the COVID-19 pandemic, HHS developed HHS Protect, a platform based on Palantir Foundry technology that was used to share and store hospital and nursing home data. HHS Protect worked well because it fulfilled three necessities for a data sharing platform: (1) access control, the ability to restrict who can access the data and track who entered data, (2) data pipelines that showed transformations and tracked data sources, and (3) a cloud-based infrastructure that enabled an Application Programming Interface (API) to retrieve data. Interviewees were agnostic toward continuing to use HHS Protect or a different platform. However, they noted that HHS Protect successfully fulfilled the requirements, and that any other system would also have to fulfill them.

In thinking about future data platforms, HHS should consider its relationships with contractors carefully. Several interviewees expressed concerns that the U.S. federal government doesn’t own the intellectual property for HHS Protect (and similar platforms). While the Palantir Foundry technology has worked well, staff were concerned that the government would no longer be able to access the data at the end of the contract period. In the future, platforms should be built so that the capabilities can be fully transferred to the U.S. federal government. This stipulation would require contractors to document their work and for the platforms to be run on government networks.

In the same vein, HHS currently uses three distinct data platforms (HHS Protect, Tiberius, and DCIPHER) that are based on the Palantir Foundry technology. Our understanding is that the government pays for these services separately. We recommend looking into the possibility of streamlining these platforms and future platforms.

3. Streamline processes for securely sharing data within the federal government for the purposes of responding to PHEs. Procedures and regulations can make data sharing slower or impossible. We recommend three steps to streamline such processes and make data sharing more efficient during a PHE. If data sharing is not possible, OpDivs and StaffDivs should have a stand-ready capacity to provide aggregated statistics or information from their data sources to the requesting agencies during a PHE.

a. Amend data use agreements (DUAs) where possible. At times, DUAs prohibit divisions from sharing data with each other. For example, restrictions on HCUP data prevent AHRQ from sharing analyses that contain hospital IDs with other OpDivs and StaffDivs. The federal government should explore whether current and future DUAs can be modified so that data can be internally shared for the purposes of responding to PHEs.
b. Develop clear governance structures and procedures for access to data-sharing platforms. For example, the governance structure of HHS Protect was not always clear. Thus, some HHS staff outside of ASPR, the CDC, and CMS did not know who to contact or how to gain access to the platform. Regardless of the data platform that HHS chooses for the future, it will require a clear governance structure and documented procedures for accessing the platform that HHS staff can easily access.
c. Streamline application request processes. The Evidence Act mandated the creation of a streamlined application process for restricted government data. An HHS Data Council workgroup has proposed a Standard Application Process for accessing data that are not publicly available. The Standard Application Process and the process for reviewing and accepting applications would be the same across OpDivs. Phase one of the pilot was completed, and phase two began in March 2022. The Data Council should support this effort and consider whether further simplifications could be made to share data within the department during PHEs.

4. Improve the expedited review process during PHEs. HHS instituted an expedited review process to approve COVID-19-related data collections. However, our interviews revealed that the current expedited approval process is poorly understood. Moreover, it only applied to surveys that exclusively asked about COVID-19 and did not apply when COVID-19 questions were added to existing surveys. Several agencies, including ACF, CDC, and HRSA, reported that the review process hindered efforts to amend existing surveys in a timely fashion and that the expedited process was slower than the regular one. For example, during the COVID-19 pandemic, ACF added questions to their Head Start Enterprise System to identify the operational status of facilities (in-person, hybrid, or closed). It took several months to get approval to add these questions. The CDC experienced similar delays when adding new COVID-19-related questions to the National Ambulatory Medical Care Survey. These delays contributed to lag times associated with these data, and during a PHE, the timeliness of data is critical.

5. Share data with stakeholders including states, Tribes, providers, and the public. HHS should create a plan to share data sources of record with stakeholders. In addition to reducing redundant data collection efforts (discussed previously), sharing data with stakeholders can also improve the quality of data. For example, interviewees noted that some HHS agencies were perceived as “black holes” for data: stakeholders submit their data without clearly understanding why the data are being collected or how their data are being used, and the final datasets are not shared with them. Interviewees suggested that the department should aim to replicate the weather service model, where stakeholders are incentivized to share high-quality data because it helps to create a valuable product that they can use. Agencies such as IHS, which does not mandate reporting for most of their data, regularly use such models to build trust and engagement.

6. Develop capabilities for rapid and preliminary data analyses. Sharing raw data is not always possible due to privacy concerns and regulations, nor is it always advisable when data must be cleaned and processed. Given these limitations, HHS should explore capabilities for rapid and preliminary data analyses.

There are several examples of successful efforts that were implemented during the pandemic. AHRQ’s HCUP team acquired quarterly hospital discharge data (instead of yearly data) from states to track hospital utilization. This data allowed it to release some data and analyze hospitalizations and ventilator use with a limited time lag. Similarly, ASPE leveraged contemporaneous Medicare Fee for Service claims data to track utilization 3-weeks after the service date. Such data were used to track ventilator utilization and develop COVID-19 risk scores. NCHS also released preliminary data from multiple provider surveys and made them available through their COVID-19 Dashboard.

Another possibility could be to create rapid-response teams within each OpDiv, which are responsible for handling information requests from senior leadership during PHEs.

7. Include human services (especially childcare) and behavioral health in the PHE-specific data plan. In the beginning months of the COVID-19 pandemic, the U.S. federal government prioritized efforts to gather timely data from hospitals and nursing homes. The effects of the pandemic on human services and behavioral health providers were largely not considered until the end of 2021. These types of providers play an important role in most PHEs. An effective response effort should include them from the start so that data gaps and policy shortcomings are not discovered down the line.

For example, 6-8 months after the COVID-19 pandemic began, senior leadership started requesting information on the pandemic’s effect on childcare. Leadership was interested in understanding the impact of COVID-19 on early childhood education to distribute relief funds and identify which providers needed support to resume services. Except for Head Start grantees, ACF largely did not have the data infrastructure to answer these requests.

Similarly, the needs of behavioral health providers were initially overlooked by senior leadership. For example, behavioral providers were not initially included among essential providers, despite providing in-person care. Therefore, they received masks and other PPE later than other providers.

<u>Data Collection Technology and Tools:</u> Modernizing data collection technologies and developing a diverse set of collection tools is essential. These efforts can help reduce data collection costs for HHS and the burden on providers. They must also consider the differences in size and resources across providers. Our study identified four recommendations in this area:

1. Automate reporting processes through investments in technology. HHS should reduce the reporting burden for large providers by incentivizing investments in technology. Many large hospital systems used automated reporting mechanisms to report data, but not all large providers had the resources to make such investments. For example, safety net hospitals typically have limited funding for making technology investments, even though they may be cost-saving in the long run. For medium-sized providers, the government should also consider working with EHR vendors and other private data management systems to help develop tools that providers can easily implement for automated reporting. Automated reporting will reduce the reporting burden while increasing the reliability and timeliness of data.

2. Provide resources to smaller providers that do not have the ability to automate reporting. To effectively track the COVID-19 pandemic, identify hot spots, and help allocate resources, HHS required daily data reporting from every hospital and weekly reporting from every nursing home in the nation. Some facilities are much smaller than others, and it is not cost-effective to automate reporting in some cases. In such cases, HHS should consider financial assistance to defray the costs of data-gathering requirements.

Some interviewees suggested deploying the national guard to help with data collection. However, interviewees who had the closest relationships with providers suggested that such efforts would strain relationships with providers and that alternative solutions, like financial assistance, would be more appropriate.

3. Develop nationally representative sentinel networks. During the COVID-19 pandemic, two of the primary data collection efforts gathered data from every hospital or nursing home. However, not every data element needs to be gathered via a census; some should be gathered via representative sentinel networks. Representative sentinel networks are useful because they are less costly and impose a lower burden on providers. They should be used to gather in-depth data, such as the number of sick providers, and to test pilot questions that may eventually be required for all providers. They should also be used to gather higher frequency data from smaller providers, like home health agencies, for which it is not feasible to gather a census. While HHS operates several sentinel networks, most are not nationally representative. HHS should expand or modify such programs so that they are statically representative, and trends can be abstracted to the whole nation. Additional funding may need to be allocated to incentivize participation in sentinel networks, especially for smaller providers.

4. Provide support to smaller OpDivs so they can use and modify data collection systems developed by larger OpDivs. Smaller OpDivs, such as IHS, do not have the resources to develop their own data collection systems and may be able to utilize modified systems that are developed by larger OpDivs. For example, all HHS OpDivs were asked to report the vaccination status of employees. IHS is a small agency, and it did not have the infrastructure to automate the collection of these data. It collected the data manually, imposing an additional burden on staff. These data were particularly valuable because IHS employs healthcare providers directly; therefore, it was able to track the vaccination status of providers in its hospitals and clinics via the HHS reporting requirements. The efficiency of valuable data collection efforts like this could be improved by modifying a data collection system developed by a larger OpDiv, like the CDC or NIH.