Legal and Ethical Framework to Use Centers for Disease Control and Prevention Data for Patient-Centered Outcomes Research

Centers for Disease Control and Prevention
National Center for Emerging and
Zoonotic Infectious Diseases

Table of Contents

Acknowledgements

The project was conducted under the technical direction and oversight of Alexa Limeres JD, Carissa Holmes MPH, and Elizabeth Mothershed MS. This report was researched and written by Denise Chrysler JD and Shweta Girish Pai MS MPH, with written contributions from Alexa Limeres JD, Carissa Holmes MPH, and other workgroup members. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC or the Department of Health and Human Services. This report was researched and compiled during September 2015-May 2017, unless otherwise indicated, and reflects the laws and regulations current at the time. This report may reflect only selected portions of laws and regulations and is not intended to be exhaustive of all relevant legal authorities and theories. This resource is for informational purposes only and is not intended as a substitute for professional, legal, or other advice.

The Centers for Disease Control and Prevention gratefully acknowledges the CDC Patient Centered Outcomes Research Workgroup and other persons listed for their participation and assistance in this project. The list of persons who contributed to this project, presented alphabetically, are:

Mary Ari

Drue Barrett

Sheree Boulet

Jeani Chang

Sara Crawford

Natalie Gonzalez

Gail Horlick

Dmitry Kissin

Mike Latham

Leonard Ortmann

Matthew Penn

Loria Pollack

Daniel Pollock

Tara Ramanathan

Mithi Sunderam

1. Executive Summary

Patient-centered outcomes research (PCOR) is a type of comparative effectiveness research (CER) that prioritizes consideration of outcomes important to individual patients: risks and potential benefits of treatment, quality of life, the range of treatment options, etc. This public health research approach offers scientists and the public the opportunity to include “real world” priorities—and therefore, drive evidence-based decisions that take patients’ views into account.

The U.S. Centers for Disease Control and Prevention (CDC) has the opportunity to contribute to this research by providing access to public health data CDC collects for PCOR researchers. These data are collected to advance public health and protect the American public from disease. Researchers with access to these data can create stronger, more scientifically sound PCOR studies, which can, in turn, fine-tune evidence-based approaches to public health and health care.

However, allowing CDC’s data to be used for PCOR introduces some specific legal and ethical challenges. As this is an emerging type of research, those legal and ethical challenges are critical for CDC to understand to inform its decisions on allowing CDC-collected data to be accessed for PCOR.

This paper offers a legal and ethical framework to navigate those legal and ethical challenges. It outlines the legal restrictions on CDC to answer the question “what can CDC do to support PCOR?” It also offers the ethical guardrails for CDC to consider the question “what should CDC do to support PCOR?” Both legal and ethical considerations are essential to CDC in this process.

This paper is not recommending the creation of a PCOR structure at CDC or advising CDC to change any of its policies and practices regarding data collection and use. Instead, this paper organizes the tools CDC needs to make decisions about how its data can best serve the public and advance the agency’s mission to support safer, healthier people.

2. Introduction to Patient Centered Outcomes Research

PCOR is a type of CER[1], which means "research evaluating and comparing health outcomes and the clinical effectiveness, risks, and benefits of 2 or more medical treatments, services, and items."[2] CER may compare the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care.

PCOR focuses on outcomes most important to patients by considering patients' needs and preferences. It seeks to answer patient-centered questions, such as:

- "Given my personal characteristics, conditions, and preferences, what should I expect will happen to me?"
- "What are my health care options, and what are the potential benefits and harms of those options?"
- "What can I do to improve the outcomes that are most important to me?"
- "How can clinicians and the health delivery systems they work in help me make the best decisions about my health and health care?"

To answer these questions, PCOR:

- Assesses the benefits and harms of preventive, diagnostic, therapeutic, or palliative health interventions to inform decision making, highlighting comparisons and outcomes that matter to people
- Takes into account an individual's preferences, autonomy, and needs, focusing on outcomes that people notice and care about such as survival, function, symptoms, and health-related quality of life
- Incorporates a wide variety of settings and diversity of participants to address individual differences and barriers to implementation and dissemination
- Investigates (or may investigate) optimizing outcomes while addressing burden to individuals, availability of services, technology, and personnel, and other stakeholder perspectives[3]

PCOR supports public health's mission by comparing interventions to reduce or eliminate disparities in health and health care, and providing research results that describe variations in health outcomes for patient subpopulations.[4] Thus, it provides a way for CDC to further improve patient care by helping to bridge public health and health care.

Building evidence on patient outcomes requires access to relevant data. For more than a decade, the federal government has worked to facilitate access to the electronic data that it collects, and to support the use of these data to advance medicine and improve the health of communities. These efforts include Congress's enactment of the E-Government Act of 2002 to guide the federal government's information technology policies and promote the use of the Internet. In particular, the E-Government Act promoted development of an integrated federal data infrastructure along with public access to high quality government information.

Building on these requirements, in 2013, Executive Order 13642 made open and machine-readable the new default for government information. Government information would be managed as an asset throughout its life cycle to promote interoperability and openness, and, wherever possible and legally permissible, to ensure that data are released to the public in ways that make the data easy to find, accessible, and usable. As a result, the federal government has built its capacity to provide the public with access to government data through its public portal, www.data.gov.

The Patient Protection and Affordable Care Act of 2010[5] continued federal efforts to improve access to federal data by requiring that the department of Health and Human Services (HHS) identify data that might advance health outcomes through research.[6] Public and private sector data would support a nationwide data infrastructure to enable patient centered outcomes research. PCOR seeks to improve the quality of health care by supporting research and disseminating research results to assist patients, clinicians, purchasers, and policymakers in making better-informed healthcare decisions.

The infrastructure of government data could facilitate access to a massive volume of health information. It could also facilitate linking and combining datasets from multiple sources to enhance the research value of individual datasets.[7] These datasets, which exhibit a variety of data types and come from various sources, include:

- Clinical data captured by electronic medical records
- Clinical trial data collected by researchers
- Claims and administrative data from health plans
- Health data collected by HHS
- Personal data generated and donated by patients
- Other data whose use patients have authorized for research

CDC collects data for public health purposes, including surveillance of disease, injury, exposure to health threats, and research to address population needs. CDC might add great value to PCOR by contributing or facilitating access to datasets to increase understanding of patient health outcomes. Secondary use of CDC's existing public health data for PCOR purposes can have a substantial impact on patient and public health, through research in areas such as epidemiology, drug safety, outcomes research, vaccines, and health services research.[8,9] The outcome results from secondary analyses of CDC data are representative of "real-world" clinical practice and may be generalizable to a wide range of patients.[10]

To support PCOR, CDC must identify and comply with all terms, both internal and external, under which it collects and maintains data. To this end, this paper first describes data that CDC collects, starting with data collection and its disclosure by its Center for Surveillance, Epidemiology, and Laboratory Services (CSELS). The paper next provides an inventory of federal laws and a discussion of ethical issues relevant to CDC data sharing.

It then covers three "data use scenarios," which CDC created to highlight unique legal and ethical implications for CDC use of data. These scenarios are hypothetical; however, they build on real instances where CDC's organizational units use reported data. By creating and working through these scenarios, CDC identified applicable laws and policies, as well as ethical implications of using CDC data for PCOR. While law and ethics often intersect, generally, law defines what an agency can do and ethics define what an agency should do. The scenarios, inventory of federal laws, and review of ethical principles, helped CDC identify key factors and gaps that CDC would need to address to expand its capacity to support PCOR.

Finally, the framework provides steps to assist CDC to systematically address legal and ethical challenges to support PCOR followed by conclusions. With this framework, CDC aims to determine how health information from a variety of data sources can be used for PCOR, consistent with ethical principles and legal requirements governing the privacy of health information and advancing CDC's mission to protect health.

Law defines what an agency can do. Ethics define what an agency should do.

3. CDC Data Overview

a. CDC Data Collection and Sharing

As the nation's public health and prevention agency, CDC depends on the collection of health information to carry out its mission of protecting Americans' health. Its different centers, offices, and institutes support scientific efforts that create the evidence base to advance this mission. These scientific efforts require collecting different categories of public health and medical data. The Public Health Service Act (PHSA) provides broad statutory authority for CDC to collect these data from healthcare and other organizations for statistical, research, and investigational purposes. Additionally, the PHSA provides CDC with broad public health authority to conduct and support research and investigations, responses to public health crises, and interventions during national emergencies.[11] CDC policy defines public health data as *"Digitally recorded factual material commonly accepted in the scientific community as a basis for public health findings, conclusions, and implementation. Public health data could be quantitative, qualitative, imaging, or genomic output. Public health data includes those from research and nonresearch activities. Public health data do not include preliminary analyses, drafts of scientific papers, plans for future research, reports, grantee progress reports, communications with colleagues, or physical objects, such as laboratory notebooks or laboratory specimens."*[12]

The majority of public health data are collected at the state and territorial levels through the respective health departments; as such, these entities control access to the data they collect in accordance with their responsibility to protect the personal information and sources from which they obtain the information. In order to support the practice of public health, CDC utilizes data use agreements with these health departments that spell out the terms and conditions under which the data will be shared with CDC and used by the various programs across the agency. These agreements also delineate how CDC shares the data with partners requiring access to the evaluations or outcomes associated with the data. States most often share aggregate data with CDC for a variety of purposes, either through providing case-specific reporting or sharing statistical data on clinical tests, vital statistics, or behavioral surveys.

Regardless of the nature of the data or how they are provided, it is incumbent on each entity in the data supply chain to ensure that the data are protected and secured in accordance with all applicable local policies, regulations, statutes, and laws.

Table 1: Selected Public Health Surveillance and Research Tools[13]

Type of Surveillance	Description
Vital statistics registries	State- and territory-level registries of live births, deaths, fetal deaths, and induced terminations of pregnancy
Disease- and condition specific registries	Collections of data (from individual case reports from healthcare professionals and laboratories or surveillance of electronic health records) regarding specific diseases and conditions, such as HIV, cancer, diabetes, lead exposure, and occupational exposures (e.g., asbestos))
Biorepositories and genetic databases	Collections of stored tissue samples and genetic information available to researchers assessing genetic associations
National Health and Nutrition Examination Survey	Data collected annually through interviews, direct physical examination, clinical laboratory tests, and related measurements to assess health and nutrition
National Health Interview Survey	Cross-sectional data collected annually on the incidence of acute illness and injury, chronic conditions and disabilities, and access to and utilization of healthcare services
National Immunization Survey	Nationally representative annual telephone survey of households with children and mail survey of immunization providers to monitor immunization coverage
National Electronic Injury Surveillance Survey	Ongoing, nationally representative survey of hospital emergency departments regarding injuries associated with consumer products
Behavioral Risk Factor Surveillance System	Nationally representative annual telephone survey of adults regarding health-related behaviors, chronic conditions, and use of preventive services

In addition to the systems discussed in the table above, there are two national surveillance systems that provide typical examples of CDC data collection which are managed by the CSELS within CDC. The National Notifiable Diseases Surveillance System is a nationwide collaboration that enables all levels of public health—local, state, territorial, federal, and international—to share notifiable disease-related health information on more than 100 nationally notifiable diseases. CSELS's Division of Health Informatics and Surveillance validates the data structures as received from state, local, and territorial health departments, processes the data, and provides it to the respective CDC programs. CDC uses the information to monitor, control, and prevent the occurrence and spread of these infectious and noninfectious diseases and conditions.[14]

The National Syndromic Surveillance Program (NSSP) promotes and advances the development of a syndromic surveillance system for the timely exchange of syndromic data, gathered primarily from emergency department visits. These data are used to improve nationwide situational awareness and enhance responsiveness to hazardous events and disease outbreaks. NSSP functions through collaboration at many levels of public health, federal agencies including the U.S. Department of Defense and the U.S. Department of Veterans Affairs, public health partner organizations, and hospitals and health professionals.[15]

CDC uses various programs to share public health data with state and local public health authorities, researchers, and the public. CDC WONDER (Wide-ranging Online Data for Epidemiologic Research) is a web application connecting users to data to help conduct research, make decisions, set priorities, assess programs, and focus resources. CDC WONDER is available to state and local health departments, researchers, healthcare providers, CDC disease-tracking programs, and the general public.[16] WONDER allows users to access statistical research data by querying numeric datasets on CDC's computers via "fill-in-the blank" web pages.[17] Public-use datasets about mortality, cancer incidence, HIV and AIDS, tuberculosis, vaccinations, natality, census data, and many other topics are available for query.[17] CDC also provides access to several types of health data through participation in Data.gov and at cdc.data.gov.

CDC's National Center for Health Statistics also developed a valuable data-sharing platform called the Research Data Centers (RDC), which allows researchers access to restricted data to advance research objectives.[17] Today, in addition to providing access to data from the National Center for Health Statistics (NCHS), the RDC also hosts restricted data from a variety of groups within HHS.[17] In order to protect the confidentiality of survey respondents, study subjects, or institutions, the RDC requires that all researchers must submit a research proposal for RDC's review to assure the RDC can share this data in a proper way.[17] The proposal provides a framework for NCHS to identify potential disclosure risks and create a data file specific to the research question.[17] Though direct identifiers (name, Social Security number, address) cannot be accessed through the RDC, indirect identifiers (geography) may be available.[18]

b. CDC Policy Regarding Data Practices

Beyond legal and ethical considerations, CDC must comply with numerous policies and guidance documents that govern access or release of data that it maintains. CDC has established internal policies regarding data practices in order to assist researchers and CDC personnel. These policies guide readers through the practical applications of federal regulations and establish agency-wide best practices for data collection and use. CDC has its own internal policies but also collects data that are governed by contractual terms set out in data use agreements, participation agreements, or other agreements between CDC and the data provider.

CDC has developed an operational policy that would apply to the release or sharing of data for PCOR, namely "Policy on Public Health Research and Nonresearch Data Management and Access" (updated 1/26/2016). This policy implements Executive Order 13642[19] and related memoranda that make open and machine readable data the new default for government information while safeguarding individual privacy, confidentiality, and national security.

"Machine readable" refers to information or data that is in a format that can be easily processed by a computer without human intervention.[20] This means that the government would provide access to the raw data with attached metadata, which underlie its reports, in a format that computers can understand and use. Researchers, innovators, journalists, businesses, and others could then examine data in ways that meet their interests and respond to their questions. For this reason, open and machine readable data directly support PCOR's vision.

CDC's goal is to ensure public access to both research and nonresearch data collected or generated using CDC funds. The "Policy on Public Health Research and Nonresearch Data Management and Access" aims to systematize the process of data collection and dissemination by requiring that Centers, Institutes, and Offices (CIOs) within CDC establish a data management plan. The goal of the policy is to ensure public access to federally funded public health data. The CIO responsible for each dataset would determine whether that dataset might be made accessible, subject to law, ethical considerations, data integrity, privacy and confidentiality concerns, and other factors.

Under the data management policy, a CIO can choose between options for release based on the type of data it collects. A CIO might make public health data accessible for public use without restrictions. Alternatively, a CIO might grant certain individuals or organizations access to data that cannot be released publicly, for example, data that contains individually identifiable or potentially identifiable information. Data sharing would need to be consistent with law and existing CDC data security requirements, meet CDC criteria, and be shared under terms and controls appropriate for the particular dataset.

Since the data management policy extends to all CDC collected data, it does not discuss the need to distinguish between research and nonresearch. CDC's "Distinguishing Public Health Research and Nonresearch Policy" sets forth guidelines on the definition of public health research. The policy interprets federal regulations meant to ensure the protection of human research participants and the effective practice of public health. Since some surveillance projects, emergency responses, and evaluations constitute research involving human participants and others do not, all CDC activities must be reviewed to make this distinction. According to the policy, "The ultimate decision regarding classification lies in the purpose of the project. If the purpose is to develop or contribute to generalizable knowledge, the project is research. If the purpose is to prevent or control disease or injury or to improve a public health program, and no research is intended at the present time, the project is nonresearch. If the purpose changes to developing or contributing to generalizable knowledge, then the project becomes research."[21] There are various guidelines and examples enumerated in the policy that explain how to apply this distinction to the various activities undertaken at CDC.

While the above policy explains how to determine whether an activity constitutes research, CDC's "Human Research Protections Policy"[22] explains the protocols CDC follows when that research involves human subjects. In the policy, there is guidance on how to follow the requirements in 45 CFR part 46[23] regarding human subjects research, and guidance regarding the Food and Drug Administration's regulations on clinical investigations. It also explains that CDC honors its ethical responsibilities beyond what is required in regulations. As the policy states, "All of CDC's human research activities, regardless of whether the research is subject to federal regulations, will be guided by the ethical principles of respect for persons, beneficence, and justice as defined in The Belmont Report."[24] Finally, the policy sets forth the requirements for institutional review board (IRB) review and approval of CDC-conducted or supported activities that are covered by human research regulations.

4. Laws That Apply to CDC in Building Capacity for PCOR

Many federal statutes and regulations apply to CDC's collection, use, and disclosure of data. CDC may obtain data from health care, individuals, and organizations for statistical, research, and investigational purposes based on broad authority set out in the PHSA.[25] CDC may also obtain data based on laws that require or authorize organizations to report data to CDC. Examples of such reporting laws are discussed in Section 6 with regard to the data use scenarios. When considering disclosure of data for PCOR, laws that provide for collection or reporting of data should be reviewed first to identify any prerequisites, conditions, or limitations on further disclosure for PCOR.

Next, privacy laws should be identified and examined to determine whether disclosure for PCOR is permitted and any restrictions on disclosure of specific datasets or data elements. Privacy laws may be broad and may protect individuals, organizations, or both. They may apply to certain providers of data, such as healthcare providers or veterans' health facilities. They may also apply to certain types of data, such as information on substance abuse diagnosis or treatment and business proprietary data. Privacy laws may apply to CDC directly, or indirectly impact CDC because they apply to the data provider or to CDC's redisclosure of data that it receives. Laws may cover data that are provided for a particular purpose, such as research, which seeks to balance ethical concerns that include the advancement of science, privacy, and other concerns. If multiple laws apply, the law that provides the greatest data protection (i.e. most restricts data disclosure) applies. Privacy laws affect the use and disclosure of data. Security laws, which seek to safeguard data from improper access, should also be reviewed.

Data use agreements and/or cooperative agreements (or analogous agreements) between data providers and the CDC must also be considered. These agreements may include terms for collecting, providing, and disclosing data that must be addressed for CDC to support access to data.

In addition to federal law, a state's law might impact CDC's ability to support researchers by providing or obtaining specific data from the state's public health surveillance systems and registries. To obtain state data, legal, ethical, and practical barriers must be identified and resolved. Since applicable laws vary by state, there are no "one size fits all" solutions. Rather, barriers must be addressed for each state. Additionally, state legislation can adopt stronger privacy protections than federal protections, such as the protections in the Common Rule for individually identifiable private data used in federally supported or conducted research. Variation among states' laws makes it difficult to develop a common protocol to support research that requires data from multiple states. It is especially challenging to review research protocols and apply standards for approval of the different IRBs in each state.

Federal laws that might impact CDC's data support for PCOR are described below, along with a brief analysis of their potential applicability and considerations that CDC needs to address to build capacity for PCOR. While this list includes key laws, it is not intended to be exhaustive.

The discussion of laws below, and their application to the data use scenarios in Section 6, illustrate challenges that must be resolved to provide identifiable data for PCOR. CDC might eliminate many legal concerns by providing de-identified or aggregate data. Generally, privacy laws either do not cover de-identified data or permit the disclosure of de-identified data without restrictions. Disclosure of de-identified data is preferred if it might meet the data need.

However, disclosure of de-identified data to PCOR may not be a simple or adequate solution.

First, while personal identifiers may be unnecessary for CDC surveillance, they may be critical for some PCOR functions. PCOR envisions the collection of massive volumes of data from multiple sources that might be linked to gather as much information as possible about individual patients. Unique identifiers are useful to link databases and create combined datasets, supply missing data, de-duplicate records, and conduct longitudinal studies. Matching techniques that do not rely on personal identifiers may be possible and useful for certain studies.[26]

Second, while federal laws allow CDC to conduct or support PCOR with de-identified data, they contain no common definition of "de-identified data;" nor do they include a uniform standard for ascertaining whether data are de-identified or criteria to measure the risk of potential re-identification of data. Appendix A shows how federal laws, described below, describe de-identified data.

Third, fully de-identified data may have limited value. Depending on the law, CDC may be able to preserve the utility of data by employing one or more of the following: coded data, limited datasets, or data that are statistically de-identified. Appendix B describes common statistical de-identification techniques.

a. Federal Laws that Apply to CDC Data Sharing

1. Federal Privacy Act

The Privacy Act of 1974[27] safeguards the public from unwarranted government collection, maintenance, use, and dissemination of personal information. The Privacy Act applies only to federal agencies; it does not apply to state or local agencies that receive federal funds. It establishes fair practices that govern information about living individuals maintained in systems of records by federal agencies. A system of records is a group of records under an agency's control from which information is retrieved by the name of the individual or other identifier. The CDC must comply with the Privacy Act[28] and with HHS regulations[29] to implement the Act. It must annually publish a System of Records Notice (SORN) in the Federal Register, describing its system for records that are covered by the Act.[1]

Under the Privacy Act, the CDC may maintain in its records only such information about an individual as is relevant and necessary to accomplish its work. It is prohibited from disclosing information that identifies an individual, without the individual's written consent, unless one of the 12 disclosure exceptions enumerated in the Act applies. These include disclosures:

- To officers and employees within the agency who have a need for the record in the performance of their duties
- Required by the federal Freedom of Information Act
- Outside the agency for a purpose that is compatible with the purpose for which the information was collected ("routine use"), as described in the agency's SORN[2]
- For use as a statistical research or a reporting record, provided that the record is to be transferred in a form that is not individually identifiable[3]
- To a person pursuant to a showing of compelling circumstances affecting the health and safety of an individual

[1] In addition to regulations, OMB provides guidelines for implementation of the Privacy Act. OMB Guidelines, posted at https://www.whitehouse.gov/omb/privacy_general.

[2] For example, a large number of projects at CDC are covered by notices for 09-20-0136, "Epidemiologic Studies and Surveillance of Disease Problems" (NCPDCID) and "Occupational Health Epidemiological Studies" (NIOSH). https://www.cdc.gov/sornnotice/; https://www.cdc.gov/SORNnotice/PrivacyFAQ/index.htm.

[3] "Statistical record" means a record in a system of records maintained for statistical research or reporting purposes only and not used in whole or in part in making any determination about an identifiable individual.

The Privacy Act applies only if data are retrieved by name or other identifying particular such as a Social Security number, or other identifying number or symbol. If data are primarily retrieved by another variable, it may not apply. The Privacy Act also contains provisions about using identifiable information for computer matching activities, for example, matching among federal datasets or between federal and state or local datasets.

To evaluate the Privacy Act's applicability and impact, the following factors should be addressed:

- Are data collected by CDC personally identifiable? The Privacy Act considers a record personally identifiable if it contains the individual's name, or the identifying number, symbol, or other identifying particular assigned to the individual, such as a finger or voice print or a photograph.[30]
- Would data be retrieved by name or other identifier?
- Will CDC disclosure of data, which are not personally identifiable, support the proposed use? If not, will disclosure be allowed by an exception?
- If an exception allows disclosure of data, is disclosure consistent with CDC's SORN?

For records covered by the Privacy Act, a federal agency must establish appropriate administrative, technical, and physical safeguards to insure the security and confidentiality of records and to protect against any anticipated threats or hazards to their security or integrity which could result in substantial harm, embarrassment, inconvenience, or unfairness to any individual on whom information is maintained.

2. Assurance of Confidentiality.

Under section 308(d) of the PHSA,[31] data collected by the National Center for Health Statistics (NCHS) as part of its authorizing legislation are automatically protected by an Assurance of Confidentiality. In addition, Assurances of Confidentiality may be issued to projects conducted by all other CDC components, after formal application to and approval by the CDC Confidentiality Review Group has been obtained.[32] At CDC, the 308(d) assurance has most often been used to protect sensitive identifiable data for non-research projects, including some surveillance projects, and for research studies collecting sensitive identifiable data. The assurance is made between CDC and the entity providing this data, allowing CDC to protect the data it collects under the assurance. Importantly, this ability to protect data under 308(d) is explicitly given to CDC under the PHSA.

If information is supplied that describes an entity or person and that information was obtained with an assurance, then that information may not be used for any purpose other than the purpose for which it was supplied unless the entity or person has consented to its use for such other purpose. An Assurance of Confidentiality provides greater privacy protection than the Federal Privacy Act because it protects a wider group of individuals, protects information about establishments, and protects data from court-compelled disclosure and Freedom of Information Act requests.

Table 2: Privacy protections under an assurance of confidentiality compared to protections set out in the Federal Privacy Act

Assurance of Confidentiality	Federal Privacy Act
Protects identifiable data about both individuals and establishments	Protects identifiable data about individuals only
Protects identifiable data about individuals who are living and deceased	Protects identifiable data about individuals who are living
Protects identifiable data about all individuals	Protects identifiable data about U.S. citizens and aliens lawfully admitted for permanent residence
Protects data from court compelled disclosure, such as subpoenas and discovery orders and freedom of information requests	Does not protect data if disclosure is ordered by a court of competent jurisdiction

For data collected under an assurance of confidentiality, CDC will need to determine how it might be able to support PCOR within the limitations of the Assurance.

3. Confidential Information Protection and Statistical Efficiency Act

The Confidential Information Protection and Statistical Efficiency Act (CIPSEA)[33] establishes a uniform policy for all federal statistical collections. This act protects the use and confidentiality of information that individuals and organizations provide to the federal government for statistical activities. “Statistical activities” means the collection, compilation, processing, or analysis of data for the purpose of describing or making estimates concerning the whole, or relevant groups or components within, the economy, society, or the natural environment. The term also includes the development of methods or resources that support those activities.

CIPSEA applies to federal agencies for data that they collect while representing that these data are being collected for statistical purposes. This act protects identifiable information provided by a person or organization. It prohibits disclosure of data in identifiable form or information acquired by an agency under a pledge of confidentiality for exclusively statistical purposes, or for non-statistical purposes, except with the informed consent of the respondent.

To evaluate CIPSEA’s applicability and impact, the following factors should be addressed:

- Did CDC collect data for statistical activities under a pledge of confidentiality?
- If so, to what extent does the proposed disclosure involve non-statistical uses?
- Would data be disclosed in identifiable form? CIPSEA defines “identifiable form” as any representation of information that permits the identity of the respondent to whom the information applies to be reasonably inferred by either direct or indirect means.

4. Common Rule

The Federal Policy for the Protection of Human Research Subjects (Common Rule) protects human subjects used in research that is conducted, supported, or subject to regulation by HHS or 16 other federal departments or agencies that have adopted it.[34, 4] To apply, the federally supported activity must be “research” that involves “human subjects.” The Common Rule defines “research” as the systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge.

[4] On January 19, 2017, HHS and 15 other federal departments and agencies issued final revisions to the Common Rule. Federal Register Volume 82, Number 12 (Thursday, January 19, 2017), available at https://www.gpo.gov/fdsys/pkg/FR-2017-01-19/html/2017-01058.htm. Most provisions in the new rule will go into effect in 2018.

Public health activities, including but not limited to certain surveillance activities, are not research. That said, because scientific principles are followed for public health activities, at times it may be difficult to determine whether or not an activity is research subject to the Common Rule.[35]

A "human subject" is a living individual about whom an investigator conducting research

- obtains data through intervention or interaction with the individual, or
- obtains "private information" that is "individually identifiable."

Private information includes information that the individual has provided for a specific purpose and can reasonably expect will not be made public (e.g. a medical record) if the information is individually identifiable. "Individually identifiable" means that the identity of the subject is or may be readily ascertained by the investigator or associated with the information. To use private data for research, generally, the investigator must demonstrate that the rights and welfare of research subjects will be protected and obtain approval of an IRB. Informed consent of a research subject is required, unless the proposed research is exempt from review or the IRB waives or modifies the informed consent requirement. When an IRB does review an activity that is research subject to the Common Rule, they are generally looking to determine that there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data.

Research using existing data, such as patient health information in medical records, might qualify under the Common Rule for a "Category 4" exemption from IRB review. Research involving collection or study of existing data, documents, and records can be exempted under Category 4 of the federal regulations if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects. This category applies in scenarios where the investigator initially has access to identifiable private information but abstracts the data needed for the research in such a way that the information can no longer be connected to the identity of the subjects. If the investigator records identifiers, IRB review might be required.

Generally, an individual's private data may not be collected or used for research absent a legally effective informed consent. "Legally effective" means that the informed consent process and elements of the informed consent meet Common Rule specifications. However, an entity might use or disclose identifiable health information without an individual's consent for research, provided it obtains documentation from an IRB that the following waiver criteria[36] were satisfied:

1) The research involves no more than minimal risk to the subjects.
2) The waiver or alteration will not adversely affect the rights and welfare of the subjects.
3) The research could not practicably be carried out without the waiver or alteration.
4) Whenever appropriate, the subjects will be provided with additional pertinent information after participation.

The Office for Human Research Protections (OHRP) provides a series of decision trees on its website.[37] These decision trees list considerations, which should be useful to determine

- whether use or disclosure of data is human subjects research that is federally supported.
- whether human subjects research is exempt from review, review may be expedited, or whether full review is required.
- for human subjects research that is not exempt, whether requirements are satisfied for an IRB to waive or modify the informed consent requirement and/or documentation of consent.

With regard to CDC, see CDC's policies on protection of human research participants (Issued 7/29/2010) that are discussed in Section 2 above.

Data might be collected at CDC for multiple purposes that include research, as well as public health surveillance. If CDC obtains private data for research, or for inclusion in a research registry, then the Common Rule may apply to it. Potentially, a research registry could be designed so that the regulations would not apply to the creation and operations of the registry through various mechanisms, including the use of codes instead of identifiers in the original release of data to a registry, or the use of computer programming to merge identifiable datasets without any person being able to view the data in identifiable form.[38]

Whether the Common Rule applies to the data contributor depends on whether the data contributor is "engaged in research" supported by federal funds.[39] OHRP has issued guidance on whether an institution is "considered to be engaged or not engaged in human subjects research conducted or supported by HHS.[40] For example, an institution, such as a healthcare provider, is not engaged in research simply by providing private data to a research repository or database. In comparison, most likely, an institution that obtains informed consent to provide an individual's private data to a research repository or database is engaged in research. If its nonexempt human-subjects research activities are supported by federal funds, the Common Rule would apply to the data contributor.

5. HIPAA Privacy Rule

The Privacy Rule,[41] adopted by HHS under the Health Insurance Portability and Accountability Act (HIPAA),[42] provides minimum federal protections for personal health information and gives patients certain rights with regard to that information. It applies to health plans, most healthcare providers, and to healthcare clearinghouses ("covered entities"). CDC is none of these. This means that the HIPAA Privacy Rule does not apply to CDC, either when it obtains identifiable data to create a research repository or database or when it provides identifiable data to an investigator or for PCOR. That said, the HIPAA Privacy Rule is relevant to identifiable data that covered entities, such as healthcare providers or health plans, provide to CDC. Consequently, CDC should consider HIPAA's limitations when making decisions to disclose data for PCOR when that data originated from covered entities.

The HIPAA Privacy Rule prohibits the disclosure of protected health information (PHI), absent the patient's authorization, unless an exception applies. PHI is distinguishable from individually identifiable information insofar as it has a specific definition in the text of the HIPAA Privacy Rule. A covered entity might disclose PHI to CDC, as described below, for certain purposes.

<u>Disclosure for a public health purpose.</u> Generally, covered entities are permitted to disclose PHI to public health authorities, such as CDC and state and local health agencies, without a patient's authorization.[43] The HIPAA Privacy Rule recognizes the legitimate need for public health authorities to have access to PHI to carry out their public health mission. Accordingly, the rule permits covered entities to disclose PHI without authorization to "[a] public health authority that is authorized by law to collect or receive such information for the purpose of preventing or controlling disease, injury or disability, including, but not limited to, the reporting of disease, injury, vital events such as birth or death, and the conduct of public health surveillance, ...investigations, and... interventions...."[44] As required by the HIPAA Privacy Rule, a healthcare provider should advise patients in its Notice of Privacy Practices that it may disclose their health information for public health activities without the patient's permission.[45]

As discussed above, CDC routinely collects data for a variety of public health purposes. Providing data for PCOR would be a "secondary use." That is, CDC data, collected for particular public health purposes, would be made available for purposes for which they were not collected. For example, if data collected by CDC for surveillance were provided to PCOR for health research, this would be a secondary use. Even if CDC collects data for a public health research project, it would be a secondary use to then make these data available to PCOR for research.

If a covered entity provides identifiable data to CDC under the public health exception, the HIPAA Privacy Rule would not impede CDC in disclosing this data for PCOR. However, several other privacy laws, which directly apply to CDC, might limit CDC's ability to provide these data to PCOR. These include the Privacy Act, the Assurance of Confidentiality, and the Common Rule, discussed above.

<u>Disclosure for a research purpose.</u> The HIPAA Privacy Rule adopts the Common Rule's definition of research. "Research" means "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge."[46] Like in the Common Rule, research includes disclosure of identifiable data to a research repository or database for future research.[47] Covered entities are permitted to use and disclose data for research with individual authorization, or without individual authorization under limited circumstances as follows.

De-identified data. As discussed above, a covered entity may provide de-identified data to CDC without authorization. Generally, de-identified data cannot include indirect identifiers, such as dates (e.g. encounter dates and dates of birth and death) and demographic data (e.g. five digit ZIP Codes and county level data). Data that include these indirect identifiers, are PHI unless an expert documents that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information.[48]

Coded data. The Privacy Rule permits a covered entity to assign to, and retain with, the de-identified health information, a code or other means of record re-identification if that code is not derived from or related to the information about the individual and is not otherwise capable of being translated to identify the individual. However, the covered entity may not (1) disclose its method of re-identifying the information to the CDC or (2) use or disclose the code for any purposes other than as a re-identification code for the de-identified data.[49]

Limited dataset. A covered entity may disclose a limited dataset to CDC for research purposes provided the parties enter into a data use agreement.[50] While a limited dataset excludes specified direct identifiers of the individual or of relatives, employers, or household members of the individual, it would permit disclosure of data with indirect identifiers, such as dates (e.g. encounter dates and dates of birth and death) and demographic data (e.g. five digit zip codes and county level data).

Identifiable data. A covered entity is permitted to use and disclose identifiable information to CDC for research if a privacy board or IRB approves a waiver of authorization.[51] The following three criteria must be satisfied for an IRB or privacy board to approve a waiver of authorization:

1) The use or disclosure of PHI involves no more than a minimal risk to the privacy of individuals, based on, at least, the presence of the following elements:

    - An adequate plan to protect the identifiers from improper use and disclosure
    - An adequate plan to destroy the identifiers at the earliest opportunity consistent with conduct of the research, unless there is a health or research justification for retaining the identifiers or such retention is otherwise required by law
    - Adequate written assurances that the PHI will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research study, or for other research for which the use or disclosure of PHI would be permitted by this subpart.

2) The research could not practicably be conducted without the waiver or alteration.

3) The research could not practicably be conducted without access to and use of the PHI.

Table 3: Criteria for waiver or modification of authorization under HIPAA Privacy Rule and consent under Common Rule

HIPAA Privacy Rule	Common Rule
1) The use or disclosure of PHI involves no more than a minimal risk to the privacy of individuals, based on, at least, the presence of the following elements; • An adequate plan to protect the identifiers from improper use and disclosure • An adequate plan to destroy the identifiers at the earliest opportunity consistent with conduct of the research, unless there is a health or research justification for retaining the identifiers or such retention is otherwise required by law • Adequate written assurances that the PHI will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research study, or for other research for which the use or disclosure of PHI would be permitted by this subpart. 2) The research could not practicably be conducted without the waiver or alteration 3) The research could not practicably be conducted without access to and use of the PHI.	1) The research involves no more than minimal risk to the subjects; 2) The waiver or alteration will not adversely affect the rights and welfare of the subjects; 3) The research could not practicably be carried out without the waiver or alteration; and 4) Whenever appropriate, the subjects will be provided with additional pertinent information after participation.

To evaluate the HIPAA Privacy Rule's applicability to the data provider and its impact, on providing data to the CDC for PCOR, the following factors should be addressed:

- Will disclosure of data be permitted? In this regard, for what purposes will a healthcare provider be disclosing identifiable data to the CDC? If data are being disclosed for research purposes, then either the patient's authorization will be required or disclosure must be permitted under the HIPAA Privacy Rule.
- Can research be accomplished with de-identified data or a limited dataset that would allow a covered entity to disclose data that includes certain identifiers for research purposes provided the parties enter into a data use agreement?

The HIPAA Security Regulations[52] apply to a covered entity's responsibility to develop and implement administrative, physical, and technical safeguards to protect electronic PHI. As discussed above, the CDC is not a covered entity that is subject to HIPAA regulations. If healthcare facilities or other covered entities electronically transfer PHI to CDC, covered entities would need to comply with its security safeguards, as required under the HIPAA Security Rule.

6. The Uniform Trade Secrets Act

A healthcare provider might report information to the CDC that it designates as trade secret or confidential business/proprietary information. The Federal Trade Secret Act[53] might affect CDC's disclosure of this information for PCOR. This act makes it a criminal offense for an officer or employee of the federal government to knowingly disclose confidential commercial and trade secret information unless he or she is authorized to do so by law. Additionally, privileged or confidential trade secrets and commercial or financial information obtained from a person are excluded from disclosure under the federal Freedom of Information Act.[54]

7. Federal Freedom of Information Act

The federal Freedom of Information Act (FOIA)[55] does not prohibit disclosure of information. Rather, it compels disclosure of reasonably described federal records or a reasonably segregated portion of the records to any person upon written request, unless one or more of nine exemptions apply to the records. Potentially, health care providers and/or states might voluntarily provide data to support CDC's efforts to build capacity for PCOR. In this situation, FOIA's disclosure requirements might need to be considered to ensure that the CDC is able to protect information regarding individuals or businesses from disclosure.

Several exemptions might apply to data maintained or to be collected by CDC. Information is exempt from disclosure to the extent that another statute requires that it be withheld.[56] As discussed above, several privacy laws require that confidentiality be maintained for identifiable data. These laws include: the PHSA, the Privacy Act, CIPSEA, and the Trade Secrets Act.

Additionally, identifiable health information might be withheld under an exclusion from the FOIA for medical files and similar files if disclosure would constitute a clearly unwarranted invasion of personal privacy; and certain information from health care providers might be withheld under an exclusion from FOIA for trade secrets and commercial or financial information obtained from a person and privileged or confidential.

8. Federal Information Security Management Act

The Federal Information Security Management Act[57] establishes a framework wherein federal agencies must inventory data, develop and implement data security requirements for data commensurate with risk, and adopt a data security plan.[5] Researchers that access and use CDC data, either on-site or remotely, must comply with federal standards.

[5] For HHS Data Privacy and Security Policies, visit http://www.hhs.gov/ocio/policy/index.html#Information

5. Ethical Implications of Using CDC Data for PCOR

a. Ethical Principles in the Context of PCOR

Decision-making regarding data sharing for research, including PCOR, should be guided by both laws that govern data sharing and ethical principles that ensure the ethical conduct of research. These ethical principles include:

- **Respect for persons:** The principle of "respect for persons" emphasizes that individuals should be treated as autonomous agents, capable of deliberating and furthering their personal goals.[24,58] In some cases, this principle may be subordinate to other ethical principles and values. In such circumstances, a waiver of informed consent requirements may apply to the public health system and be ethically acceptable. | Applying this principle to the research use of public health data gives rise to additional ethical concerns about preserving the privacy of patients, protecting the confidentiality of their data, and minimizing potential harms.[58] Some individuals may require special protection due to limited or compromised autonomy such as immaturity, imprisonment, or being in an impaired mental state. Safeguards for the confidentiality of patient data beyond applicable legal requirements may be ethically necessary to protect the privacy of those individuals.

- **Beneficence (doing good) and nonmaleficence (doing no harm):** These principles dictate that individuals participating in research should be treated in an ethical manner by protecting them from harm and ensuring their wellbeing. Entities collecting or holding data are ethically obligated to minimize potential harms to the individuals or groups contributing their data.[58] Compromising privacy, stigmatizing groups, and undermining public trust are examples of harm that might result from using data for research. However, the principle of beneficence also requires that research strive to benefit the common good by contributing to generalizable knowledge. This involves ensuring good data quality and that the study design allows the research questions to be answered. Considerations of beneficence and nonmaleficence call for researchers to optimize net benefits over harms for individuals and populations involved in research.

- **Justice:** Justice refers to giving people what they are due or owed.[24] In the context of human subjects research protection, justice usually requires fair selection of subjects and stakeholders and ensures the fair or equitable distribution of the burdens and benefits.[59] A precise definition of fairness may differ from one context to another. Some situations may call for an equal distribution while others may require an equitable distribution where the selection of research subjects, or the distribution of research burdens and benefits does not depend on the "favor" or "disfavor" of the researcher, or the "vulnerability" of the subjects.[60] In the context of research with patient health information, fair allocation may be best characterized by the equitable distribution of burdens and benefits of research.[59] It requires that any single ethnic, social, gender, racial, or socioeconomic group should not bear disproportionate burdens of research or receive disproportionate benefits of research.[59] The principle of justice in this context also requires a "fitting" match: the projected results of the research study should serve the population from which research subjects were selected. Thus, failure to include groups that might benefit from research results could represent an injustice, especially if no scientific basis exists for their exclusion.

b. Consent and Secondary Uses of CDC Public Health Data

Although secondary analyses of existing public health data satisfy the ethical principle of beneficence, they often contradict the principle of respect for persons. Respect for persons emphasizes patient autonomy that allows patients to freely dispose of property which they consider belongs to them.[61] This freedom allows them to maintain their dignity by protecting their personal information and avoiding discrimination or stigmatization.[58,62] Patients place a high value on their autonomy even though they may not fear disclosure of embarrassing information.[62]

Informed consent is an important conduit of patient autonomy. It allows a patient to assess the benefits, harms, and expectations before voluntarily agreeing or disagreeing to participate in a research study. The potential subjects of a study are the only persons who can weigh the benefits and harms of a study using their goals, priorities, and values. For instance, some privacy advocates maintain that even if the public health data used for secondary analyses are de-identified, patients should have a right to consent to the uses of their data by evaluating the risks and potential benefits of the research study in question. They argue that researchers and data collectors run the risk of violating patient autonomy by deciding for the patient how and where their data will be used.[63]

Although seeking consent would satisfy the principle of respect for persons, there are ethical concerns when linking otherwise de-identified data with identifiers to contact participants for their consent. There are also concerns regarding the practicability of obtaining consent, and the negative effects of patient consent on data quality. Obtaining consent from individuals in many cases may be expensive and impractical or could introduce bias effects in the research studies. However, consent of any form implies confidence, and confidence implies trust.[64] A patient's trust in the system affects several important outcomes such as commitment to the health care providers, adherence to treatment, and continuity of care.[65] To satisfy the principle of respects for persons, while maintaining data quality, patient trust, and minimizing harms to patients, different models of consent can be considered. Researchers must decide whether it is practical to seek consent, and if so, what form of consent will be ethically sufficient.[66]

c. Risk of Re-Identification and Associated Harms.

A patient's health information may include intimate details associated with his or her life, such as medical diagnoses, information on developmental disability, cognitive capacities, and emotional stability. Disclosure of these types of information might cause embarrassment, diminish an individual's reputation in the community, and affect his/her ability to obtain and maintain employment, insurance, and housing. Thus, strong privacy and security protections are needed to prevent unauthorized collection, use, and disclosure of private health information.

Data de-identification is one method for researchers to protect privacy while permitting other uses of personal information.[67] Research is greatly simplified by supporting PCOR with de-identified data. This is because laws that restrict data disclosure and secondary uses do not apply to de-identified data. It is often assumed that removing the direct identifiers from patient data would protect from identification. However the risk of re-identification is not eliminated if the data contains quasi-identifiers; the variables that may not directly identify individuals, but can still be used for indirect re-identification.[68] Examples of quasi-identifiers include gender, marital status, postal code, diagnosis information, profession, ethnic origin, visible minority status, income, etc.[69]

Quasi-identifiers can be used by themselves or in combination with other variables to identify an individual. Although properly de-identified data can be successfully re-identified, there is a documented low risk of re-identification. One expert in de-identification testified that only 0.04 per cent (4 in 10,000) of the individuals within datasets de-identified according to HIPPA's Safe Harbor Standard, discussed later, may be potentially re-identifiable.[70] Some rare quasi-identifiers may be more likely to result in the re-identification of an individual in a given dataset. For instance, the likelihood of the individual's identify being revealed is increased by rare quasi-identifiers such as an unusual occupation or an unusual medical diagnosis.[71] One strategy to minimize the harms associated with the use of de-identified data would be to remove or alter quasi-identifiers along with direct identifiers before the collection, use or disclosure of personal health information for secondary uses.

In spite of the low risk of re-identifying individuals, research with de-identified data may also result in group harms that do not depend on the re-identification of individual subjects. De-identified data still contains information about a patient's membership in certain groups defined by race, ethnicity, gender, religion, or other criteria.[72]
Facts or statistics about a group or a community could be used to make determinations about an individual. Such inferences can be applied to generalize qualities about all the individuals of a group.

The use of de-identified data does not eliminate existing risks of stigmatization and discrimination, especially for vulnerable populations.[72] Although there are several definitions available for the term 'vulnerable population,' the term refers to a disadvantaged sub-segment of the community requiring specific ancillary considerations and augmented protections in research.[73] In the context of research participation, vulnerable population could be defined as "any group that is unable to protect its members' self-interests in the course of being research subjects."[74] The vulnerability that these groups experience usually can be attributed to one or some combination of six traits: cognitive or communicative vulnerability, institutional vulnerability, deferential vulnerability, medical vulnerability, economic vulnerability, and social vulnerability.[75,76] The term 'vulnerable population' refers to but is not limited to the uninsured, the poor, the elderly, children, those living with mental or physical disabilities, racial and ethnic minorities, the terminally ill and special classes of subjects including students and employees. Depending on the type of research studies and conditions, vulnerability may apply to populations that are otherwise not viewed as vulnerable.[77]

In the context of data sharing, the potential loss of privacy and confidentiality makes certain populations vulnerable. Individuals or groups may be exposed to risks of stigma or discrimination due to the potential loss of privacy and confidentiality. As sensitivity to being vulnerable is relative, fears of harms due to breaches of privacy and confidentiality may express themselves differently in particular communities, ethnic groups or patient populations.[78] Such risks may not have been apparent to research participants when they provided their data. The following table provides examples of some types of public health data and the vulnerable populations associated with each data type.

Table 4: Vulnerable Populations Associated with different types of Public Health Data.

Degrees of identifiability	Explanation of terms	Examples	Possible vulnerable populations	CDC safeguards in addition to technical safeguards
Data with direct identifiers	Information that relates specifically to an individual. The inclusion of a name, Social Security number, or phone number, makes data identifiable.	US Zika Pregnancy Registry[79]	Pregnant women whose fetuses would be at high risk for complications	No access to individual-level data. Information is released to the public in an aggregate form.
Linkable or coded data	Data that is not identifiable, but can be linked to a named person with the use of a secure code.	HIV case reports[80]	Patients infected with HIV	Data is sent from the state health department to CDC using a Soundex code.
Data with indirect identifiers	Information that can be combined with other information to identify specific individuals. Information about location, race, and sex can identify an individual.	National ART Surveillance System (NASS)[81]	Egg donors, surrogate mothers, and children	Only onsite access allowed with a member of the ART team.
De-identified data	Direct and known indirect identifiers in any information are removed or obscured to minimize the risk of unintended disclosure of the identity of individuals.	National Program of Cancer Registry (NPCR) Cancer Surveillance System[82]	Individuals residing in areas with high incidence of certain cancers	Most of the de-identified datasets are publicly available.
Anonymized data	Direct and indirect identifiers have been irreversibly removed or altered so that re-identification is impossible.	NHANES Genetic Data Repository[83]	Patients belonging to minority ethnic groups	Anonymized data is available with a data use agreement.

6. Using Data For PCOR: Application of Law and Ethics to Data Use Scenarios

CDC developed three data use scenarios that demonstrate the potential for CDC to build capacity for PCOR; illustrate legal, ethical, and practical considerations; and identify challenges and gaps for implementation. Although based on actual CDC databases, these data use scenarios describe potential enhancements to the databases or practices at CDC. These enhancements might engage consumers, improve healthcare treatment or delivery, or reduce costs. These data use scenarios propose changes to existing data sources and providers to increase types and volume of data, which can be linked to create combined datasets.

Each data use scenario is summarized below and includes a map showing the flow of data among data providers and recipients. These scenarios illustrate legal and ethical issues and are governed by many of the same statutes, regulations, formal guidelines, and policies that might apply to CDC. A brief analysis discusses legal and ethical implications that apply to these scenarios. While not exhaustive, the analysis includes key legal and ethical principles that should be considered. It does not reach legal conclusions—and is not intended as legal advice—regarding what the law ultimately permits.

Data Use Scenario 1 – National ART Surveillance System (NASS)

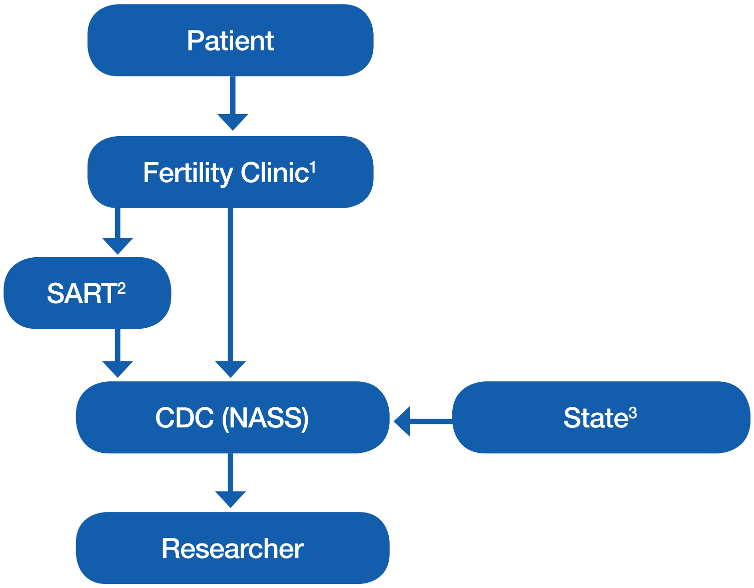
Description

The NASS data use scenario studies maternal and birth outcomes associated with assisted reproductive technology (ART). Federal law requires that all fertility clinics report data concerning each ART procedure to the CDC. These clinics enter data directly into NASS, export data from electronic records into NASS, or report data to the Society for Assisted Reproductive Technology (SART), which then reports the data to NASS on behalf of the clinic. Data include patient demographics, patient characteristics, patient obstetrical and medical history, parental infertility diagnosis, clinical parameters of the ART procedure, and information regarding resultant pregnancies and births. Data reported to NASS are used to provide consumers with information about national and clinic-specific pregnancy success rates, as well as assess infant outcomes (birth weight, plurality, maturity) and monitor trends in ART use, practice, and outcomes.

Because NASS contains only limited pregnancy outcome information, CDC initiated a collaborative project to link ART surveillance data with state surveillance systems and registries that contain more detailed information on women and infants. This project, the States Monitoring ART (SMART) Collaborative, provides a unique opportunity to establish state-based patient-centered surveillance of ART, infertility, and related issues, which aims to improve patient outcomes. Using probabilistic matching techniques, NASS data are being linked to vital records, hospital discharge data, birth defects registries, cancer registries, and other surveillance systems of participating states. Three states participate in the SMART Collaborative at this time.

Currently, state health departments and outside researchers may query NASS data remotely through the National Center for Health Statistics Research Data Center (RDC) or by working with CDC staff. De-identified cycle-specific or birth record data are maintained securely behind a firewall. States and outside researchers query NASS; aggregate data only are returned in response to the query. Researchers do not have access to cycle-specific or birth record data. For the NASS scenario, CDC could enhance data that are available for studies by increasing the number of state health departments that participate in the SMART Collaborative. In enrolling states, CDC could prioritize states that have the highest number of births from ART.

Data Use Scenario: Increased number of states reporting data to the National Art Surveillance System



1. Clinics report patient demographics and characteristics, patient obstetrical and medical history, parental infertility diagnosis, clinical parameters of the ART procedure, and information regarding resultant pregnancies and birth. Data do not include direct identifiers; do include indirect identifiers, such as date of birth.
2. A fertility clinic may report directly to the NASS or provide data to the Society for Reproductive Technology to report on its behalf.
3. States that participate in the SMART collaborative provide data from state electronic systems, including vital records, birth defects registry, cancer registry, and hospital discharge data. Data do not include direct identifiers; do include indirect identifiers. [Data are not returned to state; state may access under researcher protocol]

By adding states, and combining state data with fertility clinic data, NASS would support state-focused surveillance and research. Data from additional states would provide geographically diverse data and increase the pool of potential study subjects for both research cohorts and comparison groups. This increase might support studies of smaller geographic areas and for rare conditions that require large amounts of data. It might also improve generalizability of research results and reduce the risk of data identifiability. States and outside researchers would continue to access data through the RDC or the Division of Reproductive Health (DRH).

Legal Issues

This scenario involves both data reported to CDC by fertility clinics and data provided by states that participate in the SMART Collaborative. The Fertility Clinic Success Rate and Certification Act of 1992 (FCSRCA)[84] mandates that fertility clinics report to CDC yearly about ART cycles performed at clinics in the United States. The CDC is required to inform states and consumers by publicizing certain data for each clinic that are relevant to ART success.[85]

Unlike fertility clinics, states are not mandated to report data to CDC. However, the PHSA, described above, authorizes the CDC to collect state surveillance data. The NASS use case does not propose a change in the nature or identifiably of the data that CDC collects. Rather, it proposes that CDC enroll more states in the SMART Collaborative. Federal law does not interfere with enrolling additional states. However, each additional state would need to examine its own law to determine whether it might participate in SMART, and any prerequisites, conditions, or limitations that might apply to surveillance data that it provides to CDC.

State law and policy should be reviewed to identify opportunities for streamlining the review process, for example, through a collaborative process with a single IRB for review and a common research protocol. Several federal laws should be reviewed for their applicability and potential impact on CDC's ability to disclose NASS data for PCOR. Laws that should be reviewed include, but are not limited to, those described below.

NASS is covered by a federal assurance of confidentiality that was granted under § 308(d) of the PHSA.[86] This means that data were obtained for NASS with a guarantee that identifiable information about establishments and/or individuals will be used only for the purposes stated in the Assurance, and will not otherwise be disclosed or released without the consent of the establishments or individuals. Some of the data in NASS are sensitive because: (1) they may be used to identify an individual woman or child and (2) they relate to issues about which people may have strong ethical, religious, and/or cultural concerns. In compliance with the FCSRCA requirements, CDC releases an annual national summary report that uses information from all ART clinics as well as clinic-specific reports for each ART clinic. However, disclosure of data that might be used to identify individuals, or linkages of NASS data with other datasets, must comply with restrictions in the Assurance.

If disclosure is permitted by the Assurance, proposed disclosure of data regarding individuals should be reviewed under additional privacy laws, such as the Privacy Act and the HIPAA Privacy Rule, as well as the Common Rule. These laws protect identifiable information regarding individuals, but not entities or establishments.

The Privacy Act applies to NASS data that are retrieved by name or other identifying particular. If data are retrieved by name or identifying particular, disclosure of data identifying an individual is prohibited, without an individual's written consent, unless an exception applies. Disclosure might be permitted for statistical research or routine uses, such as "Epidemiologic Studies and Surveillance of Disease Problems," depending on the details of the specific project and the data elements that are requested.

The HIPAA Privacy Rule does not apply to CDC. However, it might apply to healthcare providers that provide data to the CDC for NASS, depending on the purpose for which data are provided. If the purpose is for public health, the Privacy Rule permits disclosure to CDC and does not apply to secondary uses by the CDC, including for research. On the other hand, if CDC obtains identifiable data from facilities for the purpose of research or to build a research repository, healthcare providers must obtain written authorization, modification, or waiver of authorization by an IRB or Privacy Board based on criteria established by the Privacy Rule. A healthcare facility might also disclose a limited dataset for research provided all HIPAA-required safeguards are satisfied. For a limited dataset, direct identifiers are removed but demographic data and dates (e.g. dates of service and birthdate) are retained.

The Common Rule establishes protections that apply to institutions that are engaged in human subjects research. If CDC obtains or discloses private data for human subjects research or for a research repository or database, the Common Rule applies to it. Unlike the HIPAA Privacy Rule, the Common Rule may apply to CDC's use or disclosure of data for human subjects research, even when it had been obtained for a non-research purpose. To use or disclose identifiable data for research, CDC must obtain IRB approval for nonexempt human subjects research to ensure compliance with the Common Rule and the protection of research subjects. Research that collects or uses existing identifiable data can be exempted from federal policy, including IRB review, provided the investigator records information in such a manner that participants cannot be identified, directly or through identifiers linked to the subjects. Nonexempt human subjects research requires that informed consent be obtained and documented, unless an IRB determines that the project satisfies criteria for modification or waiver.

Finally, if federal requirements are satisfied, all agreements that govern states' participation in the SMART Collaborative must be identified and reviewed. CDC's disclosure of data for PCOR needs to be consistent with the terms of any applicable agreements.

Ethical Issues: Balancing Research Progress with Privacy Protections

Patients who undergo ART, donors, and surrogate carriers are vulnerable to significant medical and emotional risks. Thus, in addition to extensive counseling and meticulous informed consent, there must be ironclad safeguards for protecting the privacy and confidentiality of all the people involved in this process.

Currently, NASS data does not include direct patient identifiers. A guest researcher is required to access the data on-site at CDC, rather than remotely, and a member of the ART team within CDC works in a supervisory role with the researcher, as required in CDC's Assurance of Confidentiality. These steps that are undertaken to safeguard patient privacy and confidentiality can also hinder research progress. As the CDC provides supervision and statistical support, access by a guest researcher is limited by the size of the ART team. Financial investments on the part of the researchers to travel to CDC until the completion of the project also pose challenges to research progress. This delays much needed research findings for the benefit of patients and the public. Sharing the data with states or allowing researchers remote access has the potential to improve the process of assisted reproduction and to contribute to the generalizable knowledge.

Data Use Scenario 2 – State Central Cancer Registries (CCR)

Description

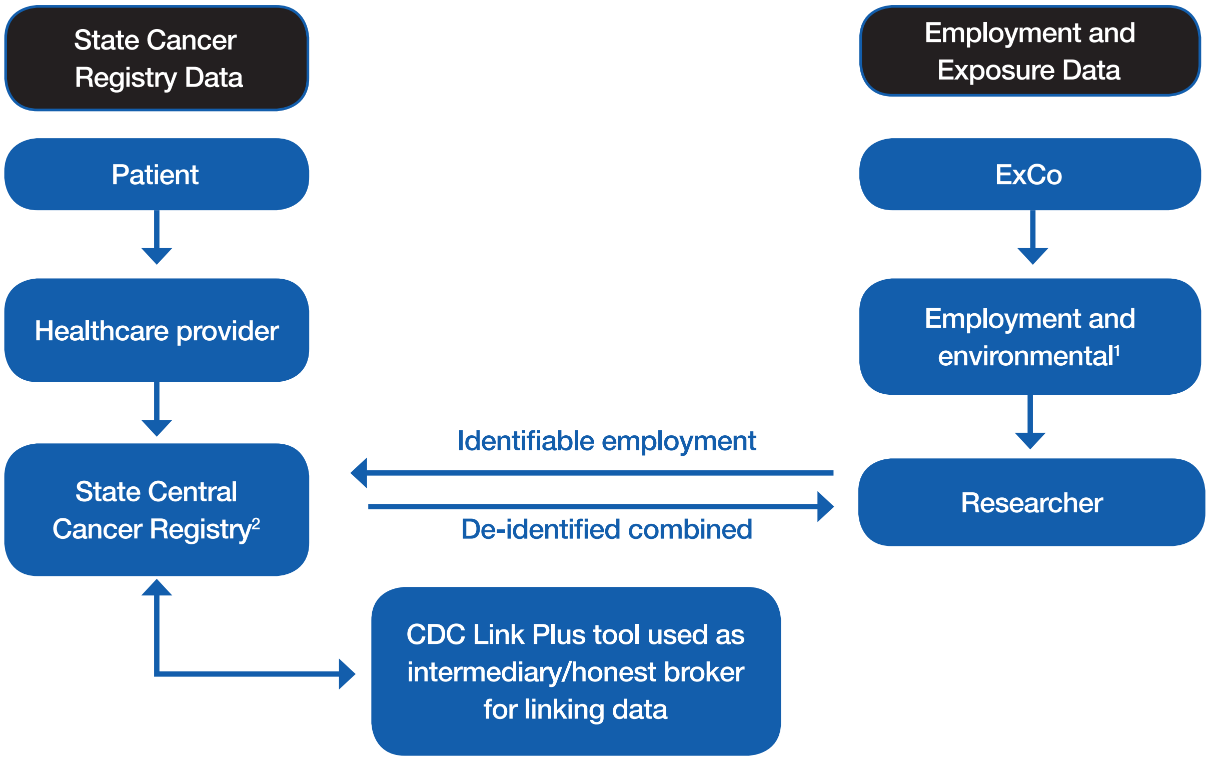
The CCR data use scenario supports understanding of health outcomes for individuals who have been exposed to carcinogens. The CDC would enhance data available to researchers by facilitating access to identifiable data collected by State Central Cancer Registries. These data are supported by CDC's National Program of Cancer Registries (NPCR) Surveillance System. Health facilities that diagnose and treat cancer report identifiable patient information to CCRs, including each patient's cancer type, stage, and treatment. CCRs then send de-identified cancer information to the NPCR Surveillance System. CDC uses these de-identified data to publish official federal statistics on U.S. cancer incidence and deaths.

Since CDC cannot link de-identified cancer data with data for individuals with known exposure to carcinogens, it is not able to provide data that are needed to study the association of exposure and cancer diagnosis. For this hypothetical scenario, CDC would assist a researcher to obtain data from CCRs to study cancer outcomes for individuals who were exposed to carcinogens while working for the fictitious employer, ExCo.[6] ExCo hired a researcher to determine if exposed employees experienced increased cancer rates compared to a comparison group of employees with no known exposure. ExCo would provide all needed employment and environmental records to the researcher. The researcher requests that CCRs match identifiers from employment records of the employees in the exposed and comparison groups with their cancer registry data. States would then provide data that would allow the researcher to determine if cancer incidence is associated with exposure.

The researcher would be responsible for preparing a protocol and obtaining approvals from appropriate IRBs. The exposure, cancers of interest, data sources, and approach to data linkage would be specified in the protocol. CCRs can perform linkages using Link Plus, a freely available record linkages tool for cancer registry programs developed and supported by the CDC. CCRs might perform linkages themselves, or use an intermediary (honest broker). While individual identifiers would be used to link employment and cancer registry data, they would be removed by the CCR or intermediary before providing the combined dataset to the researcher.

[6] A fictitious company created for the scenario

Data Use Scenario: Linkage of Employment Data and Cancer Registry Data to Create Combined Dataset to Identify Possible Adverse Outcomes from Exposure to Contaminant


State Cancer Registry Data
Patient
Healthcare provider
State Central Cancer Registry[2]
Employment and Exposure Data
ExCo
Employment and environmental[1]
Researcher
Identifiable employment
De-identified combined
CDC Link Plus tool used as intermediary/honest broker for linking data

1. Employment records identifiable to individual employees for both exposed employees and comparison employees with no known exposure. Environmental records provide data regarding potential level of exposure.
2. State Central Cancer Registry might link data and return combined dataset to researcher with personal identifiers removed, or could use an intermediary for this purpose.

It is difficult to study associations between cancer and environmental exposures because a cancer diagnosis might occur many years after exposure. Additionally, these studies require data from multiple CCRs because of the likelihood that employees changed residence between the time of exposure and diagnosis or death. Thus, the CCR scenario will explore CDC's potential role to build PCOR capacity by facilitating researcher access to data from multiple states that are essential to these studies. It raises complex legal, ethical, and practical considerations for CCRs to provide cancer data needed for the proposed study.

Legal Issues

The Cancer Registries Amendment Act (CRAA),[88] enacted in 1992, established the NPCR at the CDC to collect data on cancer occurrence, treatment, and outcomes. The CRAA authorizes the CDC to provide funding to states to support the operation of population-based, statewide registries to collect data. To receive funding, a state must have law that authorizes a statewide cancer registry and collect and disclose data to support the goals of the NPCR.[89]

While states collect identifiable cancer data, they report data to CDC with no personal identifiers. This is consistent with the CRAA, which requires that state cancer registries protect the confidentiality of all cancer case data reported to it. Generally, central registries may not disclose information that can identify an individual cancer patient. The CRAA does allow central registries to disclose identifiable information to other state cancer registries and to local and state health officials. Most states have signed a national interstate data exchange agreement,[90] which enables states to capture all cancer cases for their population and eliminate duplicate cancer reports. The national agreement allows disclosure of data for cancer research for projects approved by an IRB, unless a state has specifically restricted such disclosure in the agreement.

Central registries provide individual case reports to CDC that include details about the cancer and also demographic information such as age, race, gender, and county of residence. However, these indirect identifiers might be insufficient to enable matching of case reports to ExCo's former employees. Even if CDC could successfully match ExCo's employees to cancer reports, CDC has obtained an assurance of confidentiality for NPCR pursuant to Section 308(d) of the PHSA. Any CDC effort to determine the identity of any reported cases, or to use the information for any purpose other than statistical reporting and analysis, is a violation of the assurance.[91]

State CCRs might be able to provide relevant data for this research project by matching identifiers in employment records with identifiers in case reports. The Common Rule applies to human subjects research that is supported by the federal government. States, which receive federal funding for registries that collect cancer data, will need to consider whether they are engaged in nonexempt, federally supported human subjects research under the Common Rule.

State CCRs must consider whether providing de-identified cancer reports to the researcher for ExCo's employees that were and were not exposed to the carcinogen constitutes human subjects research. The Common Rule applies to research where the identity of the subject is or may readily be ascertained by the investigator or associated with the information.[92] While each of the cancer reports will correspond to the employee roster, it is unlikely that all employees will have a cancer report. On the other hand, employee information and cancer reports may include indirect identifiers in common. An expert might be needed to evaluate the risk of associating a particular cancer report with a particular employee.

Even if the Common Rule does not apply, each state CCR will likely be subject to state law concerning disclosure of data from the registry. In this regard, the CRAA requires that a state promulgate rules that protect the confidentiality of cancer data that provide for "a means by which confidential case data may in accordance with State law be disclosed to cancer researchers for the purposes of cancer prevention, control and research." CDC's Cancer Registry Data Access for Research Project illustrates the variety and complexity of states' cancer registry processes for gaining approval to access confidential data for research.[7] The project reports that all states require a review before providing cancer registry data for research. Depending on the state, there may be one to four levels of review that may be conducted by an IRB, another committee, the health official, or the registry director.

This scenario contemplates that the CCR will match information on former employees with case reports, and provide cancer information for these former employees without personal identifiers. The risk that individuals might be re-identified from information provided will need to be evaluated, especially since each individual would be known as an employee of ExCo during a certain time period. This study has the potential to impact the rights of ExCo's former employers who developed cancer and their families. For example, it might be used to deny compensation or defeat a liability claim against ExCo based on injuries due to the contamination. This might be a factor weighed by a state in considering this data request.

This data use scenario illustrates challenges in conducting multi-state research that complies with laws and policies that govern each participating state. Potentially, the researcher will need data from numerous registries, each subject to its own legal requirements, ethical principles, and internal policies and procedures.

[7] Described on the CDC National Program of Cancer Registries (NPCR) web page at https://www.cdc.gov/cancer/npcr/data_access/.

The Common Rule allows data providers to avoid duplication of effort by multiple-IRB reviews through cooperative research arrangements.[94] State laws may also allow for joint review arrangements. However, data providers may be reluctant to rely on other IRBs to make decisions regarding data for which they are responsible.

Ethical Issues

Variable Patient Treatment Due to Multiple IRB Review

In this scenario, the researcher must obtain approval from state cancer registries before beginning the study. IRB approval, data access, and informed consent processes are state specific and differ across state registries. Moreover, restrictions on the use and disclosure of cancer information are dictated by state laws and vary across states. Studies have demonstrated that multiple IRB reviews for the same research proposal could lead to diffusion of responsibility and potentially expose research subjects to undue harms.[98]

IRBs are often given discretion in applying and interpreting federal regulations.[99,100] For instance, different IRB systems could require changes to the informed consent form (as per the local and institutional policies). The forms designed for obtaining informed consent from individual cancer patients could vary significantly in ways they present information to patients.[101] Cancer patients could be treated differently in different states, and their autonomy may be compromised when those changes and concerns are not communicated to the other IRBs.[102] Thus, review by multiple IRBs may give rise to variability in the research practices with no ethical justification and may cause confusion among patients and researchers.[103]

Data Use Scenario 3 – National Health Safety Network (NHSN)

Description

The NHSN data use scenario supports systematic collection and analysis of healthcare outcome data for patients who have undergone surgical procedures during which medical devices were used. Currently, healthcare facilities throughout the United States report outcome data to the CDC's NHSN on healthcare-associated infections and other adverse events, such surgical site infections (SSIs). However, SSI data submitted to NHSN do not include specific identifying information about devices that are inserted or used in surgery. In 2014, the Food and Drug Administration (FDA) issued a final rule that calls for operating room staff to document intraoperative use or insertion of medical devices by recording a standard Unique Device Identifier (UDI)—located on the device label or the device itself—for each device. The FDA rule provides an opportunity for CDC to extend NHSN's surveillance scope to include UDI data and accompanying patient-identifying data. This data could track subsequent patient outcomes, including adverse events, which may be associated with prior use or insertion of a medical device intraoperatively.

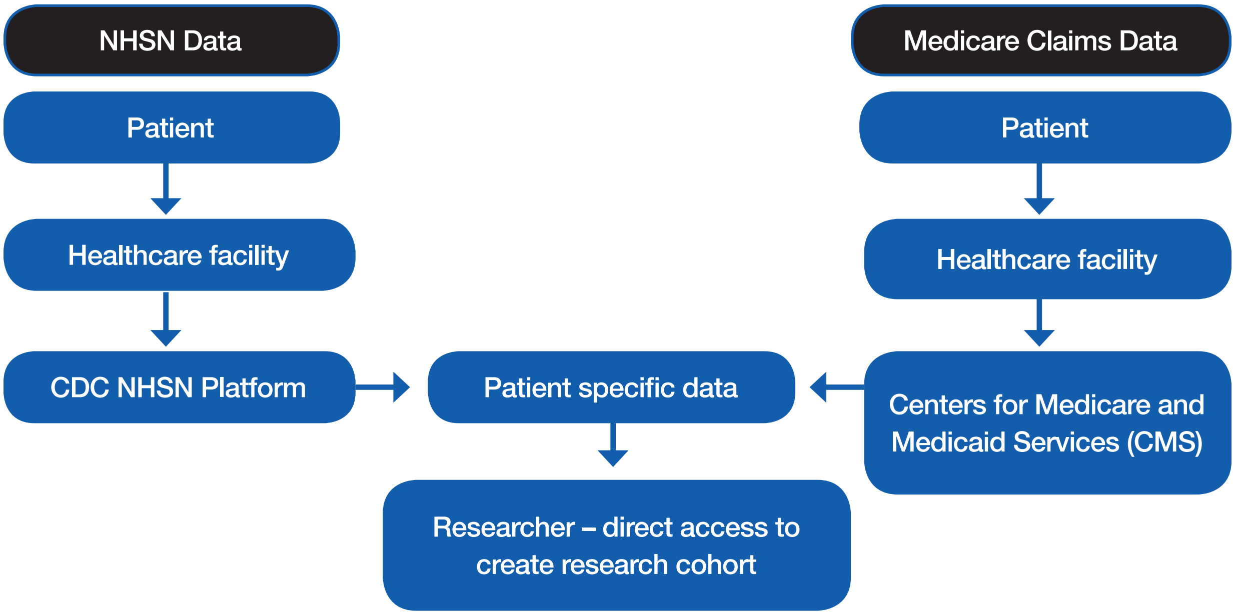
Currently, a mix of state and federal requirements and voluntary incentives provide the impetus for healthcare facilities to report data to NHSN. At times, identifiable data are reported voluntarily by facilities to NHSN to track patients at the facility level. Each healthcare facility that participates in NHSN completes an Agreement to Participate and Consent Form issued by CDC. In agreeing to participate, the facility agrees to provide certain data elements for the components it selects and consents to CDC's use of the data for a set of specified purposes.

Under this scenario, the CDC would enhance data collected through the NHSN and support their use for research by:

- Expanding data elements that NHSN collects from healthcare facilities to include UDIs for devices used intra-operatively in surgical patients

- Enhancing NHSN analytic capacity to enable linkages between NHSN's database and other databases to create combined datasets (e.g. modifying the NHSN platform to allow linkages to combine NHSN data with Medicare claims data)
- Obtaining patient identifiers from healthcare facilities that submit data to allow linkages between NHSN data and Medicare claims data
- Supporting availability of these data for research by expanding permissible uses in the Agreement to Participate and Consent between the CDC and healthcare facilities to allow reported data to be accessed for research
- Facilitating researcher direct access to these data, including remote access, to perform linkages necessary for research

NHSN Data Use Scenario:
Collection, Linkage, and Disclosure of Patient Data
to Identify Adverse Outcomes Associated with UDIs



Potentially, research using enhanced data, as proposed by the NHSN use case, could identify particular medical devices (by type and manufacturer) that are associated with adverse outcomes, including surgical complications, infection, and device replacement.

Legal Issues

This data use scenario contemplates two streams of identifiable data obtained or held by two federal agencies that might be matched by unique identifiers to create a combined dataset. Approved researchers would be able to access this dataset to create research cohorts. The scenario first raises questions about CDC's collection of identifiable data from health facilities. It then raises questions about computer matching of NHSN data and Medicare data, using unique identifiers that the NHSN and Medicare have in common, to create a combined dataset. Finally, it raises questions about providing a researcher direct access to identifiable data in the combined dataset. The legal discussion below includes several laws that might apply to CMS as a data provider. However, to thoroughly review disclosure of identifiable data from CMS to CDC, and permitted linkages, all laws and policies that govern data sharing between federal agencies must be identified and reviewed.

NHSN'S COLLECTION OF IDENTIFIABLE DATA FROM HEALTHCARE FACILITIES AND MEDICARE.

Broad authority in the PHSA authorizes the CDC to obtain data for the NHSN. The HIPAA Privacy Rule allows covered entities to provide identifiable data to the NHSN, without patient authorization, for public health purposes. Public health purposes include preventing or controlling disease, injury, or disability, including, but not limited to, the reporting of disease, injury, vital events such as birth or death, and the conduct of public health surveillance, public health investigations, and public health interventions.[104] Identifiable data includes personal identifiers, such as name and address, as well as indirect identifiers such as birthdates, dates of service, and demographics. As discussed below, a UDI is also, in part, an identifier under the Privacy Rule.

The Privacy Rule would permit healthcare facilities to expand data they report for public health purposes. Similarly, the Privacy Rule would allow CMS to provide data for public health purposes. Identifiable data requested by CDC—now and as proposed by this scenario—must represent the minimum necessary for the public health purpose.[105] In this regard, CDC should be able to articulate the public health purpose to support its data request. A covered entity may rely on a governmental agency's representation of the minimum necessary data needed for the stated purpose.

If CDC obtains identifiable data for the purpose of research, the HIPAA Privacy Rule can govern disclosure to CDC by a covered entity. In this use scenario, both healthcare providers and CMS/Medicare are covered entities. As discussed for the NASS use scenario, if CDC requests identifiable data for the purpose of research or to build a research repository, covered entities must obtain written authorization, modification, or waiver of authorization by an IRB or privacy board based on criteria established by the Privacy Rule. A covered entity might also disclose a limited dataset for research with a data use agreement, provided all HIPAA-required safeguards are satisfied. For a limited dataset, direct identifiers are removed but demographic data and dates (e.g. dates of service and birthdate) are retained.

In disclosing data for research, a covered entity must examine each data element to determine whether it is an identifier. In this regard, the Office for Civil Rights has issued guidance about UDIs, concluding that only a portion of a UDI may be disclosed as part of a de-identified dataset or limited dataset.[106]

A UDI consists of two parts:

- *Device identifier (DI).* This is the mandatory fixed portion of a UDI that identifies the model or version of a device. It is not assigned to a specific device corresponding to a particular individual. Thus, the DI portion of a device's assigned UDI may be used or disclosed as part of a limited dataset or de-identified dataset.
- *Production identifier (PI).* This is a variable portion of a UDI that corresponds to a specific device. It may include lot/batch number, serial number, expiration date, manufacturing date, and donor identification number. Thus, the PI portion of a device's assigned UDI may not be included as part of a limited dataset or de-identified dataset. This means that the Privacy Rule would permit disclosure of the PI for research only with patient authorization or if an IRB or privacy board approve a waiver of authorization.

The Common Rule applies to CDC if it obtains individually identifiable data for purposes of nonexempt human subjects research or to build a research repository. The HIPAA Privacy Rule and the Common Rule differ with regard to identifiability. For example, the PI portion of the UDI, deemed an identifier under the HIPAA privacy rule, may not render data identifiable under the Common Rule.

HIPAA VS. COMMON RULE – DISCLOSING A DE-IDENTIFIED OR LIMITED DATASET FOR RESEARCH

HIPAA Privacy Rule: Information may be de-identified by removing 18 identifiers specified in the Rule, provided that the covered entity does not have actual knowledge that the remaining information can be used alone or in combination with other reasonably available information to identify a subject (safe harbor de-identification). These identifiers include personal identifiers (such as name, address, telephone number, birth date, Social Security number, and numeric codes associated with an individual) and non-personal identifiers (such as geographic information smaller than a state and dates directly associated with an individual). Alternatively, a covered entity may rely on a determination by a properly qualified statistician using accepted analytic techniques who determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information (statistical de-identification). A limited dataset requires specific direct identifiers be excluded, including device identifiers that relate to an individual.

Common Rule: Private information is individually identifiable when the identity of the subject is or may readily be ascertained by the investigator or associated with the information. The Common Rule does not apply to research that uses data that are not individually identifiable.

COMPUTER MATCHING OF NHSN DATA AND MEDICARE DATA

The Privacy Act governs federal systems of records about individuals. It applies only if data are retrieved from records by name or other identifying particular such as a Social Security number, or other identifying number or symbol. If data are primarily retrieved by another variable, the Privacy Act may not apply. For this reason, the Office of General Counsel has opined that the NHSN, as CDC is currently utilizing it, is not a Privacy Act system of records:

> While CDC has the capability to retrieve data by personal identifier, CDC does not, as a matter of practice or policy, retrieve data in this way. Specifically, the primary practice and policy of CDC regarding NHSN data is to retrieve data by the name of the hospital or other non-personal identifier, not an individual patient, for surveillance and public health purposes. Furthermore, patient identifiers are not necessary for NHSN to operate, and CDC does not regularly or even frequently use patient names to obtain information about these individuals.

This data use scenario may require that CDC retrieve data by patient names or other identifying particular. If so, CDC's collection, use and disclosure of data for this scenario needs to be reviewed for compliance with the Privacy Act.

The Privacy Act contains provisions about using identifiable information for computer matching activities, for example, matching among federal datasets or between federal and state or local datasets. These provisions add procedural requirements for agencies to follow when engaging in computer-matching activities and require written agreements specifying the terms under which matches are to be done.[8] These are complex provisions that should be carefully reviewed to determine their applicability and scope.

[8] For example, CMS computer matching agreements are posted at https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/Privacy/ComputerMatchingAgreements.html.

DIRECT ACCESS BY EXTERNAL RESEARCHER TO IDENTIFIABLE DATA

CDC's policies would be applied concerning direct access to identifiable data. Additionally, the terms of any data sharing or participation agreements between CDC and health facilities that report to the NHSN need to be reviewed. CDC has obtained an assurance of confidentiality for the NHSN, pursuant to Section 308(d) of the PHSA. Thus, disclosure and use of identifiable information about individuals or establishments that was obtained with the assurance must comply with the terms of the assurance. Additionally, the Trade Secret Act might be relevant in the event that NHSN data, accessible to the researcher, has been designated proprietary information. If the Assurance permits disclosure, the Privacy Act should be reviewed for permissible disclosure of identifiable information.

Ethical Issues

1. Concerns Regarding Patient Autonomy

If NHSN data is linked to unique personal identifiers, it could compromise patient privacy and confidentiality. In this scenario, a third-party researcher wants to link the identifiable patient information to CMS data. Since the researcher is requesting patient-identifiable information, the patients may be exposed to unknown risks without having consented to the useage of their data.[108] Some patients may be more comfortable with authorizing the use of their data in only some cases. In this scenario, researchers are making this decision for the patients. This scenario raises a threshold question: Should researchers or data collectors notify patients when their data is being collected for public health surveillance? Should patient data collected for surveillance purposes be used for research?

2. Potential Discrimination in Insurance Coverage

Medicare and many third-party insurance payers assess the relative benefits of medical technologies to inform decisions regarding coverage, reimbursement, and pricing.[100,110] Since PCOR aims to inform value-based insurance decisions, it can play a crucial role in insurance pricing and coverage decisions.[111] If the proposed study highlights the complications caused by devices used in joint replacement surgery, an insurance company can choose to cover and reimburse patients based on the study. Devices associated with certain complications may have higher copays, lower insurance coverage, or both. The results of such studies could encourage the uptake of certain devices over others.[112] Potentially, insurance providers could discriminate between individuals who have undergone joint replacement surgery and those who haven't. Such changes could eventually be reflected in other private insurance policies as standard-of-care treatment.

In this scenario, a patient or physician preference in guiding treatment might ultimately be hampered. For instance, some patients may be benefited by rare treatment options, such as joint replacement surgeries by particular devices that may have higher co-pays or out of pocket expenses. The interests and quality of care of such patients may be compromised by coverage restrictions which don't allow to pay for older models of surgical devices. The research study results can potentially influence access to medical interventions and treatment options at the expense of the providers' discretion to make a decision on a case-by-case basis.

7. Legal and Ethical Framework: Providing CDC Data for PCOR

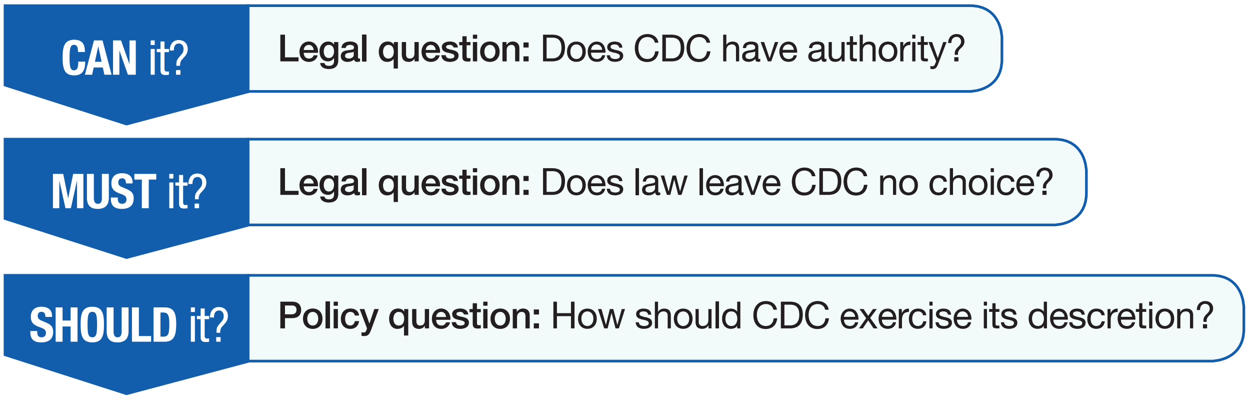
Generally, federal law and policy supports broad access to data collected and maintained by federal agencies. That said, whether a federal agency should release or share a particular dataset requires careful review, which may include the use of professional judgment and the exercise of discretion. This section describes a framework to assist the CDC—as well as its data managers—to decide whether to share data for PCOR. This framework provides a systematic approach that can be used to review data sharing policies, build the PCOR data infrastructure, and share particular data for a specific PCOR proposal.

Overview: questions to consider regarding a request for data for PCOR

Overall, the framework poses three questions to the data manager:

- “Can It?” Does CDC have the legal authority to share data as requested? This is a legal question.
- “Must It?” Is CDC required to share data as requested? For example, a federal agency must provide information requested under the Freedom of Information Act, unless an exemption to disclosure applies. This is also a legal question.
- “Should It?” If the CDC has legal discretion to share data, should it do so? In this regard, what are the ethical considerations that support or weigh against providing the requested data? The “should it” question raises concerns in addition to ethical ones, such as administrative feasibility and cost of data sharing.

Figure 1[9]: Three Basic Questions When Deciding to Share Data



CAN it? **Legal question:** Does CDC have authority?

MUST it? **Legal question:** Does law leave CDC no choice?

SHOULD it? **Policy question:** How should CDC exercise its descretion?

Determining legal authority

While legal questions are easiest when they have clear “yes” or “no” answers, these questions frequently require application of the law to the facts and the use of legal judgment in arriving at a conclusion. While potentially frustrating, the answers to legal questions often depend on the specific facts to determine permissible courses of action.

Thus, to determine the nature and scope of CDC’s legal authority to share data for PCOR, CDC must:

1) establish the facts.
2) identify applicable legal requirements.
3) apply each applicable legal requirements to the facts.

These three steps are described below.

[9] Three Basic Questions” is adapted from the Network for Public Health Law’s “Can I – Must I – Should I” framework to analyze legal authority.

a. Establish the facts

To identify and apply applicable legal requirements, factual details regarding the proposed data sharing are needed. As discussed above, identifying and applying laws to a data sharing scenario requires establishing the following facts:

1) What do the data consist of?
   i. type of data (e.g. health, substance abuse, veterans, student)
   ii. data elements (to determine whether data is identifiable, de-identified, coded)
2) From whom were the data obtained? (i.e. source of data) (e.g. public health authority, healthcare provider, specific healthcare provider such as veterans' health administration or substance use disorder programs, CMS, schools)
3) For what purpose were the data obtained? (e.g. public health, research, quality improvement)
4) Are data subject to contractual restrictions? (i.e. did CDC enter into an agreement with the data provider that sets out terms and conditions for use and disclosure of data?)
5) To whom are the data to be disclosed? (e.g. researcher, "honest broker," another public health authority)
6) For what purpose are the data to be disclosed? (e.g. public health, research, quality improvement)
7) How will the data be disclosed? (e.g. on-site access, remote web-based access)
8) Are the data being linked or combined with other datasets?

b. Identify applicable legal requirements

When reviewing a data request, it is necessary to identify legal requirements relevant to CDC data collection, protection, and disclosure to determine what is permissible. Where multiple legal requirements apply, the requirement that provides the greatest data protection (i.e. most restricts data disclosure) would govern.

Legal requirements within the following general categories might apply:

1) CDC legal authority to collect data. (e.g. general authority, such as the PHSA and/or authority that is specific to a type of data, such as the Cancer Registries Amendment Act).
2) Reporting laws. These are specific laws requiring or authorizing data reporting to CDC.
3) Privacy and confidentiality laws. Examples include:
   i. The Federal Privacy Act and the PHSA Assurance of Confidentiality, which apply directly to CDC
   ii. The HIPAA privacy regulations, which might apply to health care providers who submit data to CDC
   iii. Laws that limit redisclosure of data, i.e. the law does not directly apply to CDC, but limits redisclosure by CDC as a data recipient
4) Laws that protect business interests. (e.g. the Trade Secrets Act and the PHSA Assurance of Confidentiality).
5) Laws that protect participants in research, such as the Common Rule
6) Laws that require that CDC implement security measures to protect data, such as the Federal Information Security Management Act

Which specific laws apply depends on the requested data, the nature of the data recipient, and the proposed use of the data.

c. Apply each applicable legal requirement

The following questions might help in applying each law to the facts. Many of these questions may be already answered through the process of identifying applicable laws.

1) What does the law do?

2) To whom does this law apply?

3) How does this law apply to building capacity for PCOR?

4) Does this law allow CDC to provide the requested data for PCOR? (Note: this question might have a specific yes/no answer, or the answer might be based based on judgment and balancing competing interests if the law involves governmental exercise of discretion.)

5) How does this law support or hinder CDC in providing data?

    i. Does this law allow sharing or access to identifiable data for research?

    ii. Does this law allow release of de-identified information?

        a. Does this law define de-identification or standards to render data de-identified?

        b. Does this law establish prerequisites, conditions, or limitations not previously identified?

3) What remedies or solutions might allow CDC to maximize its ability to provide data while complying with this law?

d. Make a determination: Does law allow CDC to provide requested data for PCOR?

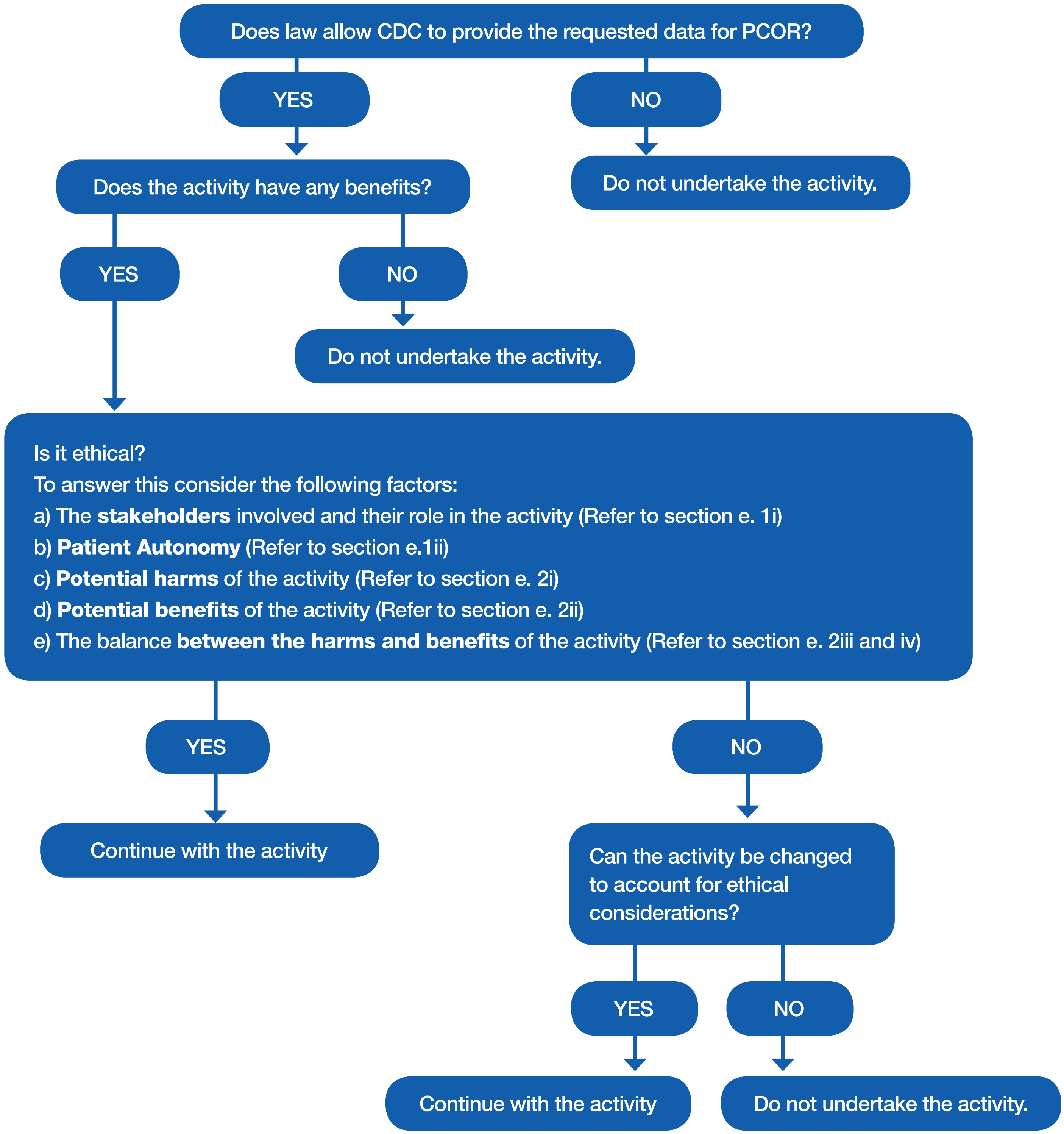
1) If no, what would law allow?

2) If yes, what are the ethical and policy implications?

e. Identify and apply ethical principles to PCOR

The ethical principles of “respect for persons,” “beneficence,” and “justice” were developed to promote the protection of human subjects in the context of research. This section expands on the applications of these principles and uses assistive questions to provide guidance on how to apply the ethical principles to evaluate data-intensive research such as PCOR.

Decision Tree for Applying Ethical Principles to PCOR



Does law allow CDC to provide the requested data for PCOR?

YES

NO

Does the activity have any benefits?

Do not undertake the activity.

YES

NO

Do not undertake the activity.

Is it ethical?
To answer this consider the following factors:
a) The **stakeholders** involved and their role in the activity (Refer to section e. 1i)
b) **Patient Autonomy** (Refer to section e.1ii)
c) **Potential harms** of the activity (Refer to section e. 2i)
d) **Potential benefits** of the activity (Refer to section e. 2ii)
e) The balance **between the harms and benefits** of the activity (Refer to section e. 2iii and iv)

YES

NO

Continue with the activity

Can the activity be changed to account for ethical considerations?

YES

NO

Continue with the activity

Do not undertake the activity.

1. Respect for persons:

To apply, first identify stakeholders who can be affected by a research study, by virtue of their interests, involvement, or relationship to the study (i.e., data collector, data recipient, researcher, or research subject).

i. Stakeholder identification

1. Primary stakeholders are entities that can be affected by the harms or benefits of any research study.[113] For instance, individuals contributing their data, certain groups or communities, and vulnerable populations.
2. Key stakeholders are entities that can "significantly influence, or are important to the success [or failure] of the project." [113] This includes the researchers, data collectors, and data stewards.

ii. Stakeholder analysis: identify the entities that[113]

1. need to provide consent
2. may bear the burdens of the research study
3. will benefit from the research activity
4. are crucial to mitigating risks and preventing harms[113]

Assistive questions[113] for evaluating stakeholders involved in a research study:

1. Is it possible or practicable to identify the rights, duties, and responsibilities of the various stakeholders involved in the research study?
2. Which stakeholders may be most at risk for harms through disclosure of sensitive information?
3. Is it possible to reasonably identify and contact research participants to obtain informed consent?
4. Is it possible to identify all vulnerable groups that may be affected?

iii. Patient Autonomy: In data-intensive studies like PCOR, the study typically involves accessing existing data and subjects are not typically in direct contact with the researcher or the data steward.[113] There are often concerns regarding the feasibility of obtaining consent from individual patients to use their data for research purposes. Researchers should be mindful that a patient's dignity is increasingly integrated with the data he/she contributes and should ensure that the confidentiality of patients' information is protected.[113]

Assistive questions for determining the need to obtain consent: [113]

1. Does the proposed study require or create data that can reveal the name, location, relations, or other identifying information of an individual?
2. Is it possible or practicable to obtain consent from patients for secondary uses of their data? If yes, what type of consent is appropriate?
3. A "broad consent" could be obtained in advance from patients authorizing future uses of their data for research purposes.
4. A dynamic consent model allows patients to tailor their data sharing preferences to a wider variety of research initiatives, in a more flexible manner.[114] Patients are contacted via a digital interface to seek their consent whenever their data are used for research purposes.

5. Is it possible for individuals to decline to participate in the research uses of collected data?

6. Are the following justifications for not obtaining informed consent present?[113]

   a. Foregoing consent is necessary to accomplish research goals.

   b. All known risks to research participants are minimal.

   c. There is an adequate plan for debriefing subjects, in case of unexpected harms or risks to research subjects.

   d. Obtaining consent has a negative impact on the data quality and is not just an inconvenience to the researcher.

2. Beneficence and Non-maleficence

i. Identify risks: In data–intensive research such as PCOR, privacy breaches or risks of re-identifications do not account for all potential harms, as there may also be risks to the integrity of a data system used for research. Therefore, one should consider different categories when identifying risks, discussed below.

Assistive questions to evaluate risks to integrity of data[113]

1. Does the research involve risks to data quality and integrity that may affect future research studies?

2. Are researchers considering risks to the integrity of data and information systems that store and process data?

Assistive questions to evaluate risks to patient privacy and confidentiality[113]

1. Does the research involve data that indirectly identifies an individual?

2. Have researchers considered the number of individuals who may be negatively affected by research activities?

3. Does the researcher plan to disclose data (with anonymization or de-identification) as part of research publication?

4. Have the risks of re-identification been considered? How accessible are secondary data sources that can be combined with published data to re-identify individuals?

5. What are the possible risks associated with the use or disclosure of research data? For instance, public disclosure, compelled disclosure, or government disclosure.

6. What is the severity of potential harms to all individuals who may be affected by research activities (e.g., use or disclosure of data, or publication of research results)?

7. Will the research study be reviewed by an IRB or an ethics review committee to account for the unintended consequences that may result from the study?

ii. Identify potential benefits: As burdens of research are borne primarily by research subjects, it is important to conduct a fair assessment of reasonably foreseeable benefits from the proposed research studies.[113]

Assistive questions to determine the benefit of the study: [113]

1. Does the research study clearly benefit society?
2. Is it possible to identify the short-term or long-term benefits for all involved stakeholders?
3. Can the research results be acted upon meaningfully by an intended beneficiary?

ii. Balance harms and benefits: It is important to weigh the burdens of research and direct and indirect harms against the benefits that the research study may yield. Because it is difficult to assess the future harms or benefits of data-intensive research, it is important to revisit the existing guidance on ethical evaluation of a research study. [113]

Assistive questions to balance risks and benefits: [113]

1. Are data de-identified where reasonably possible? Can pseudonyms or other forms of statistical controls be utilized?
2. Are data secured against threats to privacy, data integrity, or disclosure?
3. How easy is it to obtain external sources of data for linkage?
4. Should any additional factors be considered in the evaluation of and justification for certain harms?

ii. Mitigation of harms: It is important to develop procedures for mitigating the harms inherent to research involving human participants. It can be challenging to determine what type of situation requires additional protections because it exceeds a threshold of "minimal risk," especially when a study involves the use of sensitive data. [113]

Assistive questions to reduce risks and mitigate harm: [113]

1. Which organization is in the best position to mitigate harms (e.g., the data collection agency or the data recipient's organization)?
2. What warrants notification of a breach of sensitive research data or unauthorized disclosure of personal information?
3. How can harm be mitigated if it is not possible to notify individuals?

3. Justice

Assistive questions to address justice issues: [113]

1. Does the proposed study select research subjects based on gender, ethnicity, or other attribute? If so, is there a justification for this type of selection?
2. Does the proposed study treat the individuals or groups involved in an equitable manner? If not, is there a justifiable rationale for differential treatment?
3. Can the results of the study lead to social discrimination? Are there safeguards against uses of research results for social discrimination?
4. Does the research disproportionately benefit certain groups or individuals? If so, do other individuals or groups shoulder the research burdens?
5. Is there a fair system for appropriately compensating groups who are burdened?

f. Recommendations for Enhancing Benefits and Mitigating Harms

1. Techniques to mitigate risks to patient privacy and confidentiality

Generally, modified data content and restricted data access can prevent unauthorized disclosure of health data. With modified data, researchers must accept some loss of information granularity. When restricting data, data collectors retain complete data and limit access to the data for qualified users for specific purposes.[115]
The following are some of the techniques that can be considered.[116]

i. Making information available in a way that individual data items cannot be uniquely attributed to a particular individual or establishment.

ii. Distributing modified datasets whose variables have the same statistical distributions and relationships as the original data from which they are derived but do not contain any actual identifying information from the original data.

iii. Restricting access to detailed data to authorized individuals and training those individuals in confidentiality protection. Data enclaves can be set up to monitor the use of very sensitive data. An analyst can be physically present at the restricted site, or remote access can be provided to authorized users.

iv. Making information available under licensing or data use agreements that guarantee secure and confidential handling of data by trusted researchers.

v. Creating custom use files for special requests (honest broker).

2. Techniques for enhancing benefits

The following recommendations are intended to address (1) improving the transparency of the research process and (2) improving the public's awareness about secondary uses of public health data.

i. Prioritize effective communication of the research results and transparency of the research process.

Sharing study results with research subjects can make patients feel more involved in the process and could encourage more people to participate in future studies. Researchers should prepare well-written comprehensive technical summaries since incorrect interpretation of the research results can cause undue stress to study participants.[117]

ii. Educate the public about the benefits of secondary analyses of public health data.

Educating patients about how secondary research with public health data is conducted could increase trust in the research community. It is also important to stress the importance of complete and representative datasets for research studies.[117] An incomplete dataset can lead to biased results and inaccurate conclusions. Thus, conveying the negative impact of incomplete datasets on research findings may increase the public's willingness to support secondary analyses of public health data.

g. Establish and document terms for data sharing

If CDC determines that law, ethical principles, and internal policies are satisfied, it should establish and document the terms for data sharing in a data sharing and use agreement or similar document. These terms should, *inter alia*, describe the data to be shared, the purpose for data sharing, permissible uses, privacy and security requirements, monitoring compliance, and enforcement.

8. Conclusion

Sharing data for PCOR supports CDC's mission by providing data that may be used to prevent disease and improve the health of individuals and the community. PCOR focuses on the availability of data to answer questions important to individuals, healthcare providers, policymakers, and others to improve health outcomes. Consistent with CDC's mission, PCOR compares outcomes for special populations and continues the federal government's commitment to treating data as an asset.

PCOR benefits that are directly applicable to an individual or population can provide an ethical justification for CDC data collection. The knowledge generated from public health research can be made more interpretable to patients. A person may feel less reticent to be involved in research if they know that the knowledge generated from research will improve their personal health. Mistrust towards public health data collection can be ameliorated by improving patient outcomes with the use of CDC's data. Were CDC to improve outcomes for patients who use surgical devices, as hypothesized in the NHSN data use scenario, the value of public health data collection would be further defensible and recognizable. This practice strengthens the public's trust in CDC by engaging individuals who recognize the utility of data that applies to them directly.

Aggregating data for PCOR allows researchers to target patient populations without having to hypothesize a harm and test it on patient subjects. Aggregated data could help researchers to find links between negative outcomes and a patient's genetic makeup or patient history. A person might have a reaction to a drug interaction that is difficult for their provider to detect, but can be noticed by a researcher studying that particular drug combination. The use of patient data in this way can be a safe alternative to clinical trials and can improve patient outcomes without having to observe patient outcomes in person.

Multiple laws, ethical considerations, and CDC polices must be satisfied to provide data for PCOR. It takes persistence and commitment to address laws and identify pathways that support data sharing. While law envisions a robust infrastructure that provides access to CDC data, no law specifically facilitates this goal. Instead, CDC programs and their attorneys must navigate through multiple laws that may be silent on key factors, inconsistent, or ambiguous and subject to multiple interpretations. While this paper provides a framework to systematically work through these laws, the task requires substantial effort and resources. Investment of these resources may hinge on recognition of the value of identifying pathways to data sharing.

CDC identified some potential improvements to the use of public health data for PCOR purposes. It is not recommended to invest in PCOR infrastructure without further study of the burden of a CDC implemented PCOR framework. For example, CDC might consider agency-wide policies that address multiple legal standards for identifiability of data or assist staff to better weigh competing ethical considerations that balance individual and societal concerns. CDC might consider additional options for developing solutions that promote disclosure while protecting privacy and maintaining trust. For example, CDC might expand its arsenal of statistical de-identification techniques or matching techniques that do not depend on individual identifiers.

In order to further PCOR, a CDC program will have to prioritize the need for research to influence patient outcomes. This changes the use and collection of data at every stage. In order to link outcomes to populations, genders, races, and geographic locations, a research proposal might consider the use of identifiable data. There is no simple answer for how to improve the use of public health data to further PCOR at CDC. With no regulatory framework that allows the linkage of data at federal or state levels, researchers must interpret a patchwork of various federal and state laws to collect types and amounts of data needed for comparative effectiveness research and PCOR. An understanding of the process of data sharing, showing the legal and ethical factors to consider, can help data stewards and researchers identify the ways in which they can use datasets for patient-centered purposes.

9. Glossary

Anonymized data: Anonymized data are irreversibly unlinked from all patient identifiers. De-identified data that do not retain re-identification codes are considered to be anonymized.

Data linkage: This is a process of pairing records from two files and trying to select the pairs that belong to the same entity or individual. The linkage may be conducted between two distinct data sources or within a single dataset to identify multiple entries (e.g. re-admissions) for one person or record unit.[119,120]

De-identified data: De-identified data (e.g., aggregate statistical data or data stripped of individual identifiers) require no individual privacy protections and are not covered by the Privacy Rule. [Refer to Appendix A]. De-identifying can be conducted through

- Statistical de-identification: A qualified statistician using accepted analytic techniques concludes the risk is substantially limited that the information might be used, alone or in combination with other reasonably available information, to identify the subject of the information[121] [Refer to Appendix B]
- Safe-harbor method (HIPAA standard): A covered entity or its business associate de-identifies information by removing 18 identifiers, and the covered entity does not have actual knowledge that the remaining information can be used alone or in combination with other data to identify the subject.[121]

Direct identifier: Direct identifiers include information that relates specifically to an individual such as the individual's name, address, Social Security number or other identifying number or code, telephone number, e-mail address, or biometric record.

Honest broker: The role of the honest broker is to collect and provide health information to research investigators in such a manner whereby it would not be reasonably possible for the investigators to identify the subjects directly or indirectly. The "honest broker" acts as a well defined barrier between the clinical environment (in which fully identified confidential patient information is routinely exchanged as part of medical care) and the general research community (in which all information must be de-identified).[122]

Indirect identifiers: Indirect identifiers include information that do not directly identify an individual. However, they include information that can be combined with other information to identify specific individuals, including, for example, a combination of gender, birth date, geographic indicator and other descriptors. Other examples of indirect identifiers include place of birth, race, religion, weight, activities, employment information, medical information, education information, and financial information.

Individually identifiable information: Individually identifiable health information is information, including demographic data that can be used to identify the individual directly or there is a reasonable basis to believe that the that information can be used alone, or in combination with other reasonably available information to identify the individual. Examples include demographic information, medical history of a patient, the details of the health care provided to a patient, or information about the payment for the provision of health care for the patient. [121]

Limited dataset: A “limited dataset” is a limited set of identifiable patient information as defined in the Privacy Regulations issued under HIPAA. In a limited dataset, the “facial” identifiers (information that relates to the individual or his or her relatives, employers or household members) have been removed. The health information that may remain in the information disclosed includes:[123]

- dates such as admission, discharge, service, DOB, DOD
- city, state, five digits or more ZIP code
- age in numbers, days, or hours

Minimal risk: The Common Rule defines minimum risk—for non-prisoners—as risk in which “the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.”[124]

Notifiable disease: Any disease that is required by law to be reported to government authorities.

Probabilistic matching: It assigns comparison outcomes to the correct, or more likely, decision by using likelihood ratio theory. It typically assigns a percentage indicating the probability of a match. For example, probabilistic systems might check all possible name alternatives and consider variables such as nicknames, phonetics, transposed last and first names, and use of initials (Chuck L. Jones versus Chuck Lawrence Jones or C. Lawrence Jones).[125]

Protected health information: The HIPAA Privacy Rule defines “protected health information” as information:

- in any form: written, electronic or oral
- relating to past, present or future
    - physical or mental health status or condition
    - provision of health care
    - payment for provision of health care

that identifies the individual or for which there is a reasonable basis to believe can be used to identify the individual.[126, 127]

Re-identification: Re-identification, also known as identity disclosure, occurs when a person with unauthorized access to data makes a likely match between a de-identified record and the corresponding record in the identified dataset.[70]

Vulnerable population: The term “vulnerable population” refers to a disadvantaged sub-segment of the community. Their freedom and capability to protect themselves from intended or inherent harms is variably abbreviated, from decreased free will to inability to make informed choices.[128]

Appendix A. De-Identification: As Described by Select Federal Statutes

Law	Provision(s) that allow disclosure of de-identified information	Criteria or standard for determining whether information is identifiable
Health Insurance Portability and Accountability Act of 1996 (HIPAA), Public Law 104-191, implemented by the HIPAA Privacy Rule, 45 CFR Part 160 and Part 164.	The HIPAA Privacy Rule applies to protected health information (PHI). The Privacy Rule does not apply to health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual. 45 CFR § 160.103, 45 §164.500.	Information may be de-identified by removing 18 identifiers specified in the Rule, provided that the covered entity does not have actual knowledge that the remaining information can be used alone or in combination with other reasonably available information to identify a subject (safe harbor de-identification). These identifiers include personal identifiers (such as name, address, telephone number, birth date, Social Security number) and non-personal identifiers (such as geographic information smaller than a state and dates directly associated with an individual). Alternatively, a covered entity may rely on a determination by a properly qualified statistician using accepted analytic techniques who determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information (statistical de-identification). 45 CFR § 164.514.
Protection of Human Research Subjects (Common Rule), 45 CFR part 46, subpart A.	The Common Rule applies when an investigator conducting research **obtains** identifiable "private information" of a living individual (human subject) for use, study, or analysis. Private information must be **"individually identifiable"** for the Common Rule to apply. 45 CFR § 46.102(f).	Private information is individually identifiable when the identity of the subject is or may readily be ascertained by the investigator or associated with the information. 45 CFR § 46.102(f). Note: In its application of the law, the Office for Human Research Protections (OHRP) considers private information or specimens not to be individually identifiable when they cannot be linked to specific individuals by the investigator either directly or indirectly through coding systems. Examples of identifiers would include names, Social Security numbers, medical record numbers, or pathology accession numbers, or any other "code" that permits specimens or data to be linked to individually identifiable living individuals and perhaps also to associated medical information. https://humansubjects.nih.gov/from-applicants https://www.hhs.gov/ohrp/regulations-and-policy/guidance/research-involving-coded-private-information/index.html

Law	Provision(s) that allow disclosure of de-identified information	Criteria or standard for determining whether information is identifiable
Federal Privacy Act 5 U.S.C. § 552a.	The Federal Privacy Act establishes a code of fair information practices that governs the collection, maintenance, use, and dissemination of information about individuals that is maintained in systems of records by federal agencies. The Act protects a "record" of a U.S. citizen or alien lawfully admitted for permanent residence. A "record" includes any item, collection, or grouping of information about an individual that is maintained by a federal agency, including, but not limited to, his education, financial transactions, medical history, and criminal or employment history and that contains his name, or the identifying number, symbol, or other identifying particular assigned to the individual, such as a finger or voice print or a photograph. 5 U.S.C. § 552a(a)(4).	The law does not define or describe de-identification directly, but suggests that a record is de-identified by removing all "identifying particulars." 5 U.S.C. § 552a(a)(4).
Federal Assurance of Confidentiality, Section 308(d) of the Public Health Service Act, 42 U.S.C. § 242m.	This law prohibits use, release, and publication of information, if an establishment or person supplying the information or described in it is identifiable. Applies to information obtained in the course of health statistical, epidemiological, or other activities obtained in the course of certain activities undertaken or supported under the Public Health Service Act.	The law does not define or describe de-identification directly.

Appendix B. Statistical de-identification techniques

Data management and statistical controls are used to provide the most meaningful data possible while protecting privacy. Data management controls protect data from (and monitor data for) inappropriate access and use and may include administrative (e.g. data use agreements, data return or destruction policies), physical (e.g. access to data), and technical (e.g. encryption, audit trails) policies and practices. The discussion below covers statistical controls. Statistical controls are used to reduce the risk of data re-identification while minimizing loss of data to maintain data utility. These controls are also called "statistical de-identification" methods. They may be collectively called "data masking" techniques because they help to conceal identities of individuals, although "data masking" is also used to describe specific techniques to conceal identities that add "noise" or pseudo information.

Generally, the process of statistical de-identification involves applying one or more of the following techniques. These techniques overlap and may be called by different names.

1) Generalization. This may also be called "grouping." A process to reduce the precision of a data field. For example, a date can be generalized to a month and year or to a five-year interval. Data with low volumes might be grouped, such as ethnicities with small numbers. While the process maintains the truthfulness of the data, it aggregates data or replaces it with data that is less precise (reduces granularity).

    a. "Data Aggregation" is the compiling of individual data so that the totality of the information is represented by a higher-level classification group. Examples: Rolling up diagnostic codes to higher levels if considered sensitive. Representing case reports for HIV as number of cases per county.

2) Deleting. For example, the last two digits might be deleted from a five-digit zip code, which is then used as a geographic identifier. Essentially, this results in "generalization" or "grouping" of data, as described above.

3) Suppression. A process for replacing a value in a dataset with a missing or NULL value to prevent the identification of individuals in small groups or those with unique characteristics For example, a 50-year-old mother in a birth registry would be an outlier and easily identifiable, so her age value would be suppressed. In cases where re-identification is of greater concern because case numbers are low or a condition is rare, data suppression might be considered. Alternatively, data might be "grouped" as described above.

4) Subsampling. A process for releasing only a simple random sample of the dataset rather than the whole dataset. For example, a 50% sample of the data may be released instead of all of the records.

5) Masking (Perturbing) Data. Masking is a disclosure limitation method that is used to hide the original values in a dataset to achieve data privacy protection. This approach uses various techniques to replace sensitive information with realistic but inauthentic data or modifies original data values based on pre-determined masking rules. Methods include the addition of statistical noise, data swapping, or controlled rounding—resulting in "pseudo" information that reduces disclosure risk. For example:

    a. "Rounding" may involve adding statistical noise to values <10; this blurs data by adding a specific value to some case values, but simple statistics and distributions remain the same.

    b. "Data swapping" swaps information from one individual within the same sample to another individual with similar characteristics in the sample, resulting in pseudo cases. While an individual record does not represent any one individual, it still produces the same simple statistics and distributions as those produced by the original data

c. “Synthetic data” may involve substituting data that returns same rates for actual data, such as using cases from neighboring counties with similar demographics. (Similar to “data swapping”).

Resources regarding de-identification techniques:

“Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule” go to http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html.

Public Health Data Dissemination Guidelines: NAHDO Working Technical Paper Series (July 2005), available at https://www.nahdo.org/sites/nahdo.org/files/Resources/Data_Release_Access_and_Pricing/PH%20Data%20Dissemination%20Guidelines-2005.pdf.

Guidance Document on Creating and Releasing Hospital and Facility Discharge Data Public Use Files (January 2012), available at https://www.nahdo.org/sites/nahdo.org/files/publications/PUF%20Guidance%20Doc%20Final.pdf.

Endnotes

1. Establishing the Definition of Patient-Centered Outcomes Research. Patient-Centered Outcomes Research Institute website. http://www.pcori.org/establishing-definition-patient-centered-outcomes-research. Published March 13, 2012. Updated July 15, 2014. Accessed May 17, 2017.
2. 42 U.S.C. § 1320e.
3. Research & Results. Patient-Centered Outcomes Research Institute website. http://www.pcori.org/research-results/patient-centered-outcomes-research. Published May 8, 2012. Updated November 7, 2013. Accessed May 17, 2017.
4. 42 U.S.C. § 1320e.
5. Patient Protection and Affordable Care Act, 42 U.S.C. § 18001 (2010).
6. 42 U.S.C. § 1320e.
7. PCORI Board Approves $142.5 Million to Fund Expansion Phase of PCORnet, the National Patient-Centered Clinical Research Network. Patient-Centered Outcomes Research Institute website. http://www.pcori.org/news-release/pcori-board-approves-142-5-million-fund-expansion-phase-pcornet-national-patient. Published July 21, 2015. Accessed May 17, 2017.
8. Types of Surveillance Systems. The Public Health Observer website. http://publichealthobserver.com/types-of-surveillance-systems. Published January 4, 2010. Accessed May 5, 2017.
9. Williams H, Spencer K, Sanders C et al. Dynamic Consent: A Possible Solution to Improve Patient Confidence and Trust in How Electronic Patient Records Are Used in Medical Research. Eysenbach G, ed. *JMIR Medical Informatics*. 2015; 3(1):e3. doi:10.2196/medinform.3525.
10. Gliklich R, Dreyer N, Leavy M. *Registries for Evaluating Patient Outcomes*. 3rd ed. Rockville, MD: Agency for Healthcare Research and Quality; 2014. https://www.ncbi.nlm.nih.gov/books/NBK208616/. Accessed May 5, 2017.
11. 42 U.S.C. § 241(a) (Public Health Service Act, § 301(a)), pertaining to CDC's broad public health authority to conduct research and investigations, and 42 U.S.C. § 242k (Public Health Service Act, § 306), pertaining to the collection of statistical data.
12. Policy on Public Health Research and Nonresearch Data Management and Access. Centers for Disease Control and Prevention website. http://masoapplications.cdc.gov/Policy/officialPolicy.aspx?pID=385. Published April 16, 2003. Updated January 26, 2016. Accessed April 11, 2017.
13. Gostin L, Wiley L. *Public Health Law: Power, Duty, Restraint*. Oakland, CA: University of California Press; 2016.
14. CSELS statement on data management and access, Version 1.0 (2016 06 21).
15. 2017 National Notifiable Infectious Conditions (Historical). Centers for Disease Control and Prevention website. www.cdc.gov/nndss/conditions/notifiable/2017. Accessed April 12, 2017.
16. General Help for CDC WONDER. Centers for Disease Control and Prevention website. https://wonder.cdc.gov/wonder/help/main.html. Updated August 24, 2016. Accessed April 11, 2017.
17. NCHS Research Data Center (RDC). Centers for Disease Control and Prevention website. https://www.cdc.gov/rdc/index.htm. Updated December 16, 2015. Accessed April 11, 2017.
18. Restricted Data. Centers for Disease Control and Prevention website. https://www.cdc.gov/rdc/b1datatype/dt100.htm. Updated July 18, 2016. Accessed April 11, 2017.
19. Executive Order 13642, which was issued May 9, 2013, is available at https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government- Accessed Feb. 2, 2017.
20. A Primer on Machine Readability for Online Documents and Data. Data.gov website. https://www.data.gov/developers/blog/primer-machine-readability-online-documents-and-data. Published September 24, 2012. Accessed May 17, 2017.
21. CDC Policy on Distinguishing Public Health Research and Public Health Nonresearch. Centers for Disease Control and Prevention website. https://www.cdc.gov/od/science/integrity/docs/cdc-policy-distinguishing-public-health-research-nonresearch.pdf. Accessed May 17, 2017.
22. Policy on Human Research Protections. Centers for Disease Control and Prevention website. https://www.cdc.gov/od/science/integrity/docs/cdc-policy-human-research-protections.pdf. Accessed May 17, 2017.
23. 45 CFR § 46.101 et seq.
24. *The Belmont Report.* U.S. Department of *Health and Human Services website.* https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report. Published April 18, 1979. Updated March 15, 2016. Accessed May 17, 2017.
25. 42 U.S.C. § 241(a); 42 U.S.C. § 242k.

26. Dusetzina SB, Tyree S, Meyer AM, et al. *Linking Data for Health Services Research: A Framework and Instructional Guide, An Overview of Record Linkage Methods*. Rockville, MD: Agency for Healthcare Research and Quality; 2014. https://www.ncbi.nlm.nih.gov/books/NBK253312/. Accessed May 18, 2017.
27. 5 United States Code (U.S.C.) 552a.
28. Privacy Act Frequently Asked Questions. Centers for Disease Control and Prevention Website. https://www.cdc.gov/SORNnotice/PrivacyFAQ/index.htm. Updated April 11, 2012. Accessed May 18, 2017.
29. 45 C.F.R. Part 5b.
30. 5 U.S.C. § 552a(a)(4)
31. 42 U.S.C. § 242m(d)
32. Certificates and Assurances of Confidentiality. Centers for Disease Control and Prevention website. https://www.cdc.gov/od/science/integrity/confidentiality/. Accessed April 20, 2017.
33. 44 U.S.C. § 3501 *et seq*. (Title V of the E-Government Act of 2002), available at http://www.eia.gov/cipsea/cipsea.pdf.
34. 45 CFR Part 46
35. CDC Policy on Distinguishing Public Health Research and Public Health Nonresearch. Centers for Disease Control and Prevention Website. https://www.cdc.gov/od/science/integrity/docs/cdc-policy-distinguishing-public-health-research-nonresearch.pdf. Published July 29, 2010. Accessed June 15, 2017.
36. 45 CFR 46.116(d)
37. Human Subject Regulations Decision Charts. U.S. Department of Health and Human Services website. http://www.hhs.gov/ohrp/regulations-and-policy/decision-trees/index.html. Updated February 16, 2016. Accessed June 15, 2017.
38. National Health Registry Activities and 45 CFR part 46 (2011). Department of Health and Human Services website. http://www.hhs.gov/ohrp/regulations-and-policy/guidance/regarding-application-of-45-cfr-46-to-national-health-registry/index.html. Published October 14, 2015. Accessed June 15, 2017.
39. Clinical Data Registries - OHRP Correspondence (2015). U.S. Department of Health and Human Services website. http://www.hhs.gov/ohrp/regulations-and-policy/guidance/june-25-2015-letter-to-robert-portman/index.html. Published June 25, 2015. Accessed June 15, 2017.
40. Engagement of Institutions in Human Subjects Research (2008). U.S. Department of Health and Human Services Website. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/guidance-on-engagement-of-institutions/. Updated March 7, 2016. Accessed June 15, 2017.
41. 45 CFR Parts 160 and 164.
42. Pub. L. 104-191, 42 U.S.C. §300gg et seq.
43. Permitted Uses and Disclosures: Exchange for Public Health Activities. Office of the National Coordinator for Health Information Technology and U.S. Department of Health and Human Services Office for Civil Rights website. https://www.healthit.gov/sites/default/files/12072016_hipaa_and_public_health_fact_sheet.pdf. Published December 2016. Accessed June 15, 2017.
44. 45 CFR § 164.512
45. 45 CFR § 164.520
46. 164 CFR § 164.501
47. Research Repositories, Databases, and the HIPAA Privacy. U.S. Department of Health and Human Services National Institutes of Health website. https://privacyruleandresearch.nih.gov/research_repositories.asp. Published January 2004. Accessed May 5, 2017.
48. 45 CFR § 164.514. See, OCR Guidance: Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Available at http://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf ORC guidance on de-identification.
49. 45 CFR § 164.514(c)
50. 45 CFR 164.514(e)
51. 45 CFR § 164.512(i)
52. 45 CFR Part 164, Subpart C
53. 18 U.S.C. § 1905
54. 5 U.S.C. §552(a)(8)
55. 5 U.S.C. § 552
56. 5 U.S.C. §552(a)(8)
57. Title III of E-Government Act of 2002, 44 U.S.C. § 3541 *et seq*., available at https://www.gpo.gov/fdsys/pkg/PLAW-107publ347/pdf/PLAW-107publ347.pdf

58. Gliklich R, Dreyer N, Leavy M. *Registries for Evaluating Patient Outcomes*. 3rd ed. Rockville, MD: Agency for Healthcare Research and Quality; 2014. https://www.ncbi.nlm.nih.gov/books/NBK208616/. Accessed May 5, 2017.
59. Mastroianni AC, Faden R, Federman D, editors. Justice in Clinical Studies: Guiding Principles, *Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies* (Vol. 1, pp 75-83). Washington, DC: National Academy Press; 1994.
60. Responsible Conduct in Research. Human Subjects Education Module at Winona State University; 2002; Winona, Minnesota.
61. Williams H, Spencer K, Sanders C, et al. Dynamic Consent: A Possible Solution to Improve Patient Confidence and Trust in How Electronic Patient Records Are Used in Medical Research. Eysenbach G, ed. *JMIR Medical Informatics*. 2015; 3(1):e3. doi:10.2196/medinform.3525.
62. Pritts, JL. The Importance and Value of Protecting the Privacy of Health Information: The Roles of the HIPAA Privacy Rule and the Common Rule in Health Research. National Academies of Science Engineering and Medicine website. http://www.nationalacademies.org/hmd/~/media/Files/Activity%20Files/Research/HIPAAandResearch/PrittsPrivacyFinalDraftweb.ashx. Published 2008. Accessed May 5, 2017.
63. Anderson JR, Schonfeld TL. Patient consent in the era of de-identified research databases. *J Clin Oncol.* 2006; 24(4):720-1. doi:10.1200/JCO.2005.04.6151
64. Secondary Use of Health Data for Medical Research and Public Health. European Commission website. https://joinup.ec.europa.eu/community/ehealthprocurers/document/secondary-use-health-data-medical-research-and-public-health. Published November 11, 2011. Updated March 8, 2018. Accessed May 5, 2017.
65. Damschroder L, Pritts JL, Neblo MA, Kalarickal RJ, Creswell JW, Hayward RA. Patients, Privacy and Trust: Patients' Willingness to Allow Researchers to Access Their Medical Records. *Social Science & Medicine*. 2007;64:223-235. doi:10.1016/j.socscimed.2006.08.045
66. Lowrance W. Learning from experience: privacy and the secondary use of data in health research. *J Health Serv Res Policy*. 2003;8(l):2-7. doi:10.1258/135581903766468800
67. Gellman, R. The Deidentification Dilemma: A Legislative and Contractual Proposal, Fordham *Intell Media & Ent. L.J*. 2011;21(1):33 61. https://fpf.org/wp-content/uploads/2010/07/The_Deidentification_Dilemma.pdf. Published July 12, 2010. Accessed May 5, 2017.
68. Information and Privacy Commissioner of Ontario. Dispelling the Myths Surrounding De-identification: Anonymization Remains a Strong Tool for Protecting Privacy. http://www.ontla.on.ca/library/repository/mon/25006/310614.pdf. Accessed May 5, 2017.
69. Information and Privacy Commissioner of Ontario. Dispelling the Myths Surrounding De-identification: Future of Privacy Forum website. https://fpf.org/wp-content/uploads/2011/07/Dispelling%20the%20Myths%20Surrounding%20De-identification%20Anonymization%20Remains%20a%20Strong%20Tool%20for%20Protecting%20Privacy.pdf. Published June 2011. Accessed May 5, 2017.
70. Benitez K, Malin B. Evaluating re-identification risks with respect to the HIPAA privacy rule. *Journal of the American Medical Informatics Association: JAMIA*. 2010;17(2):169-177. doi:10.1136/jamia.2009.000026.
71. Information and Privacy Commissioner of Ontario. Dispelling the Myths Surrounding De-identification: Future of Privacy Forum website. https://fpf.org/wp-content/uploads/2011/07/Dispelling%20the%20Myths%20Surrounding%20De-identification%20Anonymization%20Remains%20a%20Strong%20Tool%20for%20Protecting%20Privacy.pdf. Published June 2011. Accessed May 5, 2017.
72. Rothstein MA. Is Deidentification Sufficient to Protect Health Privacy in Research? *The American journal of bioethics: AJOB*. 2010; 10(9):3-11. doi:10.1080/15265161.2010.494215
73. Shivayogi P. Vulnerable population and methods for their safeguard. *Perspectives in Clinical Research*. 2013;4(1):53-57. doi:10.4103/2229-3485.106389
74. Research Involving Vulnerable Populations. Royal College of Physicians and Surgeons of Canada website. http://www.royalcollege.ca/rcsite/bioethics/cases/section-8/research-involving-vulnerable-populations-e. Accessed May 5, 2017.
75. President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry. Focusing on Vulnerable Populations. University of North Texas website. http://govinfo.library.unt.edu/hcquality/meetings/mar12/papch08.htm. Published March 30, 1998. Accessed May 5, 2017.
76. Vulnerable and Protected Populations. Solutions Institute website.
77. http://www.solutionsinstitute.com/vulnerable-and-protected-populations/. Accessed May 5, 2017.
78. Privacy and Confidentiality. CIRE Current Issues in Research Ethics. Columbia University website. http://ccnmtl.columbia.edu/projects/cire/pac/foundation/#7_4. Accessed May 5, 2017.
79. Zika Virus. Reporting, Data Collection, and Findings. Centers for Disease Control and Prevention website. https://www.cdc.gov/zika/reporting/reporting-collection-findings.html. Accessed May 5, 2017.

80. Centers for Disease Control and Prevention. Morbidity and Mortality Weekly Report (MMWR). https://www.cdc.gov/mmwr/volumes/66/wr/mm6613a2.htm. Published April 6, 2017. Accessed May 5, 2017.
81. Assisted Reproductive Technology (ART). Centers for Disease Control and Prevention website. https://www.cdc.gov/art/nass/accessdata.html. Accessed May 5, 2017.
82. National Program of Cancer Registries (NPCR). Centers for Disease Control and Prevention website. https://www.cdc.gov/cancer/npcr/. Accessed May 5, 2017.
83. National Center for Health Statistics. Centers for Disease Control and Prevention website. https://www.cdc.gov/nchs/nhanes/biospecimens/dnaspecimens.htm. Accessed May 5, 2017.
84. 42 U.S.C. 263a *et. seq.*
85. 42 U.S.C. 263a-5
86. 42 U.S.C. § 242m(d)
87. Brezina PR, Zhao Y. The Ethical, Legal, and Social Issues Impacted by Modern Assisted Reproductive Technologies. *Obstetrics and Gynecology International*. 2012; 12:1-8. https://www.hindawi.com/journals/ogi/2012/686253/. Accessed May 5, 2017.
88. Public Law 102-515, 42 U.S.C. § 280e *et seq*.
89. National Program Of Cancer Registries Program Standards, 2012-2017 Updated January 2013. Centers for Disease Control Website. https://www.cdc.gov/cancer/npcr/pdf/npcr_standards.pdf. Updated January 2013. Accessed May 17, 2017.
90. National Interstate Data Exchange Agreement. North American Association of Central Cancer Registries (NAACCR) website. https://www.naaccr.org/national-interstate-data-exchange-agreement/. Accessed May 17, 2017.
91. National Program of Cancer Registries (NPCR) Public Use Research Database Data Use Agreement. Centers for Disease Control and Prevention website. https://www.cdc.gov/cancer/npcr/pdf/public-use/data-use-agreement_npcr-public-use-database.pdf. Accessed May 17, 2017.
92. 45 CFR 46.102.
93. 42 U.S.C. §280e.
94. 45 CFR § 46.114.
95. Questions & Answers about Cancer in the Workplace and the Americans with Disabilities Act (ADA). U.S. Equal Employment Opportunity Commission website. https://www.eeoc.gov/laws/types/cancer.cfm. Accessed May 5, 2017.
96. Genetic Information and the Workplace. United States Department of Labor website. https://www.dol.gov/oasam/programs/history/herman/reports/genetics.htm. Accessed May 5, 2017.
97. 29 CFR § 1630.2
98. Silverman H, Hull SC, Sugarman J. Variability among institutional review boards' decisions within the context of a multicenter trial. *Critical care medicine*. 2001; 29(2):235-241. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4809523/. Accessed May 5, 2017.
99. National Cancer Policy Forum; Board on Health Care Services; Institute of Medicine. Contemporary Issues for Protecting Patients in Cancer Research: Workshop Summary. http://www.ncbi.nlm.nih.gov/books/NBK247009/. Published September 19, 2014. Accessed May 5, 2017.
100. Silverman H, Hull SC, Sugarman J. Variability among institutional review boards' decisions within the context of a multicenter trial. *Critical care medicine*. 2001; 29(2):235-241. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4809523/. Accessed May 5, 2017.
101. Pyle S. Benefits of Working with a Central IRB: Improved Efficiencies and Enhanced Human Subjects Protections. *ACRP*.2013;9-12. http://www.sairb.com/IRBForms/Benefits_of_Working_with_a_Central_IRB.pdf. Accessed May 5, 2017
102. Use of Central IRBs for Multicenter Clinical Trials. Clinical Trials Transformation Initiative website. https://www.ctti-clinicaltrials.org/files/centralirbfinalreport.pdf. Accessed May 5, 2017.
103. Flynn KE, Hahn CL, Kramer JM, et al. Using Central IRBs for Multicenter Clinical Trials in the United States. Doherty TM, ed. *PLoS ONE*. 2013;8(1):e54999. doi:10.1371/journal.pone.0054999
104. 45 CFR § 164.12(b).
105. 45 CFR § 164.514.
106. Can the device identifier (DI) portion of a Unique Device Identifier (UDI) be part of a limited or de-identified data set as defined under HIPAA? U.S. Department of Health and Human Services website. https://www.hhs.gov/hipaa/for-professionals/faq/2071/can-device-identifier-di-portion-unique-device-identifier-udi-be-part-limited-or-de-identified. Accessed May 5, 2017.
107. National Healthcare Safety Network (NHSN) OMB Control No. 0920-0666 Revision Request for OMB Review and Approval. Office of Information and Regulatory Affairs Office of Management and Budget website. http://www.reginfo.gov/public/do/DownloadDocument?objectID=13196201. Published July 2009. Accessed May 5, 2017.

108. Dokholyan RS, Muhlbaier LH, Falletta JM, et al. Regulatory and Ethical Considerations for Linking Clinical and Administrative Databases. *Am Heart J*. 2009;7(6):971-82. doi: 10.1016/j.ahj.2009.03.023
109. Garber AM. Cost-effectiveness and evidence evaluation as criteria for coverage policy. *Health Aff (Millwood)*, 2004: W4-284-96. doi: 10.1377/hlthaff.w4.284
110. Sorenson C. Use of comparative effectiveness research in drug coverage and pricing decisions: a six-country comparison**.** *Issue Brief (Commonw Fund).* 2010; 91:1-14. http://www.commonwealthfund.org/~/media/Files/Publications/Issue%20Brief/2010/Jul/1420_Sorenson_Comp_Effect_intl_ib_71.pdf. Accessed May 5, 2017.
111. Sox H . The patient-centered outcomes research institute should focus on high-impact problems that can be solved quickly. *Health Aff (Millwood).* 2012;31(10):2176-82. doi: 10.1377/hlthaff.2012.0171.
112. Persad G. Priority-Setting, Cost-Effectiveness, and the Affordable Care Act. *American Journal of Law & Medicine*. 2015;41: 119-166. http://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=2521&context=facpub. Accessed May 5, 2017
113. Applying Ethical Principles to Information and Communication Technology Research: A Companion to the Menlo Report. United States Department of Homeland Security, Office of Science and Technology. https://ssrn.com/abstract=2342036. Published October 25, 2013. Revised March 30, 2014. Accessed May 5, 2017.
114. Schmietow, B. Ethical Dimensions of Dynamic Consent in Data-Intense Biomedical Research—Paradigm Shift, or Red Herring? *Ethics and Governance of Biomedical Research.*Vol.4. Switzerland: Springer International Publishing; 2016:197-209.
115. Tucker K, Branson J, Dilleen M, et al. Protecting patient privacy when sharing patient-level data from clinical trials. *BMC Medical Research Methodology.* 2016; 16(Suppl1):77. doi:10.1186/s12874-016-0169-4
116. American Statistical Association. Data Access and Personal Privacy: Appropriate Methods of Disclosure Control. http://www.amstat.org/asa/files/pdfs/POL-DataAccess-PersonalPrivacy.pdf. Published December 6, 2008. Accessed May 5, 2017.
117. Nass SJ, Levit LA, Gostin LO, editors. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington (DC): National Academies Press; 2009.
118. Dokholyan RS, Muhlbaier LH, Falletta JM, et al. Regulatory and Ethical Considerations for Linking Clinical and Administrative Databases. *Am Heart J*. 2009; 7(6):971-82. doi:10.1016/j.ahj.2009.03.023
119. Bohensky MA, Jolley D, Sundararajan V, et al. Data Linkage: A powerful research tool with potential problems. *BMC Health Services Research.* 2010;10,346. doi:10.1186/1472-6963-10-346
120. Snyder CF, Jensen RE, Segal JB, Wu AW. Patient Reported Outcomes (PROs): Putting the Patient Perspective in Patient-Centered Outcomes Research. *Medical Care*. 2013;*51*(8 0 3):S73–S79. doi: 10.1097/MLR.0b013e31829b1d84
121. Pritts JL. The Importance and Value of Protecting the Privacy of Health Information: The Roles of the HIPAA Privacy Rule and the Common Rule in Health Research. 2008; 1-61. http://www.nationalacademies.org/hmd/~/media/Files/Activity%20Files/Research/HIPAAandResearch/PrittsPrivacyFinalDraftweb.ashx. Accessed May 5, 2017.
122. Dhir R, Patel AA, Winters S, et al. A multidisciplinary approach to honest broker services for tissue banks and clinical data: a pragmatic and practical model. *Cancer*. 2008;113(7):1705–1715. doi:10.1002/cncr.23768.
123. Office of Human Subjects Research – Institutional Review Board. HIPAA and Research. Johns Hopkins School of Medicine website. http://www.hopkinsmedicine.org/institutional_review_board/hipaa_research/limited_data_set.html. Accessed May 5, 2017.
124. 45 CFR § 46.102
125. Probabilistic Versus Deterministic Data Matching: Making an Accurate Decision. HealthIT.gov website. https://www.healthit.gov/archive/archive_files/FACA%20Hearings/2010/2010-12-09%20Patient%20Linking/Probabilistic%20Versus%20Deterministic%20Data%20Matching.pdf. Published January, 2007. Accessed May 17, 2017.
126. 45 CFR § 160.103.
127. HIPAA Privacy Rule and Public Health. Centers for Disease Control and Prevention website. http://www.cdc.gov/mmwr/preview/mmwrhtml/m2e411a1.htm. Published April 11, 2003. Accessed May 5, 2017.
128. Shivayogi P. Vulnerable population and methods for their safeguard. *Perspectives in Clinical Research*. 2013; 4(1): 53-57. doi:10.4103/2229-3485.106389

CS277412