DOCUMENTS AND COMMITTEE REPORTS

Series 4
Number 22

Issues Regarding Confidentiality of Data in the Cooperative Health Statistics Systems

DHEW Publication No. (PHS) 80-1459

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Office of Health Research, Statistics, and Technology
National Center for Health Statistics
Hyattsville, Md. April 1980

Library of Congress Cataloging in Publication Data

Simmons, Walt R.
Confidentiality of data in the cooperative health statistics system.

(Vital and health statistics : Series 4, Documents and committee reports ; no. 22) (DHEW publication ; no. (PHS) 80-1459)
Bibliography: pp. 47 and 49.
1. Medical statistics—Law and legislation—United States. 2. Cooperative Health Statistics System. 3. Public health—United States—Statistical services—Access control. I. Title. II. Series: United States. National Center for Health Statistics. Vital and health statistics : Series 4, Documents and committee reports ; no. 22. III. Series: United States. Dept. of Health, Education, and Welfare. DHEW publication ; no. (PHS) 80-1459.
HA37.U1693 no. 22 [KF3827.S73] 312'.0973s
[SBN 0-8406-0175-1] [344'.73'04] 79-607109

NATIONAL CENTER FOR HEALTH STATISTICS

DIVISION OF THE COOPERATIVE HEALTH STATISTICS SYSTEM

Vital and Health Statistics-Series 4-22

DHEW Publication No. (PHS) 80-1459
Library of Congress Catalog Card Number 79-607109

PREFACE

Initial articles of agreement set forth by the National Center for Health Statistics under contract HRA 106-74-25 charged the contractor with undertaking activities intended to define and outline a program of research, development, testing, and action to resolve the major issues regarding confidentiality of data within an emerging new program—the Cooperative Health Statistics System.

Specifications provided that the project should (1) identify the nature and character of major issues having an impact on confidentiality in the Cooperative Health Statistics System; (2) delineate programs or projects directed toward the resolution of these major issues including evaluation of existing methods and practices, devising new practices that might resolve or lessen difficulties in the field, and outlining those areas in which further research or development appears necessary; (3) suggest directions for training and guidance in making either policy or operating decisions; and (4) explore the possibility that a portion of the overall confidentiality problem in the Cooperative Health Statistics System can be bypassed rather than directly confronted by using techniques and procedures that yield adequate information while avoiding the risk of disclosing privileged data.

Since initiation of the project, various events, including passage of several significant laws, persuaded the contractor (with encouragement from the Government) to expand certain aspects of the original investigation to address the central objective more fully.

In particular, the expansion takes into account new developments; includes analysis of factors collateral to central concerns relating to privacy, confidentiality, and access to data; and offers a variety of recommendations for policy positions to be implemented in the cooperative Federal-State-local health statistics system.

Many persons contributed indirectly to this report. Any brief list of acknowledgments would be unfair, for it would omit proper credit to some sources for ideas that have more than one independent origin, and perhaps imply other views to which the referenced source does not fully subscribe. The author accepts responsibility for opinions expressed in the report, with deep appreciation to all those who through written or spoken words have influenced his thoughts.

CONTENTS

ISSUES REGARDING CONFIDENTIALITY OF DATA IN THE COOPERATIVE HEALTH STATISTICS SYSTEM

Walt Simmons, former Assistant Director for Research and Scientific Development

CHAPTER I

INTRODUCTION

The Conditioning Environment

In recent years, and especially during the past decade, a complex of related and partially conflicting principles and doctrines has taxed our ingenuity. Issues of confidentiality, freedom of information, and invasion of privacy and their interactions have received extensive and steadily increasing attention from administrators, legislators, the courts, the press, students, and certain sectors of the general public. Theaters of discussion have varied: legislative bodies; the courts; public, interagency, and intraagency committees; formal commissions; conferences; articles in professional journals, magazines, and newspapers; regulations and procedural documents; and voluminous correspondence and conversation.

This complex of confidentiality, freedom of information, and invasion of privacy and its resolution are critical to society's wise handling of information. Recognition of that fact has brought forth many opinions on these matters. Yet there is no satisfactory synthesis of this outpouring as it relates to the Cooperative Health Statistics System (CHSS).

A Central Problem

A common viewpoint, which is also that of the leading Federal statistical agencies, is that statistical information is most accurate when it is secured and handled in such a manner that anonymity of persons, business establishments, and individual products is assured. The U.S. Bureau of the Census, the U.S. Bureau of Labor Statistics (BLS), the National Center for Health Statistics (NCHS), and certain other Federal agencies have a tradition of giving such assurance and faithfully holding to their promises. Typically, a respondent to these agencies is assured of anonymity by statements such as "All information that would permit identification of the individual will be held confidential, will be used only by persons engaged in and for the purposes of the survey, and will not be disclosed or released to others for any other purpose."[1] Often this guarantee is underscored by declaring explicitly that the reported information will not be used for taxation, regulation, inspection, investigation, or any other administrative purpose, and will be released or published only in the form of aggregated statistical summaries.

This policy is based on *a priori* judgments, supported by years of experience, that (1) the American public is willing and even demands that their Goverment acquire sufficient information to wisely promote the general welfare; (2) respondents do, in fact, supply acceptably accurate answers to a considerable variety of governmental statistical inquiries when they are assured that those replies are handled confidentially and

are not used in any way to make administrative decisions with respect to an identifiable person or corporate entity; and (3) there is suspicion, distrust, and a less satisfactory response when the respondent concludes that his answers will be used either overtly or covertly to his personal disadvantage.

Competing doctrines to this tradition of statistical confidentiality exist. One significant factor is the presumption that democracy works best when the public has access to information used by the Government in the administration of its programs and activities and in the making of decisions. Congress has given validity to this presumption with the passage of the Freedom of Information Act (FOIA) (5 U.S.C. section 552 as amended by Public Law 93-502), which, with certain important exceptions, provides for access to data possessed by the Federal Government.

Another qualifying consideration is fundamental to economy of effort and burden on respondents. If a datum has been reported to one governmental agency, the better course may be to permit that agency to transfer the information to another agency, under specified safeguards, rather than to allow the second agency to collect the same datum. The Federal Reports Act of 1942 (44 U.S.C. section 3501) tries to deal with this matter, although its ability to reach local or State data is limited.

Another perspective deserving special attention in the collection, processing, and dissemination of data is the concept of invasion of privacy. Definitions of the concept vary. One view is that privacy is the right to determine what information about ourselves we will share with others. Confidentiality and invasion of privacy are quite separate matters, but they have intersecting domains. Confidentiality would be much less of an issue if no topics were considered private by persons or corporations. If transmission of information from one entity to another were totally suppressed, then privacy would not be a matter for concern.

These competing doctrines—the value of assurances of confidentiality by leading statistical agencies, the principle of freedom of information, the economy of use of the same data by more than one agency for more than one purpose, and the conflict between "need to know" and "right to privacy"—constitute the central problem. They are the basic sources of the issues that confront the CHSS for which approaches to compromise are sought in this report. The economy aspect is one of the key reasons for building a cooperative statistical system among local, State, and Federal agencies in the health field. Yet the fact that operating agencies, especially at the State and local levels, need to use specific data for various administrative purposes and often identify individual persons or business establishments tremendously complicates the confidential handling of the same or allied data by statistical agencies.

Perspectives

Resolution of the competing forces is much more difficult because the problem is multidimensional. Significantly relevant considerations are found in the ethical, political, economical, legal, and administrative disciplines; and effective operational solutions require successful handling of a variety of jurisdictional, procedural, technical, and technological matters. This report discusses each of these perspectives. Analysis attempts to identify certain priorities among conflicting objectives. However, the keynote of recommendations is the concept that policies and procedures should accomplish a *balance* among competing goals that are desirable.

Structure of the Report

Chapter II outlines the background and environment that condition the privacy, confidentiality, and freedom of information complex with which the CHSS must be concerned. Chapter III presents an abstract and summary of major conclusions and recommendations. Chapters IV through XVI analyze leading issues and, in most cases, suggest resolutions of those issues. These chapters are the main body of the report and are the basis for the summary conclusions of chapter III. Chapter XVII discusses more briefly a number of other important, but less critical, issues and focuses attention on gaps in the analysis and on problems that are not fully resolved. The appendix contains a selected bibliography.

CHAPTER II

BACKGROUND

Wealth of Activity

The interlocking fields of privacy and confidentiality are receiving searching attention from various organizations. The reasons for these investigations (and subsequent pronouncements) are likewise varied, and include concern over data banks and the capacity through computers to marshall pieces of information; access to information by officials, researchers, statisticians, or citizens; doctor-patient, scientist-subject, and other provider-client relationships; police systems, insurance mechanisms, and other devices of social control; immunity of certain classes of information to subpoena by courts or legislative bodies; the need for increased volume and detail of data demanded by today's more complex social structures; physical security of confidential information; a person's inherent right to privacy; and efforts to clarify distinctions among administrative, statistical, and research uses of data.

Several hundred substantial articles and reports, including at least a score of quite prominent and influential documents, have emerged from the investigations. Some elements of consensus are developing, but the wide range of auspices and perspectives are yielding contrasting and conflicting proposed guidelines for formation of social policy. Inasmuch as statistical activities and practices are generic consequences of access to and use of data, it is highly desirable that evolving guidelines include a socially appropriate role for those activities and practices.

Immediate Stage Setting

Later chapters deal more specifically with legislative developments and with certain other formal actions that are taking place. However, it will be helpful to identify three relatively recent major events—among a number of others—that have a special significance for confidentiality issues in the CHSS.

The first event is the passage by Congress of the Privacy Act of 1974. The Act is a compromise and amalgamation of a number of other legislative proposals. It is a comprehensive measure, and its purpose is to provide safeguards against invasion of personal privacy by requiring Federal agencies to establish procedures that insure that a person can learn what personal information the Government has, why it was recorded, and how it is to be used. Furthermore, this Act gives the person some degree of control over whether he must supply information requested by the Federal Government. The Act was intended to have primary impact on data that identify individuals and are used for administrative purposes. However, through drafting ambiguities and varying interpretations, the Act and consequent executive regulations have had significant repercussions in statistical affairs.

The second situation is the fundamentally different character of certain confidentiality matters that arise as the Nation's major health statistics system changes from a highly centralized Federal operation to a more decentralized Federal-State-local cooperative enterprise. Through policy, legislation, regulation, and years of attention to the issues, NCHS has established satisfactory procedures for handling confidentiality matters in the tightly controlled centralized system. However, the expanded CHSS introduces new sovereignities, new laws, new objectives, new procedures, and a greatly enlarged arena of concern for both the rights of individuals and society's need for efficiently assembled and disseminated data.

The third factor, which has both influenced the contractor's activities and has been influenced by those activities, is the deliberations and draft reports of the Task Force on Confidentiality appointed in 1974 by the Advisory Committee on the Cooperative Health Statistics System. The Task Force and this contractor have communicated frequently. Although neither is responsible for the recommendations of the other, this report and the Task Force report do have some common ground in both analysis and conclusion.

CHAPTER III

SUMMARY OF CONCLUSIONS AND RECOMMENDATIONS

1. The interlocking domains of privacy, confidentiality, and informational requirements are receiving both extensive and intensive scrutiny. Rightfully, they are subjects of thoughtful debate and varied actions as the Nation seeks to resolve different perspectives and partially conflicting *desiderata.*

2. The central and most pervasive principle governing wise resolution is this: Appropriate *balance* must be sought between a person's fundamental right to a degree of privacy and society's acquisition of information about itself in order to guide its activities, and, indeed, to assure its freedom. This principle is consistent with the basic objective of the CHSS, which is to develop and maintain systems that provide the country with the maximum of useful and needed health information at the minimum cost in terms of both the rights of persons and corporate entities and required resources.

3. No single proposition can guide the CHSS through the confidentiality-dissemination maze. The search for balanced guidance must take simultaneous account of a multiplicity of tenets and perspectives, including ethical, political, legal, economical, administrative, and technological ones. (See chapters IV, V, X, and XI.)

4. Fundamental to productive informational policies throughout the social sciences are recognition and widespread acceptance that there are two very different kinds of informational objectives, served by two different kinds of data. One objective is the collection, processing, transfer, retrieval, and utilization of data to deal with specific persons or other entities. "Dealing with" encompasses such actions as licensing, registration, inspection, insuring, training, regulating, servicing, diagnosing, treating, charging, and paying, and, thus, conveys either benefits or penalties. This class is termed "case-action data"[a] in this report.

The other class of objectives and data is distinguished by the fact that the identity of individual elements of information—persons, corporate units, products—has no significance and accordingly is suppressed. These data are collected and disseminated for *statistical purposes only,* which means that they appear in the format of aggregates, averages, rates, ratios, percentages, distributions, and other functional relationships, and never in a manner that permits identification of individual entities. Typically, the statistical inquiry or compilation is accompanied by assurance of confidentiality given to the provider of information, which guarantees that data *will* be used for *statistical purposes only* (see chapter VII for further delineation of statistical purposes) and will not be used, in whole or in part, for regulation, inspection, taxation, or any administrative purpose or determination about identifiable individuals; or published or released in a form that would identify individuals. This report advocates the term "protected data" for such statistical information. The CHSS is concerned with systems for handling both protected data and case-action data. The two systems overlap in some areas, but are fundamentally different and must be governed by very different guidelines. (See chapter VI.)

5. Federal law recognizes essentials of the concept of protected data and gives adequate protection to data once acquired by NCHS,

[a] "Case-action data" replaces the term "micro-action data" proposed earlier by the author.

when the providers have been given assurances of confidentiality. Steps should be taken to establish further Federal and State statutory recognition of protected data, and of its distinctive characteristics throughout the CHSS. These laws and derivative regulations should include provisions that give protected data immunity from subpoena by courts or legislative bodies. Such legislation will facilitate the collection of high-quality data, contribute greatly to the production of useful health information, with minimum risks to and burden on respondents.

Following a study of existing State laws, NCHS should undertake the drafting of a model State law on CHSS confidentiality, which can serve as both a starting point and a coordinating influence on an emerging body of State law and regulations. Action on this recommendation has been taken. See Model State Health Statistics Act—a model State law for the collection, sharing, and confidentiality of health statistics.

6. Assurances of confidentiality once given must be honored without exception. However, three prior conditions should exist before assurance is given. (1) Collectors of data should exercise restraint in their inquiries; no information should be requested unless there is a definite and real need for it and an intended use that outweighs risks to the respondents and costs of processing. (2) Data that are collected should not be given the shelter of "protected data" unless such protection is judged essential to acquisition of quality information or deemed a privilege to which the respondent is clearly entitled. (3) The collector should be in a position to make certain that the promise of confidential treatment, if given, *can* be kept inviolate.

7. If a single agency, such as NCHS, operating under sheltering legislation, collects a piece of information directly from, for example, a household respondent, classifies the information as "protected data," and does not transfer it in identifiable form to any other organization, protection of confidentiality is relatively easy to maintain. The CHSS presents a different problem. It is a network of agencies, collecting a great variety of information from numerous sources and disseminating those data in various ways. In particular, certain components of the CHSS may, at times, be in possession of a datum that will be transferred through one authorized and designated channel for an administrative purpose, while along another, the same datum may be classified as "protected data" with transfer and access restricted to statistical purposes only. These differing channels must be kept distinct. The terms "protected data" and "case-action data" refer to how items of information are used, who has access to them, and what their purpose is, rather than directly relating to the items. The rules for transfer of data throughout the system must be definite, widely understood, and subject to sanctions if broken. One prevailing principle is this: If element B of the system acquires data from element A (who might be an original or secondary respondent), B is required to tell A under what authority the data are acquired, for what general purpose(s), and who, if anyone, will have further access to the data in individually identifiable form. Any further transfer of identifiable data from B to C must be in accordance with this declaration, unless a new release is obtained from A. (See chapter VI.)

8. Because the CHSS handles differing kinds of data intended for different purposes, the system should contain certain elements of flexibility that permit customized variations of procedure for controlling selected classes of data. The system should be an integrated, standardized operation, but not every component of the system should be conducted in identical fashion. (See chapter XV.)

9. Unintentional disclosure refers to any display of data that results in advertent access to individually identifiable information by parties other than the authorized custodian of the data. This situation can occur through careless publication of information in categories that are too finely classified, inadequate physical security, or linking of files that, in combination, contain too many descriptors of the individual. Operational rules must be established and enforced to minimize these risks. However, the rules should not be so constrictive that they strangle dissemination of knowledge. Because no set of precautions can give absolute protection

against sufficiently determined and sophisticated attacks on secrecy, the CHSS should not attempt a defense against every possible contingency. (See also recommendations 10, 11, and 12 and chapters IX, XII, XIII, and XIV.)

10. Linking of two data files is more difficult, less necessary, and less useful than many persons believe. Linked microdata files have utility at low cost in certain circumstances where action is to be taken with respect to an identified person or facility. The linking of micro-protected data files, however, will not often result in a benefit in CHSS that is worth the added risks and costs, or that cannot be secured by other less troublesome procedures. (See chapter IX.)

11. The public-use tape, which contains microdata for elementary units with individual identifiers removed, provides a highly flexible analytic mechanism, and is likely to become an increasingly important device for dissemination of data. The public-use tape not only can meet many of the needs for microdata, but, at the same time, should reduce the demand for individually identifiable data. Such tapes are subject to somewhat more restrictive rules for protecting against inadvertent disclosure than are necessary for published tabular material. A cell of a table usually reveals only one or two, or, at most, very few attributes of the individual units that contribute to the cell. A microtape, however, may contain 10, 20, or even more descriptors of each unit. It is axiomatic that the larger the number of descriptors, the greater the risk of positive identification.

12. Various procedures permit acquisition, transfer, or manipulation of microdata, and yet make it nearly impossible to identify individual persons or facilities. These procedures can be characterized as "avoidance techniques." They are a fertile field for development, and merit imaginative cultivation. (See chapter XIV.)

13. Whatever systems are developed under whatever controls, actions in the CHSS are taken by people, many of whom are employees in the system. The greatest safeguard the system can have is a workforce that understands and is dedicated to the conduction of a program balanced between providing useful statistics and protecting the privacy and confidence of those who supply the information. To this end it is proposed that a vigorous and continuing training program for CHSS staff be mounted. (See chapter XVI.)

14. Policy and practice must be guided by what people think the situation is, as well as by what the facts are. It is therefore critical that NCHS and CHSS vigorously promote wide understanding of the essence of statistical purposes and the role of statistical information and, in particular, maintain a public relations program that allays unjustified fear of imagined potential harm to individuals from misuse of statistical data. (See chapter XVI.)

CHAPTER IV

ETHICAL AND HUMANISTIC CONSIDERATIONS

Interaction of Disciplines

As noted in chapter I, a multidisciplinary complex that involves ethical, political, economic, legal, administrative, and other considerations is of concern. All of these are important, but the ethical and humanistic factors are preeminent. The prime objective of the CHSS is to develop a fund of information that will facilitate and enhance the planning and execution of health activities for the ultimate purpose of improving the health of the population. This objective requires securing the maximum useful information with minimum necessary infringement of the rights of any person.

Personal Rights

An unavoidable conflict between the need for information and a person's right to privacy exists. A good case can be made that a person has a "right" to expect his government to collect sufficient data from other persons to be able to promote his general welfare. Every citizen must be willing to give up a little bit of his freedom to live in a free society. In a given situation, it may not be clear precisely what ethical considerations may dictate. It should be clear, however, that rights of the individual and humanistic considerations take precedence over less than essential requirements of society and governments, a lower dollar cost for a piece of information, or the convenience of an administrator. When one is faced with choosing a course of action after analyzing a situation in terms of cost-benefit risk, high priority should be given to reducing the risk to individual persons unless the benefit to society is clear and of overriding value.

In applying this principle a reminder is in order. James B. Rule,[1] in commenting on complaints by citizens that governments ask and store too much information about them and are, thereby, in a position to exercise too much surveillance and control over them, makes a series of perceptive observations. He reminds us that in the British and American democracies, the overwhelming bulk of governmental surveillance and control exists because the majority demand services that can be provided only if appropriate record systems exist. Examples are Social Security benefits, health insurance, driver registration, tax assessments, or even oversights of credit card privileges.

Furthermore, ethical rights are not absolute. For example, it is argued in this report that an assurance not to reveal privileged information *must* be honored. Yet that promise should be weighed against an ethical responsibility to disclose the information in court if disclosure would save a person's life.

Informed Consent

A very special problem in the health field surrounds the question of informed consent. A widely held view is that when a government requests data from a respondent and a reply is not mandatory, the informed consent of the person is a prerequisite to the recording and use of a reply.

The NCHS enabling legislation, the Privacy Act of 1974, and most other guidelines in the field of privacy and confidentiality put considerable emphasis on the doctrine of informed consent. The general concept is fairly simple: A wishes to take some action that involves B; A

explains to B what this action is and what its impact on B may be, and asks B's permission to proceed; assuming that B is thus fully informed, understands, and agrees, it is said that B gives his informed consent; A is then justified in taking the proposed action.

This concept is valuable and desirable in a free society; NCHS endorses this principle. However, in medical research, medical practice, or statistical activities the exact meaning of informed consent may not be totally clear. Informed consent presumes that the explanation of a proposed action is "adequate," that the affected party "fully" understands, and that concurrence or consent is truly voluntary. The signing of a release by the affected party might satisfy some legal requirement, and still not have met the requirements of informed consent in a broader sense. For example, the release given by a Medicaid patient may involve consequences that were neither well explained to, nor understood by, the patient, and for which he really had no option if he was to receive treatment.

In statistical practice there are several special problems. If response is not mandatory, the collector should clearly inform the respondent. If too much of a point is made that the respondent need not reply unless he wishes, however, the collector may only succeed in biasing results through nonresponse, or by obtaining inaccurate response. How detailed should the explanation of intended uses of the data be? How far should all conceivable impacts on the respondent be pursued? How much pressure is justified in securing a response? Categorical answers to these questions are not possible.

There are several guidelines, however, that should result in courses of action that will be approved by a majority of reasonable people.

1. It is assumed that the inquiry is authorized by law, and that there is a definite social need for the information.

2. It is a *fact* that identifiable information will be treated as protected data (as described later in this report) when the collector has so declared, and that assurances of confidential handling will be rigidly adhered to.

3. It is a *fact* that the chances are near zero that any respondent will be harmed by his participation in the survey.

4. The burden on or discomfort to the respondent that inquiry and reply may entail are minimal compared with potential social gain.

5. The collector does explain *in general terms* the authorization for the survey, why the data are needed, and, at least, one specific way in which they will be used.

6. In a program such as the CHSS, there should be clarification of what agencies and what kinds of personnel will have access to personally identifiable data, for what purposes, and for how long.

These guidelines should be supplemented by a most significant principle. It is proposed that all persons interested in statistical systems undertake to promote and gain wide acceptance of this principle. The principle is intended to apply to protected data to be used for statistical purposes only, and means that, except during processing for a restricted time, the data will never be used, transferred, or displayed in a form that identifies individual persons or entities. The principle is that a sufficient declaration of purpose is a two-part announcement that (1) states at least one specific intended use of the data and (2) warrants that the data will be used for *statistical purposes only*. The significance of this principle is that, beyond some primary immediate justifying objective, the creation of "statistics" is an adequate summary of purpose. All possible future uses of those statistics cannot be set forth in detail. The principle does imply, however, that there are *no* purposes that will be served through the use of individually identifiable data. It is proposed that, for protected data, this principle will meet, for example, the requirement of the Privacy Act of 1974, that

agencies must "permit an individual to prevent records pertaining to him obtained by such agencies for a particular purpose from being used or made available for another purpose without his consent."

This is neither a dodge nor some devious attempt to bypass a law or other responsibility. Governmental statistics may be compiled initially for a single purpose or for a set of purposes. Once compiled, they belong to the people, and not only may, but should be used for any purpose for which they are helpful. The legal provisions such as the one just quoted are intended to relate to the use of data in microform: for example, they might prohibit the record of a person's age as reported to the Social Security Administration (SSA) from being used by a State vehicle commission to deny a driver's license. It would be ridiculous to argue that *statistics* on number of applicants for Social Security benefits by age could not be used for some previously unstated purpose without the consent of the applicants.

Inconsiderate Inquiry

A word should be said about a somewhat subtle aspect of data acquisition and use that could arise in the CHSS. Possibly an inquiry or an intended use of an excerpt could result in no real harm to the respondent but could have a dehumanizing impact in the respondent's mind, and thus be traumatic.

Respondents should never be asked to supply information for which there is no authorized use, particularly if the inquiry is one that conceivably could entail mental anguish to the respondent. Some authorized inquiries, however, probably should not be made, because the utility of the reply is not sufficient to justify the risk of possible psychic stress. For example, it would be wise to forego questioning a patient about his cancer therapy if there was any doubt about whether the patient was aware of his diagnosis. The contention that "it would be interesting to know" is not sufficient reason for asking an embarrassing or sensitive question.

CHAPTER V

POLITICAL AND ECONOMIC FACTORS

Intent of the Chapter

Ethical issues are central to the privacy-confidentiality-dissemination problem. A solution also requires consideration of other factors. The objective of this chapter is simply to underscore the importance of political and economic components of an operating system. Later chapters suggest legal, organizational, administrative, and procedural structures that also are important.

Political and Jurisdictional Concerns

In a narrow sense, terms such as "politics," "bureaucratic," "local vested interests," and "the Feds" denote a jurisdictional jungle in which the struggle for power, authority, and control is ever present. In a more enlightened context, political considerations imply that attention is paid to the structure with the best social organization. In any Federal-State-local program there will be elements of both the narrow and the more enlightened view. In the CHSS, largely a technical and service activity, the broader view should prevail. It would be shortsighted, however, to plan with total disregard to the presence of a certain amount of jurisdictional competition. And, indeed, what to one observer may appear to be petty, narrow jurisdictional pressure, is to another responsible support of legitimate political interests in the very best sense of the term. For example, the Federal Government might request permission to examine the statistical collection techniques of a professional association, or might require an audit of State use of Federal funds, believing that these actions are a necessary part of its responsibility to acquire quality data at reasonable cost. The professional association or the State agency, however, may feel such actions are unjustified questionings of their motives and capabilities. Furthermore, NCHS and a State center may disagree about which can be the most efficient primary collector of a particular data set. The point is that the preferences of local, State, and Federal agencies will not always be identical, and a successful system must reach workable compromises among those preferences. Compromises must also take into account the interests of a variety of third-party groups, including such bodies as the American Medical Association (AMA), American Hospital Association (AHA), Professional Activities Study, Medicare, Medicaid, Blue Cross, and Blue Shield.

Interaction of Political and Economic Concerns

Relationships among providers, collectors, processors, and users of data must be given attention in any framework. The peculiar circumstances of a cooperative Federal-State-local program in a very large theater of activity make the interaction of political and economic arrangements important in the CHSS.

The health industry is a multibillion dollar activity. More people are employed in the health industry than in all of the Federal Government; more than in all of the State Governments; more than in all of the local governments if schools are excluded; more than in all agricultural occupations; more, in fact, than in any other single industry in the United States, unless retail trade of all kinds is counted as a single industry.

The implication is clear: although effective planning and conduct is dependent upon the availability of relevant information, it is necessary for a significant amount of data to be collected from persons and facilities. The large volume of data required increases the dangers of invasion of privacy and risks of breaching confidentiality. The volume also emphasizes the need to not collect data that is unlikely to be used; to avoid duplication in collection, processing, and dissemination. However, avoiding duplication could mean increasing the number of persons and agencies that must handle datum beyond the initial collector who gave assurances of protection to the respondent, and this increases the risk of disclosure.

Efficiency and economy in acquiring and handling data must be sought throughout the CHSS. Yet in building this system, care should be exercised to avoid centralizing any particular function so that a rigid monopoly is created. A dynamic, vigorous system must remain flexible and must be constructed so that quality assurance and cross-checking processes are built in. This requirement means that more than one agent often must have access to certain microdata and adds one more constraint to the confidentiality problem.

CHAPTER VI

TWO CLASSES OF DATA AND TWO PURPOSES

Distinguishing Features

Governments and society need information to plan, execute, and evaluate in a rational manner. The only source of much of this information is the behavior, opinions, measurements, and records of individual persons and other entities. Because these same persons and entities have inherent rights of protection from invasion of their privacy, a significant conflict of interest results. The solution is to secure an appropriate *balance* between the competing requirements for information and protection of privacy.

Clearly, no single step will produce this balance. However, one key can open the way toward solution: the recognition that there are two quite different kinds of informational objectives, and identification of two different kinds of data that can serve those objectives. The two kinds of data have distinctive characteristics, are handled differently, and together yield high levels of needed information with minimum risk to the privacy rights of individuals.

One informational objective is the creation of bodies of statistical evidence—numerical information in the form of aggregates, ratios, percentages, indexes, and relationships—to be used for a great variety of purposes in planning, administration, and evaluation. These purposes do not, in themselves, require knowledge of identifiable individuals, establishments, or products. Indeed, such identification usually would only clutter up analysis if it were offered to the user. The user needs the aggregative tools of statistics. When he wishes to see microdata, it is only to study distributions of anonymous entities around central tendency values. (A processor will need microdata temporarily for procedural purposes and quality control, but unit identification can be removed as soon as processing has been completed, and separated from the substantive information.) We are speaking here of the familiar concept of data used for statistical purposes only, as understood and practiced by such agencies as the Census Bureau, BLS, or NCHS. For convenience, we designate such statistics as "protected data." These data have full and absolute protection of confidentiality guaranteed by law, are, in some instances, immune from subpoena, and are further sheltered by legal penalties for those who might violate the confidential status.

The CHSS is also involved with data that serve a second objective of great importance: the more coordinated, efficient collection, processing, transfer, retrieval, and utilization of data for the express objective of dealing with specific individual persons or other entities. "Dealing with" encompasses such actions as licensing, registration, inspection, insuring, training, regulating, servicing, diagnosing, treating, charging, paying, and both helping and punishing. It is suggested that this type of information be termed "case-action data." A case-action record must contain a unique and readily usable identifier of the individual entity to which it refers.

Certain classes of case-action data may be withheld from most possible consumers, yet made available to all those persons with a "need to know." The physician's patient record and certain data descriptive of employees are examples of information of this type. Other case-action data may be more widely disseminated, however, or even be entirely in the public domain—for example, name, address, size and nature of business or facilities, or name, unique

identifier, and occupation of licensed practitioners. In addition to serving their primary individualized objectives case-action data may also be aggregated and serve useful statistical objectives.

The CHSS is concerned with a national program to develop systems for both protected data and case-action data. The two systems will overlap in some areas, but fundamentally they must be very different systems with different guidelines. Recognition that they are two separate systems rather than a single system is a major step toward resolution of many perplexing issues. Legislation, policy, and procedure should carry this distinction into both planning and operation. In the CHSS, the Center is concerned primarily with protected data, and will release other data only when they have been judged to be properly in the public domain. Contrastingly, State and local agencies often will be handlers of both classes of information.

Definitions and Labels

The definitions of protected data and of other kinds of data need to be given thoughtful and precise formulation. The term "case-action data" describes a useful concept. Some may employ the term "administrative data" for a similar purpose. The label "administrative data" is, at once, both too restrictive and too encompassing. It fails to distinguish satisfactorily from protected data. *Protected data* have the fundamental attribute that they shall not be disclosed or knowingly cause to be disclosed by the collector or custodian(s) by any means, in a manner that makes it possible from such disclosure to relate the particulars obtained from any return to any identifiable individual or entity except with the consent of the provider of the information. Protected data are utilized only in statistical format for statistical purposes. *Case-action data* are all other data, whether collected in their own right as descriptive information or as byproducts of other actions, but not accorded the full nondisclosure attribute of protected data. Protected data will be displayed or released in such aggregated forms as totals, ratios, rates, and relationships; or if in microformat, only with individual identification removed; and, in all cases, for statistical purposes only, and never for purposes of taxation, regulation, investigation, or other direct action with regard to individuals. Individually identifiable items from protected data sets can only be transferred to third parties with the consent of respondents. Case action or administrative data may be used for similar statistical purposes and can also be used in microform by those with a clear need to know, or as the holders determine, if not in violation of any other law.

The above definitions should be established by both Federal and State statute, at the earliest practicable date. Pending such legislation, they should be promulgated as regulations of cognizant authority.

Another facet of these data classes must be underscored. The terms "protected data" and "case-action data" refer to how items of information are used, who has access to the items, and for what purposes they are used, rather than to the specific items. Any particular datum may be case-action data in one environment or pathway and protected data in another environment or pathway. To avoid latent prejudices about proper handling of health data, consider an illustration from another field. The circumstances in which an original source supplies information for both case-action through one channel and protected data along another are by no means restricted to health matters.

An almost classic example is the handling of wages and salaries by employers in the United States. An employer reports earnings for individuals to the Internal Revenue Service (IRS) and the SSA for case-action purposes. Both agencies use the data in restricted ways and take actions regarding individuals on the basis of the reported data. Many employers, using the same basic accounting records, also report earnings to the BLS, where the information becomes protected data, and is used for statistical purposes only. No action is taken with respect to an individual employee or employer on the basis of the BLS records, but the Nation gains valuable information on levels and trends of earnings and employment in each industry from aggregated BLS data.

Publicity

Successful publicity requires a climate in which this dichotomy of purpose is widely recognized and respected. Continuing vigorous public relations activity should have as its objective the development of public perception equal to that concerning "top secret versus other," or "lawyer-client privileged information versus ordinary testimony."

Formal Designation

The term "protected data" should be designated on all hard-copy documents and other transcriptions and coded on punchcards or magnetic tape. Forms should be printed with "PROTECTED DATA" in large (72-point) block letters. It is possible for one record to be classified "protected data" and, therefore, sheltered, and another record containing all or a part of the same information to have a different status because of its different context.

Collectors of data have the authority to designate specific information as protected data, but only within clear boundaries of authorizing legislation and regulations. The collector only assigns such a designation when there are overriding reasons for doing so. The designation must not be overused, and never used as an excuse for withholding administrative data. Once a record has been designated as "protected data," the designation can be removed only by the agency that made the assignment. (Legislation may be necessary to prevent abuse of the privilege of classifying information as "protected data.")

Scope of the Designation

Confidentiality assured to protected data should not be waived without consent of the provider of the information. Numerous existing laws, regulations, policies, and practices give, or appear to give, exemptions to promises of confidentiality in certain circumstances. Prominent among these are:

1. Court orders or subpoena for evidence.
2. Demands of legislative bodies, including committees.
3. Auditor's requirements.
4. Rights of individuals to have access to information useful in promoting their health, or in providing the best evidence of their defense in legal actions.
5. Claims of law enforcement agencies—especially in criminal affairs.

Each of these needs has a social value, as well as a potential benefit to individuals. However, these needs can be met without overriding assurances guaranteed for protected data. Protected data invariably come from a prior provider or source of the information. That provider or source should be the point at which courts, legislators, or administrators seek disclosure when social interests require identifiable data. The only exceptions should be situations where profound ethical considerations outweigh the guarantee of confidentiality. (See section "Personal Rights" in chapter IV.)

Notice to Providers

Whenever possible, the initial collector of data should inform the respondent or provider of information when information is declared "protected data." This task may not be easy; however, it is important to attempt it with strong resolve. With rare exceptions, data of an individual entity are identifiable at the point of collection. Through editing, processing, transcription, and transmission to other parties the data are still identifiable. Therefore, policy and procedures should provide for the separation of substantive data from the identifiers at the earliest practicable point and for notification to the original provider at the time of collection, if feasible.

Use of Administrative Records for Statistical Purposes

Assume that a successful distinction is made between statistical and other purposes—or between protected data and case-action data—and that this distinction is made known to an acceptable degree to all interested parties, including relevant sectors of the general

public. Focus then on that other major area of the CHSS: the situation in which information is recorded initially for an administrative purpose, but may be used additionally for statistical purposes. At this point resolution of issues of privacy and confidentiality is most difficult and subject to the greatest hazards for respondents, subjects, administrators, and statisticians. Society and the CHSS must display ingenuity, care, and wisdom to secure appropriate balance between need to know and protection of the individual. Under "Definitions and Labels," at least two major subclasses of data in this category are discussed. One subclass consists of items of information in the public domain—available to any person as public record. Except for misunderstandings, errors, or mischievous actions, they constitute no serious problem. However, instances of the other subclass are legion: They consist of items provided or recorded initially for some operational or administrative purpose with the expectation that they will be used for a particular restricted purpose and that access to them will be limited to persons with a definite need to know relevant to that purpose. When they are used for any other purpose, a potentially serious conflict arises.

Many considerations impinge on the latter subclass. Much of this report, as well as extensive literature, deals with various facets of it. Three directives are proposed:

1. When administrative data are of good quality and pertinent to a statistical or research purpose, a way *should* be found, in the interests of cost efficiency and improved knowledge, to make them available for that purpose.
2. When administrative data are made available, they must not only acquire the shelter of other protected data but also retain the full protection they enjoyed as particular kinds of administrative data. Any exception to these requirements must be stated in writing, and any transfer or use must be governed by a written protocol that has the force of a contract, if the transfer is from one agency to another.
3. Public perception of the consequence of a transfer or separate use of administrative data may be as important as the actual effects. (See chapter XVI.) Therefore, all affected parties should be clearly informed of the actions taken, the actions to be taken, and the possible impact on initial respondents; or the action and use should be restricted to courses that cannot possibly harm affected individuals.

A Policy Position

Whether data are collected originally for a statistical or an administrative purpose, confidentiality should not be promised unless there are persuasive reasons for doing so. In some situations there are compelling arguments for assurances of confidentiality. However, there are many other situations in which confidential handling of data would unnecessarily restrict use, and serve no important objective. The burden of justification for declaring a particular collection confidential is the responsibility of the collector, and is not to be taken lightly.

However, when confidentiality has been assured, it must be honored by those who assure it. Preferably, their position should be protected by shield laws—and at the very least by regulations and written procedures.

CHAPTER VII

THE STATISTICAL DISCIPLINE AND STATISTICAL PURPOSES

Introduction

The previous chapter proposes the concept of protected data. Throughout this report and elsewhere, the expression "for statistical purposes" is used frequently. These terms are closely related and already have been generally defined. Because ideas are the substance of statistical agencies, further comment is merited.

The basic function of a statistical agency is to produce information useful to planners, managers, and students. This information is produced largely by adaptation of statistical science methods that collect and display the essence of a body of evidence and do not allow distracting details to overshadow main conclusions. Statistical discipline recognizes that there is variation in nature and societies: single observations often will be unrepresentative of the classes from which they are drawn. Accordingly, attention is focused not on individual observations or entities but on the characteristics and attributes of groups of observations and relationships among different groups.

The statistician provides greater information by discovering, for example, that the *average* length of stay in a class of hospitals is 7.5 days (instead of saying that Jane Doe stayed 11 days in St. Mary's), or that the average length of stay is greater for cancer admissions to a hospital than for stroke admissions (instead of noting that Smith stayed 3 days for a cancer admission and Jones, 8 days for a stroke admission), or that the average of all lengths-of-stay was 6.5 days in 1974 compared with 7.3 days in 1970 (instead of finding that a particular patient was hospitalized in 1975 for 8 days, but was hospitalized in 1970 for only 5 days for apparently the same condition). The statistical agency, arguing from either a sampling or complete enumeration of cases, presents conclusions in averages, rates, ratios, percentages, or other mathematically expressed functional relationships.

A Fundamental Principle

The statistician *never* needs to know the identity of individual elements of data and analysis. The essence of his discipline is to treat elements as indistinguishable from one another within classifying categories. He has no wish to know the individual identities, and his work is best performed when he does not know these identities. *He must convey this fact to all interested parties.*

Extensions of Statistical Purposes

The only legitimate exceptions to this fundamental principle are in processing. It should be possible in the CHSS to delineate the exceptions, to write rules covering them, and to inform respondents concerning the exceptions so that they understand and do not disapprove.

Operational control.—Case identification is necessary for operational control to assure that processing does, in fact, carry out intended data reductions. This requirement is easily met by using a nonsense ID number that uniquely identifies the datum but does not identify a person or establishment.

Quality control.—If the main processor uses input data that are identified only by nonsense ID numbers, there must be a key some place that translates it to an identifiable entity to permit a quality check on the input. This key or crosswalk can be restricted to a subsample of all cases, and can be held by someone other than the main processor. For example, if NCHS is the

main processor, the key might be held by the State partner or by a hospital supplier of data. The Center has an important responsibility, however, for the accuracy of statistics released under its aegis. Therefore, NCHS must have access to the key when it is necessary to exercise quality control over input. This access can be limited to selected employees under oath not to divulge the identities. The key can remain in the possession of the original holder and never appear on centralized Federal records.

Duplication.—A class of operations exists for which it is highly desirable, if not essential, that the main processor—for example, NCHS—has possession of individual identifiers during a processing period. Two of the several situations of this class are mentioned briefly. One situation is in the manpower field, where the State agencies may report to NCHS numbers and descriptions of licensed persons in certain occupations. Some individuals are licensed in more than one State. Counts of persons are desired for the Nation as a whole; therefore, it is necessary to eliminate duplicates from the State reports. The most efficient way of doing this elimination is by matching individuals identified by a common denominator, perhaps a Social Security number.

The duplication problem could be handled in a different manner, and the need for a common identifier could be avoided. The essential step would be to require that the licensing document and subsequent transcriptions include the item "Number of States in which licensed." If a particular person is licensed in three States, then in statistical tabulations this person is counted as "1/3 person" each time he appears. The resulting totals will be correct unduplicated counts. In situations in which an unduplicated count is not the objective, the "1/3 weight" can be appropriately adjusted.

In another situation, NCHS might sample hospitals or physicians to get information about patients. The study may require for each sampled patient a record of his experience with hospitals and physicians. The only feasible way to assemble such data is to know, for a period, the identity of the patient.

In both of these situations, as in others, the solution is to (1) restrict access to the identifiers to a minimum number of sworn employees during the processing or matching interval, (2) physically separate the key from the substantive record as soon as it is procedurally possible, and (3) destroy the key as soon as it is no longer needed for processing or quality control.

Frame for sampling.—A major contribution of modern statistical theory is the introduction of probability sampling as a means of obtaining higher quality information at lower costs. Thus, for example, it is likely that better statistics will result from careful processing of data from a probability sampling of 500 hospitals than from a more routine tabulation of data from a universe of 7,000 hospitals—and at considerably less cost. However, sampling requires at least some measure of identification of individual units in the universe.

In the sampling of facilities or other business establishments, confidentiality is normally not a real issue. The reason is that name, address, nature of business, and "size"—the attributes usually needed for sampling—need not be given protected-data status. This information can reasonably be considered to be in the public domain.

On the other hand, within a substantial class of situations, it is improper to make a list of establishments available as a frame for sampling. Suppose the list contains—or has been constructed from—information that has been designated as "protected data"; for example, it contains a count of abortions performed in a particular hospital in the previous year. It is improper to give that list to a third party, or to use it as a frame for sampling of hospitals conditioned on that information, or to give a third party access to the sample. The immediate reason for this impropriety is that to use the list as stated is equivalent to saying: "This is a list of hospitals where abortions are performed," when the confidentiality assurance prohibits such a disclosure. The underlying reasons are that such a release might damage the hospital and had not been requested when the hospital first supplied the data.

Using lists of persons as frames for sampling has most of the characteristics of the establishment problem with additional features. The key here is to explain to the respondent, when he first replies, the extent to which his reply may

be used subsequently as the frame for further sampling or followup. Subsequent use must be within the bounds of that explanation. Any potential return to the respondent might be considered an unreasonable demand on his time. Therefore, he needs to be given the opportunity to refuse at the first inquiry.

A delicate aspect of subsequent use of protected data, for frames or other purposes, hinges around the phrase ". . .without his consent. . . ." Suppose, in a first survey, nothing is said about using the data for a given purpose. Later, the collector wishes to use the data for that purpose and considers returning to the original respondent and asking for consent to do so. This procedure is acceptable in most situations, although it would have been better to have foreseen the request when the data were first collected and to have secured consent then. In other situations, however, it is not acceptable, for example, if the respondent has to make a new decision that is in itself compromising, he should not be forced to make the decision.

A specialized instance of followup is found in "two-phase" survey designs. In the first phase, the initial measurement or inquiry is to classify persons or other entities into differing categories. In the second phase, additional measurement or inquiries are administered to subsamples of entities classified in the first phase. Clearly individual identification must be retained at least into the second phase of operations.

The necessary and sufficient action for handling these processing exceptions is the one stated in the first paragraph of the section "Extensions of Statistical Purposes" in this chapter: at the time of collection, inform respondents of intended procedures in such a way that they understand and do not disapprove.

CHAPTER VIII

NAMES AND IDENTIFICATION (ID) NUMBERS

Unit Identification

Almost every datum acquired is initially associated with a specific person, facility, product, or other entity. Usually the datum is tagged with a unique identifier that relates it to the specific entity. For persons, the likely identifier is a name or an ID number, such as a Social Security Number (SSN). Sometimes the datum is linked to the person by a nonsense number, and the key to personal identification is stored in a separate record. Facilities likewise may be labeled with their names or with an ID number. As noted in chapter VII, the need by the statistician for unit identification is not for output display but for processing purposes. The recommended guiding principle for handling protected data is to separate meaningful unit identification from substantive data at as early a stage in processing as essential requirements permit. Thus privacy infringements, either intentional or inadvertent, are minimized.

Suggested Procedural Practices

A mechanical device that facilitates data processing while offering additional protection is the assignment of a nonsense identification number for each name of a person or establishment and arranging for a separate register or key that matches on a one-to-one relationship. It is possible, and sometimes desirable, to detach the name, address, and other identifying descriptors (along with a transcription of the ID number) from original documents at an early stage in data acquisition and place them in a separate depository so that only the custodian of the depository knows the identity of any case.

For many sets of protected data in the CHSS—perhaps most—NCHS does not need to possess either the name or the key. The Center needs only the ID number and contractual assurance that the State or other provider of data (and possessor of the key) will provide identification and/or matching in a limited number of cases for purposes of editing, quality control, or followup. For such data sets, a considerable part of the Federal confidentiality problem is resolved.

An extension of this principle and its implied procedure is possible in several other situations. The State statistical agency can use it to provide protected data to other State agencies, to local agencies, and to certain other consumers, always being careful to delete from the microrecord items that might identify the individual entity.

A record that contains numerous descriptors could lead to identification of the individual, even if direct identification is removed, with sufficient desire and detective work. Although this occurrence is possible, it is doubtful that the CHSS needs to protect itself against the combination of malicious intent and scale of effort that would be required.

Social Security Numbers

One of the more controversial issues in privacy debates is the extent to which the SSN should appear on records and be used as an identifier. Thousands of words have been written pro and con on the matter. Several points stand out:

1. The SSN was not originally intended as general identification as indicated on cards issued to individuals.

2. The Social Security Administration has discouraged the use of the SSN as a general identifier.

3. The Federal Government has used the SSN for a wide range of identification purposes and, indeed, since 1943 has required its agencies to use the SSN as an identifier in any new system of personnel records.

4. The Privacy Act of 1974, however, states, in effect, that the SSN shall not be required by any Federal, State, or local agency under any record system in which it was not compulsory prior to January 1, 1975, unless it is mandated by specific Federal statute.

5. In fact, the SSN is widely used as a personal identifier in the U.S. Civil Service Commission, the military, all levels of taxing agencies, automobile and drivers' licenses, banks, insurance companies, credit card companies, private payrolls, schools, immigration authorities, and welfare agencies.

6. Despite the Privacy Act, an increasing number of students of privacy matters believe that the SSN is nearly a universal identifier and is likely to become more so.

CHAPTER IX

DATABANKS AND FILE LINKING

Introduction

Much of the public concern about confidentiality hinges on potential harmful consequences that might stem from the linking of data sets through common ID numbers or names and the building of personal dossiers in giant databanks. For example, Arthur R. Miller, professor at the Harvard Law School and a serious student of privacy matters, confessed that he, much like others, had "voiced the fear that the computer, with its insatiable appetite for information, its image of infallibility, its inability to forget anything that has been put into it, may become the heart of a surveillance system that will turn society into a transparent world in which our home, our finances, and our associations are bared to the most casual observer."[2] Many similar statements have appeared in newspapers, magazines, and books.

It is much easier to talk about such databanks than to construct them. However, if constructed, databanks could serve dangerous purposes as well as beneficial functions. Furthermore, many people advocate the building of integrated health records for individual persons, storing these records in databanks, and making access to the record relatively easy.

Most commonly the entities in databanks are thought to be persons or business establishments. However, banks in which the smallest entity or unit is a collective, such as a county or industry, are possible. In such banks most confidentiality problems disappear.

Pair of Definitions

Because this emotionally charged topic has so many facets, useful communication is likely to occur only when discussion can be focused on a few areas with recognizable boundaries. Consider first these definitions.

Suppose a file consists of records of individual units, persons, business establishments, or other entities. The record for each entity contains unit-identifying characteristics I_1, I_2,. . . (such as name or ID number) and attributes A_1, A_2, A_3, . . . (such as date of birth, marital status, or income). A second file contains unit records with the same identifying characteristics I_1, I_2, . . . and another set of characteristics B_1, B_2, B_3, . . . (perhaps State of residence, make of automobile owned, and number of times arrested in the last 3 years). The process of matching these two files in identifying characteristics and creating a new unit that contains the information, A_1, A_2, A_3, . . ., B_1, B_2, B_3, . . . is called "file linking." The new record may or may not include the common unit-identifiers I_1, I_2, ... (It may also be possible to merge the two files without using the common identifiers, but that possibility is excluded from discussion.) Either of the files might contain many entities, just a few, or even a single entity, but only files that contain a substantial number of entities are of interest.

Clearly, it is theoretically possible to link three or more files to form the new record. If three or more files are linked and *if the unit identifiers are retained in the new record,* then the collection of new records is called a "databank."

In the discussion that follows, attention will be restricted to instances where the linking is of unit records of *persons* or *facilities,* although, as noted, the process can be applied to units that are small collectives, such as a county, an industry, or an occupation.

Some Databank Cousins

Many record keeping activities have features similar to those just described, but should not be considered to be databanks—perhaps they could be called "databank cousins." The CHSS is not much concerned here with databank cousins, but for background purposes a few might be noted.

Ordinary double entry bookkeeping involves the posting of similar items to a common account; however, it is not intended that differing kinds of information about a person or facility be merged. The more extensive charge systems such as American Express or VISA are *single* files and not databanks. The same can be said for the university's file of student grades, the airline's reservation system, or the business office's records of a hospital. Merging any two of these separate files creates linked files. Merging payroll records, student records, and hospital records into a file that retained person identification constitutes a databank.

Technological Feasibility

It was always theoretically possible to compile dossiers on persons or facilities by manual techniques. Where policy, will, time, personnel, and financial resources are present, it is possible manually to merge birth, death, employment, health, military, church, financial, tax, and other records for individual persons. To some degree this merging has been done by most societies, and to a high degree by a few. However, it is difficult and expensive. Increased use of common identification numbers, such as birth or SSN's, and the capabilities of the computer have made linking of files and construction of databanks less difficult and less costly. However, this process is still neither easy nor inexpensive.

Even with the computer, constructing a databank presents many difficulties. After studying numerous record systems, Alan Westin and Michael Baker identified four major hurdles in databank building:[3]

1. Requirements for proposed databanks usually demand massive changes in component record keeping and reporting, and these changes meet with resistance from the managers of the components.

2. Conceptual problems in determining what items are truly useful and significant result in failure to establish agreed-upon data sets that are properly maintained, or that can be utilized for any significant purpose.

3. Software for appropriate edit, input, and retrieval of stored data in unforeseen format demands is rare, if existent at all. People can browse through or muddle along with imprecise records—but the computer, which must be instructed in great detail for every contingency, cannot.

4. Costs of problem solution and even routine operations are great, and top managers are most reluctant to allocate scarce resources to a low-yield databank.

Prevalence of Databanks

The construction and maintenance of large databanks are certainly technologically possible. However, except for a few special-purpose databanks, notably in the fields of credit, insurance, and law enforcement, the databank in the United States is a possibility, not a reality. In 1972, Westin and Baker found not a single general-purpose databank, either local or national, in the country.[3]

In the medical and health fields there are proponents of and experimenters with databanking information pertinent to individuals. Their long-range goal is to accelerate diagnosis and treatment by enabling physicians to retrieve a patient's medical history from a remotely located central databank. Physicians may also be able to tap stored statistical data to assign "probabilities" to specified courses of treatment. Such databanks are future goals, not present realities. A few less ambitious embryonic local medical systems are operating, for example, the multi-state information system, which links psychiatric hospitals, clinics, and outpatient centers in the New England area.

Linking Other Than for Databanks

Other demands for files do not create databanks. In various Federal Government operations, two data files can be linked for statistical purposes only. For illustration, business data from the Social Security Administration, the Internal Revenue Service (IRS), and the Census Bureau are linked to produce detailed economic information in industrial and commercial fields. The linked data are released only in aggregated form and are never used for governmental action regarding an establishment.

Another illustration is the use of one agency's data by another. The Form 1040 Personal Income Tax information for consecutive years is linked through SSN's to show an individual's migration and then fed into Census Bureau procedures to estimate local population mobility. The local population data then become input for calculation of revenue sharing. Again, such uses are strictly statistical and not for action concerning an individual. The IRS may exercise enforcement action regarding individuals, but this action has nothing to do with this linking operation. This situation illustrates an administrative record as the source of both a protected data set and a case-action datum.

In the 1970 Decennial Census, a complex evaluation project involved the linking of census data with birth registrations. The project was conducted by the Census Bureau, using NCHS birth records (with the permission of State agencies). The ultimate objective was to estimate the lack of coverage in the 1970 Census by comparing those counts with aged birth cohorts from earlier years. However, it was necessary to discover the extent of underreporting of births through a reverse record check on a sample of persons to determine if birth certificates existed for them.

In some situations, (see chapter VII), two files may be linked to form a comprehensive but unduplicated frame or directory of a universe of facilities. Such situations occur if a mailing of notices to all firms in the universe is requested, or if one wishes to draw a probability sample from a high-quality frame.

Researchers and program planners make persistent demands for linked files for multivariate analysis. The objective is to identify variables that explain or predict a dependent phenomenon of prime interest. With a single file, the analysis may be able to relate, for example, only the dependent and two independent variables. By linking two files, the analysis may be expanded by considering three additional significant predictors. The logic of this argument is persuasive. Yet aside from questions of confidentiality, three considerations that lessen the force of the demand for linked microdata files are:

1. Only in a few examples of this technique, important results that could not have been obtained by other methods were secured from linked microfiles.
2. A method that deserves more intensive study is substituting small collectives of persons for individuals as the units of analysis in multivariate equations.
3. Often the attempt to integrate two bodies of data into clean individual records faces substantial operational difficulties. Such hurdles are different classification systems in the two files, different coverages, incomplete individual records, different time references, different substantive definitions, different computer formatting, lost or incorrect documentation, and inadequate resources.

Recapitulation

Many people fear linking of two files of identifiable microdata. Even when the linking is for statistical purposes only, it does increase the risk of undesirable disclosure or exposure to invasion of privacy—albeit, nearly always to a trivial degree. Earlier sections of this chapter discuss significant features of the linking problem. The linking of microdata files is sometimes useful, but it is usually difficult and expensive and often neither necessary nor very productive. Recommendations are as follows:

1. Use the linking of microdata files infrequently because it is potentially dangerous to confidentiality and use it only when the intended product is both

highly valuable and unattainable by other means at a tolerable cost.

2. Employ available techniques to camouflage the identity of the linked microdata, and make certain that the ensuing risk of harm to individuals or facilities is inconsequential.

3. Inform the respondent at the time of original collection of a data set that a linking operation is intended.

4. *Never* use protected data in a linking operation that results in a data set used for case-action purposes.

CHAPTER X

NETWORK OF PARTICIPANTS IN THE COOPERATIVE HEALTH STATISTICS SYSTEM

Overview

The concept of the CHSS encompasses collection, processing, analysis, and use of data and implies corresponding interacting relationships among the collectors, analysts, and the ultimate users. This report cannot declare what the organizational structure of this system should be. A broad spectrum of political and economic considerations will be the major determinants of that structure. Legislation is not the least of these considerations, and certain legislative concepts will be discussed in the next chapter. This chapter discusses the impact of matters of privacy and confidentiality on organizations and the influence of organizational structure on both policy and procedures in data handling.

Entities that are or may be participants in the system are NCHS and other Federal agencies, State Centers for Health Statistics and other State departments, municipal bodies, Professional Standards Review Organizations (PSRO's), Regional Medical Programs (RMP's), the Planning Organizations under Public Law 93-641, AMA, AHA, other professional associations, Professional Activities Study (PAS), other central processors, and subcontractors. Essential elements, too, are the health services providers and recipients, which are the sources of the data: hospitals, institutions, physicians, nurses, patients, registrars, laboratories, and households. Legislators, budget authorities, schools, students, and the general public for whom the entire enterprise is undertaken also participate in the system.

A viable system should be flexible to adapt to changing situations and to take advantage of experience. However, specific and clearly understood relationships regarding confidentiality should be established among all the participants in the system. The target should be clear, and there should be a common understanding of what information will be made available by which providers and to which collectors; what degree of confidential handling of data will be exercised by those collectors; and, in particular, for whom further access to the data will be authorized and for what purposes.

Thus, guidelines and rules governing confidentiality throughout the system are needed. The organizational structure should be constructed so that the rules are known to all affected parties and can be enforced. No single agency or component of the system is autonomous—cooperative development is the mainspring of successful operation. However, every complex activity needs a coordinator, and on issues of confidentiality in the CHSS, NCHS should accept the responsibility of coordinator.

Federal Agency Relationships

Confidentiality relationships among Federal agencies, and particularly between NCHS and other Federal agencies, are largely fixed by law (see chapter XI). Although certain modifications of existing law may be desirable, confidentiality risks to which NCHS and the CHSS may be subject are minimal *within* the Federal establishment. Data acquired by NCHS under assurances of a classification equivalent to protected data have immunity from use in individually identifiable form by any agency (including all Federal Government agencies) outside NCHS, without the consent of the provider. Contrastingly, any

data acquired by NCHS but not given protected data shielding can be potentially in the public domain and available to all, including any Federal agency.

As an integral part of its role as architect and handler of protected data, the Center should ensure that other units of the Public Health Service and the Department of Health, Education, and Welfare (DHEW) stand in the same relationship to NCHS regarding dissemination of data as does any other component of the Federal Government. Legislation may be introduced to give special status to a designated set of Federal general-purpose statistical agencies among whom protected data may be shared. A somewhat controversial issue is whether NCHS might be required to permit use of identifiable protected data for statistical purposes only by other units of DHEW under provisions of the Privacy Act of 1974 or even by Federal agencies outside DHEW under provisions of the Reports Act of 1942. The Center has opposed successfully such interpretations up to the present. Even if this use were permitted, the data could certainly not be used "in whole or in part in making any determination about an identifiable individual."[4]

Relationships Between NCHS and Non-Federal Organizations

NCHS formal agreements regarding confidentiality outside Federal Government agencies should be restricted to two categories. One category is a single designated agency within each State. According to Section 306(e) of Public Law 95-623 "States participating in the System shall designate a State agency to administer or be responsible for the administration of the statistical activities within the State under the System." Preferably this agency will be a State Center for Statistics or a State Center for Health Statistics, but it could be any governmental unit chosen by the Governor.

To maximize understanding and minimize State-Federal confusion, the basic guidelines, authorities, and contracts involving health bodies within the State and regarding health statistics should be between those bodies and the State Center or between the State Center and NCHS. Special agreement might grant latitude for division of a State into two parts—for example, New York City and upstate New York or Chicago and downstate Illinois—if the State chose to do so.

For metropolitan areas that cross State lines, one State could serve as the official contact for NCHS, with local arrangements being coordinated in both States by one designated State Center.

The second category with which NCHS might have formal arrangements is original suppliers of data. These arrangements would be made with the knowledge of, and at times through, the State Centers. However, in some situations NCHS must proceed independently of State Centers and inform them of arrangements that have been made. Such situations are the Health Interview Survey, the Health Examination Survey, the Survey of Ambulatory Care, other national surveys that are not operationally a part of the CHSS, and agreements reached with certain agencies such as SSA, PAS, AHA, or the Census Bureau. A particular subclass of this type is State or local agreements in other programs in States that have not established a State Center.

Central Processors

Much can be said for the establishment of a State Center for Health Statistics or even a State Statistical Center, as a central depository and as an initial data collector, central processor, and distributor of data to Federal, State, municipal, PSRO, planning agencies, and other consumers. This idea is not fully developed here; but detailed development of the concept of the immediate question of central processing and the overall network treated in chapter IX is needed. The Statistics Center should be a permanent State agency, or a State corporation chartered specifically for this purpose. It cannot be either an ordinary private establishment or some ad hoc component of a temporary community coalition.

Metropolitan areas that cross State lines may be handled by formal State compacts.

The statistics center should be established by statute, and the nature of operation should be

watched carefully. Enabling legislation such as Statistics Canada (with much modification) gives a starting point for new enactment.

The reason for this latter recommendation and the reason that the central processor should be a governmentally chartered organization is that NCHS and the CHSS cannot avoid responsibility for the integrity of the system, and yet cannot exercise effective control without line authority based on law.

CHAPTER XI

LEGISLATION, REGULATIONS, AND RULES

Legislative Control and Protection

The privacy-confidentiality-freedom of information complex generates conflicts and antagonisms. In a society governed by law, it is essential that a legal framework be built to guide decisions made in the confidentiality realm. For the CHSS the framework is based on both Federal and State law and derivative regulations, orders, and instructions; adequate coding and indexing of these legal materials are required.

This need has been widely recognized and has resulted in a flood of statutes, regulations, rules, and court cases. More than a dozen Federal statutes, tens of thousands of words in Federal regulations, voluminous Federal rules and procedural statements, and uncounted State provisions have important bearing on the CHSS. More than 200 bills on privacy and confidentiality were introduced in the 93rd Congress, and many were reintroduced in later Congresses. Codification or even a digest of these pronouncements is beyond the scope of the present project. Here, attention is directed to some of the most significant Federal legislation in the area, a few of the emerging legislative principles, and the need for further action.

The NCHS Statutory Keystone

The keystone of NCHS policy and practice in the protection of confidential data has been restated in the revised enabling act for the Center. The Public Health Service Act (42 U.S.C. 242m) provides in section 308(d) (sections 304 and 306 refer to authorization for basic NCHS activities):

> "No information obtained in the course of activities undertaken or supported under section 304, 305, 306, 307, or 309, may be used for any purpose other than the purpose for which it was supplied unless authorized by guidelines in effect under section 306(1)(2) or under regulations of the Secretary; and (1) in the case of information obtained in the course of health statistical or epidemiological activities under section 304 or 306, such information may not be published or released in other form if the particular establishment or person supplying the information or described in it is identifiable unless such establishment or person has consented (as determined under regulations of the Secretary) to its publication or release in other form, . . ."

This statement is the current revision of the former NCHS confidentiality section 305(a) of the Public Health Service Act.

All NCHS procedural controls are consistent with this statute; equivalent Center policy antedates the passage of law. The clause beginning "(1) in the case of information" puts the sharpest teeth into the law—it limits the Secretary's discretion (as well as that of the Center) to defining "consent." This law, too, is the basis for the following explanation in DHEW regulations, which exempt NCHS data from many of the requirements of the Privacy Act of 1974 that could be troublesome to the CHSS:

> ". . . Section 308(d) of that Act requires these systems of records to be maintained and used solely as statistical records. . . . The information contained in these records is not used in whole or in part in making any determination about an identifiable individual, and as required by section 308(d) it is

not published or released in a form which would identify the individual who supplies it or the individual who [sic] it is about.... Currently, much of the information contained in these records is obtained through the voluntary cooperation of States, localities, hospitals, physicians, family planning agencies, and other organizations with the understanding that the National Center for Health Statistics will not disclose to anyone the personally identifiable information supplied by these sources."

The Privacy Act of 1974

As a result of debate over a wide range of related issues, the 93rd Congress passed a compromise omnibus bill, sponsored by Senator Ervin and Congressmen Moorhead, entitled the Privacy Act of 1974. This legislation is very complex and, in its text and accompanying explanation from the cognizant congressional committees, includes more than 9,000 words. Its precise and full meaning may not be known for several years, until supporting regulations, interpretations, and perhaps court decisions clarify ambiguous and partially conflicting provisions and objectives. Paired with the Freedom of Information Act, as amended in 1974, the Privacy Act is the most comprehensive general Federal statute on confidentiality. The following are especially notable:

1. The title of the Act and its introductory sections declare that protection of privacy is the primary objective and that the right to privacy is a personal and fundamental right protected by the Constitution of the United States. A further statement is that it is necessary and proper for Congress to regulate the collection, maintenance, use, and dissemination of information by Federal agencies.
2. The Act also declares, however, that the "use of sophisticated information technology . . . is essential to efficient operation of the government" and that regulations should permit exemptions from the protective requirements of the Privacy Act in cases "in which there is an important public policy need for such exemption as determined by specific statutory authority." (This provision produced the DHEW regulation quoted in the previous section.)
3. The Act notes the requirements of the Freedom of Information Law and does not seek to repeal any of its provisions. Therefore, the text of the Act and its interpretations are new landmarks in the search for balance between protection of privacy and confidentiality and the public's right and need to know.
4. In considerable measure the Privacy Act provides that an individual can determine what records pertaining to him are collected, maintained, used, or disseminated by Federal agencies, and can prevent those records, collected for a particular purpose, from being used or made available for another purpose without his consent.
5. The exceptions are far reaching and (subject to interpretation) may open wide quite a few doors. Some exceptions to protective control (agencies are required to formulate regulations on these matters, and the Office of Management and Budget (OMB) is their overseer) are the following:
   a. Investigative material compiled for law enforcement (with some counter exceptions).
   b. Investigative material relating to suitability for employment and related tests.
   c. Certain archival records.
   d. Data transferred to a person pursuant to disclosure of compelling circumstances affecting the health or safety of an individual.
   e. Data in "routine use" defined as "the use of a record for a purpose compatible with the purpose for which it was collected."
   f. Data given to a congressional committee on a matter within its jurisdiction.

g. Data needed for the Comptroller General of the United States.
h. Data to "a recipient who has provided advance written assurance that the record will be used solely as a statistical research or reporting record, *and the record is to be transferred in a form that is not individually identifiable.*" (Substantial ambiguity of intent results from the way this provision is phrased.)
i. Data pursuant to subpoena by the courts.
j. Material controlled by another law, such as the NCHS law quoted in the section, "The NCHS Statutory Keystone." The reason for this exemption is that the Act does not intend to repeal other laws.

6. Regarding penalties for infractions of the Privacy Act imposed upon agencies contracting with a Federal agency, the contractor and employees of the contractor shall be considered to be employees of the Federal agency and thereby subject to being declared guilty of a misdemeanor and fined up to $5,000.

7. The Privacy Protection Study Commission was established with broad powers of investigation and study and a charge to make further recommendations within 2 years. The President appointed three members of the Commission; the House and Senate appointed two members each. The Commission published its report, entitled *Personal Privacy in an Information Society,* in July 1977.

The Freedom of Information Act

The counterpoint to the Privacy Act is the Freedom of Information Act (FOIA) as amended in 1974 and 1976. The public is properly concerned not only with unwarranted invasion of privacy and improper use of privileged data but also with refusal by governments to reveal information that should be in the public domain. The CHSS must be concerned with both of these hazards.

Congress has tried to deal with the second of the two risks through the FOIA. This law declares that—with important exceptions—records possessed by the Federal Government must be made available to any person, upon demand, at cost. Among the exceptions important to the CHSS are: (1) data specifically exempted from disclosure by statute (e.g., NCHS 308(d))[b]; (2) personnel, medical and similar files, the disclosure of which would constitute a clear, unwarranted invasion of personal privacy; and (3) certain classes of privileged or confidential data. The FOIA imposes no unmanageable restraints on the CHSS, and on the contrary, is an incentive to promote wide dissemination of data in all situations in which confidentiality is unnecessary and has not been assured.

The Federal Reports Act

The Federal Statistical System is coordinated by OMB in the Executive Office of the President,[c] under the provisions of the Federal Reports Act of 1942 (44 U.S.C. 3501) and the Accounting Procedures Act of 1950 (31 U.S.C. 18b). These two Acts give authority to OMB that includes legislative review functions, allocation of budgets, and the right to withhold approval of any reporting plan. Concern has been expressed over certain sections of the Reports Act that may give the OMB Director authority to require any Federal agency to give to any other Federal agency information obtained from any person.

However, careful reading of the Act reveals restricting provisions that, in conjunction with NCHS section 308(d), leave the CHSS immune from the Reports Act regarding improper access

[b]The Government in the Sunshine Act of 1976 amended this provision by adding: "(other than section 552 b of this title), provided that such statute (A) requires that the matters be withheld from the public in such a manner as to leave no discretion on the issue, or (B) establishes particular criteria for withholding or refers to particular types of matters to be withheld."

[c]Administered in part since 1978 by the Secretary of Commerce.

to confidential data. Relevant Reports Act provisions are:

> "Information obtained by a Federal agency from any person or persons may, pursuant to this Act, be released to any other Federal agency only if (1) the information shall be released in the form of statistical totals or summaries; or (2) the information as supplied by persons to a Federal agency shall not, at the time of collection, have been declared by that agency or any superior authority to be confidential; or (3) the persons supplying the information shall consent to the release of it to a second agency by the agency to which the information was originally supplied; or (4) the Federal agency to which another Federal agency shall release the information has authority to collect the information itself and such authority is supported by legal provision for criminal penalties against persons failing to supply such information."

Furthermore, if any information is transferred from one agency to another, the data-protecting legal restraints and penalties of both agencies for improper use apply with full force to the officers and employees of the receiving agency.

The Paperwork Commission

Another act passed by Congress on December 27, 1974, created the Commission on Federal Paperwork. The charter for this Commission is to "study and investigate statutes, policies, rules, regulations, procedures and practices of the Federal Government relating to information gathering, processing, and dissemination, and the management and control of these information activities." The Commission was instructed to give a final report to Congress within 2 years of its first meeting (which was held in early Oct. 1975). Clearly, the charter was broad enough to permit recommendations that might affect the CHSS in many ways, including the issues of privacy, confidentiality, and transfer of data.[d]

[d]The Commission's report was published on Oct. 3, 1977.

Emerging Principles

Various well-supported propositions likely to condition the formulation of CHSS policies are as follows:

1. Persons have a fundamental right to privacy, and this right should not be infringed upon beyond the truly necessary requirements of society.
2. Providers of data to governmental agencies are entitled to know under what authority the data are being collected and for what purposes.
3. An individual must be able to prevent the use of personal information that was obtained for one purpose from being used for other purposes without his consent. (See chapter VII and the section "Statistical Purposes" in chapter XVII for a discussion of "purposes" relating to statistical purposes. See also the section "Informed Consent" in chapter IV for comments on "consent.")
4. Transfer of personally identifiable data from one custodian to another should occur only in accordance with carefully formulated and widely understood written rules.
5. Democratic societies and their governments agree that a wide range of data should be collected and analyzed so that planning and execution of social programs can be performed on a rational basis. (This principle is embodied in the Social Security Act, in health planning legislation, in licensing requirements, and in many other statistics.)
6. A basic difference exists between information collected and used only as statistical evidence (what this report has termed "protected data") and personally identifiable data used directly to affect the rights, benefits, privileges, responsibilities, duties, or proscriptions of individuals.
7. Freedom of access by all to data that are not privileged or confidential should be

inherent. Governments should not categorize data as privileged unless the classification is necessary to secure accurate reporting or to prevent individuals from being unjustifiably subject to harm.

8. Certain classes of privileged data should be immune from subpoena by the courts or legislative bodies.

Unresolved Legal Issues

Adequate legislation is an evolving condition. Of special concern to the CHSS are legal issues that currently are oversights such as:

1. A clear distinction between "statistical purposes" and "administrative purposes."
2. Rules that secure an acceptable balance between the statistician's need for access to data that the holders consider confidential and his legal protection from forced release of data that he has granted privileged status.
3. Guidelines for resolving situations where Federal and State law may conflict.
4. Determination of what operationally constitutes "informed consent."
5. Coding and abstracting Federal law. All relevant Federal statutes, regulations, rules, circulars, procedures, and allied documents must be compiled, coordinated, and indexed into a single printed document, and made available on a broad scale. This activity is *urgent.*
6. Regarding State law, it is unlikely that the same procedures can be accomplished. However, steps should be taken to develop guidelines, and examples should be assembled. Summaries also should be attempted as resources permit. A contract to do this for several States was a step in the right direction. Beginning with "zero draft" and progressing through many revisions, NCHS should write the model State laws, regulations, and procedures. In fact, section 306(d) of Public Law 93-353 required that this be done.[e]
7. Financial auditing procedures are an essential safeguard against malfeasance in governmental and other offices. In what manner and to what degree should auditing be allowed to infringe on privacy and confidentiality? This question is only one facet of the more general issues of legislative oversight of the executive branches.

---

[e]See *Model State Health Statistics Act.*

CHAPTER XII

UNINTENTIONAL DISCLOSURE

Intent and Consequence

Governments plan to secure needed information and to protect the affected respondents and subjects. The CHSS must recognize, however, that plans and intentions may not be fully realized, and consequently must try to incorporate safeguards that will minimize unwanted occurrences. This subject is not treated in detail, but several types of safeguards for which provision should be made are noted as follows:

1. Legislation, regulations, and rules of operation should provide sanctions and penalties for employees, officers, and other parties who fail to comply with guidelines.

2. Physical security of records should be provided, both to assure actual protection and to create a climate of recognized importance. Malicious violation of security in the CHSS is a minor problem, but the public should be informed that such violation is carefully monitored.

3. For each type of data, formal rules should be established and followed to prevent inadvertent disclosure from tabulated data, with special attention to "small cells" and unusual combinations.

4. An individual possibly may be identified from a record if the record contains multiple descriptors even if it does not show a name or an ID number. Editors must be alert to avoid such releases.

5. Another sequence that may violate security of a file—either inadvertently or intentionally—even when the primary file contains no directly identifiable name or number and only a small number of descriptors is what might be termed the "serial linking potential." Suppose Primary File I contains attributes A and C and a nonsense case code *1*. File II contains attributes C and D and cases codes *1* and *2*. File III contains attributes E, F, G, and H and case code *2*. Although Files I and III are not tied at all, it is still possible to establish the full record: A, B, C, D, E, F, G, and H and possibly the identity of the person. No universal solution exists for this problem. Fortunately, the CHSS should not be faced with it often. Recognition of the possibility of serial linking should usually provide sufficient knowledge to avoid its happening.

CHAPTER XIII

PUBLIC-USE TAPES

Traditionally, release of statistical data has been in the form of published tables. Transfer of data in unit form has occurred most often by paper on microfilm copy or punchcard. Conceptually, these transcription devices make possible every computer activity. However, potential of rapid computer transcription and association of data items are viewed as hazardous to the protection of personal information. The computer does make the transfer of data and the merging of different items of information about a given person or facility easier. The computer also has broadened the way for a flexible use of microdata that may actually reduce the risk of misuse of data for individuals—the concept is one of public-use tapes.

The public-use tape is a magnetic tape of individual records, with direct personal identification and other potentially identifying items removed so that the tape can be made publicly available. This device permits researchers or other investigators to manipulate and analyze microdata in useful ways without access to the identification of individuals. The public-use tape thus allows nearly all of the research benefits that access to identifiable records would provide, with less overall cost, and significantly reduces the demand for identifiable data.

The Center has published a report[6] that describes available NCHS public-use tapes and identifies the measures taken to protect confidentiality. Participating units of the CHSS should expand the public-use tape program, while preserving confidentiality by developing appropriate rules for necessary item suppression for the various kinds of data released.

CHAPTER XIV

AVOIDANCE TECHNIQUES

General Comment

Many of the problems and issues in the privacy, confidentiality, and freedom of information complex are difficult. For some, no direct solution that has wide acceptance can be found. Perhaps for others, no direct solution is necessary. Identified here are a few bypassing or avoidance techniques that make it possible to reach informational objectives without yielding access to microdata that identify individual persons or establishments. Some of these methodologies have special relevance to those confidentiality problems that are present in linked microdata.

Aggregated data.—The simplest avoidance techniques are restriction of access to microdata to the collectors and processors and publication or release of data only in aggregated format. Although these techniques are clearly not the answer to all problems, they may be the answer to many more than is apparent. In program planning, execution, or evaluation, aggregated data may be fully adequate or even economically more efficient than microdata.

Microaggregation.—This relatively new device seeks to retain a considerable part of the advantage of microdata, while utilizing only aggregated quantities. Many variations exist. An illustrative one sorts persons into small units of, for example, five people; then calculates an average value for each of the statistics of interest for each unit. These average-valued small units become the units of analysis. In a refinement or variant of this process, a distinct new value is assigned to each of the five persons in a small unit. The new value is the average of the five plus a random normal deviate (the size of the normal deviate was calculated from several similar groups of five persons).

Random contamination.—A different, but somewhat similar technique adds the random normal deviate to the observed value for each person. This process also preserves the overall totals and means, but camouflages the data for an individual person. It tends to preserve the distribution of cases that would result from using original data.

Random substitution.—In this variation, a first random number is selected to decide whether to leave an original microdatum untouched or to modify it; if it is to be modified, another random number is chosen to be the selector of a substitute measure for the observed measure—the substitute measure being one possessed by another person in the survey.

Range measurement.—For many purposes, an exact measure of a characteristic is unnecessary (an exact measure may be nearly impossible anyway). If the statistic is income, then instead of asking for annual income initially, the question is put in terms of "intervals" or "ranges," so that, perhaps, the income is primarily recorded as "between $5,000 and $10,000." The respondent may consider such information as nonsensitive, and, thus, the confidentiality issue is avoided.

Randomized inquiry.—This technique, called by some the random response procedure, has many possible variations. Its essential feature is that the original collector never knows how the respondent replied to a particular question, or even *if* he did. Yet summary measures are obtained, with calculable precision.

Synthetic estimates.—This general methodology also has many possible variations. The central characteristic is the use of an algorithm to calculate an "expected value" for a unit or class of units, by utilizing a weighted average of rates obtained from a larger conglomerate, with

the weights established by known, nonsensitive characteristics of the units or small class of units. Some students prefer to think of this as a special case of substituting the value obtained from a fitted regression equation for an observed datum.

Responsible constraint.—Possibly the best rule of all is the adoption of a policy of responsible constraint. This method advocates that one not collect an item at all, unless the need is clear-cut and the value of the information outweighs the risk of privacy infringement. (What a person does not know he cannot reveal.) In a similar vein, it may be better not to link two microfiles, to display tabulations of marginal value with possible unintentional disclosures, or to employ unnecessarily refined classifications of persons or facilities.

Conclusion

If the objective is a census count of hospitals or physicians by county, the avoidance techniques are not relevant. However, a large part of the CHSS output is expressed as means, averages, ratios, rates, correlations, and other statistical measures. Ingenuity in formulating avoidance techniques can produce unbiased estimates of such measures with adequate camouflage of identifiable persons. This approach may be a major pathway toward resolution of problems that are otherwise unsolvable in a climate that is increasingly sensitive to confidentiality issues.

CHAPTER XV

CUSTOMIZED VARIATIONS OF PROCEDURE

Need for Flexibility

"System" implies a structured, integrated, standardized operation that is under control. The CHSS is a system, and should govern itself accordingly. This fact does not mean, however, that every component of the CHSS should be conducted in identical fashion. Some components of the CHSS are, in effect, intended as efficient mechanisms for assembling, processing, and redistributing information that is largely a byproduct of certain administrative functions and is in the public domain. Other units, giving assurances of confidential handling, collect data whose only function is statistical. Still other units are engaged in activities that are a mixture of these other two. (See "Definitions and Labels" in chapter VI.) The policies and procedures most appropriate to the differing situations also vary. The present chapter offers examples of how the CHSS, operating under a basically uniform set of standards, can still tailor its treatment of particular components to special circumstances.

The self-standing direct collection sample survey.—This survey refers to the classical statistical sample survey in the tradition of the Census Bureau Current Population Survey, or the NCHS Health Interview Survey, in which the intent is to produce new data or newly organized information in aggregated format, with no release or disclosure of personally identifiable microdata; thus the full force of protected data applies in strictest interpretation. The survey may be continuing, intermittent, or one time. Respondent sources might be direct measurement, interview, mailed questionnaires, or transcribed records and may be obtained from persons, households, facilities, or other providers. The collector is a single agent, such as NCHS or a State Center, and assures the provider that responses will be released in anonymous form only, will not be released in identifiable microform outside the collecting agency, and will be used only for the purposes that have been described to the respondent.

If the collector is NCHS, individually identifiable data will be disclosed neither to other Federal agencies nor to any other component of the CHSS. Similarly, if a State Center is the collector, individually identifiable data will not be disclosed to NCHS.

Presumably, the respondent will have been told at least one specific purpose for which the data are being collected and that the intended use is for statistical purposes only. Furthermore, the respondent must be given a reasonable understanding of what is encompassed by the expression "statistical purposes." Thus a summary of the interpretation explained in chapters VI and VII of this report is in order. In particular, if the survey microdata may be the avenue for a subsequent contact with the respondent, he should be so informed. If the survey data are to be linked with other microdata, this fact should also be made clear. The degree to which confidentiality assurances can be enforced should also be made known to the respondent. Usually respondent identification should be removed from substantive data at the earliest feasible processing stage to minimize exposure to risk.

The cooperative protected-data sample survey.—A second class of CHSS components might include several variations. All encompass most of the features previously, but have one very important distinguishing attribute that, in turn, leads to other policy and procedural requirements. The distinguishing attribute is that, in addition to the respondent, two or possibly

three parties would have access to the microdata, rather than the single collector. The first party is the original collector—perhaps NCHS or a State Center—the second party is NCHS if the State Center is the collector, or the State Center if NCHS is the collector. The third party is a convenient label for any other specified person or agency to whom access may be granted for particular stated purposes.

The Hospital Discharge Survey (HDS) can serve as an example of the third party, but the following discussion is not meant necessarily to recommend a collection design for that particular survey. If a master HDS sample is designed in such a way that State strata or subuniverses are defined, then some State agencies will collect data from hospitals in their States and make the microdata available to NCHS. For other States, NCHS would be the collector and make microdata available to the States. In either case, all or parts of the microdata may be made available to a third party—perhaps a Public Health Service (PHS) planning agency—for comparing utilization in several inner cities.

The first two parties might well restrict their own internal uses to statistical purposes. In such cases, the general guidelines previously stated would follow, with the critical difference that neither party could guarantee the performance of the other with absolute certainty. Realistically, it must be assumed that actions by the third party are administrative in character, and may, indeed, entail results that are disadvantageous to some respondents. The justification for embarking on any activities in the CHSS that include third parties of this type is based on cost and efficiency; single collection and processing seems reasonable when the third party has the right to collect the same data that the first party already has collected. Even so, the Center may be prudent to take part in, at most, a limited number of any such activities, for they almost certainly put a strain on the NCHS image as a purely statistical agency.

When such a cooperative survey is mounted, several special steps must be taken:

1. The collector must make clear to the respondent precisely who is to have access to the microdata and for what purposes. The respondent must understand that he is authorizing transfer to the other identified parties.
2. The collector must modify his assurances of confidential handling. (See also "The Self-Standing Direct Collection Sample Survey" in this chapter.) He can continue to give very positive assurances with respect to what his immediate agency will do, but should avoid responsibility for guaranteeing that the other parties will do what they have promised.
3. Despite the precaution just stated, the collector, before agreeing to the survey, should secure, in writing, over the signature of the responsible official of the second or third party, statements setting forth the uses for the data and declaring that the data will not be used by the other party.

Manpower components.—Some manpower statistics will or may arise from enterprises that meet the conditions described earlier, but for others a differing protocol appears desirable in the CHSS. As stated previously, confidentiality should not be promised in data collection unless needed data could be secured only with such a promise, unless the absence of confidential handling constituted a clear and unnecessary invasion of privacy. The economy of single collection with multiple dissemination of microdata, is also recognized if privacy is not unreasonably invaded by such action. Many data on health manpower are in the public domain through registration and licensing bureaus, schools, professional rosters, telephone directories, and other sources. The CHSS can perform a useful function in assembling this information in convenient format and making it available to any person or organization that has a need for such information. For some purposes, statistical aggregation is sufficient. For other objectives, however, the "statistics" need to be classified into such fine categories that privacy and confidentiality cannot be assured.

In this situation, the preferred course for the CHSS is not to consider most manpower items as "protected data" and not to give assurances of confidentiality. Rather, the CHSS should

declare that it is collecting and disseminating the data as a service to the health community, not in its role as a statistical system, but purely as an agent peculiarly equipped to do an efficient job. Precedents for such actions can be found in instances in which the Census Bureau or the BLS have acted as collecting agents for the Department of Defense agencies. The exceptional procedure for manpower data can be restricted to items of information that do not infringe on privacy. Included are such items as name, address, professional classification, sex, and, perhaps, other attributes. Information on more sensitive items, such as income, age, race, nationality, number of patients, and so forth, can be secured separately through special surveys that are accorded protected data status.

Two other considerations should be noted. In one situation it is argued that for some items of manpower information, good-quality data can be secured for transmittal to Federal authorities only if the transfer is in anonymous form. Yet NCHS feels it must have a name or Social Security number not to duplicate data that may have been reported for the same person from more than one local source. It is doubtful that the first of these two premises has much validity. If NCHS wished, however, it could solve the duplication problem by simply requiring that each person's report include an item stating the number of jurisdictions from which the report might have come, if a census or complete enumeration had been taken.

The second consideration is similar to the caution that was urged in the section, "The Cooperative Protected Data Sample Survey" in this chapter, concerning access by third parties. The policy and procedure suggested here for the manpower component in the CHSS are foreign to the main thrust of the cooperative statistical system. They endanger, to some degree, public confidence in the system. Therefore, this pattern of operation should not be extended to any other components of the system.

CHAPTER XVI

TRAINING, PERCEPTIONS, AND PUBLIC RELATIONS

Training in Ethical Standards

Whatever the laws and rules may be, whatever the structure of systems, whatever the mechanics of operation—actions are taken by people, and many are employees of the CHSS. Perhaps the greatest single safeguard the system can have is a knowledgeable workforce that understands and is dedicated to conducting a program that is equally balanced between assembly and dissemination of useful statistical information and appropriate protection of the privacy and confidence of those who supply the information.

Whether it is the mark of impartiality, the devotion to the quality of a product, or the protection of confidentiality, the real strength of such Federal statistical agencies as the Census Bureau, the BLS, and the NCHS, and the better structured State and private organizations, in the last analysis, resides in the staff of those agencies. In the Center, the staff *believes* in the protection of confidentiality, and is dedicated to it.

A series of steps should be taken to indoctrinate all the Cooperative Health Statistics System personnel on the fundamental value and importance of confidentiality. These steps include (1) dissemination of written materials, (2) workshops such as the one on Privacy and Confidentiality in Atlanta, Ga., on March 3-5, 1976, (3) several regional training sessions, and (4) creation in the National Center for Health Statistics of a consultant on confidentiality to provide service to States.

The objective of this effort is to create a grass-root devotion to and advocacy of a sound policy on confidentiality, and to avoid any feeling among States that confidentiality is simply another instance of burdensome Federal regulations that must be tolerated.

Real and Perceived Situations

A continuing study of this topic points to the importance of a pervasive phenomenon: Despite significant philosophical differences between real and perceived situations in general, in the field of confidentiality issues, these situations are intermingled or even indistinguishable.

Policy and practice need to be guided almost as much by what people think the situation is as by what the facts are. This state of affairs is the consequence, in part, of deeply imbedded desires and fears that people have concerning their privacy, rights, privileges, and inhibitions being misused. It also reflects the widespread public impact of such matters as Watergate, the Ellsberg break-in, questions about the activities of the Federal Bureau of Investigation (FBI) and the Central Intelligence Agency (CIA), the Pentagon Papers, tales about credit-rating bureaus, the horror stories that have been detailed concerning computers and databanks, and the even more frightening forecasts in the book *1984.*[7] In other directions there are the cries of "freedom of the press" and of the rights of patients, students, and citizens to have access to information about them that is held by physicians, schools, and the Government.

An important consequence is the necessity for NCHS, the CHSS, and others to devote substantial energy to explaining the role of statistical information in the world, and, in particular, to a public relations program that allays unjustified fears. This objective will be enhanced by (1) making absolutely certain that a straightforward presentation of intentions is made to all concerned, and that those intentions are fully honored; and (2) assembling and disseminating only data for which there is definite need, and for which the benefits outweigh the burden on providers of the information.

CHAPTER XVII

UNRESOLVED PROBLEMS AND LESSER ISSUES

Role of This Chapter

Earlier chapters of this report directed attention to basic fundamental concepts, principles, and issues, and to derivative problems. Proposed policy positions and suggested actions have been offered for a variety of situations. This chapter consists of supplementary material of two kinds: (1) identification of related additional or pinpointed issues not fully resolved and (2) a partial catalog of specific facets of the privacy, confidentiality, and dissemination complex treated superficially or, not at all, in earlier chapters. Some problems are quite pertinent and important. Others are intrinsically small, but their resolution may have reverberations that grow into or merge with larger issues. The CHSS will probably not want to establish explicit guidelines for treating each of the items in the list. However, it is desirable to overlook none entirely when policies and procedures are formulated. Order of presentation has no particular significance.

Source data that serve both administrative and statistical purposes.—This report has emphasized the distinction between protected data and case-action data. One of the most vexing situations is when a single initial source record becomes the basis for an administrative action directly affecting an individual and an element of a statistical body of data for which confidentiality is promised. Satisfactory resolution of this matter is critical. The problem consists of the need for both sound procedures and for public acceptance of those procedures. Policy guidelines have been suggested for such situations in several parts of the report and especially in "The Self-Standing Direct Collection Sample Survey" and "The Cooperative Protected-Data Sample Survey" in chapter XV, and "Definitions and Labels" and "Use of Administrative Records for Statistical Purposes" in chapter VI. The report has not left the matter entirely unresolved. Neither is it fully resolved—particularly at the State level where a single agency, at times, performs both statistical and case-action functions. The key to a further solution will lie in the drafting of procedural rules for handling specific data sets, in taking into consideration the many principles analyzed in this report, and in promoting widespread knowledge, understanding, and acceptance of those rules.

Statistician's access dilemma.—Each special interest group—journalists, lawyers, physicians, researchers, legislators, or statisticians—tends to feel that it should have ready access to almost any data source relevant to its perceived needs and also be protected from forced disclosure to other parties. In the CHSS, this attitude constitutes an unresolved dilemma for the statistician. The essence of the statistician's position is this: He should be given access to almost any needed data because how he uses data gives society valuable information without detriment to individuals or facilities; he should be immune from compulsory release of privileged data in identifiable form because he could not acquire better evidence, guarantee the absence of individual detriment, or accomplish his mission. Not everyone accepts this position; consequently, the situation is less than ideal.

Statistical purposes.—In this report and elsewhere, the expression "for statistical purposes only" is widely used. Indeed, this concept is fundamental to much that is herein recommended. Statisticians believe they know the meaning of the expression. Yet an unequivocal

short definition, acceptable to all, is still awaited.[f] Consider the following questions, for example, to which the statistician can give answers, but answers not agreeable to everyone:

1. Is it an acceptable statistical purpose to use names or addresses collected in one enumeration as a sample in another survey? The answer is "yes," but see limiting conditions in "Frame for Sampling" in chapter VII.
2. Is it legitimate to link identifiable personal data from two surveys taken for different purposes, even if the linked data are not (easily) personally identifiable? The answer is again "yes," but with the provisos set forth in "Recapitulation" in chapter IX.
3. Is it reasonable to release "statistics" for a small area or a small class of persons if the statistics show a rate or average for the small cell that is dangerously unfavorable to members of the cell, even though data for an individual is not released? Once more the answer is "yes," but again only in special circumstances and when the value of the information clearly outweighs the rights of potentially affected individuals.

Joint partnership versus *purchase of data.*—In a fully cooperative joint partnership, all partners have equal responsibilities and in general are subject to the same governing rules. If State Centers for Health Statistics and NCHS are full partners in a joint system, then they may be subject to responsibilities for which no partner is in a position to guarantee performance because one partner does not have absolute control over the others. Is a partial solution found in an arrangement where a State collects and controls handling of certain data for its own purposes, and then cooperatively sells a product to the Federal Government? Is such a purchase of data a dodge to avoid legal requirements?

Watchdog boards.—A safe prediction is that whatever legislation is passed and whatever rules are adopted, in some instances, the laws and rules are disregarded or interpreted nonuniformly; and some situations will not be clearly covered by the laws and rules. In these circumstances a disinterested monitoring or watchdog board would be established to oversee governmental performance and to settle citizen grievances. However, if such boards were established, their charters should be very carefully drafted. They would have the potential either of suppressing information that should be readily available or of opening the gates so wide that confidentiality would have little meaning. And certainly they would contribute to time delays in resolving issues.

Additional facets.—The list of topics that have relevance to privacy, confidentiality, and dissemination of information is unending. This final section is included not in any pretense of completing a catalog of factors but in recognition that the full story extends beyond this report, and brief allusion to some further elements can underscore that fact.

Not much has been said about quality of data or quality control. This omission is not the result of doubt or about their basic importance, but is due to the fact that they are a separate field and outside the primary scope of this report. Two major intersections occur on the subjects of quality and confidentiality of data. The first is the conviction that the quality of reported data is better when the respondents and the persons to whom the data relate are anonymous. The second intersection is that most forms of quality control require access by statisticians to a sample of individually identified cases to validate input to the system. A collateral matter is the extent to which identifiable microdata should be made available to peers to permit replicate treatment and verification processes.

Physical security of data also has received little attention in this report. Much has been published elsewhere concerning this aspect. The main reason for nondetailed treatment here is

[f] One attractive comparison is made by Margaret E. Martin, "Information to be used administratively usually requires action on individual cases. . . . Statistical information, on the other hand, is intended to be aggregated or summarized in some form, and the specific identity of [individual cases] is *immaterial* to the usefulness of the results."[8]

that physical security, although not to be overlooked, is a relatively minor problem in the CHSS. Existing laws, buttressed by modest precautionary measures and rigorously enforced, are probably sufficient. Although computers open new channels for violating security, on balance they are more likely to enhance security.

There have been many suggestions and some pressure for NCHS to establish a death index; that is, a national register of all deaths so that anyone could consult the index to determine if a death certificate had ever been filed for a designated person. This is a borderline situation, in which the CHSS and the Center could perform a useful service for both researchers and administrators—but with the risk of infringing confidentiality, or at least the appearance of doing so.

Chapter XV "Customized Variations of Procedure" advocated the concept of procedural variation in certain situations and in several dimensions. It is appropriate for differing kinds of data, with such subjects as facilities, staff, patients, residents, inmates, outpatients, providers, hospitalization, ambulatory care, emergency care, births, deaths, dental data, fiscal matters, costs, expenditures, and insurance. It is relevant also to uses such as reference, planning, monitoring, control, evaluation, databanking, case treatment, workload, supply, utilization, demand, inventory, standards, levels, trends, rates, relationships, unit costs, incidence, and prevalence.

Duration of confidentiality may be an important feature of policy. Should an assurance of confidentiality extend into perpetuity? Or is 5, 10, or 50 years sufficient, or some other period?

A closely allied matter, but still distinct, is the question of retention of original or transcribed records. How long should they be kept in active files, or in archival storage?

Are there special procedures that should be invoked when data are collected by direct observation without knowledge of the subject? Is this *ipso facto* an invasion of privacy? Should the CHSS allow or outlaw such practices?

There may be legitimate differences of opinion over how completely frank interpersonal relationships should be. There are some risks in securing compliance and entirely truthful response if the statistician explains *ad nauseum* the reasons for and all conceivable uses of requested information. However, in a democracy, and especially in the currently prevailing environment in the United States, it is expected that Government will be forthright with citizens. In most situations, a straightforward approach by a collector to a provider of data, making clear the reasons for a request and how the data will be used, will result in compliance by the possessor of the data—with the assumption that there is a respectable justification for the collection. The Privacy Act of 1974, the DHEW Code of Fair Information Practice, and various conferences and professional bodies have declared a need for a policy of openness in the acquisition, handling, and dissemination of information. The CHSS should embrace this policy—not only because it must under the law, but also because it will be a productive course.

The Decennial Population Census is taken first for the purpose of apportioning congressional representation among the States, but it serves countless other purposes. The Census has become preeminently a *reference* source that describes the people who live in this country. Similarly, data collected in the CHSS serve both specific initial purposes and innumerable reference functions. Allocation of resources between these latter baseline objectives and more immediate specific purposes calls for a high order of programmatic and managerial skills.

It has been argued in this report that the CHSS should build an integrated policy and practice in the realm of privacy, confidentiality, and dissemination of data; and that the structure must embrace ethical, political, economic, legislative, and procedural considerations. However, even this broad perspective is not enough. The CHSS cannot stand alone on issues of privacy or confidentiality, no more than it can in other respects. Significant external developments and activities—currently, and undoubtedly more in the future—will have impact on the CHSS. One needs to recall only a few to be impressed with the potential consequences: regulations, rules, and court decisions under the Freedom of Information and Privacy Acts; the Study Commission under the Privacy Act; the Paperwork Commission; the Health Planning Act; possible new legislation; PSRO actions; pronouncements

of the National Commission on the Confidentiality of and Access to Health Records; the 1975 Report of the Committee on Federal Agency Evaluation Research of the National Academy of Sciences, under the title, *Protecting Individual Privacy in Evaluation Research;*[9] a report of the Committee on Privacy and Confidentiality of the American Statistical Association;[10] the report of the Research Project on Confidentiality, sponsored by the American Political Science Association and seven other organizations;[11] and an international study of how to use governmental statistics advantageously without infringing confidentiality which was conducted by investigators at the University of Western Ontario under a Ford Foundation grant.[12]

How to secure balance between the individual's right to privacy and society's need for information is no new problem. It is receiving vigorous attention on a wide front and from many perspectives. Resolution for the CHSS is a dynamic and evolutionary process that should soon reach a degree of operational stability, but for which no terminal point is foreseeable.

REFERENCES

[1]Rule, J. B.: *Private Lives and Public Surveillance.* London. Allen Lane, 1971.

[2]Miller, A. R.: *The Assault on Privacy.* Presented at a Conference on Confidentiality of Public Records, Key Biscayne, Fla., Nov. 7, 1974.

[3]Westin, A., and Baker, M.: *Databanks in a Free Society.* New York. Quadrangle Books, 1972. p. 238 ff.

[4]*Fed. Regist.* 40(158); Aug. 14, 1975.

[5]The Privacy Protection Study Commission: *Personal Privacy in an Information Society.* Washington. U.S. Government Printing Office, July 1977.

[6]National Center for Health Statistics: *Standardized Microdata Tape Transcripts.* DHEW Pub. No. (PHS) 78-1213. Public Health Service. Washington. U.S. Government Printing Office, June 1978.

[7]Orwell, G.: *1984.* New York. Harcourt Brace Jovanovich, 1971.

[8]Martin, M. E.: Statistical registration and confidentiality issues. *Int. Stat. Rev.* 42(3): 267, 1974.

[9]National Research Council/National Academy of Sciences: Protecting Individual Privacy in Evaluation Research. Washington. National Research Council/National Academy of Sciences, 1975.

[10]Report of Ad Hoc Committee on Privacy and Confidentiality. *Am. Stat.* 31(2): 59-78, May 1977.

[11]Carroll, J.: The Confidentiality of Social Science Research Sources and Data. Presented to Russell Sage Foundation, 1976. Unpublished paper.

[12]Flaherty, D. H.: *Privacy and Government Data Banks.* London. Mansell Publishing, 1979.

———— ○○○ ————

APPENDIX

SELECTED BIBLIOGRAPHY

A comprehensive bibliography on privacy and confidentiality was not attempted in this report for two reasons: First, extensive bibliographies have been and are being assembled by others; second, the monthly accretion rate to the already massive list of written materials puts the task outside the scope of the present project.

A number of documents that seem especially relevant to the issues faced in the Cooperative Health Statistics System are listed in this appendix. Many of the publications contain additional bibliographical material. Other useful references are an annotated bibliography *Ethical Issues in Health Services* released in 1970 by the National Center for Health Services Research and Development; *Psychiatry and Confidentiality—An Annotated Bibliography,* a 50-page compilation published by the American Psychiatric Association in September 1974; and a 300-item appendix to the report *Protecting Individual Privacy in Evaluation Research,* included in the following list. Special attention is directed to a computer-based bibliography on privacy and confidentiality being compiled by Professor Tore Dalenius at Brown University. This latter list in preliminary form already contains more than 800 references, and is still growing.

American Hospital Association: *Hospital Medical Records—Guidelines for Their Use and the Release of Medical Information.* American Hospital Association, Chicago, Ill., 1972.

Barabba, V. P.: The right of privacy and the need to know, in The Census Bureau, *A Numerator and Denominator for Measuring Change,* Technical Paper 37. Washington. U.S. Government Printing Office, 1975.

Bryant, E. C., and Hansen, M. H.: *Invasion of Privacy and Surveys: A Growing Dilemma.* Presented at the Smithsonian-Navy Conference on Survey Alternatives, Santa Fe, N.M., 1975.

Carrol, J. D., and Knerr, C. R.: *The American Political Science Association Confidentiality on Social Science Research Project: A Final Report.* Pol. Science Vol. 9, Fall 1976. pp. 416-419.

Commission on Human Rights: *Human Rights and Scientific and Technological Developments.* Report by the United Nations Economic and Social Council, E/CN 4/1142. New York, 1974.

Committee on Federal Agency Evaluation Research: *Protecting Individual Privacy in Evaluation Research.* Washington. National Academy of Sciences—National Research Council, 1975.

Committee on the Judiciary, United States Senate: *Freedom of Information Act Source Book: Legislative Materials, Cases, Articles.* Washington. U.S. Government Printing Office, 1974.

Department of Health, Education, and Welfare: Privacy Act of 1974—Various agencies—proposed rules and notices of systems and records. *Fed. Regist.* Aug. 27, 1975.

Duncan, J. W.: The impact of privacy legislation on the Federal statistical system. *Public Data Use* 3, 1:51-53, Jan. 1975.

Duncan, J. W.: Confidentiality and the future of the U.S. statistical system. *Proceedings of the Social Statistics Section of the American Statistical Association,* 1975.

Hulett, D. T.: Confidentiality of statistical and research data and the Privacy Act of 1974. *Stat. Rep.* June 1975. pp. 197-209.

Hulett, M.: Privacy and the Freedom of Information Act. *Admin. Law Rev.* 27(3): 275-297, 1976.

Jabine, T. B.: *The Impact of New Legislation on Statistical and Research Uses of SSA Data.* Presented at the 135th Annual Meeting of the American Statistical Association, Atlanta, Ga., Aug. 1975.

Losee, G.: Protection of privacy and confidentiality of records maintained by NCHS. National Center for Health Statistics memorandum, Oct. 22, 1975.

Martin, M. E.: Statistical legislation and confidentiality issues. *Int. Stat. Rev.*, 42(3): 267, 1974.

Metz, D. W.: Privacy legislation yesterday, today, and tomorrow. Keynote address to Federal Bar Association Conference, Washington, D.C., May 22, 1975.

Miller, A. H.: Personal privacy in the computer age: The challenge of a new technology in an information-oriented society. *Mich. Law Rev.* 67:1089, 1969.

Mugge, R. H.: Implications of Recent Confidentiality Legislation for the National Center for Health Statistics. Presented at the annual meeting of the American Public Health Association, Chicago, Ill. Oct. 1975.

Office of Management and Budget, Executive Office of the President: Privacy Act implementation—guidelines and responsibilities. *Fed. Regist.* 40(132, pt. III): 28948-28978, 1975.

Office of Management and Budget, Executive Office of the President: *Responsibilities for the Maintenance of Records About Individuals by Federal Agencies.* Circular No. A-108 directed to the heads of executive departments and establishments, 1975.

Proceedings of the Cooperative Health Statistics System Workshop on Privacy and Confidentiality, Atlanta, Ga. Mar. 3-5, 1976.

Rule, J. B.: *Private Lives and Public Surveillance.* London. Allen Lane, 1971.

Spingarn, N. D.: *Confidentiality.* Report of the Conference on Confidentiality of Health Records, Key Biscayne, Fla. Nov. 6-9, 1974.

Task Force on Confidentiality: Report by the Task Force of the Cooperative Health Statistics System Advisory Committee, 1975. Unpublished.

Ware, W. H.: *Records, Computers, and the Rights of Citizens.* Report of the Secretary's Advisory Committee on Automated Personnel Data Systems, U.S. Department of Health, Education, and Welfare, Washington, D.C., 1973.

Westin, A. F., and Baker, M. A.: *Databanks in a Free Society.* New York. Quadrangle Books, 1972.

———— ○○○ ————

*U.S. GOVERNMENT PRINTING OFFICE: 1980 311-240/12 1-3

VITAL AND HEALTH STATISTICS Series

Series 1. Programs and Collection Procedures.—Reports which describe the general programs of the National Center for Health Statistics and its offices and divisions and data collection methods used and include definitions and other material necessary for understanding the data.

Series 2. Data Evaluation and Methods Research.—Studies of new statistical methodology including experimental tests of new survey methods, studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory.

Series 3. Analytical Studies.—Reports presenting analytical or interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.

Series 4. Documents and Committee Reports.—Final reports of major committees concerned with vital and health statistics and documents such as recommended model vital registration laws and revised birth and death certificates.

Series 10. Data From the Health Interview Survey.—Statistics on illness, accidental injuries, disability, use of hospital, medical, dental, and other services, and other health-related topics, all based on data collected in a continuing national household interview survey.

Series 11. Data From the Health Examination Survey and the Health and Nutrition Examination Survey.—Data from direct examination, testing, and measurement of national samples of the civilian noninstitutionalized population provide the basis for two types of reports: (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.

Series 12. Data From the Institutionalized Population Surveys.—Discontinued effective 1975. Future reports from these surveys will be in Series 13.

Series 13. Data on Health Resources Utilization.—Statistics on the utilization of health manpower and facilities providing long-term care, ambulatory care, hospital care, and family planning services.

Series 14. Data on Health Resources: Manpower and Facilities.—Statistics on the numbers, geographic distribution, and characteristics of health resources including physicians, dentists, nurses, other health occupations, hospitals, nursing homes, and outpatient facilities.

Series 20. Data on Mortality.—Various statistics on mortality other than as included in regular annual or monthly reports. Special analyses by cause of death, age, and other demographic variables; geographic and time series analyses; and statistics on characteristics of deaths not available from the vital records based on sample surveys of those records.

Series 21. Data on Natality, Marriage, and Divorce.—Various statistics on natality, marriage, and divorce other than as included in regular annual or monthly reports. Special analyses by demographic variables; geographic and time series analyses; studies of fertility; and statistics on characteristics of births not available from the vital records based on sample surveys of those records.

Series 22. Data From the National Mortality and Natality Surveys.—Discontinued effective 1975. Future reports from these sample surveys based on vital records will be included in Series 20 and 21, respectively.

Series 23. Data From the National Survey of Family Growth.—Statistics on fertility, family formation and dissolution, family planning, and related maternal and infant health topics derived from a biennial survey of a nationwide probability sample of ever-married women 15-44 years of age.

For a list of titles of reports published in these series, write to:

Scientific and Technical Information Branch
National Center for Health Statistics
Public Health Service
Hyattsville, Md. 20782

DHEW Publication No. (PHS) 80-1459
Series 4-No. 22

NCHS
U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Office of Health Research, Statistics, and Technology
National Center for Health Statistics
3700 East-West Highway
Hyattsville, Maryland 20782

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

For publications in the *Vital and Health Statistics* Series call 301 436 NCHS

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF HEW
HEW 396

THIRD CLASS