PRIVACY ASPECTS OF HEALTH STATISTICS

Willis H. Ware

March 1976

P-5619

The Rand Paper Series

Papers are issued by The Rand Corporation as a service to its professional staff. Their purpose is to facilitate the exchange of ideas among those who share the author's research interests; Papers are not reports prepared in fulfillment of Rand's contracts or grants. Views expressed in a Paper are the author's own, and are not necessarily shared by Rand or its research sponsors.

The Rand Corporation
Santa Monica, California 90406

PRIVACY ASPECTS OF HEALTH STATISTICS*

Willis H. Ware

My discussion this morning must be understood to be my own views on the privacy matter. While I am a member of the Privacy Protection Study Commission, that body has taken no position yet; I cannot preempt it, nor would it be appropriate for me to second guess what it might decide. I want to talk about privacy from four points of view.

First, why is there a problem? Why is there suddenly a privacy issue before the country? Why are we suddenly having so much discussion about it in society? It's important to have the background in mind because the thrust and intent of privacy are often overlooked, or forgotten, or subverted by well-intended legislation. The present privacy matter before our country and, in fact, before the world, arises from an inherent conflict between the needs of government and industry for information about people and the concern of individuals about the existence and use of all such information. In particular, people have become concerned about the risk that huge amounts of information residing in data banks about individuals will be used to the disadvantage of an individual or of a group of individuals or of a part of society. Standing to the side of the conflict is computer technology; it has catalyzed the matter. However, one must not identify computer technology per se as the problem, because it is not; rather, computer technology has simply done what it so often does. It has surfaced a long-standing problem of small proportions and made it into a very acute one of major social concern. Unfortunately, many times it must be understood that the computer is a surrogate for the real problem.

There are many forces in society and in the world driving the demand for information about people. This is a big country of 225 million people, each of us leading complex lives. Each of us leaves a substantial data trail in moving from place to place and in the conduct of our lives. Making

---

*This is an edited version of a keynote talk presented to the Cooperative Health Statistics System Workshop on Privacy and Confidentiality, Atlanta, Georgia, 3 March 1975.

a country of this size even work means that government has to know a lot about people, what they do, what they own, what their behavior is, what their habits are, and what their preferences are. In addition, the social programs at federal level drive a lot of the demand for information, not only to administer them but also to carry out the Congressional requirement that such programs be evaluated and monitored. There are corresponding pressures in the private sector for having information about people. We suddenly find that facts and data about people have become an essential ingredient to the conduct of modern day society, to modern day government, and to modern day business and industry. We have also found that facts and data about people have emerged as a viable item of commerce; information per se has value, can be sold and made profitable. The classic example is, of course, the mailing list that gathers data from many sources for the purpose of solicitation.

The future is undeniably one in which large amounts of data about people will be required simply to make our country and planet work; there is no escape. What we must do is to install appropriate safeguards that allow organizations to collect and utilize that data as needed, but at the same time will guarantee each of us as individuals safety against misuse of it and legal redress in case of harm. Even so, why has the problem erupted? It's not new to use information about people for purposes of government or purposes of business. What's different? The answer lies partly in that computer technology has given us a whole new scale of record-keeping practices, and partly in that it has suddenly become apparent that there is no legally established basis for the ownership of information, nor generally any legal controls over its collection. Thus, the owner of data about people drifted into the habit of doing with it as he wishes, of using it for his own purposes, and in the case of profit-oriented industry, of using it to generate revenue.

Today information about people is sold; it is bargained for other information; it is consolidated with other information; it is used for mail solicitation or for business expediency. In the large, it is used for any activity that is favorable to the organization that happens to have it or can get it. Except for a few categories such as census data

or drug abuse data or certain health statistics data, such personal information has no legal protection and is subject to court-ordered seizure. This was underscored at Privacy Commission hearings recently in New York City; the records of credit card holders are frequently subpoened by government agencies or by private attorneys. Furthermore, an organization that holds data rarely, and probably almost never, consults the data subject before the organization uses it as it wishes. Except as provided in the Federal Privacy Act of 1974 and in a few states that now have privacy laws, the citizen generally has no way of learning what data banks exist, or of what is held on him in them. If he is fortunate enough to discover one, he has no basis on which to ask to see the record; if he's lucky enough though to do so, he has no basis on which to make sure the record is correct by requesting changes should there be mistakes.

To put it differently, the balance point between each of us as an individual and the totality of record systems that surround us is not now proper in today's world. The privacy issue, and the thrust of privacy legislation represent attempts to improve the balance point and move it toward the individual. Privacy is a social thrust to put safeguards in place to guarantee that as we give information to organizations for legitimate needs, we retain some control over its use, we are protected against its misuse or abuse, and we have a legal basis for redress of harm in case something goes wrong. While computer technology has been a central driver in creating the problem, it is also paradoxically true that computer technology is the one that will help us do something about the matter because it allows one to design and operate much tighter recordkeeping systems -- tighter in the sense of stricter control over dissemination and accountability.

The institutions of society, both public and private are being forced to reexamine recordkeeping practices that have simply grown up over a long historical period that has included dramatic social and technological change, that have never been viewed collectively as a system, but have been created with little regard for the social consequences. It's forcing our institutions to examine how their recordkeeping systems function, how their records are kept, what is done with records, who maintains them -- especially organizationally -- with whom those records are shared -- especially third parties, etc. Interestingly, privacy is also uncovering

some very subtle social issues that will need to be addressed and resolved. For example, what do we as a society consider to be acceptable uses of data about people? What do we as a society consider to be acceptable facts that we will allow to be collected? Parenthetically, it should be noted there is precedent for social control over the collection of information. In employment processes, for example, certain things may not now be asked. A collective judgment of society as represented through a legislative approach has deemed that certain things are unessential to an employment decision, and, therefore, should not be collected.

Such is the panorama of the privacy question, its scope and breadth. The thrust of privacy is not to bottle up information, nor is it to restrict information to the use of the organization that happens to have it. It is a much more diffuse and pervasive question than such a narrow view. I stress the point because the draft legislation that one sometimes sees, especially at state level, suggests that the intent is to bottle up personal data and to record every single item of activity that takes place with information about individuals.

The second aspect that I want to stress is that of definitions. Usage is far from standardized, and discussants of privacy mix up terms which leads to confusion instead of clarification. Three terms of importance are: confidentiality, security (particularly computer security), and privacy.

- o Confidentiality: a status accorded to data indicating that for some reason, it is sensitive and therefore, needs to be protected and carefully controlled vis-a-vis dissemination.
- o Security or computer security: the totality of measures that must be taken to protect a recordkeeping system in three ways. First, to protect the physical facilities, its hardware, its personnel, and its data against deliberate or accidental damage from a defined threat. I stress the phrase "from a defined threat;" normally, a system does not require protection against any threat imaginable, but only what is perceived to oppose a particular one. Secondly, security means to protect a recordkeeping system against denial of its use by rightful owners. One cannot risk the possibility that

an organized dissident group can preempt a recordkeeping system and deny it for its legitimate intended usage. Third, it is to protect the data and/or the processing programs in a system against divulgence to or use by unauthorized individuals. Unauthorized individuals cannot be allowed to access a computer-based record system and do with it inappropriate things either accidentally or deliberately.

- o Privacy: the claim of individuals or groups to determine for themselves when, how, and to what extent data about them is communicated to or used by others. This implies a sense of control, but secondly, there is the protection of an individual against harm or damage as a result of the operation of some recordkeeping -- the misuse or abuse notion. The third component is the protection of an individual or some class of individuals against unwelcome, unfair, improper, or excessive collection or dissemination of information -- the intrusive nature of data collection, or of unwarranted data collection. Health care statistics relates to each of those three aspects.

To review. Data has confidentiality as a special status; the safeguards of computer security are one part of assuring confidentiality. The same safeguards happen also to be part of assuring privacy, but in addition, there is a legal part. While there is some overlap between security and privacy, there are also basic differences.

Next in order is a discussion of the Privacy Protection Study Commission, excluding, however, the provisions and history of the Act itself. Section V of the Privacy Act of 1974 created the Privacy Protection Study Commission, which has a two year life; it began in June of 1975 and will expire in June of 1977. The Commission per se consists of seven individuals supported by a staff of approximately fifteen with offices at 2120 L Street, N.W., Washington, D. C. The members of the Commission include a rather broad scope of professional expertise -- legislative expertise at the Federal level and at the state level as well, law, business expertise, the press, and information technology. Our charter is extremely broad, and we have an enormous task before us. To paraphrase the law, we are chartered to make a study of the recordkeeping practices in the country; it is that

broad. We are to examine the recordkeeping practices of the federal government if that happens to be required, of state government, of local government, of regional government, of private industry -- recordkeeping practices anywhere except in religion. Based on such an understanding of the recordkeeping situation in the country, we are to recommend to the President and to the Congress two things: the extent to which, if any, the provisions of the Privacy Act ought to be extended to other than the federal government; and any other views that we may evolve for other kinds of legislative controls that seem appropriate or necessary to deal with the privacy issue. In addition to the above items which the law says we "shall do," there are others that we may do. We intend to do them all because the "may list" will accumulate the knowledge on which to respond to the "shall list." We will examine such diverse recordkeeping areas as medical, insurance, education, employment, personnel, credit, banking, finance, credit bureaus, commercial reporting industry, cable TV, telecommunications, travel, hotel, entertainment, and even electronic check processing. There are a few specific other things as well. We are obligated to take some position on the mailing list issue, in particular, whether an individual should have the option to have his name removed. We will also look at the matter of IRS sharing of tax data with federal and local governments, and do certain supporting studies, e.g., what laws and executive orders presently impact privacy.

We are started on this very large job. In December of last year and January of this, we held hearings on the mail list problem; the Commission knows the position it wants to take in the matter, but has not yet released a report. While it is obviously desirable to report on each completed component of our task, there is also the possibility that there exist interactions among parts of the overall issue that will make it desirable to hold some things until later. In February, we had hearings on various aspects of the credit card business. We discussed with industry leaders their use and collection of information about people. In March, we will address sharing of tax data and throughout the rest of this year, there is a scheduled hearing every month on the recordkeeping practice of some sector of industry or government. We do have a comprehensive research plan which describes each of the projects plus a comprehensive schedule of when each thing is to happen. It is all available for inspection at the Commission

offices.

Our approach is generally a matrix one. While we need to examine and document recordkeeping practices in various areas, it is also clear that certain things cut across recordkeeping issues -- the statistical use of identifiable data is one example; the Social Security Number is another. In addition to the functional area examinations, we have cross-cutting studies scheduled to be done late this year or early next year. We expect that as we go through each specific subject area, we will also acquire the data base that is essential to discuss, for example, statistical use across all record systems.

It isn't at all clear at this juncture what our end position will be; that obviously has to be so at this point in time because we simply do not know enough about the problem to have any more than individual attitudes, biases, or convictions about what the answers might be. It is important to stress that the spectrum of possible outcomes must include as an admissible answer that the situation is adequately accommodated by the present Privacy Act with perhaps slight modifications to it. That is not a statement of what will happen; I am simply stressing the point that such a result cannot be dismissed out-of-hand and a priori. At the other end of the spectrum would be a recommendation for a very broadly-based omnibus type bill as represented perhaps by HR 1984, already entered in the House. In between are possibilities for legislation targeted to specific industries or record practices.

While the private sector is tangential to the thrust of this conference, I do want to comment about it. It is not clear at this juncture what level of abuse, if any, now prevails or is latent in private industry, nor is it clear what level of cost private industry could accept vis-a-vis providing privacy and security safeguards. Each is an unclear issue, and because cost is a dominant aspect of the privacy matter, business and industry have to be of a special concern to us. In this regard, there is what I call "a state problem." Private industry sees a very ominous threat that fifty states will enact fifty different laws with fifty different enforcement dates. If that were to happen, the recordkeeping systems would be severely affected. If a state wishes to enact a privacy law that impacts only its own public institutions, an outsider cannot object; it is a matter between the legislators of that state and the taxpayers of that

state. If the taxpayers are willing to underwrite the cost for whatever their legislators enact, it is an intrastate affair. In contrast, if a state chooses to enact privacy law that addresses the private industry, then it not only impacts the affairs of that state, but it probably impacts the affairs of the whole country because most companies are in interstate commerce one way or another. Thus the law of one state will affect a company everywhere. While this scenario may seem very unlikely, industry sees it as potentially possible. Therefore, it is hoped that states will not act precipitously toward the private sector, but will await the outcome of our Privacy Commission and also state-level ones.

Finally, we need to consider the privacy matter in the context of health care statistics. My understanding of the National Center is that it collects and furnishes statistics about the health care industry in the country with several purposes in mind. First, it helps the country know how we are doing as a population and what our state of health is. It includes the vital statistics aspect of births, deaths, marriages, dissolutions, etc. The National Center also reports information about the resource availability in health care delivery. I would suppose that it collects data on such things as the number of hospitals, their capability, inventories of physicians, inventories of other health care providers, etc. It clearly must collect data and report -- or project -- demand for health services, e.g., who wants health care services, how is the need distributed nationally, and what kinds are wanted. The National Center does not make decisions, but rather provides inputs to other bureaucratic agencies that make programmatic decisions, that can make the judgments of how we're doing healthwise as a country, that can estimate the demand-versus-supply balance, and that can suggest remedial actions. As someone phrased it to me, the National Center is to provide national benchmark statistics on health matters of the country.

Given that perception, The Center clearly does deal with personal information; it obviously has to. Some of it comes from third parties; some of it comes from its own sources via commissioned surveys. Much of it is bound to be confidential, and, therefore, will have to be protected by appropriate safeguards. In my judgment, the National Center and corresponding state-level centers do have the computer security problem because

each does have to have in place certain safeguards -- for either an automated or a manual system -- that will assure the promise or the legal requirement of confidentiality. On the other hand, neither the National Center nor the corresponding state centers are using information to make determinations about people, if my perception of the role is accurate. No Center uses health care information to make decisions about such things as credit or employment or insurance or educational benefits or retirement benefits or other rights, privileges or benefits. One might argue offhand that there are no privacy issues involved, but I will argue to the contrary.

There are privacy issues, but it is important to sort them out because some of them may not be of sufficient importance to warrant attention. Since information about health care delivery was originally collected from an individual for the delivery of health care, in a precise sense, it was furnished by him for the purpose of receiving health care, and it would be his presumption that it would be used for only that purpose. Thus, if the National Center or a state center receives information about individuals that was originally given in exchange for health care, in fact it is using information for a purpose not originally intended; thus, under a strict interpretation of privacy, there is a privacy issue, and there would be a technical violation. On the other hand, it is hard to be convinced that it is a serious matter, in part because it is straightforward to remedy. One makes sure that individuals are informed of the fact that information given to physicians or to hospitals will be forwarded to appropriate state or federal agencies for consolidated assessment of health care delivery and moreover protected as confidential.

On another aspect of privacy, I would judge that the risk of abuse is minimal provided the confidentiality responsibility is taken seriously and provided that the right safeguards have been installed to assure that confidentiality. With that premise, I do not see a significant risk of abuse or misuse of personal information; on that count of privacy, I would say things look very favorable. Presumably, nothing is collected from the data subject by third parties, especially for the National Center. If true,

then there is no intrusive aspect of data collection, and on that count of privacy, one also is in good shape. However, I do understand that you commission your own surveys, and to the extent, therefore, that people might object to being asked questions about health care, health status, etc., one can imagine the intrusive-collection aspect of privacy arising; and one can imagine individuals taking the position that the government is asking questions again.

In a strict and precise interpretation of privacy -- as the word is being used today -- there is risk that privacy issues are present in your affairs. All in all, I cannot feel that there is a major privacy issue concerned with the collection of health care statistics, either at federal or state level. However, I again state my premise that confidentiality is being fully and adequately assured. In many instances, statistical conclusions will have been derived from data bases in which information is personally identifiable to people, but in doing such a derivation, it would appear certain that appropriate precautions have been taken to assure anonymity, in ways that the Census Bureau has long learned to do. Assuming that confidentiality is done properly and protected, in this circumstance, privacy would appear not to be an issue.

My understanding is that federal law stipulated information collected by the National Center cannot be used for other purposes; that negates the privacy concern that information collected for one purpose may not be used for another. However, it is important to note another characterization of the problem. The context of privacy is that of individuals, not organizations. Hospitals and clinics may well complain that the government is asking questions and wants more data, but the concern of privacy is not that of organizational privacy.

Another issue, really a collateral one, is the personal identifier. The Privacy Act speaks in regard to use of the Social Security Number. The Act prohibits use of the Social Security Number under certain circumstances. The wording of Section 7(a)(2)(b) is somewhat ambiguous; it contains the phrase "verify the identity of an individual." In many instances, it isn't clear whether the Social Security Number is an identifier of an individual or an authenticator for an identification that has been made on some other basis. The distinction is a subtle one to be

made with care; unfortunately, the legislative history does not clarify the fine difference nor the intent. It is conceivable that personal identification could be a difficulty in health statistics. For example, health care providers may be licensed in more than one state; thus, there might be duplication in the count of such people, and there may be no way to remove it. Thus, the head count can be wrong. The licensing practices of the fifty states are unknown to me, but there could readily be fifty different ways. While I can sympathize with the dilemma, I also observe that the Social Security Number is valueless in trying to deal with it, and can only wish you well.

There is a parallel to other events at the national level; one is to maintain a national driver registry for the admittedly desirably social purpose of identifying drivers whose licenses have been revoked in state A but try to get a new one in state B. Clearly, this should not be allowed to happen, but the states maintain driver records in many ways. The lack of a universal identifier is an impediment in just the way that the Center faces in trying to maintain an inventory of health care providers.

If lateral studies on individuals are done in the commissioned studies, there is obviously an identification problem so that the data base can be updated from year to year. The same matter has been faced elsewhere -- notably in the education world -- and so there are solutions that appear to be satisfactory ones. The social security one does not appear to have a pivotal role in the matter. There may be personal identifier problems that I haven't correctly perceived, but the Social Security Number does not appear to be of central concern in Center affairs.

I have not exhaustively discussed all issues. There is an obvious conflict between privacy laws, of whatever kind, and the freedom of information or public records acts in the states. It's a well recognized and unsettled point. There is also the whole question of state-federal relationship. How should the federal government relate to state agencies with regard to health care statistics? There are undoubtedly many other questions. While my discussion has not been exhaustive, I do hope it at least will provide some focusing of the discussions that you will have the next several days.