Date: May 18, 2009

To: Katie Burns
Minnesota Department of Health

From: Beth Virnig, Ph.D., M.P.H.
Associate Professor
Division of Health Policy and Management
University of Minnesota

Subject: De-identification Analysis of Minnesota Health Care Claims Reporting System

Background

This memo, pursuant to 45 Code of Federal Regulations (CFR) part 164.514 (b)(1) documents my review and determinations related to the Minnesota Health Care Claims Reporting System (MHCCRS). I am a person with "appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable." I have applied these principles and methods and have determined that subject to the limitations and qualifications set forth at the end of this letter that the MHCCRS is designed with very low risk that the information could be used alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and that this determination is justified. The design does not eliminate all risk of identification of individual subjects, but it is my professional opinion that the risk is very small.

The design of the MHCCRS uses an effective combination of the following strategies to meet security needs:

- It does not collect the most identifying of data elements (e.g., SSN, address)
- It uses a strong one-way encryption method (i.e., cannot be 'unencrypted') for the data elements used to link all of an individual's records
- Access to data is severely limited by contract— in terms of what is collected, who is allowed to use the data and which elements are transferred to the data analysis contractor
- The allowed use of the data is limited to peer grouping activities only
- Holders of the data are prohibited from linking data with other data sets
- Public reports based on the data will exclude small cell sizes
- The data collection contractor (MHIC) and the proposed analysis contractors will implement numerous information technology security protocols

This is not a legal analysis of HIPAA compliance. Rather, this assessment is based on my professional experience outlined above and pursuant to the requirements outlined in 45 part CFR 164.514 (b)(1).

Professional Experience

I am an Associate Professor in the Division of Health Policy and Management at the University of Minnesota. Much of my research and professional effort relates to the use of electronic health care (or claims) data. I first received Medicare claims data for use in 1994 and am a lead, senior or co-author on over 50 peer-reviewed publications using electronic health data from the Centers for Medicare and Medicaid Services (CMS), the linked SEER/Medicare data or other administrative sources. My responsibility as author or co-author includes making sure my work complies with CMS rules regarding data release and developing methods that keep data secure and avoid intentional or unintentional release of personal information. Since 1998, I have been an investigator on the CMS-funded Research Data Assistance Center (ResDAC) project. ResDAC is tasked with improving the number of researchers skilled in the use of Medicare and Medicaid data through assistance, teaching and outreach. My participation in these activities includes review of CMS privacy and data security policies. I am also Principal Investigator of a contract from NCI to provide technical assistance for the SEER/Medicare linked data set and to evaluate proposals for scientific and data security and to evaluate manuscripts for compliance with data release policies. I will use this background as the basis for my evaluation. While I am employed at the University of Minnesota, this opinion is not being given on behalf of the University of Minnesota.

De-Identification Analysis

The following analysis was undertaken:

1. Review of document: Minnesota Health Care Claims Reporting System: Appendices to proposed Minnesota Administrative Rules, Chapter 4653

2. Review of State of Minnesota Professional and Technical Services Contract for Encounter Data Collection and Processing; CFMS Contract No. B23794

3. Review of University of Minnesota Institutional Review Board Guidance regarding HIPAA compliance.

4. Conversations with Jonathan E. Harvell from the Maine Health Information Center, Vice President for Technology and Administration regarding encryption methods, data security and data elements

5. Conversations with Katie Burns from the Minnesota Department of Health confirming that small cell sizes will be suppressed from reports

In examining the documents mentioned above, I assumed the proposed rule that was last reviewed May 15, 2009 was the most recent iteration of the MHCCRS and should therefore be viewed as the definitive source of information about which data elements will be collected.

The MHCCRS is being created to facilitate the development of a state-wide provider peer grouping system as required by Minnesota Statutes, section 62U.04. The MHCCRS will address issues of cost and quality and incorporate risk adjustment. The State recognizes the need to collect data that can protect individuals from being identified, but still allows their records to be linked to each other over time and across care settings. The proposed system builds upon the common types of claims/administrative data and divides elements by claim type—enrollment, institutional and professional care, and pharmacy.

The strongest protection of privacy is achieved through limitations in the variables to be collected. The MHCCRS explicitly excludes most data elements that are considered to be individually identifying under the HIPAA Privacy Regulations (45 CFR part 164.514 (b)(2)).. Specifically, the MHCCRS does not collect:

1. Postal address information other than town/city, state and zip.
2. Telephone number
3. Fax number
4. Email address
5. Social security number
6. Medical record number
7. Account numbers
8. Certificate or license numbers
9. Vehicle identification/serial numbers, including license plate numbers
10. Device identification/serial numbers
11. Universal resource locators (URL)
12. Internet protocol (IP) addresses
13. Biometric identifiers, including finger and voice prints
14. Full face photographs and comparable images

The only elements that are included in the MHCCRS that are described in 45 CFR, part 164.514 (b)(2)(i), are city, 5-digit ZIP code and encrypted name and health plan number (this number identifies a specific policy, either individual or family but does not identify individuals within a policy). Given the strong one-way encryption method used for names and health plan numbers (see below), it can appropriately be argued that the only potentially identifying elements that are retained are city and 5-digit ZIP code. Thus, the MHCCRS nearly meets the safe harbor de-identification requirements described in the HIPAA Privacy Regulations.

As a second level of protection, the MHCCRS uses strong encryption to render names essentially unidentifiable. Names and health plan numbers will be transferred from health plans and third party administrators to the data collection vendor Maine Health Information Center (MHIC) in an encrypted format. MHIC will use a one-way key where each health plan company and third party administrator (data submitter) encrypts selected data elements prior to transmission to MHIC. This one-way key will be common across data submitters so that a person who moves from one policy to another can still be included in this analysis. Neither the MHIC, nor any other end users, have the ability to reverse this encryption. Thus, data that are transmitted to MHIC in encrypted form can only be used in encrypted form. Without common identifiers, outcomes of care cannot be evaluated and the process of creating peer groupings could not be achieved. However, the needed linkage can be conducted as effectively with an encrypted variable and the potential for identifying individuals is effectively eliminated.

The MHCCRS also uses contractual language to limit the risk of individual identification. The MHIC is prohibited through contract to link the data they collect with any external data and are prohibited from conducting any analysis beyond appropriate quality control activities. The MHIC system is firewalled and access to the data is severely limited and controlled.

As a further protection, the peer grouping will not be conducted by the MHIC but will be completed by a second contractor who will receive a further restricted dataset from MHIC. The analysis contractor will only receive the variables that are required to complete the peer grouping. Limiting the data elements that are transferred from MHIC to the analytic contractor will help limit risk of unintentional identification. The policy that strictly limits data elements to those that are explicitly needed is highly effective and by no means unique. Both CMS and the SEER/Medicare linkage use a similar policy—only elements that are needed are released, and elements that are associated with increased risk of identification are only released if a strong case is made for their value for the approved analysis. Other planned limits include transforming date of birth into age. This will not interfere with appropriate analysis and protects privacy of individuals.

While conducting peer grouping analysis, it is possible that pre-identified combinations of variables will result in small numbers of subjects. For example, gender, age, comorbidity and health care provider could point to a particular individual. These small groups (i.e., small cell sizes) will be suppressed in public reporting.

While there is always some risk of inadvertent release of confidential information, the system, as designed, will limit the likelihood of this event and the actual risk of identification is small. Technology and contractual requirements will reduce this risk of unintentional release of information but, perhaps more importantly, the strong restrictions on number and types of data elements collected and the number and types of elements transferred to the analysis contractor will limit the detail that could be subject to an accidental release.

Conclusion

The opinion provided in this memorandum assumes that MHCCRS is implemented as described in the above noted contracts and proposed rules and other limitations and qualifications described in this memorandum, that third parties fulfill their contractual obligations and that all protections summarized above including suppression of small cell sizes are maintained, based on my knowledge and experience, I consider that the risk is very small that the information contained in the MHCCRS could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is the subject of the information. I believe that this document outlines the basis for this determination. This opinion speaks as of the date given and is based solely on the information and documentation provided to me prior to such date.