No. 10-779

IN THE

Supreme Court of the United States

WILLIAM H. SORRELL,
ATTORNEY GENERAL OF VERMONT, et al.,

Petitioners,

v.

IMS HEALTH INC., et al.,

Respondents.

ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE SECOND CIRCUIT

BRIEF FOR DR. KHALED EL EMAM AND JANE YAKOWITZ, ESQ. AS *AMICI CURIAE* FOR RESPONDENTS

MICHAEL A. POLLARD
Counsel of Record
BRIAN L. HENGESBAUGH
AMY DE LA LAMA
LINDSAY A. PHILIBEN
BAKER & MCKENZIE LLP
130 E. Randolph Drive
Chicago, IL 60601
(312) 861-2786
michael.pollard@bakermckenzie.com

235432

COUNSEL PRESS
(800) 274-3321 • (800) 359-6859

TABLE OF CONTENTS

Page

TABLE OF CONTENTS i

TABLE OF CITED AUTHORITIES ii

I. STATEMENT OF INTEREST OF *AMICI CURIAE* 1

II. SUMMARY OF THE ARGUMENT..... 2

III. ARGUMENT 6

A. The Petitioner Amici Briefs Ignore The Effectiveness Of HIPAA De-Identification Standards Already In Place 6

B. Petitioner *Amici* Briefs Cite Situations Involving Re-Identification That Are Irrelevant To This Case 11

C. The Vermont Statute Does Not Advance Any State Interest In Regulating De-Identified Patient Data..... 18

D. The Vermont Statute Is Significantly More Restrictive Than Necessary To Serve Any State Interest In De-Identified Patient Data..... 19

E. The Vermont Statute, If Upheld, Would Open The Door For States To Restrict Many Other Beneficial Uses Of De-Identified Patient Data..... 20

CONCLUSION 22

TABLE OF CITED AUTHORITIES

Page

CASES

Southern Illinoisan v. Ill. Dep't of Pub. Health, 218 Ill. 2d 390 (Ill. 2006) 15, 16, 17

Southern Illinoisan v. Ill. Dep't of Pub. Health, 349 Ill. App. 3d 431 (Ill. App. Ct. 5th Dist. 2004) 17

STATUTES, REGULATIONS AND RULES

SUP. CT. R. 37.3 1

42 U.S.C. § 1320d (2010) 3

42 U.S.C. § 1320d-5 (2010) 9, 10

42 U.S.C. § 1320d-6 (2010) 9, 10

45 C.F.R. § 160 3

45 C.F.R. § 160.404 (2010) 9

45 C.F.R. § 162 3

45 C.F.R. § 164 3

45 C.F.R. § 164.308 (2010) 10

45 C.F.R. § 164.310 (2010) 10

Page

45 C.F.R. § 164.312 (2010) 10

45 C.F.R. § 164.316 (2010) 10

45 C.F.R. § 164.502 10

45 C.F.R. § 164.502(e)(2) 10

45 C.F.R. § 164.504(e) 10

45 C.F.R. § 164.508 (2010) 13

45 C.F.R. § 164.514 6

45 C.F.R. § 164.514 (a) (2010) 3, 6

45 C.F.R. § 164.514 (b)(1) (2010) 7

45 C.F.R. § 164.514 (b)(2)(i) (2010) 7

45 C.F.R. § 164.514 (b)(2)(ii) (2010) 7

18 VT. STAT. ANN. § 4631 (2010) 2

18 VT. STAT. ANN. § 4631(d) (2010) 3

18 VT. STAT. ANN. § 4631(e) 19

Health Information Technology for Economic and Clinical Health Act of 2009, Pub. L. No. 111-005 (2009) 3

Page

Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (2000) (codified as amended in scattered sections of 18, 26, 29, 42 U.S.C.), *amended by* the Health Information Technology for Economic and Clinical Health Act of 2009, Public Law 111-005 (2009) 3

Illinois Freedom of Information Act, 5 ILCS 140/1 *et. seq.* (West, 1998). 15

OTHER AUTHORITIES

HHS.gov, http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/deidentificationagenda.html. 7

Deborah Lafky, *The Safe Harbor Method of De-Identification ONC Presentation*, October 9, 2009, www.ehcca.com/presentations/HIPAAWest4/lafky_2.pdf 9

National Committee on Vital and Health Statistics Report to the Secretary of the U.S. Department of Health and Human Services, *Enhanced Protections for Uses of Health Data: A Stewardship Framework for 'Secondary Uses' of Electronically Collected and Transmitted Health Data*, December 19, 2007, available at http://www.ncvhs.hhs.gov/071221lt.pdf 9

Page

Michael A. Stoto, J. Domingo-Ferrer, L. Franconi, eds., *The Identifiability of Pharmaceutical Data: A Test of the Statistical Alternative to HIPAA's Safe Harbor; CD-only annex Privacy in Statistical Databases, Lecture Notes in Computer Science 4302* (2006), *see* http://explore.georgetown.edu/publications/index.cfm?Action=View&DocumentID=25940 14

Douglas Sylvester & Sharon Lohr, *Counting on Confidentiality: Legal and Statistical Approaches to Federal Privacy Law After the USA Patriot Act*, 2005 Wis. L. Rev. 1033, 1112 (2005) 18

I. STATEMENT OF INTEREST OF *AMICI CURIAE*[1]

Dr. Khaled El Emam, Ph.D., is a senior investigator at the Children's Hospital of Eastern Ontario Research Institute and heads the Electronic Health Information Laboratory. Dr. El Emam also holds the Canada Research Chair in Electronic Health Information at the University of Ottawa, where he is an Associate Professor in the Faculty of Medicine and the School of Information Technology and Engineering. Dr. El Emam holds a B.Eng. (Honors) in Computer Systems and Electronics, and a Ph.D. from the Department of Electronic and Electrical Engineering, King's College, at the University of London (UK).

Dr. El Emam's main area of research is privacy of personal health information, which includes developing techniques for the secure anonymization of health information and assessing the re-identification risk of health datasets. Prior to his work at the Electronic Health Information Laboratory, Dr. El Emam was a Senior Research Officer at the National Research Council of Canada and head of the Quantitative Methods Group at the Fraunhofer Institute in Kaiserslautern, Germany. Dr. El Emam is the founder of Privacy Analytics, Inc., a company that utilizes his research and expertise. In 2003 and 2004, he was ranked as the top systems and software engineering scholar worldwide by the Journal of Systems

1. This brief is filed with the written consent of all parties. Consent letters are on file with the Clerk of the Court. Sup. Ct. R. 37.3. No counsel for a party authored this brief in whole or in part, nor did any person or entity, other than *amici* or their counsel, make a monetary contribution to the preparation or submission of this brief. *Id.* 37.6.

and Software based on his research on measurement and quality evaluation and improvement, and ranked second in 2002 and in 2005.

Jane Yakowitz, Esq., is a member of the faculty of Brooklyn Law School as a Visiting Assistant Professor, where she teaches Information Privacy Law. Professor Yakowitz's research focus includes privacy law and empirical legal studies. She previously served as the Director of Project SEAPHE (Scale and Effects of Admissions Preferences in Higher Education) at UCLA School of Law, which investigates the effects that admissions preferences based on factors such as race, socioeconomic status, and athletics have on their intended beneficiaries. Professor Yakowitz received her *Juris Doctorate* degree from Yale Law School, and also earned her *Bachelor of Science in Mathematics, with Distinction,* from Yale University.

II. SUMMARY OF THE ARGUMENT

Amici submit this brief to address Petitioner's arguments, as supported by *amici* briefs filed by the Electronic Privacy Information Center ("EPIC"), the Electronic Frontier Foundation ("EFF"), the AARP and the National Legislative Association on Prescription Drug Prices ("AARP"), and the Vermont Medical Society ("VMS") (collectively, "Petitioner *Amici* Briefs"), on the limited issue of the purported privacy risk to the de-identified patient data that is incidentally regulated by the Vermont statute, 18 VT. STAT. ANN. § 4631 (2010), at issue here ("the Vermont Statute"). In short, Petitioner *Amici* Briefs overstate the risk of re-identification of the de-identified patient data in this case. Not only do Petitioner

Amici Briefs advocate privacy interests "even greater than what the legislature expressly recognized in the findings" (EPIC Brief, at 15), they also attempt to support their arguments by referencing examples that stray far afield from explaining how the use and disclosure of the data at issue, which consists of certain information about the *prescriber*, such as the prescriber's name and address, the name, dosage, and quantity of the drug prescribed, and certain data about the prescriber's patients such as age and gender ("Prescriber Data"), could lead to the risk of patient re-identification. If accepted by the Court, such arguments would have significant unintended consequences, as they would open the door for states to drastically restrict the beneficial uses of de-identified data in health care and other settings. Without addressing what level of scrutiny this Court should apply, *Amici* support affirming the decision below.

By way of background, the Vermont Statute imposes a ban, absent prescriber authorization, on the use and disclosure of Prescriber Data for the purpose of marketing prescription pharmaceuticals to the prescriber. 18 Vt. Stat. Ann. § 4631(d) (2010). Federal privacy law, pursuant to the Health Insurance Portability and Accountability Act, 42 U.S.C. § 1320d *et seq.*, 45 C.F.R. Parts 160, 162, and 164 ("HIPAA"), as amended by the Health Information Technology for Economic and Clinical Health Act of 2009, Pub. L. No. 111-005 (2009) ("HITECH"), provides that any patient data contained in Prescriber Data must be de-identified before such patient data can be used or disclosed for marketing purposes (i.e., the requirement is that there must be no reasonable basis to believe that the data can be used to identify the patient). 45 C.F.R. § 164.514(a). This means that the effective scope of the Vermont Statute with

respect to patient data is limited to de-identified patient data. The real risks of re-identification of any such data are very small, such that Vermont has no recognizable interest in regulating de-identified patient data.

The Petitioner *Amici* Briefs overstate the risk of re-identification of de-identified patient data by ignoring the protective privacy effect of the strict de-identification requirements under HIPAA. Instead, the Petitioner *Amici* Briefs provide numerous citations to risks of re-identification in various unrelated contexts. Some of these examples relate to data that is outside the scope of HIPAA (e.g., movie rentals and Internet search queries), and therefore is not protected by the HIPAA de-identification requirements. Other examples may be within the scope of HIPAA, but do not have any relation to the Prescriber Data at issue in this case (e.g., genetic data which is highly specific and carries its own inherent risks of re-identification, or hypothetical situations involving nude photos or detailed psychological notes that could be embarassing even without direct identifiers). Still other examples relate to situations that pre-date HIPAA (e.g., direct consumer marketing based on pre-2003 purchases of fertility treatment pharmaceuticals) and would not be permitted today. Another citation relates to an unpublished study where estimates of re-identification risks using prescription data were made using a complex multi-stage approach; however, the accuracy of the estimated risks is unknown and such risks were still found to be very small in most cases. Moreover, there is no publicly known example of an actual re-identification attack that has taken place in the real world by a real adversary to defeat the HIPAA de-identification requirements (as opposed to attempts in research settings). Consequently, the examples cited by the Petitioner *Amici* Briefs do not

demonstrate how the use or disclosure of Prescriber Data for marketing to prescribers (i.e., the scope of restriction in the Vermont Statute) could compromise HIPAA patient re-identification protections.

Moreover, even if Vermont were to have any residual interest in protecting de-identified patient data, the Vermont Statute does not directly advance that interest in any material way. Among other concerns, for one class of patients (i.e., those whose prescribers authorize the disclosure of their own Prescriber Data and the related de-identified data about such patients), the Vermont Statute makes no difference whatsoever, because the de-identified data about all such patients will be used and disclosed for marketing purposes. At the same time, the Vermont Statute is excessively broad with respect to another class of patients (i.e., those whose prescribers do not authorize such disclosure of their own data and the related patient data), because there is an outright ban on any use or disclosure of such data for marketing purposes, whereas less restrictive alternatives (such as a simple restriction on re-identification) would be sufficient to meet any purported state interest in protecting de-identified patient data. As such, the Vermont Statute is a "poor fit" to achieve any alleged state interest in protecting de-identified patient data.

It is important in this case for the Court not to expand the privacy interests implicated by the Vermont Statute beyond those actually identified by the Vermont legislature on the basis of overstated risk of patient re-identification. To do so is quite likely to lead to unintended consequences through the destruction of beneficial uses of de-identified data in health care and other settings.

III. ARGUMENT

A. The Petitioner Amici Briefs Ignore The Effectiveness Of HIPAA De-Identification Standards Already In Place

The patient privacy arguments advanced in the Petitioner *Amici* Briefs rely heavily on the notion that "[c]hanges in technology and in the overall information environment mean that traditional privacy safeguards such as de-identification are no longer reliable." (EFF Brief at 8). Amici disagree with that statement. It is contradicted by the plain language and applicable rigorous standards for de-identification that already exist under federal HIPAA privacy regulations. *See* 45 C.F.R. § 164.514.

HIPAA establishes a high standard that patient information is "de-identified" only when "there is no reasonable basis to believe that the information can be used to identify an individual." 45 C.F.R. § 164.514 (a). There are two methods to comply with this high standard. The first is more common in the context of the pharmaceutical industry. It requires a formal determination by a qualified statistician who, applying statistical and scientific principles and methods for rendering information not individually identifiable, determines that the risk is *very small* that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information. The statistician must document the methods and results of the analysis that justify such determination ("Statistical Method"). In performing this analysis, the statistician must take into account available

technology and the overall information environment. 45 C.F.R. § 164.514(b)(1). Thus, even as technology and the overall information environment change over time, the statistician must account for such changes in the analysis, and reach a conclusion that the risks of re-identification are "very small" in order for the patient information to be properly "de-identified."

The second method is less common in the context of the industry in this case. It involves the removal of eighteen specified patient identifiers, including but not limited to, patient name, location (other than state or 3-digit ZIP codes with populations greater than 20,000), email address, telephone number, Social Security Number, and the like ("Safe Harbor Method"). 45 C.F.R. § 164.514(b)(2)(i). Significantly, the eighteenth identifier that must be removed is "any other unique identifying number, characteristic, or code." This is a broad definition that will change over time as technology and the overall information environment change. Moreover, even after removing all of the designated identifiers, the covered entity must *not* have actual knowledge that the remaining data could be used alone or in combination with other information to identify the individual. 45 C.F.R. §164.514(b)(2)(ii). HHS also is undertaking a review to confirm whether the specifics of the Safe Harbor Method should be updated to reflect any developments in the marketplace. *See* HHS.gov, http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/deidentificationagenda.html (*last visited* Mar. 28, 2011). For all of these reasons, the de-identification standards established by HIPAA properly address changes in technology and the overall information environment, and provide strong protections against re-identification.

Patient data that has been properly de-identified pursuant to HIPAA poses very small risks of re-identification. Recently, the HHS Office of the National Coordinator for Health Information Technology ("ONC") conducted an empirical test of the Safe Harbor Method of de-identification. Specifically, ONC posed the question: "Can a Safe Harbor de-identified data set be combined with readily available outside data to re-identify data set subjects?" ONC established a well-qualified team to conduct the project, including a Principal Investigator, a Project Director, a Project Manager, a Lead Statistician, a Statistician, a Research Scientist, and a Research Analyst. The team began with a set of approximately 15,000 patient records (all related to an ethnic minority group seen at an academic health center serving a multi-country region) that had been de-identified in accordance with the Safe Harbor Method. It then sought to match the de-identified records with identifiable records in a commercial data repository (notably, a repository that had been considered reliable enough to be used by the US Census to verify and cross-check its household data). The team also conducted manual searches through other external sources (e.g., InfoUSA) to determine whether any of the records would align with any of the "uniques" in the de-identified data set. After all of these activities, and after cross-checking the results of the text with the health center that originally supplied the Safe Harbor de-identified data, the team determined that it was able to accurately re-identify *only two of the fifteen thousand individuals, for a match rate of 0.013%*. Notably, the choice to focus this assessment on a particular ethnic minority was made specifically to "increase the likelihood of an 'easy' match by using subjects who had self-identified as part of the minority ethnic group." This means that

this estimate of re-identification risk would likely have been even lower if a general population sample had been used instead. ONC's observation following this study indicated that: "[s]ome researchers and others have stated that increased personal data availability, e.g. on the Internet, makes re-identification easy, but there has been little empirical evidence to support that claim." Deborah Lafky, *The Safe Harbor Method of De-Identification ONC Presentation*, October 9, 2009, www.ehcca.com/presentations/HIPAAWest4/lafky_2.pdf (*last visited* Mar. 28, 2011).

By the estimates reported in the August 23, 2007 public testimony of one privacy researcher, Dr. Latanya Sweeney, only 0.04% (4 in 10,000) of the individuals within data sets de-identified using the Safe Harbor Method might be potentially re-identifiable. *See* National Committee on Vital and Health Statistics Report to the Secretary of the U.S. Department of Health and Human Services, *Enhanced Protections for Uses of Health Data: A Stewardship Framework for 'Secondary Uses' of Electronically Collected and Transmitted Health Data*, December 19, 2007, *available at* http://www.ncvhs.hhs.gov/071221lt.pdf (*last visited* Mar. 28, 2011). This means that, relative to the use of individually identified data, the use of de-identified data under HIPAA has been shown to reduce re-identification risks by at least *2,500 fold* over the identification risks that would result from allowing direct access to health information including protected patient information.

HIPAA also carries strict penalties for non-compliance. 45 C.F.R. § 160.404; 42 U.S.C. §§ 1320d-5 and 6. Specifically, amendments to HIPAA under Section

13410 of HITECH provide for significant civil penalties of up to $1.5 million for all violations of a single requirement in a calendar year, and for criminal penalties ranging from fines of $50,000 to $250,000 and imprisonment for terms ranging from one (1) year to ten (10) years. *Id.* HHS has primary enforcement authority under HIPAA, and refers potential criminal cases to the US Department of Justice. *Id*. HITECH also confers various powers to enforce HIPAA on the State Attorneys General by empowering them to obtain damages on behalf of state residents or to enjoin HIPAA violations. *Id*. This penalty regime helps to ensure rigorous enforcement of the HIPAA de-identification standards and other requirements.

HITECH also expanded the reach of such penalties to apply not only to pharmacies and other covered entities but also to business associates (service providers and similar entities that handle individually identifiable patient data on behalf of covered entities). 45 C.F.R. §§ 164.308, 164.310, 164.312, 164.316; 42 U.S.C. §§ 1320d-5 and 6. In the context of this case, the pharmacies or other organizations that disclose Prescriber Data are generally HIPAA-covered entities that have direct duties to ensure that all patient data is de-identified *before* it can be used or disclosed for marketing purposes. 45 C.F.R. § 164.502. In addition, the Respondents in some situations act as business associates that perform the de-identification of patient data on behalf of the covered entities pursuant to specific, HIPAA-mandated privacy contracts, called business associate agreements. Such agreements contain required provisions regarding HIPAA privacy and security, and prohibit the business associates from engaging in any use or disclosure of HIPAA-protected data that goes beyond what the covered entity could do. 45 C.F.R. § 164.502(e)(2); 45 C.F.R. § 164.504(e). As such, even though HIPAA

does not apply to every entity that may ever access or use de-identified data, HIPAA establishes a rigorous and enforceable privacy legal framework to prevent the use and disclosure of HIPAA-protected data for marketing purposes unless such data is properly de-identified.

B. Petitioner *Amici* Briefs Cite Situations Involving Re-Identification That Are Irrelevant To This Case

The Petitioner *Amici* Briefs have provided many citations that they allege support their contentions that "[Prescriber Data] at issue in this case presents grave re-identification issues" (EFF Brief at 12) and that "Patient records are at risk of being re-identified" (EPIC Brief at 24). While the citations provided highlight general concerns that re-identification of supposedly de-identified data can be accomplished by highly trained experts, virtually all of the citations have *no relevance whatsoever* to the immediate situation, because they do not address how the use or disclosure of Prescriber Data specifically compromises patient re-identification protections. By way of example, the EPIC Brief describes how an adversary can determine the identity of a consumer if the adversary has information about "six precise ratings a person in the Netflix video-rental database has assigned to obscure movies." (EPIC Brief at 32-33). Similarly, the EPIC Brief describes how researchers "have been able to reidentify anonymized records in databases as varied as Social Security records, Internet search queries, and video rentals." The data detailed in these examples is irrelevant to the patient data at issue here. Therefore, even if such examples are accurate, they provide no support to the claim that patient data de-identified in accordance with HIPAA is at risk for re-identification.

The EPIC Brief and EFF Brief also describe in detail the re-identification risks posed by *genetic* information. (EPIC Brief at 29-31; EFF Brief at 12). Genetic information is highly specific data that poses significant risks from a re-identification standpoint. Nevertheless, although genetic information often is subject to HIPAA, it is certainly not the type of Prescriber Data that is subject to regulation under the Vermont Statute (because it would not be provided to the pharmacy or other covered entity that would disclose Prescriber Data and therefore would not be subject to further disclosure), nor otherwise be at issue in this case.

The Petitioner *Amici* Briefs also set forth various citations that involve inherently sensitive topics and issues, but again do not in any way explain how the use or disclosure of Prescriber Data compromises patient de-identification procedures. For example, the Petitioner *Amici* Briefs raise generalized privacy concerns in an imaginary situation where "a psychiatrist publishes verbatim counseling notes in a best-selling book, but in a way that the specific identity of the patient cannot be determined … [where] the patient protests at having her story chronicled in agonizing detail to the public…" (EPIC Brief at 18). They also cite an imaginary situation where a person is uploading "nude pictures of a woman . . . to the Internet without her consent though without identifying her by name." (EPIC Brief at 19; *see also* EFF Brief at 23). Again, there is no link between the data at issue in these imaginary scenarios and the Prescriber Data covered by the Vermont Statute.

The Petitioner *Amici* Briefs also cite various situations that involve activities that pre-date HIPAA. For example,

the EFF Brief cites an instance where a woman who received unsuccessful fertility treatments later received marketing materials for products and services that would have been appropriate if she had had the child. (EFF Brief at 8 – 9). The marketing in this example pre-dates the adoption of HIPAA in 2003, and such marketing would indeed now be prohibited now under HIPAA. *See* 45 C.F.R. § 164.508.

The Petitioner *Amici* Briefs also reference an unpublished working paper by Dr. Sweeney (EPIC Brief at 27, citing Sweeney, Patient Identifiability in Pharmaceutical Marketing Data, Cambridge, Data Privacy Working Paper No. 1015(2011)), which appears to describe analyses conducted in 2003 with prescription data. Sweeney reports estimates of re-identification risks for data in two of the nine states examined, New York (with 2.3% uniquely re-identifiable) and Arizona (with 1.24% uniquely re-identifiable), that are higher than the Safe Harbor Method re-identification risk estimate previously made by Dr. Sweeney of 0.04% (i.e., 4 possible re-identifications in 10,000, as cited above). In the other seven states at issue in the study (Texas, Pennsylvania, Illinois, California, Florida, Michigan, and Massachusetts), Dr. Sweeney's re-identification risks were found to be less than the Safe Harbor Method estimate.

It is important to look critically at the methods used in this study. Dr. Sweeney uses a complex, multi-stage re-identification attack which involves first matching the prescription data to other medical databases to acquire further quasi-identifiers (QIs) that did not exist in the data for the patients, and then matching this data with the additional QIs to a population register (like a

voter list). Dr. Sweeney also specifies that: "[i]n order to facilitate the linking, we used additional prescription data to construct models for inferring patient ZIP from pharmacy ZIP, service date from the date of fill and prescription information, and we used databases that relate medications to diseases." Executing such a complex attack requires an extremely advanced skill set. Use of databases to relate medications to diseases or diagnoses requires a sophisticated re-identification attacker with knowledge of how to find and properly employ such medical informatics databases. Even more critical, however, is the statement that "we used additional prescription data to construct models for inferring patient ZIP from pharmacy ZIP." Another scientist working on the same project, Michael Stoto, Ph.D., reported in a separate publication on this study that "the calculations assume that the ZIP code of the pharmacy where the prescription was filled is the same as the ZIP code of the residence of the individual," and then concludes in his discussion of the study that "we were not able to ascertain the validity or reliability of the matching methods used." Michael A. Stoto, J. Domingo-Ferrer, L. Franconi, eds., *The Identifiability of Pharmaceutical Data: A Test of the Statistical Alternative to HIPAA's Safe Harbor; CD-only annex Privacy in Statistical Databases, Lecture Notes in Computer Science 4302* (2006), *see* http://explore.georgetown.edu/publications/index.cfm?Action=View&DocumentID=25940 (*last visited* March 28, 2011).

In order for such a re-identification attack to reliably occur and have any reasonable certainty of truly re-identifying individuals, complex statistical modeling requiring highly advanced skills and training would be required in order to reliably infer the likely patient ZIP

Codes from the pharmacy ZIP Codes. Indeed, a supposed re-identification attacker would further need access to prescription data which contained the actual patient's ZIP codes in order to construct any such statistical models capable of quantifying the likelihood that the patient actually resides in any particular ZIP code based on the pharmacy ZIP code information. Importantly, patient 5-digit ZIP codes are regulated patient data under HIPAA, and would generally only be available to HIPAA-regulated entities (i.e., covered entities or business associates) in today's environment. Consequently, even if another sophisticated re-identification attacker existed and wished to replicate this attack now, it is highly unlikely in today's environment that she would have access to the necessary HIPAA-protected patient data required to design such an attack without being subject to the rest of HIPAA as either a covered entity or a business associate. It is also highly implausible that a person with ill-intent who had access to HIPAA-protected patient data as a covered entity or a business associate would bother with the considerable effort involved in mounting a re-identification attack with de-identified data when such individual could simply directly examine the HIPAA-protected patient data.

The Illinois Supreme Court has gone so far as to conclude that Dr. Sweeney's testimony concerning the risk of re-identification, offered in a bench trial arising under the Illinois Freedom of Information Act, 5 ILCS 140/1 et. seq. (West, 1998), failed to establish that a significant number of individuals in the general public could re-identify the data at issue in that case. *Southern Illinoisan v. Ill. Dep't of Pub. Health*, 218 Ill. 2d 390 (Ill. 2006). In *Southern Illinoisan*, an Illinois newspaper requested

the Illinois Department of Public Health to release from the Illinois Health and Hazardous Substances Cancer Registry (the "Cancer Registry") copies of documents relating to the incidence of neuroblastoma in Illinois during a specified period of time. The plaintiff newspaper requested release of the information in a format showing type of cancer, ZIP code, and date of diagnosis, but not including patients' names or addresses.

In opposition, the government agency relied initially on an affidavit, and ultimately, on sealed testimony in a bench trial, of Dr. Sweeney. The affidavit attested that she could:

> '[S]how how persons can be re-identified from the Illinois Cancer Registry when the combination of data elements that includes only type of cancer, date of diagnosis, and ZIP code is provided.' According to Dr. Sweeney, her experiment 'established that a significant number of individuals in the general public with access to a personal computer, using traditional database software, who purchase or acquire public data sets will be able to reidentify individuals in the Illinois Cancer Registry,' as this 'seemingly anonymous information can be re-identified by linking the information to databases that are made available to the public.'

Id. at 397.

In affirming the lower courts' orders requiring production, the Illinois Supreme Court noted that "although Dr. Sweeney stated that the equipment and data sets that she used during her experiment would be readily

available to the general public, the methodology she used during her experiment was unique to her education, training and experience, and not easily duplicated by the general public." *Id.* at 425. Therefore, the Illinois Supreme Court held that the Department of Public Health failed to demonstrate that the release of the Cancer Registry information in the form requested by the plaintiff would lead to the re-identification of the persons described in the data. *Id.* at 426-27. On this point, the Illinois Supreme Court quoted the Illinois Appellate Court in its earlier decision as to why the unrebutted testimony of Dr. Sweeney was insufficient:

> Had the defendants, with the many and varied resources available to a state agency, wished to present specific evidence on the extent to which other individuals possess the unique knowledge, experience, and analytical skills necessary to replicate Dr. Sweeney's work, they were free to do so. Had they chosen to bring in other witnesses who also had been able to identify the subjects from the data in question, they were free to do that as well. Had they done one or both of these things, this court would be in a better position to evaluate the threat of which the defendants complain. However, the defendants did not do so, and they now must stand by the evidence they actually presented, not by alarmist conjecture about the resounding policy implications of that somewhat limited evidence.

Id. at 426, *quoting and citing Southern Illinoisan v. Dept. of Pub. Health*, 349 Ill. App. 3d 431, 435 (Ill. App. Ct. 5th Dist. 2004). So too this Court should avoid making

sweeping policy judgments on the basis of the written submissions of the Petitioner *Amici* Briefs.

Beyond the significant weaknesses described above in the various citations provided in the Petitioner *Amici* Briefs, it is also worthwhile to note that there is no publicly known example of an actual re-identification attack that has taken place in the real world by a real adversary to defeat the HIPAA de-identification requirement (as opposed to in a research setting). *See* Douglas Sylvester & Sharon Lohr, *Counting on Confidentiality: Legal and Statistical Approaches to Federal Privacy Law After the USA Patriot Act*, 2005 Wis. L. Rev. 1033, 1112 (2005). This simply reinforces that the risks of re-identification raised by the Petitioner *Amici* Briefs are largely theoretical concerns.

Based on the foregoing, the Petitioner *Amici* Briefs have failed to demonstrate that Prescriber Data specifically creates re-identification risks. Additionally, if any such increased risks from Prescriber Data existed, it is a requirement of the HIPAA statistical de-identification requirements that such risk must be properly controlled in order for the data to qualify as de-identified.

C. The Vermont Statute Does Not Advance Any State Interest In Regulating De-Identified Patient Data

Even if Vermont were found to have some form of recognizable interest in regulating the de-identified patient data that is incidentally contained within Prescriber Data, the Vermont Statute does not advance such interest in any meaningful way. First, the Vermont Statute permits the use and sharing of Prescriber Data for a variety of

purposes, such as: pharmacy reimbursement; prescription drug formulary compliance; patient care management; utilization review by a health care professional, the patient's health insurer, or the agent of either; or health care research. 18 VT. STAT. ANN. § 4631(e). Simply put, this de-identified patient data is *already available* in the marketplace, and by restricting marketing to physicians, the Vermont Statute does not substantially change that availability. Second, even if there were some marginal difference due to the physician marketing restriction, such a difference would be entirely inapplicable to one class of patients (i.e. those whose prescriber authorized the marketing). For this entire class of patients, the statute has absolutely no protective effect on the patients' de-identified data.

D. The Vermont Statute Is Significantly More Restrictive Than Necessary To Serve Any State Interest In De-Identified Patient Data

Finally, the "poor fit" of the Vermont Statute to address any recognizable state interest in de-identified patient data is also illustrated by the situation for patients whose prescriber has not authorized the use of Prescriber Data for marketing. For this class of patients, there is an entire ban on the disclosure of de-identified patient data for purposes of marketing to physicians. This is significantly more restrictive than necessary to address any patient privacy interest in de-identified data. For example, Vermont could have required pharmacies and others to establish "data use" or "restricted access" agreements to prevent re-identification of the patient data. Vermont could also have simply prohibited re-identification of de-identified patient data. The outright ban on the use and

disclosure of de-identified patient data for this class of patients plainly fails to satisfy the requirement that the law must be a "good fit" to achieve the stated claim that the Vermont Statute protects patient privacy interests.

E. The Vermont Statute, If Upheld, Would Open The Door For States To Restrict Many Other Beneficial Uses Of De-Identified Patient Data

The use of de-identified patient data clearly results in powerful and meaningful improvements in protections for individuals' health privacy. The challenge facing public policy makers is to more clearly communicate to the public that, rather than posing newly appreciated and troubling privacy risks, the use of de-identified data routinely results in vast improvements in the privacy protections for individuals compared to the use of identified data. The full range of uses of de-identified patient data to improve our health system has not been well-understood by the general public or by some privacy advocates. Considerable investments have been made in the collection, management and use of de-identified administrative health information. In fact, such use of de-identified administrative claims data has supported a great majority of the health care systems improvements achieved in recent decades.

Specifically, the use of de-identified health data has helped to support an immeasurable number of vital health systems improvements such as quality improvement, health systems planning, health costs monitoring, healthcare fraud detection, and waste and abuse detection, all of which have been increasingly conducted using de-identified data. Additionally, it should be noted that as the promise of Electronic Health Records, including

both Electronic Medical Records and Personal Health Records advances, and a wealth of richer, more detailed, de-identified clinical data grows, the progress of the nation's numerous envisioned goals for health care reform (such as adverse drug event monitoring, patient safety improvements and reducing health disparities) will all be built on a foundation of such de-identified patient data. Taken together, the provisions within HIPAA and HITECH greatly expand the already existing incentives for using statistically de-identified data and provide strong protections for both personal privacy and for assuring the continued availability of de-identified patient data as a public good supporting a broad and increasing range of improvements to the quality and efficiency of the U.S. healthcare system.

CONCLUSION

For all the foregoing reasons, Amici Curiae, Dr. Khaled El Emam and Jane Yakowitz, Esq., respectfully suggest that the decision of the Court of Appeals for the Second Circuit below be affirmed.

Respectfully submitted,

Michael A. Pollard
Counsel of Record
Brian L. Hengesbaugh
Amy de La Lama
Lindsay A. Philiben
Baker & McKenzie LLP
130 E. Randolph Drive
Chicago, IL 60601
(312) 861-2786
michael.pollard@
bakermckenzie.com

Counsel for Amici Curiae
Dr. Khaled El Emam
Jane Yakowitz, Esq.