



August 19, 2013

Lek Kadeli  
Principal Deputy Assistant Administrator  
Office of Research and Development  
Environmental Protection Agency  
Washington, DC 20460  
VIA E-MAIL

Dear Mr. Kadeli:

Thank you for your letter of July 8, inquiring about the permissibility of sharing research data used in certain epidemiological studies focusing on the health effects of particulate matter and ozone pollution. The following is the American Cancer Society's (the Society's) response to your questions.

For 100 years, the Society has worked tirelessly to save lives and create a world without cancer. Along with millions of supporters—over one million of whom volunteered to participate in our research studies—we have committed ourselves to eliminate cancer as a major public health problem. We have been able to lead the way in cancer research by building a foundation of trust with the public and by always placing the public good at the forefront of our mission.

Your inquiry appears to focus on Cancer Prevention Study II (CPS-II) data that were used in four of the studies listed in your letter: Krewski et al (2000),<sup>1</sup> Pope et al (2002),<sup>2</sup> Jerrett et al (2009),<sup>3</sup> and Krewski et al (2009).<sup>4</sup> CPS-II data were not used in the other studies you identified.

**What Is CPS-II and  
Why Are the Data So Valuable?**

The Society established CPS-II in 1982. Over the last 31 years, through the recruitment of nearly 1.2 million male and female participants by approximately 77,000 volunteers in 50 states, the District of Columbia and Puerto Rico, the Society has amassed this data set as a powerful tool to identify the risk factors for cancer and, ultimately, learn how to prevent it. CPS-II data contain comprehensive demographic information as well as health, personal habit history, and economic information. Mortality follow-up of the entire CPS-II cohort continues today with biennial linkage to the National Death Index. The Society has also followed up with subgroups of the larger cohort in a variety of ways, including through repeat questionnaires for assessing cancer incidence and other information and the collection of blood samples and buccal cells for genetic analysis. In addition, Society epidemiologists recently began the retrospective and

---

stay well | get well | find cures | fight back | cancer.org | 1.800.227.2345

**National Home Office**  
250 Williams Street, Atlanta, GA 30303-1002  
404.329.7740 f) 404.329.7530

prospective collection of breast, colorectal, hematopoietic and prostate cancer tumor specimens. In short, the CPS-II data set is one the most comprehensive longitudinal data sets in existence.

CPS-II data and corresponding follow-up studies using the data have played a major role in cancer prevention both nationally and internationally over the past several decades. More than 500 scientific articles have been published and the findings have significantly contributed to our understanding of the health effects of tobacco use, obesity, diet, physical activity, hormone use, and various other exposures in relation to cancer and other diseases.

The value to science and the public of the CPS-II data is incalculable. It is a very large snapshot of human information as it existed and evolved over a period of time, and it continues to be extremely relevant to scientific inquiry. It is a medical treasure built with the commitment of our donors, volunteers, staff, and, most importantly, CPS-II participants.

### **Responses to EPA's Specific Questions**

- 1. Who owns and/or holds the data necessary to replicate the relevant studies and what are the concerns, if any, associated with making such data publicly available?**

#### **A. Control of data**

The Society owns, holds and is entrusted with the stewardship of the individual-level CPS-II data. The Society funded and oversaw the collection of the data, and now directs and controls their dissemination. We obtained some of the mortality data in the CPS-II data set from the Centers for Disease Control and Prevention, which manages the nation's National Death Index (NDI). As we explain below, the Society's use and subsequent disclosure of NDI data is limited to those uses and disclosures permitted under NDI's implementing regulations.

The CPS-II data have since been linked, using participant zip codes or other location information, to ecological information about the area in which the subjects lived (the "Linked Analyses"). These Linked Analyses are conducted by Dr. Daniel Krewski at the R. Samuel McLaughlin Centre for Population Health Risk Assessment at the University of Ottawa, under an agreement with the Society to ensure that he and the University handle our individual level data from CPS-II responsibly and ethically.

#### **B. Concerns associated with publicizing data**

The Society has a number of serious legal, ethical, and policy concerns regarding disclosure of both the individual level CPS-II data and the Linked Analyses. At the core of our concern is the Society's ethical obligation as steward of personal and highly confidential information. Accordingly, we follow prevailing privacy norms with respect to the data, and we made assurances to participants, the NIH, and the NDI. To provide identifiable data to Congress under these circumstances would violate these legal obligations and commitments. Moreover, the Society's decades-long investment of resources made the collection of CPS-II data possible, and today the data are priceless.

i. *The Society's duty to maintain confidentiality*

a) Certificate of Confidentiality and the National Death Index

The CPS-II data are protected by a Certificate of Confidentiality issued by the NIH to the Society. Under section 301(d) of the Public Health Service Act (42 U.S.C. 241(d)) the Secretary of Health and Human Services may authorize persons engaged in biomedical, behavioral, clinical, or other research to protect the privacy of individuals who are the subjects of that research. This authority has been delegated to the NIH. 42 U.S.C. 241(d). The statute prohibits involuntary disclosure of protected research data:

Persons authorized by the NIH to protect the privacy of research subjects may not be compelled in any Federal, State, or local civil, criminal, administrative, legislative, or other proceedings to identify them by name or other identifying characteristic. 42 U.S.C. 241(d)

If the Society were forced to provide CPS-II data to Congress in direct violation of this statute, the Society would not only breach its Certificate of Confidentiality, but the entire concept of the Certificate and the protection it provides could be in doubt.

Moreover, under these circumstances the Society could not release the information it has received about CPS-II participants' cause of death from the National Death Index, a necessary component of the data to reanalyze the studies in question. The NDI regulations include protections against releasing identifiable information. As we describe in response to Question #2, we are not aware of any way to create a de-identified version of the CPS-II data set sufficient to protect the confidentiality of the participants while at the same time allowing a true replica of the studies.

b) Privacy Policies

The Society is sensitive to and understands the important role of Congress in oversight of environmental policy, but we are concerned that the House of Representatives Committee on Science, Space and Technology's authorization to issue a subpoena for our CPS-II data may put the Society in a position that is inconsistent with prevailing privacy and security standards. Since at least the mid-20<sup>th</sup> century, confidentiality has been a central tenet of ethical protections for research participants. Individuals share confidential information about themselves to make biomedical and public health research possible and, in exchange, researchers and the public at large assure these volunteers that their confidential data will only be used and disclosed in certain, limited ways. In recent years, these privacy and security protections have become enshrined in various forms, for example in the Health Insurance and Portability and Accountability Act and its implementing regulations, confidentiality protections set forth in the National Death Index regulations, state law, and "privacy by design" principles set forth by the Federal Trade Commission. Although these privacy and security frameworks differ in some respects, core commonalities persist, suggesting a converging set of expectations pertaining to privacy and security.

For example, prevailing privacy norms recognize the need for individuals to be informed about possible permissible uses and disclosures of their data. A closer look at HIPAA is instructive as to legal and public expectations as to privacy. The central tenet of HIPAA is that all uses and disclosures of identifiable data are prohibited, unless they are expressly permitted. Permitted disclosures include those made pursuant to carefully worded authorizations, to *bona fide* researchers under certain, controlled and monitored circumstances, and for public health purposes to health care oversight agencies. HIPAA does not contain any exception to these principles for general congressional curiosity.

Although the Society itself is not directly regulated by HIPAA, most research institutions, such as hospitals and academic medical centers, must comply. The Society is committed to extending the same privacy protections to its research participants as the law would empower institutional providers to extend to their research participants. CPS-II participants deserve no less.

c) Protocols for maintaining confidentiality

Every voluntary participant was assured that their identity and the information they provided, often of a very personal nature, would be kept confidential and used only in connection with research. Volunteers who participated in CPS-II were motivated by a desire to help the fight against cancer and were assured that their commitment and generosity of time and candor would be protected. The confidentiality protections that the Society has in place are vital to the success of research participant recruitment efforts. To balance our promise to the CPS-II participants with our commitment to scientific inquiry, we have a rigorous process to allow outside investigators to request access to CPS-II data subject to confidentiality protections, as explained in our answer to Question #3 below.

ii. *Negative effect on future research*

Violating our legal obligations and breaking the promises we made to participants could damage not only the Society's reputation, but also the next phases of our scientific and public health work. For example, we are currently recruiting participants for our third cancer prevention study ("CPS-3"), and we are concerned that even the threat that Congress might appropriate and possibly make participants' information publicly available could negatively impact our recruitment efforts. More importantly, if research participants believe that confidentiality protections might be limited in circumstances such as these, individuals' willingness to participate in research in all areas may be eroded.

The rationale for the Federal government's acquisition of the CPS-II data appears to be that these underlying data were used in studies that the EPA cited to justify regulatory action. But this sets a dangerous precedent for scientific research: organizations will have reason to fear that any research data cited in connection with a government rulemaking might be subject to confiscation and distribution to the public. This kind of precedent could create a disincentive to researchers to share data, especially if there is a connection to a government rulemaking. Moreover, research entities might limit their own work, choosing to conduct only research that would not be used for government rulemaking to ensure their underlying data are protected. The result could be a breakdown in the collaborative process between scientists necessary to scientific advancement and an impediment to scientific inquiry, particularly in areas of interest for the government. In

addition, this introduces a fundamental disparity in the ethical protections and safeguards for participants in research depending on whether the research is used to inform government policy. What a tragic and ironic disincentive it would be to inform the public that when they give of themselves to support research identified as being of national importance, they must sacrifice basic confidentiality protections.

iii. *Congress cannot properly order EPA to 'take' this data*

The Society's individual level CPS-II data at issue here were funded and collected by the American Cancer Society, and, to the best of our knowledge, without the use of Federal funds. As it is a longitudinal, nationwide study dating from 1982, it is unique and not replicable, and its value cannot be measured. If we were forced by a Committee of the U.S. Congress or by any agency of the executive branch of the federal government to make public this privately created and privately funded resource, it could be akin to taking our property without just compensation in violation of the Fifth Amendment.

iv. *Uncertainty about dissemination caused by Congress is a concern*

Our concerns about confidentiality, the adverse effect on research, and the acquisition of our private property are compounded by statements made about how Congress might disseminate our participants' information. It is our understanding that the House of Representatives Committee on Science, Space and Technology has authorized the Committee Chairman to acquire the CPS-II data by subpoena, if necessary, with the intention of making the data set available "*on the Internet*," as the Chairman stated in an August 1, 2013 public hearing on the subject. The idea that Congress would publish our participants' information online only magnifies our concerns.

- 2. What are the technical options for making these data publicly available, taking into account any concerns about the release of confidential personal health information or other confidential data? What are the implications of these options for replicating these studies? What level of effort in terms of time and resources would be required for these options?**

In order to accurately replicate the studies, Congress will need data and statistical programs that the Society does not hold or control in addition to the raw data in CPS-II. First, Congress will need access to the National Death Index to link the CPS-II data to death records, and to do that, Congress would need the Society to provide participants' name, social security number, date of birth, and state of residence. Then, Congress or others would have to link the appropriate ecological variables to our CPS-II data. Otherwise, Congress will need access to the Linked Analyses, which are maintained by Dr. Daniel Krewski at the University of Ottawa, under an agreement with the Society.

With respect to the Linked Analyses, we do not currently have the internal expertise to determine definitively whether it is possible to code or otherwise modify them in such a way as to protect the confidentiality of our CPS-II participants and also allow for true replication of the studies. To determine what might be possible, we would have to engage outside experts, at considerable expense. This is likely to be a time-consuming and long-term effort with uncertain resolution.

Regarding the CPS-II data, it appears impossible to create a public version that would protect the confidentiality of the CPS-II participants while at the same time allowing a true replica of the studies. To enable study replication, we would have to include individual level information, including participants' location, such as zip code or partial zip code, to enable others to link ecological information. The zip code or partial zip, along with updated zip codes for a portion of the participants, would be listed with a wide variety of personal information, including age, race, gender, education, marital status, height, weight, alcohol consumption, smoking history, exposure to environmental tobacco smoke, occupational history and exposures, and, if applicable, cause of death and death date. Using HIPAA as our guide, we note that zip code *alone* is, in some cases, considered an identifier. Accordingly, the residual zip code information, which is necessary to facilitate the linking with ecological data, combined with other information about each participant, such as race, ethnicity and other data points, would heighten the risk of re-identification. In fact, in light of explosion of publicly available data that can be used to re-identify individuals with data otherwise appearing to be de-identified, regulators continue to expand the single data fields that are classified as "identifiers." While the Society might be able, with sufficient time and resources, to remove all of the confidential or identifying information so that individual CPS-II participants could not be identified, such a data set would be so limited and generic that it would not enable a researcher to replicate the studies in question.

- 3. If there are no feasible options for making all of the data publicly available, how would a researcher gain access to the full set of underlying data in order to replicate these studies? Please provide any documentation you believe would be helpful in understanding this process.**

The Society recognizes the value of externally-proposed studies that are of general interest and high scientific merit. We welcome outside investigators to request access to our data following our application process, the details of which are available on our website.<sup>5</sup> We only grant access to well-qualified researchers who have demonstrated that their proposed research is well-designed and has the potential to significantly contribute to scientific discourse, and who have the requisite knowledge, qualifications, and experience to conduct the analysis and protect our data.

Once a proposal is accepted, we take various measures to protect our data. Each researcher who is granted access to the data has restrictions on the use and publication of the data and must conduct the research consistent with applicable legal and ethical requirements. Further, a deep understanding of the history of CPS-II and the complexity of the database is needed to conduct scientifically valid research using CPS-II data. Therefore, we require external researchers to work collaboratively with Society investigators, including co-authorship on any resulting publications, and the researchers and their institutions must sign the Society's "Collaboration Agreement," which includes requirements designed to protect the confidentiality of the participants in the research. Moreover, we only give the investigator access to the data that are necessary to conduct the analysis.

The Society may choose to deny requests from individuals sponsored by interest groups who have demonstrated they are not interested in independent and objective scientific research. For example, we have on occasion refused to provide access to scientists who were publicly linked to

sponsorship by tobacco companies. These data are a public trust. We take that responsibility seriously.

We are currently engaged in more than 30 collaborations with outside investigators. With respect specifically to the CPS-II data used for the studies referenced in your letter, I am sure you are aware that the Krewski (2000) study was a replication of original studies precisely because some were concerned about the objectivity related to the results and conclusions of these original studies. As a result of those concerns, the Society shared the necessary data under a confidentiality agreement to ensure our data were properly protected and the reanalysis was done under the auspices of the Health Effects Institute and conducted by a neutral third party.

In summary, the Society has a number of concerns regarding the potential disclosure of our CPS-II data. To compile the CPS-II data set, we assured the 1.2 million individuals who provided personal information to help us understand what causes and prevents cancer that we would maintain the confidentiality of this information. We also applied for and were awarded a National Institutes of Health-issued Certificate of Confidentiality that protects the entire data set, from the date of its inception from disclosure. At the same time, we value the contributions that outside investigators can make using our CPS-II data, which is why we have a process to allow them to apply to access our data subject to confidentiality protocols. Producing CPS-II data to the Federal government outside of our standard process, when we can be given no assurances of how it will be used, by whom, and how widely it would be disseminated, would cause the Society to betray its own policies, the promises it made to participants, covenants with both the NIH and the National Death Index, and prevailing privacy norms. Moreover, the Society has invested countless resources to collect and analyze the CPS-II data, including three decades of work, tens of millions of dollars, and the dedication of 77,000 volunteers. Leaving aside the Society's critical concerns about confidentiality for the citizens who provided personal data, it would be improper for the Federal government to imply appropriate this privately created data set and make it publicly available.

The Society has engaged outside counsel to assist it in protecting the integrity of our CPS-II data. Please include Mr. Stephen M. Ryan of McDermott Will & Emery, LLP and the Society's General Counsel, Mr. Timothy B. Phillips, on all future correspondence. They are the only persons authorized to respond for the Society to any EPA need for further information.

Thank you for your careful consideration of the issues we have raised.

Sincerely,

A handwritten signature in blue ink that reads "Otis Brawley". The signature is written in a cursive style with a large, looped "y" at the end.

Otis Brawley, MD, FACP  
Chief Medical and Scientific Officer

<sup>1</sup> Krewski D, Burnett RT, Goldberg MS, Hoover K, Siemiatycki J, Jarret M, Abrahamowicz M, White WH. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. Special Report. Health Effects Institute, Cambridge MA, 2000.

<sup>2</sup> Pope CA III, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD. Lung cancer, cardiopulmonary mortality and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association* 2002;287:1132-1141.

<sup>3</sup> Jerrett M, Burnett RT, Pope CA III, Ito K, Thurston G, Krewski D, Shi YL, Calle E, Thun M. Long-term ozone exposure and mortality. *New England Journal of Medicine* 2009;360:1085-1095.

<sup>4</sup> Krewski D, Jerrett M, Burnett RT, Ma R, Hughes E, Shi Y, Turner MC, Pope CA III, Thurston G, Calle EE, Thun MJ. Extended follow-up and spatial analysis of the American Cancer Society Study linking particulate air pollution and mortality. HEI Research Report 140, Health Effects Institute, Boston MA. 2009.

<sup>5</sup> <http://www.cancer.org/acs/groups/content/@research/documents/document/aCPSc-039148.pdf>