



HARVARD UNIVERSITY
Office for Sponsored Programs

Catherine Breen
Director, Office for Sponsored Programs
catherine_breen@harvard.edu

Holyoke Center, Suite 600
1350 Massachusetts Avenue
Cambridge, MA 02138

t.617.495.9047
f.617.496.2524

March 8, 2012

Jill Young
Branch Chief
U.S. Environmental Protection Agency

RE: EPA Grant Number R-827353-01-0: Congressional Freedom of Information Act Request

Dear Ms. Young:

Thank you again for granting Harvard University ("Harvard") additional time to respond to your letter to me dated January 9, 2012. Enclosed with this letter are the research data responsive to your request, provided in accordance with 2 CFR 215.36(d)(1). The research data are described in more detail below.

As your letter made clear, you requested research data related to a study conducted by Professor Francine Laden and other Harvard researchers, which received funding from the EPA via the grant referenced above, and which was described in an article entitled: "Reduction in Fine Particulate Air Pollution and Mortality," published in Volume 173 of the American Journal of Respiratory and Critical Care on January 19, 2006 (the "Article"). According to your letter, when promulgating the Cross State Air Pollution Rule (CSAPR), an agency action that has the force and effect of law within the meaning of 2 CFR 215.36(d), the U.S. Environmental Protection Agency (EPA) used and cited this study.

There are two types of research data that were produced with EPA Grant Number R-827353-01-0 and used in the analysis described in the Article: a set of "Air Pollution" data, and a set of "Vital Statistics" data. These data are combined in a SAS file labeled "Harvard_6cities_030912.sas7bdat" and are described in the word file "6CITIES data dictionary".

- (1) The "Air Pollution" data consist of a number of fields all relating to fine particulate air pollution levels. These fields consist either of air quality monitoring data that had been

collected prior to 2000 as part of the Harvard Six Cities study (“Six Cities Data”) or of air quality data derived from the EPA’s own publicly available air quality monitoring data, which can be found on the EPA’s website (currently at <http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqdata.htm>) (“EPA Data”) in combination with the Six Cities Data.¹ The Six Cities Data is not subject to mandatory disclosure, because the “Shelby Amendment,” which was published in the Federal Register on March 16, 2000 as a Final Rule, and later codified in 2 CFR 215.36(d), “is applicable only to data first produced under new or competing continuing grants awarded after April 17, 2000, the regulation’s effective date.” *Salt Inst. v. Thompson*, 345 F.Supp.2d 589, 595, *aff’d sub nom Salt Inst. v. Leavitt*, 440 F.3d 1565 (4th Cir. 2006). Nonetheless, Harvard is voluntarily providing the “Air Pollution” data to you in its entirety, including without limitation the fields consisting of the Six Cities Data and information derived in part from the Six Cities Data.

- (2) The “Vital Statistics” data consist of a number of fields that provide aggregated information about deaths in the Six Cities between 1989 and 1998. These data were obtained by the researchers from the National Death Index (NDI), which as you know is part of the National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). As stated on its website, the NDI is a central computerized index of death record information that was established as a resource for epidemiologists and other health and medical investigators. The NDI is available to investigators “solely for statistical purposes in medical and health research.” http://www.cdc.gov/nchs/data_access/ndi/about_ndi.htm

To obtain data from the NDI, the Harvard researchers signed a confidentiality agreement, promising not to publish or release data in any form to any party if a particular individual (or establishment) is identifiable.² In their aggregated form, we believe that the data we are providing cannot be used to identify individuals. However, we note that, as stated in the National Center for Health Statistics Data Use Agreement, a copy of which is enclosed as Attachment A, “it may be possible in rare instances, through complex analysis and with outside information to ascertain from the data set the identity of particular persons or establishments,” and that “[c]onsiderable harm could ensue if this were done.” Thus, we leave it to you to consider whether it might be appropriate for any ultimate recipients to sign the NCHS Data Use Agreement before receiving this data set.

¹The authors’ derivation methods are described in detail both in the Article and in its online supplement.

²Indeed, the confidentiality agreement that is part of the current NDI application form states explicitly: “In accordance with Section 308(d) of the Public Health Service Act, such identifiable data will specifically not be provided in response to a direct order from an official of any government agency, the Administration or Congress, nor in response to an order from a court of justice.” http://www.cdc.gov/nchs/data/ndi/NDI_Application.pdf

Ms. Jill Young
March 8, 2012
Page 3

I thank you again for your courtesy in allowing us extra time to respond to your request. If you have any questions or comments about the submission, please do not hesitate to contact me

Sincerely,

A handwritten signature in black ink, appearing to be 'CB', with a long, sweeping horizontal line extending to the right.

Catherine Breen

Attachment A
National Center for Health Statistics
Data Use Agreement (DUA) for Vital Statistics Data Files

Requesting Organization:

Requested Data File(s):

Proposed Use(s):

Vital statistics data are provided to NCHS by vital statistics jurisdictions with the understanding that the data are protected under the provisions of the Public Health Services Act (42 U.S.C. 242m(d)), and that any file released under a data use agreement requires both NAPHSIS and NCHS review and approval of proposed use.

The Public Health Service Act (42 U.S.C. 242m(d)) provides that the data collected by the National Center for Health Statistics (NCHS) may be used only for the purpose for which they were obtained; any effort to determine the identity of any reported cases, or to use the information for any purpose other than for health statistical reporting and analysis, would violate this statutory restriction and the conditions of this data use agreement. NCHS does all it can to assure that the identity of data subjects cannot be disclosed; all direct identifiers, as well as characteristics that might lead to identification, are omitted from the data set. Nevertheless it may be possible in rare instances, through complex analysis and with outside information to ascertain from the data set the identity of particular persons or establishments. Considerable harm could ensue if this were done.

Therefore, the undersigned gives the following assurances with respect to all NCHS data sets:

- I will not use nor permit others to use the data in these data sets in any way except for statistical reporting and analysis and for the purposes described in the data request.
- I will not release nor permit others to release the data sets or any part of them to any person who is not a member of this organization, except with the approval of NCHS. Under section 308(d) of the Public Health Service Act, the only persons to be allowed access to these data sets will be staff members of this organization or its contractor(s) who have been authorized to work with the data and have, prior to being granted access to the data, read and signed this DUA Statement in the space provided below.
- I will not attempt to link nor permit others to attempt to link the data set with individually

identifiable records from any other NCHS or non-NCHS data set;

Data Use Agreement (DUA) for Vital Statistics Data Files—Cont.

- I will not attempt to use the data sets nor permit others to use them to learn the identity of any person or establishment included in any data set;
- If I should inadvertently discover the identity of any person or establishment, then (a) I will make no use of this knowledge, (b) I will immediately advise the Director of the Division of Vital Statistics of the incident, (c) I will safeguard or destroy the information that would identify an individual or establishment, as requested by NCHS, and (d) I will inform no one else of the discovered identity.

In addition, I will make every effort to release all statistical information in such a way as to avoid inadvertent disclosure. For example:

- **No figure, including totals, should be less than 10 in tabulations for sub-national geographic areas, regardless of number of years combined.**
- No data on an identifiable case should be derivable through subtraction or other calculation from the combination of tables in a given publication.
- No data should permit disclosure when used in combination with other known data.

I will secure identical written assurances from every individual within this organization who will have access to this data set.

My signature below indicates my agreement to comply with the above-stated statutorily-based requirements with the knowledge that deliberately making a false statement in any matter within the jurisdiction of any department or agency of the Federal Government violates 18 USC 1001 and is punishable by a fine of up to \$10,000 or up to 5 years in prison.

Further conditions for data use:

NAPHSIS and NCHS have reviewed and approved the use of the data provided under this agreement for purposes described in the requestor's application for one year from the date of receipt of the data. The data files listed under "Requested Data Files" above are the property of the National Center for Health Statistics (NCHS), Division of Vital Statistics (DVS). Permission is granted to use these data files for one year from the date of receipt. At the expiration of the one year period, the data files and any copies of the data files must be destroyed. Users must notify DVS in writing that the file(s) have been destroyed. This policy will be strictly enforced; however, extension of this usage period will be given consideration under appropriate circumstances, when requested in writing. Requests for extensions must be sent to the Director, Division of Vital Statistics, 3311, Room 7311 Toledo Road, Hyattsville, MD, 20782.

Citation of NCHS: Users of these data are asked to acknowledge NCHS and the vital statistics jurisdictions as the data source in published reports and studies for which the files were used. NCHS and the vital statistics jurisdictions should also be cited in reports, articles, and news releases in electronic and print media describing the studies or results of the studies. The following is the recommended citation:

National Center for Health Statistics. [*Name of data file(s)*] (*[year(s)]*), as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program.

Return completed and signed form:

Joyce Arbertha
Division of Vital Statistics
National Center for Health Statistics
3311 Toledo Road - Room 7324
Hyattsville, MD 20782

Print Name of Official authorized to execute agreements*, or Faculty Advisor (if applicable):

Title: _____

Organization: _____

Signature: _____ Date: _____

*The type of "official authorized to execute agreements" will, of course, vary among organizations. Whenever possible, this official should be at a higher level of authority than the principal investigator or other person responsible for the study or project; for example, the dean of a college, a vice president of a company, the director of a government division or bureau.

Under section 308(d) of the Public Health Service Act, the only persons to be allowed access to these data sets will be staff members of the above organization or its contractor(s) who have been authorized to work with the data and have, prior to being granted access to the data, read and agreed to the data-use requirements specified above in this DUA. By their signatures below, the following individuals affirm that they have read the DUA and will abide by the specified DUA data-use requirements:

Print: _____ Signature _____ Date _____

Print: _____ Signature _____ Date _____

Print: _____ Signature _____ Date _____

Print: _____ Signature _____ Date _____

Print: _____ Signature _____ Date _____

Print: _____ Signature _____ Date _____

Data dictionary

Variable	Description
Vital Statistics Data	
myearon2	Years of follow-up – from initiation date through death or end of follow-up (December 31, 1998)
period	1 if ended follow-up before 1/1/1990; 2 if ended follow-up later
dead	Identified as dead during the entire course of follow-up (1/0)
ndead	Identified as dead from the NDI search performed in the year 2000 (1/0)
cvddead	Dead from cardiovascular disease (ICD9 400-440) (1/0)
lungca	Dead from lung cancer (ICD9 162) (1/0)
respdead	Dead from nonmalignant respiratory disease (ICD9 485-496) (1/0)
othdead	Dead from other causes (1/0)
Air Pollution Data (Combination of 6CITIES monitoring data and AQS data)	
cityc	WAT=Watertown; KH=Kingston/Harrimon; STL=St Louis; STU=Steubenville; POR=Portage; TOP=Topeka
pm2_5n	PM2.5 value used in the NEJM paper (6CITIES monitoring data)
aire1*	PM2.5 value used for study year 1
aire2	PM2.5 value used for study year 2
aire3	PM2.5 value used for study year 3
aire4	PM2.5 value used for study year 4
aire5	PM2.5 value used for study year 5
aire6	PM2.5 value used for study year 6
aire7	PM2.5 value used for study year 7
aire8	PM2.5 value used for study year 8
aire9	PM2.5 value used for study year 9
aire10	PM2.5 value used for study year 10
aire11	PM2.5 value used for study year 11
aire12	PM2.5 value used for study year 12
aire13	PM2.5 value used for study year 13
aire14	PM2.5 value used for study year 14
aire15	PM2.5 value used for study year 15
aire16	PM2.5 value used for study year 16
aire17	PM2.5 value used for study year 17
aire18	PM2.5 value used for study year 18
aire19	PM2.5 value used for study year 19
aire20	PM2.5 value used for study year 20
aire21	PM2.5 value used for study year 21
aire22	PM2.5 value used for study year 22
aire23	PM2.5 value used for study year 23
aire24	PM2.5 value used for study year 24
aire25	PM2.5 value used for study year 25

avg_allf	Average PM2.5 using all years of follow-up
avg_p1	Average PM2.5 1980-1985 (6CITIES monitoring data)
avg_p2	Average PM2.5 1990-1998
avg_pd	Difference avg_p2 minus avg_p1

*the aire series is derived from both the 6CITIES monitoring data and the AQS data