

**Commonwealth of Massachusetts
Center for Health Information & Analysis (CHIA)
Non-Government APCD Request for Data**

This form is to be used by all applicants, except Government Agencies, as defined in 957 CMR 5.02.

NOTE: *In order for your application to be processed, you must submit the required application fee. Please consult the fee schedules for APCD data for the appropriate fee amount. A remittance form with instructions for submitting the application fee is available on the CHIA website.*

I. GENERAL INFORMATION

APPLICANT INFORMATION	
Applicant Name:	Dr. Helen Suh
Title:	Professor
Organization:	Northeastern University
Project Title:	Assessing Susceptibility to Air Pollution Using Hybrid Data Mining and Epidemiological Techniques
Mailing Address:	316 Robinson Hall Department of Health Sciences Northeastern University Boston, MA 02115
Telephone Number:	617-373-5925
Email Address:	h.suh@neu.edu
Names of Co-Investigators:	
Email Addresses of Co-Investigators:	
Original Data Request Submission Date:	
Dates Data Request Revised:	
Project Objectives (240 character limit)	The purpose of this project is to evaluate the impact of air pollution on clinical morbidity indicators using novel techniques that combine data mining and epidemiological methods and to identify factors that affect susceptibility to these impacts. We will do so by linking CHIA data to a variety of air pollution measures, including fine particulate air pollutants (PM2.5), ozone, PM2.5-associated metals and black carbon, and PM2.5 sources.
Project Research Questions (if applicable)	<ol style="list-style-type: none"> 1. What is the association between air pollution and clinical morbidity indicators, including cause-specific hospital admissions, emergency department visits, and out-patient visits? 2. How do these impacts vary by air pollutant and health outcome? 3. What population groups are most susceptible to air pollution’s harmful effects? How does susceptibility vary with age? 4. Does roadway density, green space, and socioeconomic conditions of neighborhoods modify air pollution-health

	impacts?
--	----------

II. PROJECT SUMMARY

Briefly describe the purpose of your project and how you will use the requested CHIA data to accomplish your purpose.

This project will help to predict air pollution-mediated health risks for Massachusetts residents and examine how these risks change as people age and how they may vary with neighborhood of residence. To do so, we will create a new database that links CHIA data with neighborhood data and with daily, monthly and yearly air pollution concentrations measured at multiple sites within Massachusetts and estimated using GIS-based spatio-temporal models. We will analyze these linked data using traditional epidemiological approaches and data mining approaches.

The ultimate goals of our analyses will be to examine the relationship of air pollution (fine particles (PM2.5), ozone, PM2.5-associated metals and black carbon, and PM2.5 sources) and cause-specific hospital admissions, emergency department visits, and outpatient visits. In so doing, we will identify populations at greatest risk, as well as critical time periods and/or conditions when health risks may change.

III. FILES REQUESTED

Please indicate the databases from which you seek data, and the year(s) of data requested.

ALL PAYER CLAIMS DATABASE	Single or Multiple Use	Year(s) Of Data Requested Current Yrs. Available 2009 – 2013
<input checked="" type="checkbox"/> Medical Claims	<input type="checkbox"/> Single Use <input checked="" type="checkbox"/> Multiple Use	<input checked="" type="checkbox"/> 2009 <input checked="" type="checkbox"/> 2010 <input checked="" type="checkbox"/> 2011 <input checked="" type="checkbox"/> 2012 <input checked="" type="checkbox"/> 2013
<input checked="" type="checkbox"/> Pharmacy Claims	<input type="checkbox"/> Single Use <input checked="" type="checkbox"/> Multiple Use	<input checked="" type="checkbox"/> 2009 <input checked="" type="checkbox"/> 2010 <input checked="" type="checkbox"/> 2011 <input checked="" type="checkbox"/> 2012 <input checked="" type="checkbox"/> 2013
<input type="checkbox"/> Dental Claims <input checked="" type="checkbox"/> Member Eligibility <input type="checkbox"/> Provider <input type="checkbox"/> Product	<input type="checkbox"/> Single Use <input type="checkbox"/> Multiple Use <input type="checkbox"/> Single Use <input type="checkbox"/> Multiple Use <input type="checkbox"/> Single Use <input type="checkbox"/> Multiple Use <input type="checkbox"/> Single Use <input type="checkbox"/> Multiple Use	<input type="checkbox"/> 2009 <input type="checkbox"/> 2010 <input type="checkbox"/> 2011 <input type="checkbox"/> 2012 <input type="checkbox"/> 2013 <input checked="" type="checkbox"/> 2009 <input checked="" type="checkbox"/> 2010 <input checked="" type="checkbox"/> 2011 <input checked="" type="checkbox"/> 2012 <input checked="" type="checkbox"/> 2013 <input type="checkbox"/> 2009 <input type="checkbox"/> 2010 <input type="checkbox"/> 2011 <input type="checkbox"/> 2012 <input type="checkbox"/> 2013 <input type="checkbox"/> 2009 <input type="checkbox"/> 2010 <input type="checkbox"/> 2011 <input type="checkbox"/> 2012 <input type="checkbox"/> 2013

IV. REQUESTED DATA ELEMENTS [APCD Only]

State and federal privacy laws limit the use of individually identifiable data to the minimum amount of data needed to accomplish a specific project objective. Please use the [APCD Data Specification Workbook](#) to identify which data elements you would like to request and attach this document to your application.

V. FEE INFORMATION

Please consult the fee schedules for APCD data) and Case Mix data, available at http://chiamass.gov/regulations/#957_5, and select from the following options:

APCD Applicants Only

- Academic Researcher
- Others (Single Use)
- Others (Multiple Use)

Are you requesting a fee waiver?

- Yes
- No

If yes, please submit a letter stating the basis for your request. Please refer to the [fee schedule](#) for qualifications for receiving a fee waiver. If you are requesting a waiver based on the financial hardship provision, please provide documentation of your financial situation. Please note that non-profit status alone isn't sufficient to qualify for a fee waiver.

VI. MEDICAID DATA [APCD Only]

Please indicate here whether you are seeking Medicaid Data:

- Yes
- No

Federal law (42 USC 1396a(a)7) restricts the use of individually identifiable data of Medicaid recipients to uses that are directly connected with the administration of the Medicaid program. If you are requesting Medicaid data from Level 2 or above, please describe in detail why your use of the data meets this requirement. Applications requesting Medicaid data will be forwarded to MassHealth for a determination as to whether the proposed use of the data is directly connected to the administration of the Medicaid program. MassHealth may impose additional requirements on applicants for Medicaid data as necessary to ensure compliance with federal laws and regulations regarding Medicaid.

It is not directly related to the administration of the Medicaid program. Rather our work will help to investigate whether health risks associated with air pollution and neighborhood characteristics are higher or otherwise differ for Medicaid recipients as compared to individuals with other types of insurance. Further, by inclusion of Medicaid recipients, we will be able to conduct air pollution epidemiological investigations that include populations that are typically thought to be at greatest health risks.

VII. FILTERS

If you are requesting APCD elements from Level 2 or above, describe any filters you are requesting to use in order to limit your request to the minimum set of records necessary to complete your project. (For example, you may only need individuals whose age is less than 21, claims for hospital services only, or only claims from small group projects.)

APCD FILE	DATA ELEMENT(S) FOR WHICH FILTERS ARE REQUESTED	RANGE OF VALUES REQUESTED
Medical Claims		

Pharmacy Claims		
Dental Claims		
Membership Eligibility		
Provider		
Product		

IX. PURPOSE AND INTENDED USE

1. Please explain why completing your project is in the public interest.

Air pollutant exposures have consistently been linked to a variety of adverse health conditions, including numerous clinical morbidity indicators such as as increased hospital admissions and emergency department visits. Most of these findings have been demonstrated for elderly individuals, given the lack of available administrative health information for individuals under 64 years of age. Further, most of these findings have been examined for one health outcome at a time, making it difficult to examine air pollution-mediated health impacts comprehensively a ross multiple measures of health. As a result, remarkably little is known about how air pollution impacts health at younger ages, how these impacts change as people age, and whether these impacts are consistent across clinical health indicators. Studies that further our understanding of how air pollution impacts overall health are critical to our ability to identify modifiable exposures and to develop comprehensive policies that reduce health risks and foster well-being in Massachusetts residents. These new studies should be broad in temporal scope, with multiple measures of health and inclusive of entire cohorts, such as those living in Massachusetts, with zip code-specific and longitudinal measures of exposure, health and potentially modifying factors, such as socio-economic factors (SES). These features will allow us to estimate risks accurately, precisely, and comprehensively, to assess how risks change as people age, and to identify at-risk populations. Such research is essential to our ability to develop policies that reduce the risks of air pollution on health, particularly for populations that are particularly susceptible to these risks.

2. **Attach** a brief (1-2 pages) description of your research methodology. (This description will not be posted on the internet.)
3. Has your project received approval from your organization’s Institutional Review Board (IRB)? Please note that CHIA will not review your application until IRB documentation has been received (if applicable).
 - Yes, and a copy of the approval letter is attached to this application.
 - No, the IRB will review the project on _____.
 - No, this project is not subject to IRB review.
 - No, my organization does not have an IRB.

X. APPLICANT QUALIFICATIONS

1. Describe your qualifications to perform the research described or accomplish the intended use of CHIA data.

Professor Suh is a Professor in the Department of Health Sciences in the Bouve College of Health Sciences at Northeastern University and is the Director of the Population Health doctoral program at Bouve, adjunct faculty at the Harvard School of Public Health and a Senior Fellow at NORC at the University of Chicago. Dr. Suh earned a Sc.D. and an MS in Environmental Health from the Harvard School of Public Health, and a SB in Biology from the Massachusetts Institute of Technology.

An internationally recognized expert in air pollution health effects in the areas of environmental epidemiology, exposure assessment and air pollution, Dr. Suh has led multidisciplinary teams in environmental exposure assessment and epidemiology for over 20 years. Her research focuses on three general areas within air pollution health effects, including: 1) assessment of the impact of lifestyle and neighborhoods on air pollutant exposures and human health; 2) examination of multi-pollutant impacts on human health; 3) development of GIS-based spatio-temporal modeling tools for epidemiological research. Dr. Suh is the lead investigator in numerous research projects, including an NIH-funded study investigating the impacts of air pollution and lifestyle of cognitive and cardiac health and an EPRI-sponsored study examining the association of chronic air pollution exposures and mortality. As part of these studies, Dr. Suh develops and uses innovative analytic tools, such as GIS-based spatio-temporal models that are able to predict air pollution exposures and consider their diverse sources and properties. Results from Dr. Suh's studies have already and will continue to advance our understanding of air pollution health impacts and the development of appropriate policies and regulations to reduce their health risks.

Dr. Suh's research has been published in leading journals, including *Environmental Health Perspectives* (EHP), *Epidemiology*, and *Circulation*. Her papers have had significant impact, as evidenced by her h-index of 50 and her almost 8000 citations. Dr. Suh's expertise is well recognized. Dr. Suh is an Associate Editor of the *Journal of Exposure Science and Environmental Epidemiology* (JESEE), a leading journal that publishes exposure assessment and exposure-motivated environmental epidemiology research. In addition, Dr. Suh has been appointed by the US EPA Administrator to the Clean Air Scientific Advisory Committee (CASAC), a seven-member committee that provides independent advice to the EPA Administrator on the technical bases for EPA's national ambient air quality standards. She is also a member of the CASAC subcommittees for ozone, nitrogen oxides, and sulfur dioxide and previously served as a member of the CASAC Committee for Particulate Matter. She was recently a member of the National Academy of Sciences (NAS) Committee on Scientific Tools and Approaches for Sustainability, charged with an evaluation of scientific tools and approaches for incorporating sustainability concepts into assessments used to support EPA decision-making. Previously, Dr. Suh served on several Institute of Medicine and National Academy of Science committees and has been invited to speak at numerous workshops and conferences.

2. Attach résumés or curricula vitae of the applicant/principal investigator, key contributors, and of all individuals who will have access to the data. (These attachments will not be posted on the internet.)

XI. DATA LINKAGE AND FURTHER DATA ABSTRACTION

Note: Data linkage involves combining CHIA data with other databases to create one extensive database for analysis. Data linkage is typically used to link multiple events or characteristics that refer to a single person in CHIA data within one database.

1. Do you intend to link or merge CHIA Data to other datasets?
 - Yes
 - No linkage or merger with any other database will occur
2. If yes, will the CHIA Data be linked or merged to other individual patient level data (e.g. disease registries, death data), individual provider level data (e.g., American Medical Association Physician Masterfile), facility level (e.g., American Hospital Association data) or with aggregate data (e.g., Census data)? [check all that apply]

Individual Patient Level Data

What is the purpose of the linkage:

N/A

What databases are involved, who owns the data and which specific data elements will be used for linkage:

N/A

Individual Provider Level Data

What is the purpose of the linkage:

N/A

What databases are involved, who owns the data and which specific data elements will be used for linkage:

N/A

Individual Facility Level Data

What is the purpose of the linkage:

N/A

What databases are involved, who owns the data and which specific data elements will be used for linkage:

N/A

Aggregate Data

What is the purpose of the linkage:

We will link the CHIA cause-specific hospital admissions, emergency department visit, outpatient visit, and medication use data from 2009-2013 to daily air pollution, behavioral risk factor, and census data. We will use this linked data set to examine the impacts of air pollution on health and to assess whether behaviors and neighborhood characteristics modify these impacts.

What databases are involved, who owns the data and which specific data elements will be used for linkage:

All data linked to the CHIA data are publicly available, with air pollution data obtained from the US EPA’s Air Quality System (AQS), behavioral risk factor data from the Behavioral Risk Factor Surveillance System (BRFSS) Selected Metropolitan/Micropolitan Area Risk Trends (SMART), and census data from the US Census American Community Survey. In addition, we may supplement the air pollution data with that measured as part of the Harvard-EPA Particle Center.

3. If yes, for each proposed linkage above, please describe your method or selected algorithm (e.g., deterministic or probabilistic) for linking each dataset. If you intend to develop a unique algorithm, please describe how it will link each dataset .

We will link CHIA, air pollution, behavioral and census data by CHIA date of admission and zip code. Each CHIA record will be linked deterministically.

4. If yes, please identify the specific steps you will take to prevent the identification of individual patients in the linked dataset.

CHIA data will be provide by zipcode, with each admission, visit, and medication usage accompanied by information regarding, gender, race/ethnicity, date of admission or visit, and zipcode of residence. As a result, linked CHIA data will contain no patient identifiers.

Regardless, all CHIA data files will be stored and maintained in a secure environment, with access only to approved NEU employees and approved third parties. The complete inventory of CHIA data files will be stored on the large data server, with only designated users allowed to read or execute data. Groups will be created to limit access to specific data files as necessary. For this project, designated users will not be allowed to download or write data outside the project’s large data server and compute node.

Note that for this project, designated users will be allowed to link CHIA data files to other publicly available data files, such as those containing data on air pollution, behavioral and census variables. These linked data files will also be stored on the project’s large data server, with access again restricted to designated users and no downloading or writing of these data sets outside the projects large data server and compute node. Ability to link data sets will be restricted to a minimal number of approved parties.

Northeastern’s privacy and security program is compliant with federal government regulations. Our IT systems have completed the full federal C&A process for a number of government agencies.

Northeastern has established protocols that are multilayered to secure computer systems and the data they contain. At MGHPCC we have a hardware firewall for all access to all NEU networks (management, storage and compute) in the data center. In addition physical access is restricted to only approved NEU employees via a dual key card and pod key entry with security personnel that check credentials before entry and pod keys are provided. Northeastern cluster accounts are validated using the Northeastern Windows Active Directory for only those users that have been approved to access the cluster. These accounts terminate automatically if a user is no longer affiliated to Northeastern University. They cannot use cluster resources when this happens. For third party access, a Northeastern faculty member has to sponsor the account, which in this case is Dr. Helen Suh, the PI of the study. Once sponsored by Dr. Suh, the account will then be enabled in the Northeastern Windows Active Directory for cluster access. Again once the sponsorship period ends, the account will be disabled in the Northeastern Windows Active Directory and cluster access is no longer possible. Northeastern currently has other government projects that require similar compliance with privacy and security of data, including census data from

the Bureau of Census. The cluster can enforce "Secure" capability in accordance with [IAW FIPS 200 \(http://csrc.nist.gov/publications/fips/fips200/FIPS-200-final-march.pdf\)](http://csrc.nist.gov/publications/fips/fips200/FIPS-200-final-march.pdf) and [NIST 800-53 \(http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf\)](http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf) for example. To remotely connect to the compute environment a two factor authentication is used (user has to be enabled in Northeastern University Active Directory and user has to have clearance on login nodes for access).

Information Technology Services – Research Computing (ITS – RC) maintains applications, data and security standards on the cluster that support management of research projects (define permissions, users, projects at any point in time), software, user and group accounts, software and compute schedulers and related storage. Data between these systems may be exchanged on a real time basis.

The following are in place for security of non-anonymized or non-neutralized data that contain sensitive user or other information. These data are typically made available to specific researchers in fully disclosed format. The researcher can then decide what level of access users of the data can get.

- 1) Data in storage are in lock down with only designated users allowed to read, write or execute data. For this project, we will allow designated users to read and execute data, but not to download or write data outside the project's large data server and compute node.
- 2) Groups are created such that users in particular groups have structured and limited access to data as decided by the researcher who is the owner of the data.
- 3) When data are moved into the cluster encryption is used and when data are deleted secure deletion is employed.
- 4) When data are staged for run on the compute nodes the staging location is also in lock down.
- 5) Individual jobs can be restricted in their resource utilization, based on user/group/project membership.
- 6) Researchers can provide their own applications that are installed via restricted modules so that only designated persons can use them on the cluster.
- 7) All software is evaluated for stability and viruses before installing on the cluster.
- 8) The Federal Government and USCB have very strict access and security requirements. These security requirements are outlined in Census Bureau IT Security Program Policy, the DOC IT Security Program Policy and NIST SP 800-53r4I. The cluster adheres to these.
- 9) Specifically:
 - a) Users can only access data that they have authorized access to via group and user specific memberships.
 - b) Intermingling of data sets can be locked down. So any user cannot have access to or merge or query two separate data sets. These can be enforced by the researcher that owns the data sets easily on requesting what is required to ITS – RC.
 - c) Separate backups if needed can be enforced without comingling with other data.
 - d) Security tools are in place to monitor intrusions to the cluster like "ssh black lists", "tripwire", and "iptables".
 - e) The cluster sits behind a firewall. All access to and out of the cluster is monitored and information is stored historically.

Both ITS – Research Computing, and the Office of Information Security periodically review cluster security and make recommendations to ensure continued compliance.

5. If yes, and the data mentioned above is not in the public domain, please attach a letter of agreement or other appropriate documentation on restrictions of use from the data owner corroborating that they agree to have you initiate linkage of their data with CHIA data and include the data owner's website.

XII. PUBLICATION / DISSEMINATION / RE-RELEASE

1. Describe your plans to publish or otherwise disclose CHIA Data, or any data derived or extracted from such data, in any paper, report, website, statistical tabulation, seminar, conference, or other setting.

We will publish results from our analyses in peer-reviewed publications in relevant scientific journals and will present these results at scientific conferences.

2. Will the results of your analysis be publicly available to any interested party? Please describe how an interested party will obtain your analysis and, if applicable, the amount of the fee.

Results from our analyses will be made available through journals and scientific conferences. We will not charge any fees for analyses involving CHIA data.

3. Will you use the data for consulting purposes?

- Yes
- No

4. Will you be selling standard report products using the data?

- Yes
- No

5. Will you be selling a software product using the data?

- Yes
- No

6. Will you be reselling the data?

- Yes
- No

If yes, in what format will you be reselling the data (e.g., as a standalone product, incorporated with a software product, with a subscription, etc.)?

7. If you have answered “yes” to questions 3, 4 or 5, please describe the types of products, services or studies.

XIII. USE OF AGENTS AND/OR CONTRACTORS

Third-Party Vendors. Provide the following information for all agents and contractors who will work with the CHIA Data.

Company Name:	
Contact Person:	
Title:	
Address:	
Telephone Number:	
E-mail Address:	
Organization Website:	

8. Will the agent/contractor have access to the data at a location other than your location, your off-site server and/or your database?

- Yes
- No

If yes, please provide information about the agent/contractor’s data management practices, policies and procedures in your Data Management Plan.

9. Describe the tasks and products assigned to this agent or contractor for this project.

10. Describe the qualifications of this agent or contractor to perform such tasks or deliver such products.

11. Describe your oversight and monitoring of the activity and actions of this agent or subcontractor.

XIV. ASSURANCES

Applicants requesting and receiving data from CHIA pursuant to 957 CMR 5.00 (“Data Recipients”) will be provided with data following the execution of a data use agreement that requires the Data Recipient to adhere to processes and procedures aimed at preventing unauthorized access, disclosure or use of data, as detailed in the DUA and the applicant’s CHIA-approved Data Management Plan.

Data Recipients are further subject to the requirements and restrictions contained in applicable state and federal laws protecting privacy and data security, and will be required to adopt and implement policies and procedures designed to protect CHIA data in a manner consistent with the federal Health Insurance Portability and Accountability Act of 1996 (HIPAA).

By my signature below, I attest to: (1) the accuracy of the information provided herein; (2) my organization’s ability to meet CHIA’s minimum data security requirements; and (3) my authority to bind the organization seeking CHIA data for the purposes described herein.

Signature:	HS
Printed Name:	Helen Suh
Title	Associate Professor
Original Data Request Submission Date:	08/11/2015
Dates Data Request Revised:	