Commonwealth of Massachusetts
Center for Health Information & Analysis (CHIA)
Non-Government MA 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 MA 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:	David Dranove/Soheil Ghili
Title:	Dranove: Walter McNerney Distinguished Professor of Health Industry Management, Professor of Strategy Ghili: PhD Candidate
Organization:	Kellogg School of Management, Northwestern University
Project Title:	**Narrow Netowrks in the Health Insurance Market**
Mailing Address:	2001 Sheridan Rd, Evanston, IL 60208 Office#617
Telephone Number:	847-682-3430
Email Address:	d-dranove@kellogg.northwestern.edu s-ghili@kellogg.northwestern.edu
Names of Co-Investigators:	
Email Addresses of Co-Investigators:	
Original Data Request Submission Date:	
Dates Data Request Revised:	
Project Objectives (240 character limit)	I examine the provider-networks in for health insurance plans offered at the MA exchanges. The goal is to identify the determinants of Narrow Networks and to quantify their effects on consumer welfare. I use econometrics and simulation.
Project Research Questions (if applicable) or Business Use Case(s):	1.What will be the consumer-welfare consequence of policies intended to broaden the provider-networks of health insurance plans offered at the exchanges? In particular: 1.a) what will be the consequence of implementing the "Any Willing Provider" rule? 1.b) what will be the consequence of imposing minimum network-breadth requirements, and penalizing plans who violate?

	2. How is the breadth of the network shaped by some characteristics of the market. In particular: average income of the consumers, plan characteristics (e.g. co-insurance rates), etc

II. PROJECT SUMMARY

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

The purpose of this project is to gain more insights into how much regulation is optimal to ensure network adequacy in the ACA's exchange markets while not running the risk of excessively increased premiums. The reason why "network-adequacy-assuring" requlations may increase premiums is that they could take away insurers' bargaining leverage when negotiating over re-imbursement rates with hospitals. The increased re-imbursement rates, arising from the forgone insurer bargaining leverage, can in turn be passed on to consumers in the form of increased premiums.

The goal of my paper is to quantify the positive consumer-welfare effects of having better networks and negative consumer-welfare effects of having higher premiums, and then weight the two effects against one another to see where the best balance lies.

In order to be able to quantify the effects of "network-adequacy-assuring" regulations, I'll need to have a model of insurer-provider* bargaining. I need to know what re-imbursements look like now, between each insurer and each provider per unit of treatment, in order to be able to predict what they would look like as the outcome of a regulation. The APCD data will allow me to measure those reimbursements.

III. FILES REQUESTED

Please indicate which MA APCD file(s) you are requesting, the year(s) of data requested, and your justification for requesting *each* file. Please refer to the MA APCD Release 4.0 Documentation Guides for details of the file contents.

MA ALL PAYER CLAIMS DATABASE FILES	Year(s) Of Data Requested Current Yrs. Available ☒ 2010 ☒ 2011 ☒ 2012 ☒ 2013 ☒ 2014
☒ **Medical Claims**	**Please provide justification for requesting Medical Claims file:** The main purpose the MC file serves in my project is I use the claims data to infer the insurer-provider bargaining over the reimbursement rates. This is central to my project as I'd like to predict how much bargaining power insurers would lose if a regulation prevented them from leveraging the possibility of dropping providers out of their network when they bargain with those providers over reimbursement terms.

	The second use for the MC file is "demand estimation." I will use the claims file to measure customers' decisions on which hospital(s) to refer to for care when they get sick. This will enable me to infer their prefernces for different hospitals, and, hence, for different hospital networks. Note: to be able to implement my intended analysis, I will need to have the unencrypted NPIs. I will use the NPIs to identify different hospitals and hospital campuses. I will NOT use the NPIs to identify individual providers. For more details on this, please refer to four rows below in the current table.
☒**Pharmacy Claims**	**Please provide justification for requesting Pharmacy Claims file:** The reason why I need this file is in the same spirit as the argument for the MC file. Pharmacutical drugs are another form of health care provision and I'd like to investigate the factors determining the leverage insurers have when bargaining over the rembusement rates for those drugs. Note that the current version of the project aims at analyzing hospital networks, and doesn't examine pharmacy networks. But if the analysis of hospital networks is not computationally burdensome, I might add pharmacy networks later on. Analyzing providers at the individual level, though, is not and will not be on the scope of this project.
☐ **Dental Claims**	**Please provide justification for requesting Dental Claims file:** **N/A**
☒ **Member Eligibility**	**Please provide justification for requesting Member Eligibility file:** I need the ME file to infer the basic demographics of the patients who use different providers through different insurance plans. This will enable me to Identify the effects of those demographics on preferences for hospitals, and also, preferences for insurance plans. Knowing these preferences is of crucial importance, as it has direct implications on how much leverage a hospital has when bargaining with an insurance plan over reimbursement terms.

☐ **Provider** (encrypted NPI) Standard or ☒ **Provider*** (unencrypted NPI)	**Please provide justification for requesting Provider file:** I need the provider file to be able to link the APCD to external datasets regarding hospital characteristics using the NPIs. For instance, I'll need to infer what hospital characteristics lead to more frequent network presence, or to the bargaining leverage to charge a higher reimbursement rate. In order to be able to make such inferences, I'll need hospital NPIs to link reimbursements to hospital characteristics.
	***Please provide justificaiton for requesting unencrpted NPI (if requested). Refer to specifics in your methodology:** I need the unencrypted file because I'll need to link each hospital that patients refer to in the dataset to some aggregate hospital characteristics (like the total number of beds, case mix index, whether it has an academic affiliation, whether it's for profit, etc). I need to make this linkage because an important part of my analysis is trying to understand which hospital characteristics (including but not limited to those mentioned in the previous paragraph) are good predictors of hospitals' network presence and hospitals' marginal costs of providing care. I need to identify these predictors in order to then be able to predict which hospitals will be adde to/removed from insurance plans' networks of providers in response to different regulations. I will NOT use the unencrypted NPIs to identify individual providers. Studying individual providers is not part of my project.
☒ **Product**	**Please provide justification for requsting Product file:** I need the product file to be able to link the claims to insurance plans. For instance, I need to infer how much (average per patient) Celticare, as a "narrow network plan", pays to Cooley Dickinson Hospital annually in reimbursement terms. Comparing this annual reimbursement to that of a broader network plan (say NHP) for the same

	hospital will enable me to infer if, and by how much, a narrow network plan can negotiate lower reimbursement terms with hospitals. To be able to make this linkage, I need to provider file.

IV. GEOGRAPHIC DETAIL

Please choose *one* of the following geographic options for MA residents:

☐ 3 Digit Zip Code (MA)	☒ 5 Digit Zip Code (MA)

*****Please provide justification for requesting 5 digit zip code. Refer to specifics in your methodology:**

One of the basic demographics whose effect I'd like to investigate on hospital preference is the distances of each consumer from different hospitals. Of course I won't need to know the exact address for any patient but I need to address measure to be fine enough to capture that, for example, a patient living in West Roxbury might have different preferences (in terms of distances) for hospitals in the Boston area than one living in Charlstown does. The 3 digit zip-code does not provide this fine measure. But the 5 digit zipcode is, at the same time, fine enough to give me an adequately accurate measure and coarse enough to not increase the risk of identifying individual patients.

More specifically, the way the 5 digit zip-codes will be used are as follows:
The market that I'm interested in is the market of Boston greater area, which I define as places with 30 miles of downtown Boston. I partition this area into 20 "representative locations". Most representative location contain multiple 5-digit zip codes. I use the 5 digit zip codes for members to assign them to one of these representative areas. This way, give that I observe in the APCD what hospital the member has chosen to receive care from, I will be able to measure how hospital choices vary across representative areas.

I then measure the distance from each of the representative areas to each of the 35 general and acute care hospitals within Boston Greater Area manually using google maps (no automation or GIS will be used). The outcome will be a small (35 by 20) table of distances in miles.

At the end of this process, I will have the following pieces of information about each representative area:

1) The distances from that representative area to each hospital
2) Hospital choice patterns of people from that representative area

Combining the above two, I will be able to infer how much distance matters in hospital choice. I will do that by measuring how less often patients from choose hospitals farther from their representative areas than they choose hospitals within or close to their areas.

V. DATE DETAIL

Please choose *one* option from the following options for dates:

☐ Year (YYYY) (Standard)	☒ Month (YYYYMM) ***	☐ Day (YYYYMMDD) *** [for selected data elements only]

***** If requested, lease provide justification for requesting Month or Day. Refer to specifics in your methodology:**

I use the month data to consturct a different "admission events". That is, for claims from the same patient occurred in the same month, I lump them together in one "admission event" with a modal primary diagnosis and one modal hospital.

VI. FEE INFORMATION

Please consult the fee schedules for MA APCD 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 refer to the Application Fee Remittance Form and submit a letter stating the basis for your request (if required). 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.

VII. 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.

N/A

VIII. PURPOSE AND INTENDED USE

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

The first research question stated above, pertains directly to consumer welfare. The ultimate goal of this research is to shed light on what plicy regarding provider networks government should take in order to realize the highest value of consumer welfare.

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.

IX. APPLICANT QUALIFICATIONS

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

David Dranove is the Walter McNerney Distinguished Professor of Health Industry Management at Northwestern University's Kellogg School of Management, where he is also Professor of Strategy and Chair of the Department of Strategy. He was previously Director of the Health Enterprise Management program. He has a PhD in Economics from Stanford University.
Professor Dranove's research focuses on problems in industrial organization and business strategy with an emphasis on the health care industry. He has published nearly 100 research articles and book chapters and written five books, including *The Economic Evolution of American Healthcare* and *Code Red.* His textbook, *The Economics of Strategy*, is used by leading business schools around the world. Professor Dranove regularly consults with leading healthcare organizations in the public and private sector and serves on the Executive Committee and Board of Directors of the Health Care Cost Institute. He has also served as an economics expert in several high profile healthcare antitrust cases.

Soheil Ghili is a PhD candidate in Managerial Economics and Strategy at Kellogg School of Management, Northwestern University. He has worked as a research assistant at Northwestern Univeristy with sensitive insurance enrollment data. He has also co-authored an academic paper on pharmaceutical patent settlements, which involved using sensitive data (the paper will be revised and resubmitted to Journal of Law and Economics).

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.)

X. 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:

NA

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

NA

   ☐ Individual Provider Level Data

   What is the purpose of the linkage:

NA

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

NA

   ☒ Individual Facility Level Data

   What is the purpose of the linkage:

1) Linking to Individual Hospitals

The APCD data will be linked to the American Hospital Association data to get data on characteristics of the hospitals studied. This is necessary as I need to be able to compare hospitals within the provider-network of any given insurance plan to those out of its network in order to infer what types of hospitals are excluded by narrow-network plans. This has direct implications for my consumer welfare analysis.

2) Linking to individual Health Insurance Plans

The APCD data will also be linked to data on insurance plan characteristics such as name, network of hospitals providers, market share, premium, etc. It's essential to my analysis to know what hospitals each plan is covering in its network, as one of the main objectives of my modelling process is to capture how networks are form and why some of them are narrow. The dataset that I'll use to get insurance plan characteristics from will be public. For the network of providers for each plan, I'll either look them up on Mass Connector or the insurers' websites. For data on insurance plan market shares, premiums, deductibles and other characteristics, I'll use public datasets with that information.

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

1) Linking to Individual Hospitals

I will need to merge the APCD data with the American Hospital Association (AHA) survey data to get hospital characteristics. The linkage will occur using MC024 through MC035 as well as the provider NPI. I use these fileds to get hospital names and locations. Then I use the name-and-location information to match the data with the AHA data.

2) Linking to individual Health Insurance Plans

The linkage will be through PR003 through PR005 and HD002 to link to data from Mass Connector on features of the plans.

☒ Aggregate Data

3. What is the purpose of the linkage:

I'll link tha APCD data to data from census. The purpose of this linkage is to account for the effect of patient characteristics on the choice of providers (and, indirectly, of insurance plans).

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

To implement the linkage, I need data on the geographical location where the patient lives. I won't need the exact address; thus, no patient will be identified. I use data on zip, and

state/province from the ME dataset (Derived-ME9,ME016). I will then use this geographic indicators to link the APCD data to census data.

I will also use Derived-ME15 to link to census data in order to infer different preferences (in particular, different price sensitivities) by different age groups.

5. 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 .

The linkage will be a deterministic one. Details follow for each linkage:

1. Linking the MC file to the AHA data

I will use the NPI as well as the provider zip code to construct an identifier at the facility level (that is, an ID that would, for instance, tell apart NorthShore Hospital's Lynn and Salem campuses.
A similar identifier will be built for the AHA data off of the "aha_id" variable and the locations in that dataset. Then, I'll merge the two datasets using that variable.

Two important notes regarding the above merger:

Note1: As explained above, I will NOT analyze the claims at the individual provider level. The "provider" in this always a hospital.

Note2: the AHA data will NOT be merged to the raw MC file. Before being linked to the AHA data, the MC file will be aggregated so that for each hospital we know the number of admissions per month, the number of admissions with each diagnosis category, and the average reimbursement by each payor to each hospital for each of those categories. Once this aggregation has been completed, the aggregated file will be merged with the AHA data and used for analysis.

2. Linking the PR file to public data on insurance plan characteristics

In the market that I'm currently planning to analyze in this project (i.e. the CommCare market), there were only 5 insurance plans. I exclude one of them and am left with only 4. So, no particular ID will be constructed for the public dataset. For the PR file, I use the Payor ID to identify the insurer (in Commcare, the 4 plans are owned by 4 different insurers). This way, I will be able to merge the two files.

3. Linking the MC data to census data

The variable linking the MC file to census data will be a "customer segment identifier." I will partition the population of the market customers (in the case of CommCare, the current market that I'm planning to examine, the below 300% FPL population the the greater Boston), into various segments. The segmenting will be based on location and (if I will have the computational capacity) age group. This ID will be constructed for both the MC file and the census data. The linking will take place using this segment ID.

Note: even though I'm asking for the 5 digit zip codes for the members, the location measure that I will use to construct the market segments will be coarser than 5-digit zip codes (but finer than 3-digit, that's why I need to 5-digit). My expectation is that in the final analysis, the greater Boston area will be partitioned into 20 to 30 geographic segments.

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

The hospital and insurance plan linkages do not increase the likelihood of the identification of individual patients.
As for linkages between patient information to geographic information, the level at which I link the two datasets (i.e. zip-code level) does not allow for identifying individual. Of course any small cells will be furher aggregated if necessary.

7. 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.

The American Hospital Association (AHA) annual reports are the only non-public-domain dataset that the APCD will be linked to in my project. The corresponding data use agreements have been attached.

XI. 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.

The goal is to submit the paper coming out of this research to peer reviewed academic journals. This paper might also serve the paper as the co-investigator's "job market paper" (That is, he'll be presenting the paper at those academic departments, if any, that will invite him to their campuses to give a "job talk")

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.

Once ready, the paper will be uploaded to co-investigator's academic website (not constructed yet). I could also provide the manuscript upon request by email.

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.)?

N/A

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

N/A

XII. 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.

XIII. 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:	
Printed Name:	DAVID KOVARIK
Title	Director, Information & Systems Security/ Compliance
Original Data Request Submission Date:	11/19/15
Dates Data Request Revised:	2/24/16