

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:	Mathijs de Vaan	Peter S. Bearman
Title:	Assistant Professor	Professor
Organization:	Columbia University	Columbia University
Email Address:	md2999@columbia.edu	psb17@columbia.edu
Telephone Number:	510-612-4697	212-854-3094
Project Title:	The effects of Network Structure on Provider Learning, Health Outcomes, and Costs of Care	
Mailing Address:	606 W 122 nd street	
Names of Co-Investigators:		
Email Addresses of Co-Investigators:		
Original Data Request Submission Date:	October 20 th 2015	
Dates Data Request Revised:	February 25 th 2016	
Project Objectives (240 character limit)	The goal of this project is twofold. First, it aims to examine the co-evolution of interactions between providers and variation in care. And second, it examines the effect of this interactive process on health outcomes and costs of care. The methods used in this research include network analysis, simulation, and regression techniques.	
Project Research Questions (if applicable)	1. What are the mechanisms by which provider networks emerge and change? 2. Why are some providers embedded in dense clusters of between-provider interactions while the interactions of others are much more dispersed? 3. Do stable interactions between providers cause them to provide better, less costly, and more predictable care? 4. Do dense networks of providers cause patients to receive better care (as proxied by better health outcomes) at lower costs?	

II. PROJECT SUMMARY

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

The goal of this project is to examine how provider networks form, how they affect provider behavior, and how these dynamics can account for variation in health outcomes and costs of care. The APCD offers a unique context to measure interactions between providers and link the frequency and structure of these interactions to the outcomes of the care that patients receive.

The first question that this research project addresses is how referral networks between providers form. Providers typically have multiple options when referring their patients, but little is known about the selection and change in

referral partners. Of particular interest is the effect of the introduction of Accountable Care Organizations on the referral patterns of providers affected by these organizational changes. In answering this question we build on recent advances in network methodology to identify the sources of selection and change. The second research question further explores variation in the stability of referral patterns between providers. How do some providers become embedded in dense clusters of referral patterns that persist over time while the referral patterns of others are much more fragmented? If the argument holds that effective communication between providers can result in better care at lower cost levels, identifying variation in referral patterns can help Accountable Care Organizations to develop strategies for effective and affordable care provision. The third research question examines how interactions between providers shape their behavior in terms of the care that they provide. When providers share patients and discuss optimal care, providers are argued to influence each other and align the care they provide. This specific study aims to understand the conditions under which they do so and the behaviors that are likely to be influenced. In answering this question, we will focus especially on the role of physician networks in the adoption of specific treatments (including for example the adoption of off-label uses of prescription drugs). The fourth research question uses the findings from the first three research questions to identify how the structure of the referral network and the co-evolving behavior of providers causes patients to receive better care (as proxied by better health outcomes) at lower costs.

In conducting our research we will build on a large number of variables available in the APCD. For example, we will use both medical and pharmacy claims to construct the provider networks that we have hypothesized about. Moreover, to uniquely identify the actors in the network we will use the provider and patients EIDs. Finally, using data on patient visits, price, treatment decisions, physician group affiliation, and diagnosis codes, we are able to model the complex relationship between network structure on the one hand and health outcomes and cost of care on the other.

In sum, using recently introduced network models, this research exploits the richness of the APCD to test whether patients treated by providers embedded in dense clusters are characterized by better health and financial outcomes such as appropriate medication use, low hospital readmission rates, and lower overall costs, than providers that embedded in sparse networks.

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 <input checked="" type="checkbox"/> 2010 <input checked="" type="checkbox"/> 2011 <input checked="" type="checkbox"/> 2012 <input checked="" type="checkbox"/> 2013 <input checked="" type="checkbox"/> 2014
<input checked="" type="checkbox"/> Medical Claims	<p>Please provide justification for requesting Medical Claims file: Medical Claims data is needed to identify diagnosis, treatment dates, and admission type. These data are needed to capture the set of specialists to which the patient can potentially be referred given a diagnosis. Moreover this data is needed to subset referral patterns by diagnosis.</p> <p>We are requesting the Medical Claims data with decrypted NPIs. The analyses that we propose build on relational data (i.e. which physician refers to which physician?). Diagnosis, treatment dates, and admission type are likely to be predictors of the formation of these relations. Therefore, having access to the unencrypted NPIs is essential.</p>
<input checked="" type="checkbox"/> Pharmacy Claims	<p>Please provide justification for requesting Pharmacy Claims file: Pharmacy Claims are essential for our research and specifically for the third research question. These data allow us to categorize the types of treatment associated with</p>

	<p>specific diagnoses. By doing so the Pharmacy Claims provide the data needed to study the spread of treatment decisions through the physician referral network.</p> <p>We are requesting the Pharmacy Claims data with decrypted NPIs. The analyses that we propose build on relational data (i.e. which physician refers to which physician?). Treatment type is likely to be a strong predictors of the formation of these relations. Therefore, having access to the unencrypted NPIs is essential.</p>
<input type="checkbox"/> Dental Claims	Please provide justification for requesting Dental Claims file:
<input checked="" type="checkbox"/> Member Eligibility	<p>Please provide justification for requesting Member Eligibility file:</p> <p>Patient demographics, insurance coverage, and product information are needed to understand the patient mix that physicians see. The patient mix is likely to be a strong predictor of the referral relations between physicians that do emerge.</p>
<input type="checkbox"/> Provider (encrypted NPI) Standard or <input checked="" type="checkbox"/> Provider* (unencrypted NPI)	<p>Please provide justification for requesting Provider file:</p> <p>The provider data is the backbone of our proposed analyses. These data are needed to identify the physicians in our study and their interactions.</p> <p>*Please provide justificaiton for requesting unencrpted NPI (if requested). Refer to specifics in your methodology:</p> <p>To make sure that our analyses build on correctly specified models we will include variables from the AMA Masterfile. To link the AMA Masterfile to the APCD, having acces to the unencrypted NPIs is essential.</p>
<input checked="" type="checkbox"/> Product	<p>Please provide justification for requesting Product file:</p> <p>The insurance product data will help us to determine and identify the different insurance types and insurance types vary by physicians. This data is essential to capture the coverage structures under which physicians operate.</p>

IV. GEOGRAPHIC DETAIL

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

<input type="checkbox"/> 3 Digit Zip Code (MA)	<input checked="" type="checkbox"/> 5 Digit Zip Code (MA)
<p>***Please provide justification for requesting 5 digit zip code. Refer to specifics in your methodology:</p> <p>Referral networks may emerge as a result of several factors -- one of which is spatial distance between physicians. A very well established literature on social networks has shown that many social relationships are primarily driven by the co-location of individuals. Moreover, this literature shows that interactions between individuals decay exponentially with distance. We assume that physicians referrals are organized according to similar principles and detailed data on the location of physicians is therefore essential. To be precise, variation in behavior between physicians within the 3 digit zip code may be caused by spatial variation. To capture this variation and not make false inferences, the 5 digit zip code needs to be observed. We will use ArcGIS to find the longitude and the latitude of each address and to calculate the distances between these spatial points.</p>	

V. DATE DETAIL

Please choose one option from the following options for dates:

Year (YYYY) (Standard)	Month (YYYYMM) ***	<input checked="" type="checkbox"/> Day (YYYYMMDD) *** [for selected data elements only]
<p>*** If requested, please provide justification for requesting Month or Day. Refer to specifics in your methodology:</p> <p>It is essential for the analyses that we have proposed to capture the precise dates at which patient – physician interactions take place as well as the precise dates at which treatments are administered and health outcomes are observed. Without the precise dates the temporal order of events will be obscured and we will therefore not be able to construct meaningful referral networks and link them directly to health outcomes.</p>		

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.

The unique set of patients in the Medicaid population provides a source of salient variation that can help explain patterns in referral networks. Both the inclusion of low income patients and the inclusion of children will allow us to explore in more detail how demographic factors shape referral patterns. Answering questions about how reimbursement schemes affect referral patterns — and subsequently cause variation in health outcomes — will assist the administration of the Medicaid program and MassHealth in evaluating the potential benefits of introducing constraints in referral options.

VIII. PURPOSE AND INTENDED USE

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

This project will identify patterns in referral networks and examine how variation in these patterns is related to differences in health outcomes and costs of care. The past decade has been characterized by a stark increase in the probability that an ambulatory visit will lead to a referral, but very little is known about the causes of this increase and the types of referrals that have accounted for this increase. Moreover, given the Affordable Care Act's proposals to provide patients with better and affordable care, it is important to answer questions about efficiency in referral patterns. Using newly developed techniques in the fields of network science, causal inference, and econometrics this project will address the complexity of the relational nature of the APCD. One of the newly developed methods that this project will build on allows us to use detailed referral data, patient and provider characteristics, and other external factors to model how referral networks emerge and change. By using these newly developed techniques, this project will go beyond simply describing the data; it will use models that allow us to make causal inferences which will inform market regulators and administrators to improve care and critically think about the financial consequences of certain choices in health care plan design.

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.**

Mathijs de Vaan is an Assistant Professor at the Haas School of Business at UC Berkeley. He obtained his Ph.D. in Sociology from Columbia University and is currently a Fellow at the Interdisciplinary Center of Innovative Theory and Empirics (INCITE). He also holds a Ph.D. and an M.Sc. in economics from Utrecht University. Mathijs has extensive research experience in the fields of social network analysis, health economics, medical sociology, and econometrics and he has programming experience with R, Python, Stata, and SQL. Mathijs has experience analyzing large healthcare databases including IMS health data and data from the Adverse Event Reporting System maintained by the FDA. Mathijs has authored and co-authored several academic studies that use sensitive and restricted-use data. **Peter Bearman** is the Director of INCITE, the Cole Professor of Social Science, and Co-Director of the Health & Society Scholars Program, the Mellon Interdisciplinary Training Program, and OHMA at Columbia University. A specialist in network analysis, he co-designed the National Longitudinal Study of Adolescent Health.

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:

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

Individual Provider Level Data

What is the purpose of the linkage:

We will link providers to the American Medical Association Physician Masterfile (AMA Masterfile) to obtain information on provider specialty and demographic data.

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

The AMA Masterfile NPI will be linked to the CHIA owned NPI.

Individual Facility Level Data

What is the purpose of the linkage:

We will use the American Hospital Association Annual Survey Database for hospital characteristics. Linking CHIA data to this data will help us understand how provider referral choices are informed by organizational structures.

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

We will link the NPI in the American Hospital Association Annual Survey Database to the CHIA owned NPI.

Aggregate Data

What is the purpose of the linkage:

We will link the ZIP code in CHIA's APCD data to U.S. Census data to enable us to explore information on socioeconomic status, and regional characteristics. Moreover, the U.S. Census data is needed to construct a patient risk set. We will not identify individual patients or individual providers.

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

ZIP codes from the U.S. Census data will be linked to ZIP codes in the CHIA data.

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 .

Deterministic matching will be used to link the NPI in the AMA masterfile to CHIA's NPI, the NPI in the American Hospital Association Annual Survey Database, and the ZIP code in CHIA's APCD data to the ZIP code in the US Census data.

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

No individual level data will be involved in this matching process.

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.

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 results of this research project will be submitted for publication in academic peer-reviewed journals. The findings will also be presented at academic conferences and workshops. In these publications, identification of individual patients will not be possible and any cell with fewer than 11 observations will not be disclosed. The effect sizes presented in our publications are average effects and the averages are taken over large numbers of individuals.

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.

The results of the study will be available in print and online from the publishers of the journals in which the research will be published. The results will also be directly available from the investigator's website and can be downloaded at no costs.

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.

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.

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:	<i>Juliana Powell</i>
Printed Name:	JULIANA POWELL, Columbia University
Title	ASSOCIATE DIRECTOR OF OPERATIONS, MS
Original Data Request Submission Date:	October 20 th 2015
Dates Data Request Revised:	February 25 th 2016