

Commonwealth of Massachusetts  
Center for Health Information & Analysis (CHIA)  
Non-Government Agency Application for Data

**NOTE:** This application is to be used by all applicants, except Government Agencies, as defined in 957 CMR 5.02.

**I. GENERAL INFORMATION**

| <b>APPLICANT INFORMATION</b>             |   |
|--|---|
| Applicant Name:                          | Keith Marzilli Ericson (primary applicant)  |
| Title:                                   | Assistant Professor of Markets, Public Policy, & Law and Faculty Research Fellow  |
| Organization:                            | Boston University School of Management and National Bureau of Economic Research   |
| Co-Investigator:                         | Jim Rebitzer  |
| Title:                                   | Professor of Management, Economics and Public Policy; Everett J. Lord Distinguished Scholar; Research Associate   |
| Organization:                            | Boston University School of Management and National Bureau of Economic Research   |
| Co-Investigator:                         | Benjamin Lubin  |
| Title:                                   | Assistant Professor of Information Systems  |
| Organization:                            | Boston University School of Management  |
| Co-Investigator:                         | Brigham Frandsen  |
| Title:                                   | Assistant Professor of Economics  |
| Organization:                            | Brigham Young University  |
| Co-Investigator:                         | Kimberley Geissler  |
| Title:                                   | Research Associate  |
| Organization:                            | Boston University School of Management  |
| Co-Investigator:                         | Amanda Starc  |
| Title:                                   | Assistant Professor of Health Care Management   |
| Organization:                            | University of Pennsylvania  |
| Project Title:                           | <b>Understanding Insurance, Provider Networks, and Outcomes</b>   |
| Date of Application:                     | September 2013  |
| Project Objectives (240 character limit) | We examine characteristics of provider networks and insurance policies and relationships with patient outcomes to better understand provider and enrollee behaviors. We use regression techniques, network analysis, and simulation.  |
| Project Research Questions               | We investigate how consumers value insurance plan designs and provider networks, and examine the links between plan design, network structure, and outcomes. Specifically, we ask the following questions:<br>1. What are the consequences of broader versus narrower insurance plan choice |

|  |  |
|--|--|
|  | <p>set (i.e. variation in deductibles, actuarial value, etc.)?</p> <p>2. Can more complex insurance contracts improve outcomes and patient welfare by linking cost-sharing to more information (e.g. provider quality, patient health status)?</p> <p>3. What is consumer willingness-to-pay for additional network access from their employer’s plan menu, and how does this affect insurance plan design?</p> <p>4. Do consumers with greater medical utilization gravitate towards certain kinds of plans or networks of providers?</p> <p>5. What do professional networks of shared patients among physicians look like, and how do such networks vary by type of insurance plan (e.g. HMO v PPO vs. Medicaid)?</p> <p>6. What is the relationship between professional networks of shared patients among physicians, resource use, and patient outcomes?</p> |
|--|--|

Please indicate if you are a Researcher, Payer, Provider or Provider Organization and you are seeking data pursuant to 957 CMR 5.04 (De-Identified Data) or 957 CMR 5.05 (Direct Patient Identifiers for Treatment or Coordination of Care).

|  |  |
|--|--|
| <p><input checked="" type="checkbox"/> Researcher</p> <p><input type="checkbox"/> Payer</p> <p><input type="checkbox"/> Provider / Provider Organization</p> | <p><input checked="" type="checkbox"/> 957 CMR 5.04 (De-Identified Data)</p> <p><input type="checkbox"/> 957 CMR 5.05 (Direct Patient Identifiers)</p> |
|--|--|

All other requests are subject to 957 CMR 5.06.

**II. PROJECT SUMMARY**

Briefly describe the purpose of your project and how you will use the CHIA data?

This project investigates how consumers value insurance plans and networks. This requires we establish links between plan design, network structure, and health/utilization outcomes. We will use the CHIA data to first examine the link between characteristics of insurance plans and the utilization of health care, including particular procedures and total spending. We then examine the provider networks available to consumers in each plan/insurance type. We estimate models of demand for particular providers (e.g. hospitals, using measures of geographical distance) and then for insurer-specific networks. We also examine how network structure is associated with outcomes. We associate network-breadth measures with price levels, utilization patterns, and enrollee composition. We also model the structure of provider-provider connections (e.g. referral networks) using insights from network theory (e.g. concepts of connectedness, centrality, etc). We examine how this structure varies by insurance plan design, and how these structures are associated with measures of health outcomes and process quality.

These analyses will use standard forms of regression analysis, hazard models, simulated method of moments, models of consumer choice (e.g. differentiated product demand models) and welfare (e.g. expected utility models), and network structure modeling (e.g., clustering coefficients, betweenness, and spectral analysis).

**III. FILES REQUESTED**

Please indicate the databases from which you seek data, the Level(s) and Year(s) of data sought.

| DATABASE  | Level 1 <sup>1</sup> or 2 <sup>2</sup>   | Single or Multiple Use                                   | Year(s) Of Data Requested<br>Current Yrs. Available<br>2009 - 2011   |
|---|--|--|--|
| X Medical Claims  | <input type="checkbox"/> Level 1<br>X Level 2  | Select... Single   | X2009 X2010 X2011  |
| X Pharmacy Claims   | <input type="checkbox"/> Level 1<br>X Level 2  | Select... Single   | X2009 X2010 X2011  |
| <input type="checkbox"/> Dental Claims<br>X Member Eligibility<br>X Provider<br>X Product | <input type="checkbox"/> Level 2<br>X Level 2<br>X Level 2<br>X Level 2  | Select... Single<br>Select... Single<br>Select... Single | <input type="checkbox"/> 2009 <input type="checkbox"/> 2010 <input type="checkbox"/> 2011<br>X2009 X2010 X2011<br>X2009 X2010 X2011<br>X2009 X2010 X2011 |
| <b>CASEMIX</b>  | <b>Level 1 - 6</b>   |  | <b>Fiscal Years Requested</b>  |
| <b>Inpatient Discharge</b>  | <input type="checkbox"/> Level 1 – No Identifiable Data Elements<br><input type="checkbox"/> Level 2 – Unique Physician Number (UPN)<br><input type="checkbox"/> Level 3 – Unique Health Information Number (UHIN)<br><input type="checkbox"/> Level 4 – UHIN and UPN<br><input type="checkbox"/> Level 5 – Date(s) of Admission; Discharge; Significant Procedures<br><input type="checkbox"/> Level 6 – Date of Birth; Medical Record Number; Billing Number |  | <u>1998-2012 Available</u><br>(limited data available 1989-1997)   |
| <b>Outpatient Observation</b>   | <input type="checkbox"/> Level 1 – No Identifiable Data Elements<br><input type="checkbox"/> Level 2 – Unique Physician Number (UPN)<br><input type="checkbox"/> Level 3 – Unique Health Information Number (UHIN)<br><input type="checkbox"/> Level 4 – UHIN and UPN<br><input type="checkbox"/> Level 5 – Date(s) of Admission; Discharge; Significant Procedures<br><input type="checkbox"/> Level 6 – Date of Birth; Medical Record Number; Billing Number |  | <u>2002-2011 Available</u>   |

<sup>1</sup> Level 1 Data: De-identified data containing information that does not identify an individual patient and with respect to which there is no reasonable basis to believe the data can be used to identify an individual patient. This data is de-identified using standards and methods required by HIPAA.

<sup>2</sup> Level 2 (and above) Data: Includes those data elements that pose a risk of re-identification of an individual patient.

|                      |   |                            |
|----------------------|---|----------------------------|
| Emergency Department | <input type="checkbox"/> Level 1 – No Identifiable Data Elements<br><input type="checkbox"/> Level 2 – Unique Physician Number (UPN)<br><input type="checkbox"/> Level 3 – Unique Health Information Number (UHIN)<br><input type="checkbox"/> Level 4 – UHIN and UPN; Stated Reason for Visit<br><input type="checkbox"/> Level 5 – Date(s) of Admission; Discharge; Significant Procedures<br><input type="checkbox"/> Level 6 – Date of Birth; Medical Record Number; Billing Number | <u>2000-2011 Available</u> |
|----------------------|---|----------------------------|

**IV. REQUESTED DATA ELEMENTS [APCD]**

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. REQUESTED DATA ELEMENTS [CASE MIX]**

Please use the CASE MIX DATA SPECIFICATION WORKBOOK to identify which deniable data elements (from Level 2 or above) you would like to request and attach this to your application.

**VI. MEDICAID DATA**

Federal law (42 USC 1396a(a)7) restricts the use of individually identifiable data of Medicaid recipients to uses that benefit 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 benefits the administration of the Medicaid program.

This project will identify how insurance plans can be designed to improve individuals' health outcomes and welfare, and how to increase health system efficiency. The results of our analyses will inform market regulators as they evaluate policies, particularly those that affect network coverage in MassHealth. Our results will identify patterns of provider networks that patients value and that deliver effective care, giving more information to the Medicaid program about efficient allocation of providers. Additionally, many studies have examined the effects of limited provider networks for Medicaid – we will look at the structure of these networks and their relationships with patient outcomes, particularly as compared to other types of insurance including HMOs. This will potentially inform Medicaid as to the value of expanding networks (if any) in terms of outcomes including cost and utilization measures such as hospitalizations or emergency department visits.

**VII. MEDICARE DATA**

Medicare data may be disseminated to state agencies and/or entities conducting research projects that are directed and partially funded by the state if such research projects would allow for a Privacy Board or an IRB to make the findings listed at 45 CFR 164.512(i)(2)(ii) if the anticipated data recipient were to apply for the data from CMS directly. If you are requesting Medicare data, please explain how your research project is directed and partially funded by the state and describe in detail why your proposed project meets the criteria set forth in 45 CFR 164.512(i)(2)(ii). Applicants must describe how they will use the data and inform CHIA where the data will be housed. CHIA must be informed if the data has been physically moved, transmitted, or disclosed.

N/A

**VIII. DIRECT PATIENT IDENTIFIERS<sup>3</sup>**

State and federal privacy laws may require the consent of Data Subjects prior to the release of any Direct Patient Identifiers. If you are requesting data that includes Direct Patient Identifiers, please provide documentation of patient consent or your basis for asserting that patient consent is not required.

**IX. REQUESTS PURSUANT TO 957 CMR 5.04**

Payers, providers, provider organizations and researchers seeking access to Level 1 (de-identified) data are required to describe how they will use such data for the purposes of lowering total medical expenses, coordinating care, benchmarking, quality analysis or other administrative research purposes. Please provide this information below.

N/A

**X. 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                |   |                           |

**XI. PURPOSE AND INTENDED USE**

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

This project will identify how insurance plans can be designed to improve individuals' health outcomes and welfare, and how to increase health system efficiency. The results of our analyses will inform market regulators as they evaluate policies that affect network coverage and plan generosity in the health insurance exchanges, as well as in the individual and group markets. Our results will identify patterns of provider networks that patients value and that deliver effective care, which will be of use to providers and insurance plans/carriers that desire to improve care.

2. **Attach** a brief (1-2 pages) description of your research methodology. (This description will not be posted on the internet.)

<sup>3</sup> Direct Patient Identifiers. Personal information, such as name, social security number, and date of birth, that uniquely identifies an individual or that can be combined with other readily available information to uniquely identify an individual.

3. Has your project received approval from your organization's Institutional Review Board (IRB)?

- 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.

## XII. APPLICANT QUALIFICATIONS

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

Keith Marzilli Ericson holds a PhD in Economics from Harvard University and a BA in economics and political science from Williams College. He is an Assistant Professor of Markets, Public Policy, and Law in the Boston University School of Management, teaching courses in econometrics and optimization theory. He is also a Faculty Research Fellow at the National Bureau of Economics Research.

Jim Rebitzer holds a Ph.D. in economics from the University of Massachusetts at Amherst and a BS in biology from the University of Illinois in Urbana Illinois. He is a professor of Management, Economics and Public Policy at the Boston University School of Management where he is also an Everett J. Lord distinguished scholar as well as being a research associate at the National Bureau of Economic Research. He has considerable experience working with confidential commercial insurance records in empirical, health services research.

Benjamin Lubin holds a PhD in Computer Science from Harvard University and an AB in Computer Science from Harvard University. He is an Assistant Professor of Information Systems in the Boston University School of Management, teaching courses in computer science. He is also a Hariri Institute junior faculty fellow.

Brigham Frandsen holds a PhD in Economics from the Massachusetts Institute of Technology and a BS in economics and physics from Brigham Young University. He is an Assistant Professor of Economics at Brigham Young University, teaching courses in econometrics. He also was a Robert Wood Johnson Scholar in Health Policy Research at Harvard University.

Kimberley Geissler holds a PhD in Health Policy and Management from the University of North Carolina Gillings School of Global Public Health and a BA in chemistry and economics from Williams College. She is a Research Associate in the Boston University School of Management. She is also an Adjunct Assistant Professor of Health Policy and Management at the University of North Carolina.

Amanda Starc holds a PhD in Business Economics from Harvard University and a BA in Economics from Case Western Reserve. She is an Assistant Professor of Health Care Management in the Wharton School, University of Pennsylvania.

Ericson and Starc have examined the Massachusetts Health Insurance Exchange (Connector) in a series of papers (2013 American Economic Review, 2013 Inquiry, 2012 National Bureau of

Economic Research Working Paper, 2013 Working Paper). All investigators have worked with sensitive health data previously (Ericson: Marketscan; Rebitzer: numerous datasets including commercial insurance claims data; Frandsen: confidential medical and drug claims data from a large private insurer; MEPS Insurance Component; Census Long Form (confidential 1:6 sample); Lubin: Marketscan; Starc: Medigap; Geissler: State Hospital Discharge Datasets, Marketscan).

2. Attach résumés or curriculum 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.)

### XIII. DATA LINKAGE AND FURTHER DATA ABSTRACTION

1. Does your project require linking the CHIA Data to another dataset? YES X NO
2. If yes, will the CHIA Data be linked to other patient level data or with aggregate data (e.g. Census data)?  
Patient Level Data  Aggregate Data X
3. If yes, please identify all linkages proposed and explain the reasons(s) that the linkage is necessary to accomplish the purpose of the project.

We propose to link CHIA data to the following aggregate datasets, described here:

- 1) Hospital linkages – We will link hospitals to the American Hospital Association Annual Survey Database for hospital characteristics; to the Medicare Hospital Compare dataset for quality and aggregate health outcome data. This is needed to describe provider networks and model patient choice of hospital.
- 2) Provider linkages – We will link providers to the American Medical Association Physician Masterfile for provider specialty and demographic data; to tiering (quality/cost-efficient care) measures for specialist providers participating in the GIC UniCare plans; and to the Massachusetts Health Quality Partners (MHQP) provider dataset to accurately link providers to practices for determination of practice level measures and to link to quality data (ambulatory physician group practice linkages). To link to provider information, we need a number of provider identifiers. In published analyses and reports, we will not identify providers or report information where deductive disclosure would be possible (e.g., we will mask small cells, etc.). This is needed to characterize provider networks (e.g. are more efficient doctors more likely to be in the same network) and model patient choice of provider.
- 3) Ambulatory physician group practice linkages – We will link provider data to MHQP quality data on clinical and patient experience measures for primary care physicians. This is needed to characterize provider networks and model patient choice of provider.
- 4) Geographic area linkages – We will link member geographic data (ZIP, city, county) to the Area Resource File and the American Communities Survey/Census data to get information on healthcare supply, socioeconomic status, and regional characteristics. This is needed to account for variation in patient characteristics that might affect patient use of medical care or

outcomes; we do not identify individual patients, merely link to characteristics of their ZIP code.

5) Carrier and/or insurance plan linkages – We will link carrier and/or insurance plan data to market share and premium data from Mass Connector, as well as to the network definition of plans in the Connector (by looking up particular providers to determine if they are in-network for a given plan). This is needed to model insurer price setting (which is jointly determined with patient demand for insurance), how prices move with plan generosity, and to model consumer choice of insurance plan.

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

As these datasets do not increase the ability or likelihood of identification of individual patients in the linked datasets, these linkages would not jeopardize patient confidentiality. As discussed in the data security and integrity section, we will take great care to ensure the confidentiality of the data.

**XIV. 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, or similar document.

We will submit the results of our study for academic publication in peer-reviewed journals. We will have summary statistics and analyses completed using the data, but no identification of patients will be possible. If there are small cells in the analysis (<10 patients), we will censor these cells to maintain confidentiality.

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 will be available for no fee at the researchers' websites or upon email request.

- |   |     |                          |    |   |
|---|-----|--------------------------|----|---|
| 3. Will you use the data for consulting purposes?               | YES | <input type="checkbox"/> | NO | X |
| 4. Will you be selling standard report products using the data? | YES | <input type="checkbox"/> | NO | X |
| 5. Will you be selling a software product using the data?       | YES | <input type="checkbox"/> | NO | X |

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



|  |
|--|
|  |
|--|

**XV. 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:         | N/A |
| Contact Person:       |     |
| Title:                |     |
| Address:              |     |
| Telephone Number:     |     |
| E-mail Address:       |     |
| Organization Website: |     |

1. Will the agent/contractor have access to the data at a location other than your location or in an off-site server and/or database?                      YES                         NO

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

|  |
|--|
|  |
|--|

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

|  |
|--|
|  |
|--|

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

|  |
|--|
|  |
|--|