Commonwealth of Massachusetts
Center for Health Information & Analysis (CHIA) Request for Data
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 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	Jennifer Kwok
Title	PhD Candidate in Economics
Organization	University of California, Berkeley
Project Title	Influence of the Provider-Patient Relationship on Patient Health-Behavior and Outcomes
Mailing Address	530 Evans Hall #3880 Department of Economics University of California, Berkeley Berkeley, CA 94720-3880
Telephone Number	201-315-2586
Email Address	jenkwok@berkeley.edu
Names of Co-Investigators	Benjamin Handel (Faculty Supervisor)
Email Addresses of Co-Investigators	handel@berkeley.edu
Original Data Request Submission Date	November 16, 2015
Dates Data Request Revised	
Project Objectives (240 character limit)	The project studies the provider-patient relationship with a focus on the effects of continuity in the provider-patient relationship and how factors within the provider-patient relationship contribute to healthcare variation.
Project Research Questions (if applicable) or Business Use Case(s)	The project includes two main research questions, each with multiple objectives. 1. What are the effects of provider continuity on patient health-related behavior and outcomes? a. Determine the causal effects of provider discontinuities that occur with individual physicians. b. Compare the causal effects of provider discontinuities that occur with (i) individual physicians who are part of multi-physician group practices, (ii) individual physicians in single-physician practices, and (iii) entire multi-physician group practices. 2. What are the contributions of factors in the provider-patient relationship to healthcare variation? a. Decompose healthcare variation into provider-specific factors, patient-specific factors, and idiosyncratic provider-patient match effects. b. Identify (i) organization and physician characteristics that explain provider-specific effects and (ii) provider-patient relationship

	characteristics that explain idiosyncratic provider-patient match effects. c. Characterize provider-patient sorting based on patient-specific and provider-specific factors. d. Determine effects of healthcare provider organizations on referrals.

II. PROJECT SUMMARY

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

The provider-patient relationship impacts healthcare access and production. In healthcare access, the relationship affects whether a patient has a usual source of care, successfully seeks care, and receives timely care. In healthcare production, the relationship determines the quality, effectiveness, and efficiency of care. This project will examine the influence of the relationship on patient health-related behavior and outcomes.

Question 1: What are the effects of provider continuity on patient health-related behavior and outcomes?
My research will investigate the effects of provider continuity on patient health-related behavior and outcomes. Provider continuity occurs when a patient is treated by the same provider over time. Provider discontinuities, defined as breaks in continuity, could affect the quantity and quality of healthcare. Loss of access to providers results in loss of relationship capital, including knowledge and trust. Patients face search costs (e.g., finding in-network provider) and transaction costs (e.g., transferring medical records) in obtaining new providers. Provider continuity likely benefits many patients, such as through better preventive care and chronic conditions management. On the other hand, some patients who switch providers could receive valuable second opinions, which could counteract the negative effects of provider discontinuities.

Question 2: What are the contributions of factors in the provider-patient relationship to healthcare variation?
My research will investigate the contributions of organization-specific, provider-specific, and patient-specific factors to variation in healthcare. The remaining variation is due to idiosyncratic provider-patient match effects, which may be correlated with match-specific characteristics (e.g., relationship length) that impact health-related outcomes. I will characterize provider-patient sorting, which I define to be based on patient-specific and provider-specific factors (e.g., high-spending patients to high-spending providers). I will also measure the correlation of provider-specific effects across patient demographic groups and outcome measures. For example, I will examine (a) whether physicians with the greatest estimated physician-specific effects for female patients also have the greatest estimated physician-specific effects for male patients and (b) whether physicians with the greatest estimated physician-specific effects on healthcare spending also have the best estimated physician-specific effects on patient health outcomes. Furthermore, I will explore how referrals contribute to provider-patient sorting and matching.

Use of requested CHIA data. My project requires matched provider-patient administrative claims, provider characteristics, and patient characteristics. Administrative claims include physician information (e.g., identifiers, specialty, location), patient information (e.g., demographics, location), diagnoses, medical claims, pharmacy claims, and dental claims. I plan to use the Massachusetts APCD, currently available for years 2010-2014. I will also separately use CMS Research Identifiable Files that include 20% of Original Medicare beneficiaries during years 1999-2012, and I have already received CMS approval under Data Use Agreement 28157.

The requested CHIA data – the Massachusetts APCD – will be particularly valuable. APCDs have a few key features: (a) they include individuals under age 65 (to complement my sample of Original Medicare beneficiaries), (b) variation in provider networks yields a type of provider discontinuity, and (c) they may include some demographic information (e.g., gender, age, race/ethnicity, subscriber employment characteristics). The Massachusetts APCD is ideal for a few reasons. Massachusetts individuals have been uniquely less likely to make labor market decisions based on health insurance, since the 2006 health care reform law created an individual mandate and the Massachusetts Health Connector for individual health insurance. Also, the Massachusetts APCD has high population coverage and high-quality data relative to other state all payer claims databases.

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 Medical Claims file is required to (a) identify potential provider discontinuities (i.e., breaks in provider continuity) and other provider-related changes (e.g., merger of physician practice groups), and (b) measure health-related outcomes (e.g., primary care, specialty care, emergency care, healthcare spending). Please see the Research Methodology for additional information.
☑ **Pharmacy Claims**	**Please provide justification for requesting Pharmacy Claims file:** The Pharmacy Claims file is required to (a) help identify potential provider discontinuities and other provider-related changes as a supplemental source of information on medical visits, and (b) measure health-related outcomes (e.g., drug adherence). Please see the Research Methodology for additional information.
☑ **Dental Claims**	**Please provide justification for requesting Dental Claims file:** The Dental Claims file is required to (a) identify potential provider discontinuities and other provider-related changes, and (b) measure health-related outcomes (e.g., dental checkups). Please see the Research Methodology for additional information.
☑ **Member Eligibility**	**Please provide justification for requesting Member Eligibility file:** The Member Eligibility file is required to (a) select member samples based on their characteristics (e.g., demographics), (b) control for member characteristics, and (c) examine heterogeneity by member characteristics. Please see the Research Methodology for additional information.
☐ **Provider** (encrypted NPI) Standard or ☑ **Provider*** (unencrypted NPI)	**Please provide justification for requesting Provider file:** The Provider file is required to (a) identify provider facility and office information (e.g., physician practice group, office locations), (b) control for provider characteristics, and (c) examine heterogeneity by provider characteristics. Please see the attached research methodology for additional information.
	***Please provide justification for requesting unencrypted NPI (if requested). Refer to specifics in your methodology:** The Provider file with unencrypted NPI is required to match providers in the Massachusetts All Payer Claims Database files with individual provider level data from other files using the unencrypted NPIs. Please see Section X. Data Linkage and Further Data Abstraction and the Research Methodology for additional information.
☑ **Product**	**Please provide justification for requesting Product file:** The Product file is required to (a) determine patient healthcare covered services and costs, (b) control for plan characteristics, and (c) examine heterogeneity by plan characteristics. Please see the Research Methodology for additional information.

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:** My methodology requires the 5 digit zip code because it requires the precise locations of providers and members/patients. Precise information on provider location and member/patient location will be used to determine (a) distances between member/patient home locations and provider office/facility locations and (b) distances between subscriber/patient work and provider office/facility locations. Please see the Research Methodology for additional information.	

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, please provide justification for requesting Month or Day. Refer to specifics in your methodology:** My methodology requires day-level dates because it requires the precise timing of events. The following events should be measured at the day level if possible (and otherwise, at the month level): (a) patient plan-related changes; (b) provider location and other employment-related changes; (c) plan provider network changes; and (d) medical, pharmacy, and dental claims. Precise information on a patient's plan(s), plan provider networks, and claims will be used to determine a patient's provider network on a given date. Precise information on provider location changes and other provider employment-related changes (e.g., retirement) – as well as information on a patient's set of in-network providers – will be used to determine the timing of potential provider discontinuities (i.e., breaks in provider continuity) and other provider-related changes (e.g., merger of physician practice groups) that could affect patient health-related behavior and outcomes. Please see the Research Methodology for additional information.		

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.

I am requesting the Medicaid data to (a) to identify heterogeneity in results by patient demographics and (b) to study features of the MassHealth insurance plans. Medicaid data will enable me to examine heterogeneity by income and other Medicaid eligibility criteria as well as certain chronic conditions that may be more prevalent among the Medicaid population. Using Medicaid data, I can also fully assess whether the provider-patient relationship is more important to certain socioeconomic and racial/ethnic minority groups, since cultural competence is likely important to healthcare. Furthermore, Medicaid data will allow me to study features of MassHealth insurance plans (e.g., provider networks) and how they influence the provider-patient relationship.

VIII. PURPOSE AND INTENDED USE

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

   The provider-patient relationship is a central component of healthcare access and healthcare production. Healthcare access depends on having health insurance, having a usual source of care, successfully seeking care, and receiving timely care, among other factors. Beyond healthcare access, the relationship determines healthcare production, and thus, the quality, effectiveness, and efficiency of healthcare. The provider-patient relationship is particularly relevant to recent developments in the US healthcare system. Narrow and limited network health plans cause more patients to switch providers, whereas accountable care organizations (ACOs) and managed care plans increase provider responsibility and promote coordination of care. A better, in-depth understanding of the provider-patient relationship will enable researchers to accurately assess these, as well as other, healthcare system features. Please see the Appendix on Public Interest for additional information.

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.

I am a PhD candidate in the Department of Economics at the University of California, Berkeley. My dissertation focuses on the provider-patient relationship and will include papers from this project. I received Bachelors in Science degrees in Biology and Chemical-Biological Engineering from the Massachusetts Institute of Technology. Prior to graduate school, I worked as an Associate Consultant at the Boston Consulting Group, focusing on the pharmaceutical and medical device industries, and as a Research Professional at the University of Chicago, focusing on health economics and development economics. As a PhD student, I completed the Department's field courses in industrial organization, labor economics, and psychology and economics and gained extensive programming experience. My background in health-related fields and experience in economics provide me with valuable expertise for empirical health economics research.

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:

   Not applicable.

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

   Not applicable.

   ☑ Individual Provider Level Data

   What is the purpose of the linkage:

   The purpose of linking MA APCD data to individual provider level data is to obtain additional information on provider characteristics, which will be used to (a) select provider samples based on their characteristics, (b) control for provider characteristics, and (c) examine heterogeneity by provider characteristics. For example, I plan to use physician characteristics, including age, gender, medical school, medical school graduation year, primary and secondary specialty, practice type, group practice affiliations, hospital affiliations, patient volume, insurance plans accepted, acceptance of new patients, and hours of operation. Please see the Research Methodology for additional information.

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

Databases

- CMS National Plan and Provider Enumeration System (NPPES) Files: Publicly available datasets
- American Medical Association (AMA) Physician Masterfile: Publicly available datasets to be licensed from AMA Physician Masterfile licensee
- SK&A Physician Database: Publicly available datasets to be licensed from SK&A
- State Boards and License Verification (e.g., Commonwealth of Massachusetts Board of Registration in Medicine, Massachusetts Health Care Safety & Quality License Verification Site): Publicly available information from websites
- CMS Unique Physician Identification Number (UPIN) Directory: Publicly available datasets to be purchased from Centers for Medicare & Medicaid Services (CMS); supplemental since data are available for only 2003-2007
- CMS Unique Physician Identification Number (UPIN) Group File: Publicly available datasets to be purchased from Centers for Medicare & Medicaid Services (CMS); supplemental since data are available for only 2003-2007
- Other similar, publicly available datasets

Data Elements for Linkage

- Medical Claims: National Provider ID (MC026), Provider Characteristics (MC027, MC032, MC036-MC037), Provider Location (MC034-MC035)
- Pharmacy Claims: National Provider ID (PC021, PC048), Pharmacy Location (PC023-PC024), Provider Location (PC054-PC055)
- Dental Claims: National Provider ID (DC020), Provider Characteristics (DC021, DC026, DC030), Provider Location (DC028-DC029)
- Member Eligibility File: Health Care Home National Provider ID (ME038), Attributed PCP Provider ID (ME124), Organization ID (ME125)
- Provider File: National Provider ID (PV039-PV040), Provider Characteristics (PV015)

☑ Individual Facility Level Data

What is the purpose of the linkage:

The purpose of linking MA APCD data to individual facility level data is to obtain additional information on healthcare provider organization characteristics, which will be used to (a) control for healthcare provider organization characteristics, and (b) examine heterogeneity by healthcare provider organization characteristics. For example, I plan to use hospital characteristics including size, services offered, inpatient and outpatient utilization, and physician arrangements. Please see the Research Methodology for additional information.

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

Databases

- CMS National Plan and Provider Enumeration System (NPPES) Files: Publicly available datasets
- CMS Provider of Services (POS) File: Publicly available datasets
- American Hospital Association (AHA) Annual Survey Database: Publicly available datasets to be licensed from AHA
- SK&A Physician Database: Publicly available datasets to be licensed from SK&A
- CMS Unique Physician Identification Number (UPIN) Group File: Publicly available datasets to be purchased from Centers for Medicare & Medicaid Services (CMS); supplemental since data are available for only 2003-2007
- Other similar, publicly available datasets

Data Elements for Linkage

- Medical Claims: National Provider ID (MC026), Provider Characteristics (MC027, MC032, MC036-MC037), Provider Location (MC034-MC035)

- Pharmacy Claims: National Provider ID (PC021, PC048), Pharmacy Location (PC023-PC024), Provider Location (PC054-PC055)
- Dental Claims: National Provider ID (DC020), Provider Characteristics (DC021, DC026, DC030), Provider Location (DC028-DC029)
- Member Eligibility File: Health Care Home National Provider ID (ME038), Attributed PCP Provider ID (ME124), Organization ID (ME125)
- Provider File: National Provider ID (PV039-PV040)

☑ Aggregate Data

What is the purpose of the linkage:

The purpose of linking MA APCD data to aggregate data is to obtain additional information on characteristics of patient geographic areas (e.g., income, education) and characteristics of physician practice markets and geographic areas (e.g., based on residents in health care markets). Information on patient geographic areas will be used to (a) control for patient characteristics, and (b) examine heterogeneity by patient characteristics. Information on physician practice markets and geographic areas will be used to (a) control for provider and healthcare provider organization characteristics, and (b) examine heterogeneity by provider and healthcare provider organization characteristics. Please see the Research Methodology for additional information.

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

Databases

- US Census: Publicly available datasets
- Internal Revenue Service (IRS) Statistics of Income (SOI) Individual Income Tax Statistics ZIP Code Data: Publicly available datasets
- American Community Survey (ACS): Publicly available datasets
- Current Population Survey (CPS): Publicly available datasets
- Consumer Expenditure Survey (CE): Publicly available datasets
- National Association of Insurance Commissioners (NAIC): Publicly available datasets
- External Source Codes listed in the Massachusetts All Payer Claims Database Documentation Guide Appendix, which include Countries (American National Standards Institute); States and Other Areas of the US (US Postal Service); National Provider Identifiers (National Plan and Provider Enumeration System; Department of Health and Human Services; Centers for Medicare and Medicaid Services (CMS)); Provider Specialties (CMS); Health Care Provider Taxonomy (Washington Publishing Company); North American Industry Classification System (US Census Bureau); Language Preference (US Census Bureau); International Classification of Diseases (ICD) 9 and 10 (American Medical Association); Healthcare Common Procedure Coding System (HCPCS), Current Procedural Terminology (CPT), and Modifiers (American Medical Association); Dental Procedure Codes and Identifiers (American Dental Association); Logical Observation Identifiers Names and Codes (Regenstrief Institute); National Drug Codes and Names (US Food and Drug Administration); Standard Professional Billing Elements (CMS); Standard Facility Billing Elements (National Uniform Billing Committee); Diagnosis-Related Group (DRG), Ambulatory Payment Classification (APC), and Present on Admission (POA) Codes (CMS); Claim Adjustment Reason Codes (Washington Publishing Company); and Race and Ethnicity Codes (Centers for Disease Control and Prevention): Publicly available datasets
- Other similar, publicly available datasets

Data Elements for Linkage

- Medical Claims: Member Address (Derived-MC6-Derived-MC7, MC015); Member Gender (MC012), Member Age (Derived-MC18), Provider Location (MC034-MC035), Submission Date (Derived-MC1-Derived-MC2), CHIA Incurred Date (Derived-MC9), Date of Service (MC059-MC060), Admission Date (MC018), Discharge Date (MC069)
- Pharmacy Claims: Member Address (Derived PC6-Derived-PC7, PC015); Member Gender (PC012),

Member Age (Derived-PC14), Pharmacy Location (PC023-PC024), Provider Location (PC054-PC055), Submission Date (Derived-PC1-Derived-PC2), CHIA Incurred Date (Derived-PC10), Prescription Written Date (PC064), Prescription Filled Date (PC032)

- Dental Claims: Member Address (Derived-DC6-Derived-DC7, DC015), Member Gender (DC012), Member Age (Derived-DC13), Provider Location (DC028-DC029), Submission Date (Derived-DC1-Derived-DC2), CHIA Incurred Date (Derived-DC9), Date of Service (DC035-DC036)
- Member Eligibility File: Member Address (Derived-ME6, Derived-ME8, ME016), Subscriber Address (Derived-ME7, Derived-ME9, ME109), Member Gender (ME013), Member Age (Derived-ME15), Benefit Status (ME063, ME066), Coverage (ME018-ME020, ME029-ME031, ME081, ME118), Submission Date (Derived-ME1-Derived-ME2), Product Enrollment Dates (ME041-ME042), Member Benefit Plan Contract Enrollment Dates (ME129-ME130), Member PCP Dates (ME047-ME048)
- Product File: Product Characteristics (PR003-PR006), Product Dates (PR009-PR010); NAIC Code (PR017)

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.

Individual Provider and Facility Level Data Linkages

- CMS National Plan and Provider Enumeration System (NPPES) Files: Deterministic using unique provider identifiers (e.g., National Provider ID (NPI)) and then probabilistic for additional matches using potentially non-unique provider identifiers (e.g., healthcare provider organization name, provider name)
- CMS Provider of Services (POS) File: Deterministic using unique provider identifiers (e.g., provider number) and then probabilistic for additional matches using potentially non-unique provider identifiers (e.g., healthcare provider organization name, healthcare provider organization address)
- CMS Unique Physician Identification Number (UPIN) Directory: Deterministic using unique provider identifiers (e.g., Unique Physician Identification Number (UPIN)) and then probabilistic for additional matches using potentially non-unique provider identifiers (e.g., provider name, provider address)
- CMS Unique Physician Identification Number (UPIN) Group File: Deterministic using unique provider identifiers (e.g., Group Unique Physician Identification Number (UPIN)) and then probabilistic for additional matches using potentially non-unique provider identifiers (e.g., healthcare provider organization name, healthcare provider organization address)
- American Hospital Association (AHA) Annual Survey Database: Deterministic using unique provider identifiers (e.g., Medicare ID) and then probabilistic for additional matches using potentially non-unique provider identifiers (e.g., healthcare provider organization name, healthcare provider organization address)
- American Medical Association (AMA) Physician Masterfile: Deterministic using unique provider identifiers (e.g., National Provider ID (NPI), Unique Physician Identification Number (UPIN)) and then probabilistic for additional matches using potentially non-unique provider identifiers (e.g., provider name, provider address)
- SK&A Physician Database: Deterministic using unique provider identifiers (e.g., National Provider ID (NPI), Unique Physician Identification Number (UPIN)) and then probabilistic for additional matches using potentially non-unique provider identifiers (e.g., provider name, provider address)
- State Boards and License Verification (e.g., Commonwealth of Massachusetts Board of Registration in Medicine, Massachusetts Health Care Safety & Quality License Verification Site): Deterministic using unique provider identifiers (e.g., license state and number) and then probabilistic for additional matches using potentially non-unique provider identifiers (e.g., provider name)

Aggregate Data Linkages

- US Census: Deterministic if possible (and otherwise, then probabilistic) for subscriber/member/patient geographical area (e.g., subscriber ZIP code, member ZIP code, employer ZIP code) and demographics (e.g., gender, age) linkage; Deterministic if possible (and otherwise, then probabilistic) for provider geographical area (e.g., provider ZIP code, provider census tract, provider census block) linkage

- Internal Revenue Service (IRS) Statistics of Income (SOI) Individual Income Tax Statistics ZIP Code Data: Deterministic if possible (and otherwise, then probabilistic) for subscriber/member/patient geographical area (e.g., subscriber ZIP code, member ZIP code, employer ZIP code) and demographics (e.g., gender, age) by time (e.g., year) linkage; Deterministic if possible (and otherwise, then probabilistic) for provider geographical area (e.g., provider ZIP code) by time (e.g., year) linkage
- American Community Survey (ACS): Deterministic if possible (and otherwise, then probabilistic) for subscriber/member/patient geographical area (e.g., subscriber ZIP code, member ZIP code, employer ZIP code) and demographics (e.g., gender, age) by time (e.g., year) linkage; Deterministic if possible (and otherwise, then probabilistic) for provider geographical area (e.g., provider ZIP code, provider census tract, provider census block) by time (e.g., year) linkage
- Current Population Survey (CPS): Deterministic if possible (and otherwise, then probabilistic) for geographical area (e.g., state and metropolitan statistical areas (MSAs)) by time (e.g., year) linkage; Probabilistic for demographics (e.g., gender, age) by time (e.g., year) linkage
- Consumer Expenditure Survey (CE): Deterministic if possible (and otherwise, then probabilistic) for geographical area (e.g., state) by time (e.g., year) linkage and probabilistic for demographics (e.g., gender, age) by time (e.g., year) linkage

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

I will not link CHIA data with individual patient level data, so linked datasets will not substantially increase the risk of identification of individual patients. The linked individual provider level data, individual facility level data, and aggregate data do not provide additional patient level, visit level, or procedure level information. Furthermore, as discussed in the Data Management Plan, I will establish and maintain appropriate administrative, technical, and physical safeguards to protect the confidentiality of the data.

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.

   I plan to share my research findings at multiple stages of my project. Unpublished results may be shared in seminars, conferences, workshops, and working papers. I will later include papers in my dissertation and publish papers in peer-reviewed academic journals in economics. All disclosures of CHIA data will not include any output with cell sizes of less than 11 observations.

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 my analysis will be publicly available. My dissertation will be freely available from the UC Berkeley Libraries, and my publications will be available from peer-reviewed academic journals.

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.)?

   Not applicable.

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

   Not applicable.

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:	None
Contact Person:	None
Title:	Not applicable
Address:	Not applicable
Telephone Number:	Not applicable
E-mail Address:	Not applicable
Organization Website:	Not applicable

1. 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.

2. Describe the tasks and products assigned to this agent or contractor for this project.
   Not applicable.

3. Describe the qualifications of this agent or contractor to perform such tasks or deliver such products.
   Not applicable.

4. Describe your oversight and monitoring of the activity and actions of this agent or subcontractor.
   Not applicable.

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	
Title	
Original Data Request Submission Date	
Dates Data Request Revised	