

# **Massachusetts Division of Health Care Finance and Policy**

## **Application for All-Payer Claims Database (APCD) Data**

Applications for APCD data must meet the requirements set forth in regulation **114.5 CMR 22.00: Health Care Claims Data Release** and any Administrative Bulletins promulgated under this regulation.

The regulation and bulletins are available online at

<http://www.mass.gov/eohhs/gov/departments/hcf/regulations.html>.

Information provided on pages 1-4 of this application will be posted on the internet for public comment.

<b>A. APPLICANT INFORMATION</b>	
Principal Investigator:	Jonathan T. Kolstad (primary applicant)
Title:	Assistant Professor of Health Care Management and Faculty Research Fellow
Organization:	University of Pennsylvania and the National Bureau of Economic Research
Co-Investigator:	Amanda E. Kowalski
Title:	Assistant Professor of Economics and Faculty Research Fellow
Organization:	Yale University and the National Bureau of Economic Research
Co-Investigator:	Michael Punzalan
Title:	Doctoral Student
Organization:	University of Pennsylvania
Project Title:	The Effects of Fragmentation in Health Care
Date of Application:	August 31, 2012
Date of Revision:	December 12, 2012
Brief Description of Project (240 character limit)	The project will evaluate the extent to which organizational fragmentation affects the division of labor among providers by affecting incentives.

**Note 1:** Revisions in this application were made in response to Committee suggestions from the November monthly meeting. Following prior instructions from APCD staff, we have highlighted our responses in yellow.

**Note 2:** For ease of reading, the prior set of responses to Committee feedback from the October monthly meeting has been removed from the body of the application. This text has been collected in an Addendum and is posted at the end of our application. Since the extra formatting is unnecessary, we have removed the yellow highlighting that originally marked this text.

## B. DATA REQUESTED

### 1. PUBLIC USE

File	SINGLE USE* '08 – '09 – '10	REPEATED USE* '08 – '09 – '10	MULTIPLE USE* '08 – '09 – '10
Medical Claims	x x x	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Pharmacy Claims	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Dental Claims	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Membership Eligibility	x x x	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Provider	x x x	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Product	x x x	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

### 2. RESTRICTED USE

File	SINGLE USE* '08 – '09 – '10	REPEATED USE* '08 – '09 – '10	MULTIPLE USE* '08 – '09 – '10
Medical Claims	x x x	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Pharmacy Claims	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Dental Claims	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Membership Eligibility	x x x	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Provider	x x x	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Product	x x x	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

\* The Division reserves the right to change proposed “use level” after review of this application.

#### Definitions:

- **Single Use:** Use of the data for a project or study.
- **Repeated Use:** Use of the data as an input to develop a report or product for sale to multiple clients or customers provided that it will NOT disclose APCD data. Examples include: development of a severity index tool, development of a reference tool used to inform multiple consulting engagements where no APCD data is disclosed.
- **Multiple Use:** Use of the data to develop a product or service that will be sold in the marketplace and will disclose APCD data. Examples include: a benchmark report produced by analyzing APCD data, a query tool to ease access to APDC data.

3. **Filters:** If you are requesting data elements from the Restricted Use dataset, 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 products.)

File	Data Element(s)	Range of Values Requested
Medical Claims		
Pharmacy Claims		
Dental Claims		
Membership Eligibility		

Provider		
Product		

4. **Restricted data elements:** If you are requesting Data Elements from the Restricted Use dataset, list each restricted data element you are requesting on the attached Data Element List and explain why you need access to EACH Restricted Use data element for your project. Limit your request to the minimum data elements necessary to complete the project and be specific as to how each element relates to your proposed model/analytic plan. Add rows to this table as needed.

Notes:

- 1.) To our knowledge, formal data element descriptions are only available in the data submission guides.<sup>1</sup> We have copied these descriptions from the guides to our application to comply with the application instructions and to maintain consistency with APCD documentation. We understand that each restricted data element we would receive may differ from this description, e.g. MC013 is entitled “Member Birth Month” but the formal description is “Member/Patient’s date of birth.” Though variable masking is not indicated in variable names, we understand which variables would be received as masked, e.g. PV035 “SSN Id.” We will revise our application if there is a set of descriptions that more precisely matches the restricted data element list.

Restricted Data Element Name	Restricted Data Element Description	Data File (Medical, Pharmacy, Dental, Eligibility, Provider, Product)	Justification (reason this data element is necessary for your project)
Payer (MC001)	Carrier Specific Submitter Code as defined by APCD	Medical	Provider incentives vary by payer. Since we are studying the interaction between organizational form and provider incentives, it is important to distinguish among payers in the data.
National Plan ID (MC002)	CMS National Plan Identification Number (PlanID)	Medical	Provider incentives vary by plan. Since we are studying the interaction between

---

<sup>1</sup> <http://www.mass.gov/eohhs/researcher/physical-health/health-care-delivery/hcf-data-resources/apcd/user-resources.html>

			organizational form and provider incentives, it is important to distinguish among plans in the data.
Insurance Type Code/Product (MC003)	Type / Product Identification Code	Medical	Provider incentives vary by insurance type. For example, patients with HMO coverage present different incentives than patients with PPO coverage. Since we are studying the interaction between organizational form and provider incentives, it is important to distinguish among plans in the data.
Payer Claim Control Number (MC004)	Payer Claim Control Identification	Medical	Since we want to measure activity by claims, it is important to be able to distinguish unique claims in the data.
Insured Group or Policy Number (MC006)	Carriers group or policy number	Medical	Patient populations may vary systematically across groups. It is important to be able to distinguish unique groups in the data.
Individual Relationship Code (MC011)	Member/Patient to Subscriber Relationship Code	Medical	Member incentives will vary with relationships to subscribers, e.g. individual versus out-of-pocket maximum, and these incentives will influence the outcome of patient visits. Hence, it is important to account for relationships between members in the data.

Member Birth Month (MC013)	Member/Patient's date of birth	Medical	Choice behavior varies significantly with age and is a frequently-used conditioning variable included in social science analysis. We would like to condition our estimates on this variable as well.
Member city name (MC014)  The DRC recommended we request 5-digit ZIP code. However, APCD staff informed us that 5-digit ZIP code is not available due to HIPAA. Therefore our request for member city name still stands.	City name of the Member/Patient	Medical	Empirical studies show that travel time affects hospital choice by patients (all else equal, patients prefer hospitals closer to their homes). The same preference may hold for choice of specialist. This is a potential confounder in our study. This fact strongly indicates that we need to control for patient location.
Member ZIP Code (MC017)  The DRC recommended we request 5-digit ZIP code. However, APCD staff informed us that 5-digit ZIP code is not available due to HIPAA. Therefore our request for 3-digit ZIP code still stands.	Zip Code of the Member/Patient	Medical	Empirical studies show that travel time affects hospital choice by patients (all else equal, patients prefer hospitals closer to their homes). The same preference may hold for choice of specialist. This is a potential confounder in our study. This fact strongly indicates that we need to control for patient location.
Admission Date (MC018)	Inpatient Admit Date	Medical	Our study examines successive patient-provider encounters. It is essential that we accurately observe service dates.
Admission Hour (MC019)	Admission Time	Medical	Our study examines successive patient-

			provider encounters. It is essential that we accurately observe service dates.
Discharge Hour (MC022)	Discharge Time	Medical	Our study examines successive patient-provider encounters. It is essential that we accurately observe service dates.
Service Provider Number (MC024)	Service Provider Number	Medical	It is critical for our study that we match providers to practices. Therefore distinguishing among providers in the data and linking observations to the Provider data file is essential.
Date of Service - From (MC059)	Date of Service	Medical	Our study examines successive patient-provider encounters. It is essential that we accurately observe service dates.
Date of Service - To (MC060)	Date of Service	Medical	Our study examines successive patient-provider encounters. It is essential that we accurately observe service dates.
Patient Control Number (MC068)	Patient Control Number	Medical	Our study examines successive patient-provider encounters. It is essential that we accurately observe service dates.
Discharge Date (MC069)	Discharge Date	Medical	Our study examines successive patient-provider encounters. It is essential that we accurately observe service dates.
Billing Provider Number (MC076)	Billing Provider Number	Medical	We value this variable to the extent that it aids in distinguishing

			among providers and practices.
Product ID Number (MC079)	Product Identification Number	Medical	Member incentives will vary by product. Hence, it is important to account for variation in products among otherwise similar members in the data.
Reason for Adjustment (MC080)	Reason for Adjustment Code	Medical	We are using claims data to observe activity. If the claim has been adjusted it may be important for us to account for this fact.
Allowed amount (MC098)	Allowed Amount	Medical	The allowed amount is informative for provider incentives. Therefore it is important we account for it.
Referring Provider ID (MC112)	Referring Provider Number	Medical	Referral patterns reveal the incentive effects of organizational fragmentation. This could be a critical variable for our work (though the study could be implemented if this variable is not well populated).
Denial Reason (MC124)	Denial Reason Code	Medical	Since we want to measure activity by claims, it is important to be able to include the proper claims.
Attending Provider (MC125)	Attending Provider	Medical	It is critical for our study that we match providers to practices. Therefore distinguishing among providers in the data and linking observations to the

			Provider data file is essential.
Plan Rendering Provider Identifier (MC134)	Plan Rendering Number	Medical	It is critical for our study to distinguish among providers in the data.
Provider Location (MC135)	Location of Provider	Medical	It is critical for our study to distinguish among providers and locations in the data.
CarrierSpecificUniqueMemberID (MC137)	Member/Patient Carrier Unique Identification	Medical	It is critical for our study to distinguish among members in the data.
Former Claim Number (MC139)	Former Claim Number	Medical	Since we want to measure activity by claims, it is important to be able to include the proper claims.
CarrierSpecificUniqueSubscriberID (MC141)	Member/Patient Carrier Unique Identification	Medical	It is critical for our study to distinguish among members in the data.
Payer (ME001)	Carrier Specific Submitter Code as defined by APCD. This must match the Submitter Code reported in HD002.	Eligibility	Provider incentives vary by payer. Since we are studying the interaction between organizational form and provider incentives, it is important to distinguish among payers in the data.
National Plan ID (ME002)	CMS National Plan Identification Number (PlanID)	Eligibility	Provider incentives vary by plan. Since we are studying the interaction between organizational form and provider incentives, it is important to distinguish among plans in the data.
Insured Group or Policy Number (ME006)	Carriers group or policy number	Eligibility	Patient populations may vary systematically across

			groups. It is important to be able to distinguish unique groups in the data.
Individual Relationship Code (ME012)	Member/Patient to Subscriber Relationship Code	Medical	Member incentives will vary with relationships to subscribers, e.g. individual versus out-of-pocket maximum, and these incentives will influence the outcome of patient visits. Hence, it is important to account for relationships between members in the data.
Member Birth Month (ME014)	Member's date of birth	Eligibility	Choice behavior varies significantly with age and is a frequently-used conditioning variable included in social science analysis. We would like to condition our estimates on this variable as well.
Member City Name (ME015)  The DRC recommended we request 5-digit ZIP code. However, APCD staff informed us that 5-digit ZIP code is not available due to HIPAA. Therefore our request for member city name still stands.	City name of the member	Eligibility	Empirical studies show that travel time affects hospital choice by patients (all else equal, patients prefer hospitals closer to their homes). The same preference may hold for choice of specialist. This is a potential confounder in our study. This fact strongly indicates that we need to control for patient location.
Member ZIP Code (ME107)  The DRC recommended we request 5-digit ZIP code. However, APCD staff informed us that 5-digit ZIP code is not	Zip Code of the Member	Eligibility	Empirical studies show that travel time affects hospital choice by patients (all else equal, patients prefer hospitals closer to

available due to HIPAA. Therefore our request for 3-digit ZIP code still stands.			their homes). The same preference may hold for choice of specialist. This is a potential confounder in our study. This fact strongly indicates that we need to control for patient location.
Special Coverage (ME031)	Special Coverage Code	Eligibility	Patient behavior may vary systematically with type of health insurance purchase and, therefore, affect encounter outcomes. It is important we account for this in our estimation.
Health Care Home Number (ME036)	Health Care Home Number	Eligibility	Since we are studying behavior across types of organizational form, it is important to distinguish entities among this important type of organization.
Product ID Number (ME040)	Product Identification Number	Eligibility	Member incentives will vary by product. Hence, it is important to account for variation in products among otherwise similar members in the data.
Product Enrollment Start Date (ME041)	the date the member was enrolled in the product	Eligibility	Member incentives will vary by product. Hence, it is important to know when a member was covered by a given product.
Product Enrollment End Date (ME042)	Enrollment Date	Eligibility	Member incentives will vary by product. Hence, it is important to know when a member was covered by a given product.
Member PCP ID (ME046)	Member's PCP Identification	Eligibility	The theory we would like to test particularly

	Number		concerns primary care physicians. Linking members to PCPs is critical to our project.
NewMMISID (ME075)	NewMMIS Identification Number	Eligibility	It is critical for our study to distinguish among patients in the data.
Member rating category (ME076)	Member Rating Category Code	Eligibility	Choices may differ systematically among individuals with different (estimated) risk types and may affect encounter outcomes. Therefore, it is important we account for these differences.
Recipient Identification Number (ME079)	MassHealth RID Number	Eligibility	It is critical for our study to distinguish among patients in the data.
Recipient Historical Number (ME080)	MassHealth RHN Number	Eligibility	It is critical for our study to distinguish among patients in the data.
CarrierSpecificUniqueMemberID (ME107)	Member/Patient Carrier Unique Identification	Eligibility	It is critical for our study to distinguish among members in the data.
Subscriber City Name (ME108)  The DRC recommended we request 5-digit ZIP code. However, APCD staff informed us that 5-digit ZIP code is not available due to HIPAA. Therefore our request for subscriber city name still stands.	City name of the Subscriber	Eligibility	Empirical studies show that travel time affects hospital choice by patients (all else equal, patients prefer hospitals closer to their homes). The same preference may hold for choice of specialist. This is a potential confounder in our study. This fact strongly indicates that we need to control for patient location.
Subscriber ZIP Code (ME110)	Zip Code of the Subscriber	Eligibility	Empirical studies show that travel time affects

The DRC recommended we request 5-digit ZIP code. However, APCD staff informed us that 5-digit ZIP code is not available due to HIPAA. Therefore our request for 3-digit ZIP code still stands.			hospital choice by patients (all else equal, patients prefer hospitals closer to their homes). The same preference may hold for choice of specialist. This is a potential confounder in our study. This fact strongly indicates that we need to control for patient location.
CarrierSpecificUniqueSubscriberID (ME117)	Subscriber Carrier Unique Identification	Eligibility	It is important for our study to distinguish among subscribers in the data.
Product ID number (PR001)	Product Identification Number	Product	Provider incentives vary with product. Since we are studying the interaction between organizational form and provider incentives, it is important to distinguish among products in the data.
Carrier License Type (PR003)	Carrier License Type	Product	Both provider incentives and patient characteristics may vary systematically with license type. Therefore it is important to distinguish payers by license type in the data.
Product Line of Business Model (PR004)	The Line of Business / Insurance Models the Product relates to	Product	Both provider incentives and patient characteristics may vary systematically with product line. Therefore it is important to distinguish product

			lines in the data.
Insurance Plan Market (PR005)	Insurance Plan Market Code	Product	Both provider incentives and patient characteristics may vary systematically with insurance plan market. Therefore it is important to distinguish markets in the data.
Other Product Benefit Description (PR007)	Benefit Description	Product	It is important to observe all product attributes that affect incentives or are correlated with behavior.
Product Start Date (PR009)	Product Start Date	Product	Timing of enrollment could be indicative of underlying risk type, a common concern in empirical studies in economics. Hence, knowing when an individual could have enrolled in a product and when he actually enrolled in a product could be important for estimation.
Product End Date (PR010)	Last date on which members could be enrolled in this product	Product	Timing of enrollment could be indicative of underlying risk type, a common concern in empirical studies in economics. Hence, knowing when an individual could have enrolled in a product and when he actually enrolled in a product could be important for estimation.
Payer (PV001)	Carrier Specific Submitter Code as defined by APCD. This must match the	Provider	Since we are primarily interested in providers and practices, it is essential we can match

	Submitter Code reported in HD002.		claims to providers.
Plan Provider ID (ME002)	Carrier Unique Provider Code	Provider	It is critical for our study to distinguish among providers in the data.
Provider DOB Year (PV015)	Provider's date of birth	Provider	Practice patterns have been shown to correlate with age. This strongly suggests we need this demographic variable.
Street Address1 Name (PV016)	Street address of the Provider	Provider	A critical component of our study is matching providers to practices. Detailed geographic information is an important component of our matching strategy.
Street Address2 Name (PV017)	Secondary Street Address of the Provider	Provider	A critical component of our study is matching providers to practices. Detailed geographic information is an important component of our matching strategy.
SSN Id (PV035)	Provider's Social Security Number	Provider	It is critical for our study to distinguish among providers in the data.
Medicare Id (PV036)	Provider's Medicare Number	Provider	It is critical for our study to distinguish among providers in the data.
Medical/Healthcare Home ID (PV054)	Medical Home Identification Number	Provider	Provider incentives may vary with organizational structure. It would be important to account for this particular organizational form.
Provider Affiliation (PV056)	Provider Affiliation Code	Provider	Provider incentives may vary with affiliations. It would be important to

			account for this during estimation.
--	--	--	-------------------------------------

Members of the DRC suggested at the November monthly meeting that we request the following elements in our revised application. We have set apart these requests for readability.

Based on the documentation, our request consists of all fields containing NPI.<sup>2</sup> However, please advise if there are other fields containing NPI that we are not aware of and we will revise the application.

Data Element Name	Description	Data File (Medical, Pharmacy, Dental, Eligibility, Provider, Product)	Justification (reason this data element is necessary for your project)
National Provider ID (PV039)	<p>Based on DRC feedback, we request an <b>UNMASKED</b> version of this variable.</p> <p>This variable is <b>MASKED</b> according to the "All Payer Claims Dataset (APCD) Data Element List" (accessed December 12, 2012).</p>	Provider	Following a DRC member suggestion, inclusion of this variable will substantively improve the quality of our physician matching procedure.
National Provider2 ID (PV040)	<p>Based on DRC feedback, we request an <b>UNMASKED</b> version of this variable.</p> <p>This variable is <b>MASKED</b> according to the "All Payer Claims Dataset</p>	Provider	Following a DRC member suggestion, inclusion of this variable will substantively improve the quality of our physician matching procedure.

<sup>2</sup> According to <http://www.mass.gov/chia/docs/p/apcd/apcd-provider-file-submission-guide-2012-12-01.pdf> and <http://www.mass.gov/chia/docs/p/apcd/apcd-medical-claim-file-submission-guide-12-2010.pdf>.

	(APCD) Data Element List" (accessed December 12, 2012).		
National Service Provider ID (MC026)	<p>Based on DRC feedback, we request an <b>UNMASKED</b> version of this variable.</p> <p>This variable is <b>MASKED</b> according to the "All Payer Claims Dataset (APCD) Data Element List" (accessed December 12, 2012).</p>	Medical	Following a DRC member suggestion, inclusion of this variable will substantially improve the quality of our physician matching procedure.
National Billing Provider ID (MC077)	<p>Based on DRC feedback, we request an <b>UNMASKED</b> version of this variable.</p> <p>This variable is <b>MASKED</b> according to the "All Payer Claims Dataset (APCD) Data Element List" (accessed December 12, 2012).</p>	Medical	Following a DRC member suggestion, inclusion of this variable will substantially improve the quality of our physician matching procedure.

## C. PURPOSE AND INTENDED USE

1. Please describe the purpose of your project and how you will use the APCD.

Our goal is to determine how and in what measure organizational fragmentation affects the division of labor among providers. Any study of this topic requires observations of patient visits, diagnoses, and service providers. The MA APCD captures this information – as well as much other important information – and will be the primary source of data for our statistical analysis. We provide a more detail-oriented discussion in our response to item C3.

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

Organizational fragmentation in health care has long been a subject of concern for health care policy. It was recently addressed in the Patient Protection and Affordable Care Act (PPACA). The PPACA mandates that the federal executive branch design incentive programs to encourage providers to form Accountable Care Organizations (ACOs). The hope is that formation of these ACOs will lead to better patient outcomes and more efficient use of services. To the best of our knowledge, there is little evidence on what the actual effects of these provider integration incentives will be, and there certainly are few if any estimates based on claims data from recent years. The goal of our project is to fill gaps in this evidence base and provide sound empirical guidance for future policy.

3. **Attach** a brief (1-2 pages) description of your research methodology. (This description will not be posted on the internet.)
4. 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 N/A.
  - No, this project is not subject to IRB review
  - No, my organization does not have an IRB

## D. APPLICANT QUALIFICATIONS

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

Jonathan T. Kolstad holds a PhD in Health Policy (Economics) from Harvard University and a BA in Economics from Stanford University. He is an Assistant Professor of Health Care Management at The Wharton School of the University of Pennsylvania, teaching courses in health economics. He is also a faculty research fellow at the National Bureau of Economic Research.

Amanda E. Kowalski holds a PhD in Economics from the Massachusetts Institute of Technology and an AB in Economics from Harvard University. She is an Assistant Professor of Economics at Yale University, teaching courses in public finance and health economics. From June 2011 to August 2012 she served as the Okun Model Early Career Fellow at the Brookings Institution, a think tank in Washington, DC. She is also a faculty research fellow at the National Bureau of Economic Research.

Michael Punzalan holds an MSc in Economics from the London School of Economics and Political Science and a BS in Economics and Finance from New York University. He is a Doctoral Student in the Department of Health Care Management at The Wharton School of the University of Pennsylvania.

Jointly our research experience and advanced theoretical training in health economics qualify us to execute the project described above. We are trained in cutting-edge econometric methodologies, knowledgeable of the important institutional details of health care provision, and possess vast experience working with de-identified medical claims data from several sources, public and private.

2. Describe the software you plan to use to analyze the data and the experience that the applicant's team members have in using that software.

We will analyze the data primarily with Stata. If custom statistical coding is necessary we will use a program such as Matlab. We have extensive experience with these programs. Articles based on our work with these programs have been published in peer-reviewed economics journals.

3. 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.)

For reference, the requested information is attached for Jonathan T. Kolstad (PI), Amanda E. Kowalski (Co-PI), Michael Punzalan (Co-PI), Mohan Ramanujan (IT administrator), and Daniel Feenberg (IT administrator).

Research assistants may be hired as the project progresses and they will be required to submit signed confidentiality statements before they can access the data.

## E. DATA LINKAGE AND FURTHER DATA ABSTRACTION

1. Does your project require linking the APCD to another dataset?  
YES  NO
2. If yes, will the APCD be linked to other patient level data or with aggregate data (e.g. Census data)?  
Patient Level Data  Aggregate Data  Other
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.

To complete this project, we need to link providers in the APCD and the SK&A Office-Based Physicians database. The SK&A database is a privately collected census of US practices and providers. The linkage is necessary because the theory we would like to investigate requires extremely credible observations of about each provider's membership in a single- or multi-specialty practice. In theory it may be possible to construct this information from the APCD data. But, the APCD "User Submission Guides" suggest that it will be very difficult to do this using the APCD alone. The recent Quality Assurance webinar suggests that further data validation and cleaning may be necessary before researchers forego certain external sources of data. As a matter of feasibility and because the SK&A data is specifically collected for documenting health care organizations and associated providers, we believe it is most sound to link to this external data set.

Per the recommendation given by a DRC member at the November monthly meeting, we have amended our prospected match procedure from "address only" to "unmasked NPI + provider last name + provider street address."

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

Linking to the provider database does not increase the chance that individual patients will be identified. We are only adding provider membership in practices and characteristics of those practices to the data.

## F. RE-RELEASE OF DATA

Applicants must obtain prior approval from the Division to publish reports that use APCD files. Applicants must provide the Division with a copy of any report at least 30 days prior to release to outside parties, including peer review and prepublication analysis by anyone other than the individuals named in this Application. The Division will review the report to ensure that the publication will not permit identification of an individual patient or permit identification of a specific payment by individual payer. The Division may prohibit release of reports that may permit identification of individual patients or specific payment by individual payer.

1. Describe your plans to publish or otherwise disclose any APCD data elements, or any data derived or extracted from such data, in any paper, report, website, statistical tabulation, or similar document.

Our analysis will be contained in an academic paper. The paper will follow the norms of our academic discipline. Therefore, tables of summary statistics of the most critical variables are always provided. Test statistics for hypothesis testing will also be provided. Simulations based on a stylized model may also appear, depending on the results of the hypothesis testing.

Hence, the lowest level of aggregation for statistical reporting will be among large subgroups in the data, making identification of patients or payments almost certainly impossible to identify.

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

Yes, the results of this research will be publicly available with no fee. We will make the paper and the statistical code available through our professional websites or email request.

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. If you have answered “yes” to questions 3, 4 or 5, please (i) describe the types of products, services or studies; (ii) estimate the number and types of clients for which the data will be used and (iii) describe any rerelease of data by your clients.

## **G. USE OF AGENTS OR CONTRACTORS**

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

Company Name:	
Contact Person:	
Title:	
Address:	
Telephone Number:	
Fax 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.