

Application for Massachusetts All-Payer Claims Data (Non-Government) [Exhibit A – Data Application]

I. INSTRUCTIONS

This form is required for all Applicants, Agencies, or Organizations, hereinafter referred to as “Organization”, except Government Agencies as defined in [957 CMR 5.02](#), requesting protected health information. All Organizations must also complete the [Data Management Plan](#), and attach it to this Application. The Application and the Data Management Plan must be signed by an authorized signatory. This Application and the Data Management Plan will be used by CHIA to determine whether the request meets the criteria for data release, pursuant to 957 CMR 5.00. Please complete the Application documents fully and accurately. Prior to receiving CHIA Data, the Organization must execute CHIA’s [Data Use Agreement](#). Organizations may wish to review that document prior to submitting this Application.

Before completing this Application, please review the data request information on CHIA’s website:

- [Data Availability](#)
- [Fee Schedule](#)
- [Data Request Process](#)

After reviewing the information on the website and this Application, please contact CHIA at apcd.data@state.ma.us if you have additional questions about how to complete this form.

The Application and all attachments must be uploaded to IRBNet. All Application documents can be found on the [CHIA website](#).

Information submitted as part of the Application may be subject to verification during the review process or during any audit review conducted at CHIA’s discretion.

Applications will not be reviewed until the Application and all supporting documents are complete and the required application fee is received.

A [Fee Remittance Form](#) with instructions for submitting the application fee is available on the CHIA website. If you are requesting a fee waiver, a copy of the [Fee Remittance Form](#) and any supporting documentation must be uploaded to IRBNet. Please be aware that if your research is funded and under that funding you are required to release raw data to the funding source, you may not receive CHIA Data.

II. FEE INFORMATION

1. Consult the most current [Fee Schedule](#) for All-Payer Claims Database data.
2. After reviewing the Fee Schedule, if you have any questions about the application or data fees, contact apcd.data@state.ma.us.
3. If you believe that you qualify for a fee waiver, complete and submit the [Fee Remittance Form](#) and attach it and all required supporting documentation with your application. Refer to the [Fee Schedule](#) (effective Feb 1, 2017) for fee waiver criteria.
4. Applications will not be reviewed until the application fee is received.
5. Data for approved Applications will not be released until the payment for the Data is received.

III. ORGANIZATION & INVESTIGATOR INFORMATION

Project Title:	Measuring Risk, Predictors, and Impact of Medical Errors and Malpractice Over Time
IRBNet Number:	2044500-1
Organization Requesting Data (Recipient):	Medical Professional Mutual Insurance Company (“Coverys”)
Organization Website:	https://www.coverys.com/
Authorized Signatory for Organization:	Scott Weber
Title:	Head of Strategy and Corporate Development
E-Mail Address:	sweber@coverys.com
Telephone Number:	443-836-6973
Address, City/Town, State, Zip Code:	One Financial Center, 13 th Floor, Boston, MA 02111
Data Custodian: (Individual responsible for organizing, storing, and archiving Data)	Christian Galvin
Title:	Manager, Network Administration
E-Mail Address:	cgalvin@coverys.com
Telephone Number:	(800) 225-6168
Address, City/Town, State, Zip Code:	One Financial Center, 13 th Floor, Boston, MA 02111
Primary Investigator (Applicant): (Individual responsible for the research team using the Data)	Thaochau Phan
Title:	Data Scientist
E-Mail Address:	tphan@coverys.com
Telephone Number:	408-250-4053
Address, City/Town, State, Zip Code:	One Financial Center, 13 th Floor, Boston, MA 02111
Names of Co-Investigators:	
E-Mail Addresses of Co-Investigators:	Chris Coulter (ccoulter@coverys.com)
	Sam Greenblatt (sgreenblatt@coverys.com)
	Michelle Fritz (mfritz@coverys.com)
	Will Eggbeer (weggbeer@coverys.com)

IV. PROJECT INFORMATION

IMPORTANT NOTE: Organization represents that the statements made below as well as in any study or research protocol or project plan, or other documents submitted to CHIA in support of the Data Application are complete and accurate and represent the total use of the CHIA Data requested. Any and all CHIA Data released to the Organization under an approved application may ONLY be used for the express purposes identified in this section by the Organization, and for no other purposes. Use of CHIA Data for other purposes requires a separate Data Application to CHIA **or** written request to CHIA, with approval being subject to CHIA’s regulatory restrictions and approval process. Unauthorized use is a material violation of your Organization’s Data Use Agreement with CHIA.

1. What will be the use of the CHIA Data requested? [Check all that apply]

Epidemiological Health planning/resource allocation Cost trends

- | | | |
|---|--|--|
| <input checked="" type="checkbox"/> Longitudinal Research | <input checked="" type="checkbox"/> Quality of care assessment | <input type="checkbox"/> Rate setting |
| <input checked="" type="checkbox"/> Reference tool | <input type="checkbox"/> Research studies | <input checked="" type="checkbox"/> Severity index tool (or other derived input) |
| <input type="checkbox"/> Surveillance | <input type="checkbox"/> Student research | <input checked="" type="checkbox"/> Utilization review of resources |
| <input type="checkbox"/> Inclusion in a product | <input type="checkbox"/> Other (describe in box below) | |

[Click here to enter text.](#)

2. Provide an abstract or brief summary of the specific purpose and objectives of your Project. This description should include the research questions and/or hypotheses the project will attempt to address, or describe the intended product or report that will be derived from the requested data and how this product will be used. Include a brief summary of the pertinent literature with citations, if applicable.

Coverys is a national medical malpractice liability insurer focused on helping the medical community address challenges around care delivery, medical errors, and risk management in our rapidly changing healthcare landscape. We seek to use the CHIA MA APCD data to study whether quality, utilization, and other metrics derived from medical and pharmacy claims data can be used to predict the likelihood and severity of medical malpractice claims and medical errors. Our aim is to better understand this relationship to 1) better assess providers' medical malpractice risk during underwriting and to 2) design and deliver more effective and targeted risk management programs for provider organizations.

1) Improving Underwriting of Medical and Insurance Risk: Accurate underwriting of medical liability is essential for creating a proper system of checks and balances on physicians, and incentivizing them to adopt best practices or proactively engage in risk management. Providers with a history of medical errors or indicators of poor patient safety could be priced higher in terms of medical liability coverage, thus creating an impetus for said provider to improve existing practices and operations to limit future increases in liability premiums. Conversely, providers with low risk of medical errors and high patient safety could be rewarded with lower premiums, motivating them to continue to engage in best practices regarding patient safety. Our research will enable us to better understand the factors associated with medical errors and to gain a more holistic view of current and historical provider performance and liability risk. This in turn will allow us to have a more nuanced and thorough view from which to price medical liability policies so as to limit the rate of false positives (i.e. accidentally punishing a well-performing provider with high liability cost) and false negatives (i.e. not pricing a poor-performing provider high enough) in our underwriting.

2) Advancing Risk Management Capabilities: Coverys has a long history of providing risk management services to our insured healthcare organizations. A typical risk management engagement involves analyzing an organization's medical liability claims in an effort to identify signals, or patterns, that can be used by healthcare organizations to improve patient safety and reduce malpractice risk. Because it can take many years for a malpractice claim to be processed and adjudicated, analyzing medical malpractice claims data results in a helpful, though dated, assessment of medical liability risk. Moreover, malpractice claims are only filed after a medical error has occurred whereas the goal of risk management is to preempt such errors before they can arise, so there are natural limits to how much information and action we can derive from malpractice claims alone. Our teams believe that medical claims can serve as an invaluable supplementary data source to overcome this issue and provide us with a more prospective means of managing risk, rather than addressing that risk with a provider only after the fact. We intend to research predictors of patient safety in order to proactively identify earlier on high-risk providers who can benefit from risk management as well as to monitor for new or growing clinical areas of concern in terms of medical error rates. These insights will enable risk managers to better collaborate with providers to address potential patient safety concerns and develop more customized risk management solutions for their specific healthcare operations.

Lastly, we provide a high-level overview of some of the questions and problems our research hopes to investigate below. A more detailed description can be found in the attached Research Methodology document that our team provided.

- 1) Comparison of risk-adjusted medical cost and adverse events rates between patients who have incurred a potential medical error with those without medical errors
- 2) Estimating the impact of diagnostic delays on patient mortality rates

- 3) Comparison of cost and quality performance between physicians with history of malpractice claims and risk management engagement versus those without such claims or engagements
- 4) Measuring the effect of provider's individual clinical proficiency on malpractice risk
- 5) Measuring the effect of service/procedure complexity on malpractice risk
- 6) Measuring organizational or physician network effects on an individual physician's medical liability risk

Note: Outputs of the study will not be used to set rates for individual or group health insurance products. Our research focuses on medical professional liability insurance coverages.

3. Has an Institutional Review Board (IRB) reviewed your Project?

- Yes [*If yes, a copy of the approval letter and protocol must be included with the Application package on IRBNet.*]
 No, this Project is not human subject research and does not require IRB review.

4. **Research Methodology:** Applications must include either the IRB protocol or a written description of the Project methodology (typically 1-2 pages), which should state the Project objectives and/or identify relevant research questions. This document must be included with the Application package on IRBNet and must provide sufficient detail to allow CHIA to understand how the Data will be used to meet objectives or address research questions.

See attached "MA-APCD Research Methodology" document.

V. PUBLIC INTEREST

1. Briefly explain why completing this Project is in the public interest. Use quantitative indicators of public health importance where possible, for example, numbers of deaths or incident cases; age-adjusted, age-specific, or crude rates; or years of potential life lost. *Uses that serve the public interest under CHIA regulations include, but are not limited to: health cost and utilization analysis to formulate public policy; studies that promote improvement in population health, health care quality or access; and health planning tied to evaluation or improvement of Massachusetts state government initiatives.*

Medical errors and their associated liability claims impose a heavy burden on the healthcare system. Beyond simply putting a patient's life in danger, medical errors have the following secondary effects:

- 1) **Increased Medical Utilization:** Physician errors lead to increased medical costs to patients as patients may now need additional clinical care to manage their resulting injuries or disabilities. One economic study conducted by the Society of Actuaries (SOA) estimated that medical errors imposed roughly \$19.5 billion in cost to the U.S in 2008, of which 87% or \$17 billion was a result of the increased medical utilization of patients who were affected by the medical errors.¹
- 2) **Loss of Worker Productivity and Income:** Errors also lead to a loss of worker productivity as patients may need to miss work to recover from their injuries or may have suffered a disability that prevents them from continuing their prior occupation. In the same economic study, the SOA estimated that medical errors resulted in \$1.1 billion worth of loss productivity based on an evaluation of short-term disability claims.¹ As a result, patients are placed in a financially precarious situation in terms of income due to medical errors.
- 3) **Loss of Trust in Healthcare Institutions:** Errors result in a greater distrust of the healthcare system as patients may suffer anxiety, depression, or emotional trauma from the medical error that could prevent them from seeking medical care in the future. Surveys conducted by the Betsy Lehman Center indicate that roughly one-third of respondents who have experienced a medical error 3-6 years prior will continue to sometimes or always avoid medical care even post the experience².

In addition, the associated malpractice/liability claims that result from these medical errors impose further legal, administrative, and financial tolls on the healthcare system. A study by the American Medical Association (AMA) estimates that roughly 34% of

physicians will face a malpractice claim at some point during their career³. A separate analysis of 2006 to 2015 malpractice claims, also conducted by the AMA, found that indemnity payments average around \$365,503. More interestingly though, the study found that a small fraction of malpractice claims—the 11.2% of claims involving payouts of at least \$1 million—contributed a disproportionate share (41.9%) of total indemnity payments.⁴ This suggests that a better understanding of the medical errors associated with these high payment claims and even a moderate level of risk management in just a few key areas of medical errors can have an outsized impact on containing liability costs overall.

Aligned with our mission to support and protect the healthcare community, Coverys believes that developing more advanced methods for predicting medical malpractice risk has several potential public benefits. As stated in the previous section (Section IV Question #2), this research will be used to improve underwriting of risk and to promote early detection/intervention of high-risk providers by risk managers. The former is necessary for creating an incentive structure that can push providers to adopt better risk management practices. The latter is critical for creating more proactive risk mitigation strategies in the long run. Overall, this research has the potential to have wide-ranging impact for both Massachusetts and the national healthcare community. According to the 2022 State Physician Workforce Data Report, Massachusetts ranks number one in terms of states with the most active physicians per resident, at 466 providers per residents or roughly 32,116 total providers⁵. Thus, as a leader, Massachusetts can serve as an example of the improvements that could be brought about for other states to follow.

Sources:

1. <https://www.soa.org/globalassets/assets/Files/Research/Projects/research-econ-measurement.pdf>
2. <https://betsylehmancenterma.gov/assets/uploads/Cost-of-Medical-Error-Report-2019.pdf>
3. <https://www.ama-assn.org/sites/ama-assn.org/files/corp/media-browser/public/government/advocacy/policy-research-perspective-medical-liability-claim-frequency.pdf>
4. <https://www.ama-assn.org/sites/ama-assn.org/files/corp/media-browser/public/government/advocacy/policy-research-perspective-liability-insurance-claim.pdf>
5. <https://www.beckershospitalreview.com/workforce/this-state-has-the-most-physicians-per-capita.html>

VI. DATASETS REQUESTED

The Massachusetts All-Payer Claims Database is comprised of medical, pharmacy, and dental claims and information from the member eligibility, provider, and product files that are collected from health insurance payers licensed to operate in the Commonwealth of Massachusetts. This information encompasses public and private payers as well as data from fully-insured and self-insured plans. APCD data are refreshed and updated annually and made available to approved data users in Release Versions that contain five calendar years of data and three months of run-out. For more information about APCD Release Versions, including available years of data and a full list of elements in the release please refer to release layouts, data dictionaries and similar documentation included on [CHIA's website](#).

Data requests are typically fulfilled on a one time basis, however; certain Projects may require future years of data that will become available in a subsequent release. Projects that anticipate a need for future years of data may request to be considered for a subscription. Approved subscriptions will receive, upon request, the same data files and data elements included in the initial Release annually or as available. Please note that approved subscription requests are subject to the Data Use Agreement, will require payment of fees for additional Data for Non-Government Entities, and subject to the limitation that the Data can be used only in support of the approved Project.

1. Please indicate below whether this is a one-time request, or if the described Project will require a subscription.

One-Time Request **OR** Subscription

2. Select Release Version and years of data requested (Release Versions and years not listed may not be available).

ANNUAL RELEASE 2020

- 2016
 2017
 2018
 2019
 2020

ANNUAL RELEASE 2021

- 2017
 2018
 2019
 2020
 2021

3. Specify below the data files requested for this Project, and provide your justification for requesting *each* file.

 Medical Claims**Describe how your research objectives require Medical Claims data:**

Medical claims will be used to:

- 1) **Identify potential cases of medical errors for analysis:** Diagnosis codes on the claims will be used to flag cases of hospital-acquired conditions (HACs), preventable external injuries (i.e. E-diagnosis codes), potential diagnostic errors (i.e. ICD code Z71.1), and/or conditions associated with AHRQ's Patient Safety Indicators (PSIs).
- 2) **Develop cost, quality, and patient outcome metrics:** These metrics include, for instance, preventable readmission rates, ER rates, length of stay, or overall cost of care which will be used to measure the impact of a medical error or to be tested as predictors in our risk models.
- 3) **Create clinical factors for risk-adjustment and benchmarking with peers:** These factors can include, for instance, age of patient at time of service, gender, complications and comorbidities flags (i.e. CC/MCC flags), indicators of chronic conditions (i.e. HCC flags), or prior utilization flags (i.e. patient has a prior 30-days IP stay flag). Risk-adjustment of previously mentioned cost, quality, and patient outcome metrics by these patient risk factors is necessary for producing proper comparison and benchmarking of medical error risk among physicians. For instance, elderly patients naturally have a greater risk for adverse events because they tend to have more chronic conditions/comorbidities that can increase the complexity of performing a given procedure. Their age and number of chronic conditions also mean that elderly patients will interact with the healthcare system more frequently than the average adult in order to properly manage their conditions, naturally putting them at greater exposure or potential opportunity for an adverse event to occur. Because providers treat different patient case mixes, if age along with other patient characteristics are not properly accounted for in the risk-adjustment of cost and quality metrics, we may inadvertently penalize certain providers as "underperformers" on certain metrics simply because they may be dealing with a more elderly or complex cohort of patients.

 Pharmacy Claims**Describe how your research objectives require Pharmacy Claims data:**

Pharmacy claims will be used in conjunction with medical claims for the same purposes as those listed above. For instance, diagnosis codes on pharmacy claims or usage of certain drugs can be used to flag certain chronic conditions or comorbidities for

risk-adjustment. Drugs also play a crucial role in the overall care management of patients, so pharmacy claims are needed to ensure a comprehensive picture of the impact of a medical error and the cost/quality metrics that will be used in our liability risk models.

Dental Claims

Describe how your research objectives require Dental Claims data:

Click here to enter text.

Member Eligibility

Describe how your research objectives require Member Eligibility data:

Member eligibility data is needed to derive demographic factors (e.g. gender, Medicare/Medicaid enrollment status, drug coverage status, etc.) for use in risk-adjustment of cost, quality, and patient outcome metrics. As previously alluded to under “Medical Claims” section, lack of risk-adjustment by said factors may lead to improper comparison or benchmarking of clinical performance between providers or biased model estimates of liability risks.

Member eligibility is also required for calculating the total eligible member population so that cost and quality metrics can be “standardized” by this number and made comparable across different groupings or periods in time. For instance, we cannot simply compare the total raw claim cost between say 2020 and 2021 to conclude if there has been a change in spending because the number of members and months of enrollment may have changed from year to year. Rather we must standardize the cost (i.e. divide the total cost by the total eligible member population in each year) to derive the cost per member per month (PMPM) and compare based on the risk-adjusted PMPM cost.

In addition, we are interested in exploring whether cost, quality, or medical error rates differ significantly between different types or levels of insurance coverage (e.g. Medicare, commercial, with/without drug coverage, etc.). For instance, if the same provider display large differences in their risk-adjusted medical error rate between Medicare vs commercial patients, it may suggest bias in the quality and delivery of care being provided.

Provider

Describe how your research objectives require Provider data:

Provider data is used to derive physician characteristics (e.g. physician age, specialty, etc.) that will be use as predictors in liability risk models and for proper peer-to-peer comparison of medical error risk.

Product

Describe how your research objectives require Product data:

Click here to enter text.

VII. DATA ENHANCEMENTS REQUESTED

State and federal privacy laws limit the release and use of CHIA Data to the minimum amount of data needed to accomplish a specific Project objective.

All-Payer Claims Database data is released in Limited Data Sets (LDS). All Organizations receive the “Core” LDS, but may also request the data enhancements listed below for inclusion in their analyses. Requests for enhancements will be reviewed by CHIA to determine whether each represents the minimum data necessary to complete the specific Project objective.

For a full list of elements in the release (i.e., the core elements and additional elements), please refer to [release layouts, data dictionaries](#) and similar documentation included on CHIA’s website.

1. Specify below which enhancements you are requesting in addition to the “Core” LDS, provide your justification for requesting each enhancement.

a. Geographic Subdivisions

ZIP code and state geographic subdivisions are available for Massachusetts residents and providers only. Small population ZIP codes are combined with larger population ZIP codes. One ZIP Code per person (MEID) per year has been assigned based on the ZIP code/state reported in the member eligibility record’s earliest submission year month. If the record does not have an MEID, assignment is based on distinct OrgID/Carrier Specific Unique Member ID.

Non-Massachusetts ZIP codes and state codes except for CT, MA, ME, NH, NY, RI, and VT are suppressed.

Select one of the following options.

<input checked="" type="checkbox"/> 3-Digit Zip Codes (standard)	<input type="checkbox"/> 5-Digit Zip Codes***
***If requested, provide justification for requesting 5-Digit Zip Code. Refer to specifics in your methodology:	

b. Date Resolution

Select one option from the following options.

<input type="checkbox"/> Year (YYYY) (Standard)	<input type="checkbox"/> Month (YYYYMM) ***	<input checked="" type="checkbox"/> Day (YYYYMMDD) *** [for selected data elements only]
*** If requested, provide justification for requesting Month or Day. Refer to specifics in your methodology:		
<p>Specific date values will be needed to construct certain quality and patient outcome metrics for use in our liability risk models or to measure the impact of medical errors. For instance, the specific month and day are necessary to identify if an IP stay occurred within 30 days of another IP hospitalization or SNF stay for constructing readmission rates and SNF bounceback rates as well as to measure overall length of stay.</p> <p>In addition, one type of medical error our team seeks to study include diagnostic delays as determined by the amount of time until the first occurrence of a given diagnosis code. As such, the exact month and day is necessary to properly measure the amount of delay in question.</p>		

c. National Provider Identifier (NPI)

Select one of the following options.

<input type="checkbox"/> Encrypted National Provider Identifiers (standard)	<input checked="" type="checkbox"/> Decrypted National Provider Identifiers***
*** If requested, provide justification for requesting decrypted National Provider Identifier(s). Refer to specifics in your methodology:	
<p>The decrypted NPI is necessary in order to link APCD data with physician-level data acquired from other sources, such as National Plan and Provider Enumeration System (NPPES) data, Coverys’ internal malpractice claims data, or Coverys’s licensed Definitive Health data. This linking is necessary in order to:</p>	

- 1) **Study network effects on liability risk:** For instance, NPDES and Definitive Health data provides information on a physician's associated organizations, affiliated facilities, and network. One of our team's research question surrounds the impact of a physician's peer or organizational network on their individual liability risk. As such, linking on associated organization information to APCD data is necessary for examining network effect on medical error risk.
- 2) **Flag physicians with history of medical errors:** Coverys's internal malpractice data will be used to flag NPIs in the APCD data who have incurred a malpractice claim. Our study seeks to compare the differences in cost and quality performance between physicians with history of malpractice claims versus those without in order to develop predictors of liability risk.

VIII. MEDICAID (MASSHEALTH) DATA

1. Please indicate whether you are seeking Medicaid Data:

- Yes
 No

2. Federal law (42 USC 1396a(a)7) restricts the use of individually identifiable data of Medicaid recipients to uses that are **directly connected to the administration of the Medicaid program**. If you are requesting MassHealth Data, please describe, in the space below, why your use of the Data meets this requirement. *Your description should focus on how the results of your project could be used by the Executive Office of Health and Human Services in connection with the administering the MassHealth program.* Requests for MassHealth 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 MassHealth program. CHIA cannot release MassHealth Data without approval from MassHealth. This may introduce significant delays in the receipt of MassHealth Data.

Click here to enter text.

3. Organizations approved to receive Medicaid Data will be required to execute a [Medicaid Acknowledgment of Conditions](#). MassHealth may impose additional requirements on applicants for Medicaid Data as necessary to ensure compliance with federal laws and regulations regarding Medicaid.

IX. DATA LINKAGE

Data linkage involves combining CHIA Data with other data to create a more extensive database for analysis. Data linkage is typically used to link multiple events or characteristics within one database that refer to a single person within CHIA Data.

1. Do you intend to link or merge CHIA Data to other data?
- Yes
 No linkage or merger with any other data will occur

2. If yes, please indicate below the types of data to which CHIA Data will be linked. [Check all that apply]

- Individual Patient Level Data (e.g. disease registries, death data)
- Individual Provider Level Data (e.g., American Medical Association Physician Masterfile)
- Individual Facility Level Data (e.g., American Hospital Association data)
- Aggregate Data (e.g., Census data)
- Other (please describe):

3. If yes, describe the dataset(s) to which the CHIA Data will be linked, indicate which CHIA Data elements will be linked and the purpose for each linkage.

<p>1) National Plan & Provider Enumeration System (NPPES): a public-use provider registry data bank</p>
<p>a. Linked on: NPI</p> <p>b. Purpose: to identify additional provider characteristics (e.g. specialty, practice location, organizational associations, etc.) that may be relevant for risk-adjustment and peer-to-peer benchmarking</p>
<p>2) National Practitioner Data Bank (NPDB): a public-use data repository for medical malpractice payments and adverse actions by health care practitioners and suppliers.</p>
<p>a. Linked on: Practitioner's specialty, medical license type, and/or practice location</p> <p>b. Purpose: to generate aggregate prevalence and likelihood measures for adverse medical errors at the specialty, medical license type, and/or geographic location level to test correlation with aggregate cost/quality metrics calculated from claims also at the specialty, license type, and/or geographic location level</p>
<p>3) Definitive Health Care Data: a licensed data repository of aggregated physician-level and hospital-level characteristics/performance metrics</p>
<p>a. Linked on: NPI and/or Hospital Name (or CCN Identifier if available)</p> <p>b. Purpose: to acquire additional provider characteristics for physicians and hospital (e.g. MIPS Quality Scores, bed size, ACO participation status, teaching hospital status, etc.) for use in risk-adjustment and peer-to-peer benchmarking</p>
<p>4) Census Data: a public-use repository of census or geographical data maintained by the U.S Census Bureau</p>
<p>a. Linked on: Zip Code</p> <p>b. Purpose: to identify possible geographic factors and differences that may influence cost/quality of care for use in risk-adjustment and benchmarking</p>
<p>5) Medicare's Physician Fee Schedule (PFS) Relative Value Files: a public-use data file managed by the Center for Medicare and Medicaid Services on the Relative Value Units (RVUs) used to construct annual reimbursement rates for physician services</p>
<p>a. Linked on: HCPCS codes</p> <p>b. Purpose: When setting annual physician reimbursement rates, CMS adjusts its rates to account for the premiums that physicians pay for professional liability insurance (PLI) through the PLI RVU. By merging on the PLI RVU to the HCPCS codes, we can identify, in the APCD claims, high-frequency procedures that are also associated with high PLI RVUs and that, as a consequence, may be causing Medicare an outsized burden to cover in terms of medical liability cost. This will aid our research into estimating the cost impact of medical malpractice for certain procedure categories and identifying specific high-risk procedures as focus areas for risk management.</p>
<p>6) Coverys Malpractice Claims Data: an internal data repository of medical malpractice claims managed by our organization</p>
<p>a. Linked on: NPI</p> <p>b. Purpose: to supplement aggregate data provided by the NPDB in measuring adverse medical error rates and malpractice cost as well as to flag NPIs with history of medical error in APCD data for further analysis</p>

4. 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.

- | |
|--|
| <ol style="list-style-type: none"> 1) National Plan & Provider Enumeration System (NPPES): exact matching on NPI 2) National Practitioner Data Bank (NPDB): exact matching on license type and/or U.S state or based on internal clinical expertise, creating a custom mapping of physician specialties to license fields to crosswalk between physicians under a given license field in the NPDB and physicians with the associated specialties in the APCD. 3) Definitive Health Care Data: exact matching on NPI and/or Hospital Name/CCN 4) Census Data: exact matching based on zip code 5) PFS Relative Value Files: exact matching on HCPCS codes 6) Coverys Malpractice Claims Data: exact matching on NPI |
|--|

5. If yes, attach or provide below a complete listing of the variables from all sources to be included in the final linked analytic file.

1) National Plan & Provider Enumeration System (NPPES):
--

- | |
|--|
| <ol style="list-style-type: none"> a. Provider's practice location b. Provider's gender c. Provider's primary and secondary taxonomy codes d. Associated organizations |
|--|

2) National Practitioner Data Bank (NPDB):

- | |
|---|
| <ol style="list-style-type: none"> a. Claim report type b. Date of claim report c. Provider's work and home state and country d. Provider's state of license e. Provider's field of license f. Provider's age group g. Provider's graduation year group h. Provider's malpractice allegation group and specific allegation description i. Allegation severity measure j. Malpractice payment amount k. Patient Type (e.g. Inpatient, Outpatient) l. Date of adverse event m. Length of adverse action penalty n. Entity Type o. Practitioners' Number of Malpractice Payment Reports p. Practitioners' Number of Licensure Reports q. Practitioners' Number of Clinical Privileges Reports r. Practitioners' Number of Professional Society Membership Reports s. Practitioners' Number of DEA Reports t. Practitioners' Number of Exclusion Reports u. Practitioners' Number of Government Administrative Reports |
|---|

- v. Practitioners' Number of Contract Termination Reports
- w. State's patient compensation
- x. Medical license type

3) Definitive Health Care Data:

- a. Physician primary and claims-based specialty
- b. Physician clinical activity level and percentile compared to peers
- c. Physician propensity to prescribe brand name drugs
- d. Affiliated facility name
- e. Affiliated network name
- f. Network prevalence ranking
- g. Physician gender
- h. Physician age
- i. Physician medical school graduation year
- j. Physician medical school name
- k. Physician medical credentials (MD, DO, DPM, etc.)
- l. Physician is a hospitalist flag
- m. Physician is an executive flag
- n. Physician primary practice location
- o. Physician MIPS Performance Category and Rate
- p. Physician MIPS Quality, Interoperability, and Improvement Activities Score
- q. Physician's participation status in ACO, BPCI, CJR, VBP
- r. Hospital taxonomy code and description
- s. Hospital location
- t. Hospital number of bed group
- u. Facility type
- v. Hospital's serious complication rate
- w. Hospital's associated medical school/teaching status
- x. Hospital's quality violation category
- y. Hospital's referral network
- z. Hospital's average length of stay

4) Census Data:

- a. Income/poverty level
- b. Program income and public assistance level
- c. Education level
- d. Housing affordability level
- e. Housing vacancy level
- f. Homeownership rate
- g. Race
- h. Gender
- i. Age group

5) PFS RVU Data:

- a. Relative Value Units (RVUs)
- b. Geographic Practice Cost Index (GPCIs)
- c. Conversion Factors (CFs)

6) Coverys Malpractice Claims Data:

- a. Number of malpractice claims
- b. Avg expense per claim
- c. Avg indemnity per claim
- d. Claim allegation code and description

- e. Risk management code and description
- f. Area of occurrence code
- g. Surgery/Procedure and Medication Code
- h. Treatment Area Code
- i. Comorbidities code
- j. Injury severity code
- k. Readmission status and diagnosis code
- l. Injury code
- m. Patient Status (e.g. Inpatient, Outpatient, ER, Observation, Unknown, Other)

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

1) NPPES:

This data file does not directly contain any personally identifiable information about a patient, only physician-level characteristics, and so will pose no risk to the identification of any individual patient.

2) NPBD:

The only PHI-related fields contained in the NPBD are the age and gender of the affected patient, and there is no member name or any member identifier field to enable linking with APCD. In addition, the only provider identifier field in the NPBD is an arbitrary number that is unique to this file and cannot be used to link to any other provider identifiers such as NPI that are commonly used in claims data. As a result, patients cannot be identified either through any direct linkages or through their association with any given provider in the NPBD. As an added precaution, we will also remove any PHI-related files (e.g. patient age, gender, etc.) from the NPBD prior to linking to the APCD.

3) Definitive Health Care Data:

All data contained in Definitive have been pre-aggregated by Definitive to either the physician-level or hospital-level and contains no patient-specific information to possibly merge on. Definitive does not provide Coverys with any existing access to patient-level data or to data cells that fall below the 11-member threshold. As such, any aggregated performance metrics Coverys may seek to merge on about a given physician or hospital is already guaranteed to be based on a large enough patient population as to not single out any individual patient.

4) U.S Census Data:

The census files have been pre-aggregated by the U.S Census Bureau to a level such that linking on a member identifier field would not be possible. In addition, as our team is only requesting the first 3-digits of the zip code from APCD to be able to merge on census data, these first 3-digits will only enable to us identify, at the lowest geographical layer, a region or perhaps a large city within the state. Thus, the geography covered by the first 3 digits will tend to be very wide and encompass a relatively large population size, such that any associated census data will likely be too high-level to identify or single out any patient within that region.

5) PFS RVU Data:

The data in this file is at the HCPCS level and contains only information relevant to a procedure. Thus, there is no risk in identifying any individual patient by merging on this file.

6) Coverys Malpractice Claims Data:

For the Coverys malpractice claims data, Coverys intends to take the following steps to prevent identification of individual patients:

I. Steps with regards to the general storage/maintenance of data files:

- 1) When the APCD is received, all fields on the medical and pharmacy claims tables related to patient demographics (e.g. age and gender) will be removed from those individual files, if present, and will be consolidated into a separate table containing only a Member ID, Claim ID, and the removed patient demographics columns. As a significant portion of our research is run at a claims-level, separating the PHI-fields from the other claims-level fields that are more frequently used for our analysis will reduce the likelihood that PHI is shared to invalid parties querying said claims or unnecessarily shown in research scenarios where PHI may not be even be needed to conduct the analysis at hand.
- 2) Next, Coverys will apply user permissioning controls on both the table produced from Step 1 above and APCD's Member Eligibility table in our data warehouse/servers. Specifically, permissioning will be set such that only the Coverys' research investigator(s) and data warehouse engineer on this project will be allowed to view/read/write/edit information in these tables. If PHI data is required by other internal Coverys analysts during the course of research, data from these tables will be queried only by Coverys' lead investigator(s) or data warehouse engineer. Limiting user access to the PHI data tables will reduce the likelihood of PHI information being inadvertently spread to invalid parties and ensures that scenarios requiring linking of PHI between APCD's and Coverys's data have been thoroughly reviewed and vetted by Coverys's lead investigator(s) or data engineer as fully necessary and secure for the research at hand.
- 3) For analysis that requires linking of data between APCD's and Coverys' data and that results in a combined table with PHI fields in it, Coverys will apply similar user permissioning controls as mentioned above on the combined, derived table so that only the lead investigator(s) or data warehouse engineer will have view/read/write/edit access to the combined table. Requests from other internal analysts who may need to access information from the combined table will similarly be evaluated and carried out by either the lead investigator(s) or data warehouse engineer after vetting the necessity of the request.
- 4) Finally, all columns or fields containing PHI information in either the table produced from Step 1, APCD's Member Eligibility table, or the combined, derived table from Step 3 will be encrypted with AES 256 bit. This level of encryption will be applied to all mentioned data sources housed in either our on-premise data warehouse or our cloud storage databases. Additional data masking and tokenization can be applied to PHI fields to provide further security. In general, all original source files received from CHIA (i.e. Medical Claims, Pharmacy Claims, Member Eligibility, and Provider files) will be housed in our cloud storage databases, both due to the nature of their large file size requirements and to provide the highest level of protection available.

II. Steps with regards to the presentation of any outputs during the course of the research process:

- 1) Data cells referencing linked dataset and containing data on fewer than 11 patients will be censored from reporting in any externally published outputs or data visualization software (e.g. Tableau, Power BI, etc.).
- 2) No patient-level data tables will be made available in any externally published outputs.
- 3) If data visualization softwares (e.g. Tableau, Power BI, etc.) are used during the course of presenting results, downloading and editing permissions on workbook files will be disabled for all external parties to prevent inadvertent distribution of PHI. Downloading and editing permissions on workbook files may be given, on a case-by-case basis, to internal organization employees once use case and internal users have been vetted by either Coverys' lead investigators or data engineer as necessary for the purpose of the research.

Finally, with regards to Section I and II above, it should be noted that Coverys has a Director for Information Security/Technology, Director of Compliance, and Director for Legal to ensure that our analytics team adheres to best practices for storage, management,

and communication of sensitive data across all our data sources. These departments will conduct, as needed, additional rounds of reviews on our team's outputs to ensure compliance with our data usage agreement.

X. PUBLICATION / DISSEMINATION / RE-RELEASE

1. Do you anticipate that the results of your analysis will be published or made publically available? If so, how do you intend to disseminate the results of the study (e.g.; publication in professional journal, poster presentation, newsletter, web page, seminar, conference, statistical tabulation)? Any and all publication of CHIA Data must comply with CHIA's cell size suppression policy, as set forth in the Data Use Agreement. Please explain how you will ensure that any publications **will not disclose a cell less than 11**, and percentages or other mathematical formulas that result in the display of a cell less than 11.

Coverys may publish summary level results of the study on our own organizational website. To ensure adherence to CHIA's cell suppression policy, the publishing of any study results will be subject to review by Coverys' compliance and/or legal counsels prior to release.

2. Describe your plans to use or otherwise disclose CHIA Data, or any Data derived or extracted from such Data, in any paper, report, website, statistical tabulation, seminar, or other setting that is not disseminated to the /public.

Internal use of study results at Coverys include: 1) development of risk management programs to mitigate the identified factors associated with increased medical malpractice risk and 2) improving underwriting models for medical malpractice insurance. For the majority of these internal use cases, study results will typically be provided in an aggregated format such as at an NPI level or county level so that even among internal parties, PHI data can be protected or provided in only scenarios reviewed and vetted by either the Principal Investigator, co-investigator, and data engineer as fully necessary.

We also recognize that APCD does not contain claims for all populations in Massachusetts (i.e. MassHealth is excluded from this study) and as such, our study results may not be fully representative of all patients or providers. However, we believe our study can still act as a signal, even if a rough one, of possibly larger issues afflicting the state's healthcare system. For instance, if a provider is determined to have a high hospital-acquired infection (HAC) rate based on the APCD data, it might be suggestive of larger organizational or process failure in the hospital. And that failure has the potential to put current and future MassHealth patients at risk, even if we may not have the data to quantify the impact. It is not often the case that larger organizational or process failure affects solely one specific sub-population in a hospital, but rather all patients receiving care in the hospital are placed at risk. Thus, even with the data limitation, our study can still be a productive step forward in meeting one of our main research goals—identifying early warning signals for creating a proactive, prospective means of risk management. As a precaution though to ensure that our study results are appropriately interpreted and used, we intend also to fully detail these data limitations in a written format and to accompany any internal or external disclosure of study results or associated product tools with this documentation.

3. What will be the lowest geographical level of analysis of data you expect to present for publication or presentation (e.g., state level, city/town level, zip code level, etc.)? Will maps be presented? If so, what methods will be used to ensure that individuals cannot be identified?

Zip codes (up to the requested 3-digits) will be lowest level of granularity if geographic maps are used. Similar precautions referenced above will be applied here as well when analyzing data geographically. Specifically:

- 1) Censoring of data cells with fewer than 11 members in geographical maps
- 2) If data visualization software are used during the course of research or presentation, disabling downloading/editing permissions on workbook files for all external parties.
- 3) If map visualizations are externally published or disseminated, they will be subject to review by Coverys' compliance and/or legal counsels to ensure adherence to CHIA's Data Usage Agreement prior to publication.

4. Will you be using CHIA Data for consulting purposes?

- Yes
 No

5. Will you be selling standard report products using CHIA Data?

- Yes
 No

6. Will you be selling a software product using CHIA Data?

- Yes
 No

7. Will you be using CHIA Data as in input to develop a product (i.e., severity index tool, risk adjustment tool, reference tool, etc.)

- Yes
 No

8. Will you be reselling CHIA Data in any format not noted above?

- Yes
 No

If yes, in what format will you be reselling CHIA Data?

N/A

9. If you have answered “yes” to questions 5, 6, 7 or 8, please provide the name and a description of the products, software, services, or tools.

N/A

10. If you have answered “yes” to questions 5, 6, 7 or 8, what is the fee you will charge for such products, software, services or tools?

N/A

XI. APPLICANT QUALIFICATIONS

1. Describe your previous experience using claims data. This question should be answered by the primary investigator and any co-investigators who will be using the Data.

The investigators who will be involved with using the APCD dataset, if provided, are Thaochau Phan (M.B.A), Chris Coulter (F.S.A., M.A.A.A., M.Sc.), Sam Greenblatt (M.Sc. M.A), Michelle Fritz (Ph.D), and Will Eggbeer (M.B.A). As employees of Coverys, all investigators have received HIPAA compliance training on safeguarding protected health information, consistent with company’s security policies. These researchers also have extensive prior experience in handling sensitive medical malpractice data, APCD data acquired from other states, and Medicare claims from CMS. Additionally, a few of the investigators—specifically, Thaochau Phan,

Chris Coulter, and Will Eggbeer—were previously employed by Archway Health Advisors, a Massachusetts based company recently acquired by Coverys. In 2015-2016, Archway Health Advisors was approved to receive MA APCD data to study value-based care in the state, and as such, these investigators have specific experience in the management of MA APCD data. In total, our team consists of members with broad, proven experiences in the proper handling and analysis of sensitive data.

Thaochau Phan, M.B.A.,

Thao is a Data Scientist at Coverys, and has performed analyses using large commercial medical claims datasets, the 100% Medicare claims datasets, various APCD files, Coverys' internal malpractice claims data, and the National Practitioner Data Bank data. Thao received a B.A. from UCLA, and an M.B.A. from Boston University, specializing in Health Administration, Operations, and Technology Management. Thao previously worked as a lead programmer/policy analyst for a health research/consultancy organization (Acumen, LLC) that operates as a collaborator and contractor to CMS on its claims operations and value-based care analytics. Past projects worked on by the principal investigator include using claims to develop cost standardization, episode grouping, risk-adjustment, and benchmarking logic for CMS's MACRA, Hospital VBP, and the BPCI programs. The investigator has also worked as an analyst for an accountable care organization (Tufts Medicine) managing back-end table/schema development for claims data in SQL servers, constructing secure, automated front-end reporting tools for physicians and medical directors, and producing performance metrics for ACO's participation in Medicare Shared Savings and Next Generation Programs.

Chris Coulter, F.S.A., M.A.A.A., M.Sc.

Chris is Coverys' Senior Vice President and Actuary. He leads the analytics teams that supports Coverys' value-based care division and malpractice pricing division. He serves as an in-house expert on medical claims data. Chris is a Fellow of the Society of Actuaries, a member of the American Academy of Actuaries, and a Chartered Enterprise Risk Analyst. Chris received a B.Sc. from Simon Fraser University and an M.Sc. from the University of British Columbia. Chris has close to 15 years of experience with health care insurance in the US and in particular medical claims. Previous analytics projects led and overseen by Chris include using Medicare claims to evaluate physician performance and benchmarks under CMS's Qualified Entity program, and modeling insurance pricing for providers participating in CMS's Medicare Shared Savings Program, CMS's BPCI Program, and CMS's OCM program.

Samuel Greenblatt, M.Sc., M.A.

Sam is a Data Scientist at Coverys, and he has performed analyses using multiple large commercial medical and pharmaceutical claims datasets, Medicare claims data, and Coverys' internal malpractice claims data. Sam received a B.A. and an M.A. from Brandeis University and an M.Sc. from the London School of Economics and Political Science, specializing in economics and historical research. Prior to joining Coverys he performed analyses related to questions of drug market definition, causation within public health crises, impact of the Covid-19 pandemic on property insurance claims, and impacts of faulty medical devices while working at Analysis Group. Sam began his career performing macroeconomic forecasts for the U.S. Bureau of Labor Statistics. In total he has 15 years of experience performing analysis on large and complex data assessing causation, studying diverse cohorts, and forecasting future impacts (while working for Coverys, Analysis Group, Compass Lexecon, and the BLS). This work includes over four years focused primarily on healthcare, public health, and insurance markets.

Michelle Fritz, Ph.D.

Michelle is a Data Scientist at Coverys, specializing in developing analytical models using Coverys' malpractice claims data, Medicare claims data, and data from APCD files to analyze risk levels of insureds and support pricing decisions. She holds a Ph.D. in Condensed Matter Physics and Nanotechnology from the Universidad Autónoma de Madrid, with a focus on Computational Condensed Matter. Before joining Coverys, she worked at Qunnect, a Quantum Internet startup, where she developed automation models for optics systems. Prior to that, she worked as a consultant, utilizing diverse data sets to build models for automation, forecasting, and analytics across various industries. With over 7 years of experience in data science, Michelle brings a strong scientific background and a track record of leveraging data-driven insights to drive business decisions.

Will Eggbeer, MBA

Will is the Senior Director of Data Science at Coverys and leads the Data Science team. Will oversees the risk modeling and analytic product development for Coverys' medical malpractice and value-based-care insurance lines. Prior to joining Coverys, Will led new

product development at Archway Health Advisors, where he focused on developing products and services to help patients identify high quality providers in their areas of need. He also previously led product strategy for Athenahealth's patient engagement business. Will earned a B.S in Computer Science & Engineering from Bucknell University and an MBA from the Tuck School of Business at Dartmouth. In addition to his focus on product development, will has experience with health information technology, health policy, value-based payment programs including the CMS BPCI-A and MSSP ACO programs, healthcare data privacy and security, and the analysis of medical claims data.

2. **Resumes/CVs:** When submitting your Application package on IRBNet, include résumés or curricula vitae of the principal investigator and co-investigators. (These attachments will not be posted on the internet.)

XII. USE OF AGENTS AND/OR CONTRACTORS AND/OR COLLABORATORS

By signing this Application, the Organization assumes all responsibility for the use, security and maintenance of the CHIA Data by its agents, including but not limited to contractors. The Organization must have a written agreement with the agent or contractor limiting the use of CHIA Data to the use approved under this Application as well as the privacy and security standards set forth in the Data Use Agreement. CHIA Data may not be shared with any third party without prior written consent from CHIA, or an amendment to this Application. CHIA may audit any entity with access to CHIA Data.

Provide the following information for **all** agents, contractors, and collaborators who will have access to the CHIA Data. *[Add agents or contractors or collaborators as needed.]*

COLLABORATOR INFORMATION	
Company Name:	
Company Website	
Contact Person:	
Title:	
E-mail Address:	
Address, City/Town, State, Zip Code:	
Telephone Number:	
Term of Contract:	

1. Describe the tasks and products assigned to the collaborator for this Project and their qualifications for completing the tasks.

2. Describe the Organization's oversight and monitoring of the activities and actions of the collaborator for this Project, including how the Organization will ensure the security of the CHIA Data to which the agent or contractor has access.

Click here to enter text.

3. Will the collaborator have access to and store the CHIA Data at a location other than the Organization's location, off-site server and/or database?

- Yes
 No

4. If yes, a separate Data Management Plan **must** be completed by the collaborator.

AGENT/CONTRACTOR #1 INFORMATION	
Company Name:	
Company Website	
Contact Person:	
Title:	
E-mail Address:	
Address, City/Town, State, Zip Code:	
Telephone Number:	
Term of Contract:	

1. Describe the tasks and products assigned to the agent or contractor for this Project and their qualifications for completing the tasks.

N/A

2. Describe the Organization's oversight and monitoring of the activities and actions of the agent or contractor for this Project, including how the Organization will ensure the security of the CHIA Data to which the agent or contractor has access.

N/A

3. Will the agent or contractor have access to or store the CHIA Data at a location other than the Organization's location, off-site server and/or database?

- Yes
 No

4. If yes, a separate Data Management Plan **must** be completed by the agent or contractor.

[INSERT A NEW SECTION FOR ADDITIONAL AGENTS/CONTRACTORS AS NEEDED]

XIII. ATTESTATION

By submitting this Application, the Organization attests that it is aware of its data use, privacy and security obligations imposed by state and federal law *and* confirms that it is compliant with such use, privacy and security standards. The Organization further agrees and understands that it is solely responsible for any breaches or unauthorized access, disclosure or use of CHIA Data, including, but not limited to, any breach or unauthorized access, disclosure or use by any third party to which it grants access.

Organizations approved to receive CHIA Data will be provided with Data following the payment of applicable fees and upon the execution of a Data Use Agreement requiring the Organization to adhere to processes and procedures designed to prevent unauthorized access, disclosure or use of data.

By my signature below, I attest: (1) to the accuracy of the information provided herein; (2) this research is not funded by a source requiring the release of raw data to that source; (3) that the requested Data is the minimum necessary to accomplish the purposes described herein; (4) that the Organization will meet the data privacy and security requirements described in this Application and supporting documents, and will ensure that any third party with access to the Data meets the data use, privacy and security requirements; and (5) to my authority to bind the Organization.

Signature: (Authorized Signatory for Organization)	DocuSigned by: <i>Scott Weber</i> 715D08718F3C49E...
Printed Name:	Scott Weber
Title:	Head of Strategy and Corporate Development
Date:	6/6/2024

Attachments:

A completed Application must have the following documents attached to the Application or uploaded separately to IRBNet:

- 1. IRB approval letter and protocol (if applicable), or research methodology (if protocol is not attached)
- 2. Data Management Plan (including one for each agent or contractor that will have access to or store the CHIA Data at a location other than the Organization's location, off-site server and/or database);
- 3. CVs of Investigators (upload to IRBNet)

APPLICATIONS WILL NOT BE REVIEWED UNTIL THEY ARE COMPLETE, INCLUDING ALL ATTACHMENTS.



center
for health
information
and analysis