

**Commonwealth of Massachusetts  
Center for Health Information & Analysis (CHIA)  
Non-Government Application for MA APCD Limited Data Set**

*This form is required by all Applicants, except Government Agencies as defined in [957 CMR 5.02](#). All Applicants must also complete the Data Management Plan, attached to this Application. The Application and the [Data Management](#) Plan must be signed by an authorized signatory of the organization. This Application and the Data Management Plan will be used by CHIA to determine if your organization may receive CHIA data. Please be sure the documents are completed fully and accurately. You may wish to consult the Evaluation Guide that CHIA will use to review your documents. Prior to receiving CHIA Data, the organization must execute the [Data Use Agreement](#). You may wish to review that document as you complete these forms. This application should be completed by the Primary Investigator, and must be signed by a party with authority to bind the organization seeking CHIA Data for the purposes described herein.*

***NOTE: In order for your Application to be processed, you must submit the required application fee. Please consult the fee schedules for MA APCD data for the appropriate fee amount. A [remittance](#) form with instructions for submitting the application fee is available on the CHIA website.***

*All attachments must be uploaded to IRBNet with your Application. All applications documents can be found on the [CHIA website](#) in Word and/or PDF format.*

**I. GENERAL INFORMATION**

APPLICANT INFORMATION	
Applicant Name: (Primary Investigator)	Belen Fraile
Title:	Senior Director of Population Management and Decision Support
Organization:	Dana-Farber Cancer Institute
Project Title:	Understanding Variations in Cancer Care in the State of Massachusetts, Measuring Value
Mailing Address:	450 Brookline Ave. Boston, MA 02215
Telephone Number:	617-632-5350
Email Address:	<a href="mailto:Belen_fraile@dfci.harvard.edu">Belen_fraile@dfci.harvard.edu</a>
Names of Co-Investigators:	Yichen Zhang; Stephanie Gottsch; Shaobin Dong
Email Addresses of Co-Investigators:	<a href="mailto:Yichen_zhang@dfci.harvard.edu">Yichen_zhang@dfci.harvard.edu</a> <a href="mailto:Stephanie_gottsch@dfci.harvard.edu">Stephanie_gottsch@dfci.harvard.edu</a> <a href="mailto:Shaobin_Dong@dfci.harvard.edu">Shaobin_Dong@dfci.harvard.edu</a>
Original Data Applicant Submission Date:	6/3/2016
Dates Data Application Revised:	7/14/2016
Project Objectives (240 character limit):	1) Understand variations in outcomes of cancer care delivery among different providers and cancer types across the State. 2) Understand full spectrum of care of DFCI's patients to identify potential care gaps and opportunities to improve.
Project Research Questions (if applicable) or Business Use Case(s):	Section 1-

	<p>a) Is there variation in cancer-specific efficiency among provider peer groups?                  b) Is there variation in cancer-specific efficiency among different types of cancer?                  c) Can a combination of these efficiency indicators formulate a value metric for cancer care?                  Section 2-                  a) Are there difference in outcomes (total cost of care and utilization) driven by how are patients share care with other non-DFCI providers?</p>
--	--

**II. PUBLIC INTERST & PROJECT SUMMARY**

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

Anticipating and responding to evolving health care policies, using “big data” to guide the treatment of patients, as well as adapting to the changing professional and financial aspects of Oncology will be addressed in “Understanding variations in cancer care in the state of Massachusetts. Measuring Value”  
 This study has two sections:

Section 1-  
 The objectives in this section are to understand the variations in cancer care delivery among different providers, different cancer types and their outcomes through the identification of efficiency indicators collected using CHIA administrative database.

Section 2-  
 The objective in this section is to understand the full spectrum of care and outcomes (cost and utilization) for Dana-Farber Cancer Institute’s patients before and after they visit and/or receive treatment at our center, to identify potential gaps within the care that we provide and opportunities for improvement.  
 The analytic methodology would be similar, if applicable, to both sections:  
 Providers will be assigned to peer groups of hospitals including: Comprehensive Cancer Centers (Dana-Farber is the only comprehensive Cancer Center in the State), Academic teaching Institutions and community hospitals.  
 The indicators that would be used to assess the variation between providers and between cancers may include (but are not limited to):

- Number of dates of service (visits, inpatient admissions, diagnostic tests)
- Repeated diagnosis tests
- Inpatient days
- Cost per day of life (from diagnosis to death)
- Cost per episode of care (different definitions of episodes may apply)
- Cost of the end of life (last 6 months of life)

To assess the variability, the population will be risk adjusted by all available information, such as demographic information (i.e. gender, age, median household income), location of services (i.e. institution) and overall health status (applying 3M (Clinical Risk Groups/ CRG (see methodology sections for specifics)

The primary research questions to be addressed by this study include:

Section 1:

- a) Is there variation in cancer-specific efficiency among provider peer groups?
- b) Is there variation in cancer-specific efficiency among different types of cancer?
- c) Can a combination of these efficiency indicators formulate a value metric for cancer care?

Section 2:

a) Are there difference in outcomes (total cost of care and utilization) driven by how are patients share care with other non-DFCI providers?  
 CHIA data offers a unique opportunity to not only understand better the care and outcomes for our patients but also to truly understand oncology care variations and measure value across providers in the state of Massachusetts.

Contrary to what Medicare database offers us, we should be able to capture a more complete picture of oncology patients not only the elderly, but also the younger and minorities.

These efficiency indicators, possible only by the use of an administrative database like CHIA data, would allow a consistent data collection and measurement methodology design that could be used as oncology care measures of performance across the state.

Findings could be escalated to inform policies or measurement development at the national level.

2. Has an Institutional Review Board (IRB) reviewed your project?

- Yes, a copy of the approval letter and protocol must be **attached** to this Application
- No, this project is not human subject research and does not require IRB review.

3. If your project has not been reviewed by an IRB, please **attach** a brief (1-2 page) description of your project including the methodology, objectives, and research questions.

**III. DATA FILES REQUESTED**

1. Please indicate the MA APCD databases from which you seek data, the year(s) of data requested, and your justification for requesting *each* file. Please refer to the [MA APCD Release 5.0 Data Specifications](#) for details of the file contents.

MA ALL-PAYER CLAIMS DATABASE FILES	Year(s) Of Data Requested Current Yrs. Available <input checked="" type="checkbox"/> 2011 <input checked="" type="checkbox"/> 2012 <input checked="" type="checkbox"/> 2013 <input checked="" type="checkbox"/> 2014 <input checked="" type="checkbox"/> 2015
<input checked="" type="checkbox"/> <b>Medical Claims</b>	<p><b>Please describe how your research objectives require Medical Claims data:</b></p> <p>The Medical claims file will be used to identify diagnostic codes, providers delivering care, procedures, charges, payments and date of death variables. Claims data files will be used in analysis to determine various efficiency measures including cancer-specific vs. all patient care results, variation among provide peer groups, variation among different types of cancer, and whether or not a combination of these efficiency indicators can formulate a value metric for cancer care.</p>

<input checked="" type="checkbox"/> <b>Pharmacy Claims</b>	<b>Please describe how your research objectives require Pharmacy Claims data:</b>  The Pharmacy claims file will be use to complete the charges and utilization from oncology patient, in particular paying attention to the growing oral antineoplastic agents from the past years
<input type="checkbox"/> <b>Dental Claims</b>	<b>Please describe how your research objectives require Dental Claims data:</b>
<input checked="" type="checkbox"/> <b>Member Eligibility</b>	<b>Please describe how your research objectives require Member Eligibility data:</b>  The Member file will be utilized to determine the basic demographics of the patient population, ultimately used for risk adjustment and building comparable samples for the analysis.
<input checked="" type="checkbox"/> <b>Provider</b>	<b>Please describe how your research objectives require Provider data:</b>  The provider file will be used to identify the location of the providers where patients receive care during the different stages of the disease and will be use to not only identify DFCL and its affiliates but also define peer provider groups that are the core basis for comparison in our analysis.
<input checked="" type="checkbox"/> <b>Product</b>	<b>Please describe how your research objectives require Product data:</b>  The product file will be used to identify insurance and coverage overall to estimate if these factors drive part of the variability.

**IV. GEOGRAPHIC DETAIL**

Please choose one of the following geographic options for MA residents. *For releases with 5 digit zip code, CHIA will apply a substance abuse filter which will remove all claims that include a substance abuse diagnosis.*

<input checked="" type="checkbox"/> 3 Digit Zip Code (MA) (standard)	<input type="checkbox"/> 5 Digit Zip Code (MA)***
<b>***Please provide justification for requesting 5 digit zip code. Refer to specifics in your methodology:</b>	

**V. DATE DETAIL**

Please choose one option from the following options for dates:

<input type="checkbox"/> Year (YYYY) (Standard)	<input type="checkbox"/> Month (YYYYMM) ***	<input checked="" type="checkbox"/> Day (YYYYMMDD) *** <a href="#">[for selected data elements only]</a>
<b>*** If requested, please provide justification for requesting Month or Day. Refer to specifics in your methodology:</b>		
Our analysis is based on encounter level claim information and will be accumulated to monthly and yearly level. Refer to		

methodology for more details

**VI. NATIONAL PROVIDER IDENTIFIER (NPI)**

Please choose one of the following options for National Provider Identifier(s):

<input type="checkbox"/> Encrypted National Provider Identifier(s) (standard)	<input checked="" type="checkbox"/> Unencrypted National Provider Identifier(s)***
<p><b>*** If requested please, provide justification for requesting unencrypted National Provider Identifier(s). Refer to specifics in your methodology:</b></p> <p>The unencrypted provider file will be use to not only identify DFCI and its affiliates but also define peer provider groups that are the core basis for comparison in our analysis</p> <p>We will connect the NPI numbers to our internal NPI database in order to link our providers to the providers found in CHIA’s file.</p> <p>We will also link the NPI number to the CMS NPI Registry in order to identify important attributes of specific facilities, such as primary taxonomy.</p>	

**VII. MEDICAID DATA**

Please indicate here whether you are seeking Medicaid Data:

- Yes
- No

Federal law (42 USC 1396a(a)7) restricts the use of individually identifiable data of Medicaid recipients to uses that are directly connected to the administration of the Medicaid program. If you are requesting Medicaid data from Level 2 or above, please describe, in the space below, why your use of the data meets this requirement. Requests for Medicaid data will be forwarded to MassHealth for a determination as to whether the proposed use of the data is directly connected to the administration of the Medicaid program.

The financial reserves required to manage a cancer diagnosis are substantial. (Wagner 2004) From diagnosis through survivorship to end-of-life care, the direct and indirect costs of cancer care can cause significant hardship for families without the financial resources to buffer the additional expenses. (Jagsi, et al. 2014) (Ramsey, et al. 2013) In fact, some studies have found that financial strain at the time of diagnosis has a consistent independent association with the well-being of patients with cancer, indicating an association between high levels of financial strain and increased symptom burden as well as a poorer quality of life at the time of diagnosis. (Lathan 2016) The role of financial strain and its effects on health and health behaviors has been evaluated previously by Hudson et al, (2013) who examined the impact of race and socioeconomic position on self-rated health and depression and found that both racial discrimination and lower socioeconomic position over the life course were related to increased depressive symptoms and poorer self-reported health status. Thus, in order to truly understand variations in outcomes of cancer care delivery among different providers and cancer types across the State and the full spectrum of care of DFCI’s patients to identify potential care gaps and opportunities to improve, access to Masshealth claims data will be crucial.

**VIII. DATA LINKAGE AND FURTHER DATA ABSTRACTION**

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

1. Do you intend to link or merge CHIA Data to other datasets?

- Yes
- No linkage or merger with any other database will occur

2. If yes, please indicate below the types of database to which CHIA Data 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 level (e.g., American Hospital Association data)
- Aggregate Data (e.g., Census data)
- Other (please describe):

3. If yes, describe the data base(s) to which the CHIA Data will be linked, which CHIA data elements will be linked; and the purpose for the linkage(s):

The data will be merged with the Census ZIP Code Tabulation Areas in order to obtain demographic attributions.

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.

The linkage method is deterministic.

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

Confidentializing data is a core element of this grant and data management plan, particularly given the fact that DFCI manages most of the rare Cancers of the state and small numbers could trigger some risks of identification. There are a few techniques that our team uses frequently to keep the confidentiality at the same time that we maximize the usefulness of the data.

A) Data aggregation with cell suppression by:

- a. Frequency rule: Those Tables with counts less than 10 would be considered to pose a disclosure risk, hence will be protected and would be shared.
  - b. Primary and Secondary suppression: Some of the summarization Tables could use Data suppression techniques by deleting individual records or data items from the data file. This rule shall apply to Primary and secondary and consequential suppression elements.
- B) Combining or collapsing categories: The amount of classification detail available in a table shall be reduced, particularly in those areas in which the Table might get very detailed and the counts might be too small.
- C) Sampling and Random rounding are additional methods that shall be used when presenting aggregated data in those areas in which by combining Tables with common elements, it may be possible to identify certain individuals.

#### **XI. PUBLICATION / DISSEMINATION / RE-RELEASE**

1. Describe your plans to publish or otherwise disclose CHIA Data, or any data derived or extracted from such CHIA Data, in any paper, report, website, statistical tabulation, seminar, conference, or other setting. 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 display a cell less than 11, and no percentages or other mathematical formulas will be used if they result in the display of a cell less than 11.

We most likely will be publishing our results in Medical Oncology Journals such as the New England Medical Journal (NEMJ), in peer reviewed journals from professional organizations such as the American Society of Clinical Oncology (ASCO), in clinical journals such as Journal of Clinical Oncology (JCO) or Journal of National Comprehensive Cancer Network, and/or in policy journals such as Health Affairs.

2. Do you anticipate that the results of your analysis will be published and/or publically available to any interested party? Please describe how an interested party will obtain your analysis and, if applicable, the amount of the fee, that the third party must pay.

An interested party in obtaining our analysis will be able to contact one of the four investigators directly for more information. No fee will be attached to accessing our results.

3. Will you use CHIA Data for consulting purposes?

- Yes
- No

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

- Yes
- No

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

- Yes

No

6. Will you be reselling CHIA Data in any format?

Yes

No

If yes, in what format will you be reselling CHIA Data (e.g., as a standalone product, incorporated with a software product, with a subscription, etc.)?

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

8. If you have answered “yes” to questions 4, 5, or 6, what is the fee you will charge for such products, services or studies?

## X. APPLICANT QUALIFICATIONS

1. Describe your qualifications (and the qualifications of your co-investigators) to perform the research described.

**Principal Investigator: Belen Fraile, MD, MSc**

Dr. Fraile serves as Senior Director of Population Health Management at Dana-Farber Cancer Institute (DFCI). She leads the Oncology Population Health Management program at DFCI and is part of the leadership team focused on health care delivery innovation and responsible for the development and support of system-wide programs to improve value and performance such as assessing the quality of care and efficiency of a shared care model or piloting alternative payment options for Oncology. She holds expertise in Health Reform implementation in Massachusetts and the US, Health Care Delivery Models, payment models in Health Care and value measurement in Health Care

**Co-Investigator: Yichen Zhang, MS, PhD**

Ms. Zhang is a PhD in Health Services Research with a concentration in health economics, and has dual-masters in Applied Economics and Management Science. Serving as a health economist at the Dana Farber Cancer Institute, she is proficient in Cost-Effectiveness analyses (CEA), econometrics modeling, and statistical programming. She has skills with experimental design, large data management and analysis with key statistical principles and statistical software including SAS, Stata, and Microsoft Excel/Access. Her research experience includes both qualitative and quantitative studies, such as multivariate analysis, longitudinal and dependent variable analysis, survival analysis, propensity score matching (PS), instrument variable and structurally equation model (SEM).

She has participated in Economics Evaluation of Oncology Clinical Pathway, Treating Obesity in Underserved Rural Areas Weight Control Program; P-20 Diabetes Self-Management; Effects of Children Case-mix Adjustments on Home

Care Payments for Medicaid Personnel Care Services; EMCAP Capacities of VA Emergency Management and Evaluation; The Perioperative Surgical Home with American Society of Anesthesiologists; and the East Texas Medical Center Emergency Department Referral Evaluation project.

**Study Manager: Stephanie Gottsch, MPH**

Ms. Gottsch is the Program Manager for Dana-Farber’s Department of Population Management and Value. She holds a Master’s of Public Health in Health Policy and Management and has held various financial management roles during her time at DFCI. She currently serves as the project manager for multiple projects within her department. She will be responsible for coordination and collection of data as well as assisting in the analytic and measurement strategy of this project.

**Consultant: Shaobin Dong, MBA**

Mr. Dong is a consultant for Dana-Farber’s Department of Population Management and Value. He has over 14 years of experience in the health care delivery system. He holds expertise in areas including medical informatics, population risk adjustment and statistical programming. He was one of the key contributors to several pilot programs, including bundled payment, patient center specialty program, aimed to reduce medical cost while improving quality. Mr. Dong holds a MBA and is a SAS Certified Advanced Programmer.

2. **Attach** résumés or curricula vitae of the Applicant/principal investigator, and co-investigators. (These attachments will not be posted on the internet.)

**XI. USE OF AGENTS AND/OR CONTRACTORS**

**Please note: 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.**

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

Company Name:	
Contact Person:	
Title:	
Address:	
Telephone Number:	
E-mail Address:	
Organization Website:	

1. Will the agent have access to the CHIA Data at a location other than your location, your off-site server and/or your database?

- Yes, a separate Data Management Plan **must** be completed by each agent who will store CHIA Data
- No

2. Describe the tasks and products assigned to this agent for this project; their qualifications for completing the tasks; and the Organization’s oversight of the agent, including how the Organization will ensure the security of the CHIA Data to which the agent has access.

--

**XII. FEE INFORMATION**

Please consult the [fee schedules](#) for MA APCD Data and select from the following options:

- Researcher
- Others (Single Use)
- Others (Multiple Use)

Are you requesting a fee waiver?

- Yes
- No

If yes, please refer to the [Application Fee Remittance Form](#) and submit a letter stating the basis for your request (if required). Please refer to the [fee schedule](#) for qualifications for receiving a fee waiver. If you are requesting a waiver based on the financial hardship provision, please provide documentation of your financial situation. Please note that non-profit status alone isn't sufficient to qualify for a fee waiver.

By submitting this Application, the Data Applicant attests that it is aware of its data use, privacy and security obligations imposed by state and federal law *and* is compliant with such use, privacy and security standards. The Data Applicant further agrees and understands that it is solely responsible for any breaches or unauthorized access, disclosure or use of any CHIA Data provided in connection with an approved Application, including, but not limited to, any breach or unauthorized access, disclosure or use by its agents.

Applicants requesting data from CHIA will be provided with data following the execution of a Data Use Agreement that requires the Data Applicant to adhere to processes and procedures aimed at preventing unauthorized access, disclosure or use of data.

**By my signature below, I attest to: (1) the accuracy of the information provided herein; (2) that the requested data is the minimum necessary to accomplish the purposes described herein; (3) the Data Applicant will meet the data privacy and security requirements describe 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 (4) my authority to bind the organization seeking CHIA Data for the purposes described herein.**

Signature: (Authorized Agent)	
Printed Name :	Belen Fraile
Title:	Senior Director of Population Management and Decision Support
Signature: (Applicant/Primary Investigator)	
Name:	Belen Fraile
Title:	Senior Director of Population Management and Decision Support
Original Data Request Submission Date:	6/3/16
Dates Data Request Revised:	8/9/16

Attachments. Please indicate below which documents have been attached to the Application and uploaded to IRBNet:

- 1. IRB approval letter or summary of project (if applicable)
- 2. Resumes of Applicant and co-investigators
- 3. Data Management Plan (for each institution that will store CHIA Data)