

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

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

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

**I. GENERAL INFORMATION**

APPLICANT INFORMATION	
Applicant Name:	Dr. Turgay Ayer
Title:	Assistant Professor
Organization:	Georgia Institute of Technology H. Milton Stewart School of Industrial and Systems Engineering
Project Title:	Upcoding Bias and Prediction Accuracy of Affordable Care Act (ACA) Risk Adjustment Models
Mailing Address:	417 Groseclose Bldg. 765 Ferst Dr. Atlanta, GA 30332
Telephone Number:	(404) 385-6038
Email Address:	ayer@isye.gatech.edu
Names of Co-Investigators:	Zhaowei She
Email Addresses of Co-Investigators:	zshe3@gatech.edu
Original Data Request Submission Date:	April 25, 2016
Dates Data Request Revised:	
Project Objectives (240 character limit)	We want to investigate the potential dilemma between upcoding bias and prediction accuracy in current HHS-HCC ACA risk adjustment models and the possibility of designing better risk adjustment models.
Project Research Questions (if applicable) or Business Use Case(s):	<ol style="list-style-type: none"> <li>1. Do current ACA risk adjustment models minimize the upcoding incentives within the current level of prediction accuracy?</li> <li>2. Do the current ACA risk adjustment models achieve the level of prediction accuracy such that major prediction biases in both individual and group levels are ruled out?</li> <li>3. Optimize the risk adjustment models to balance</li> </ol>

	effects of upcoding bias and prediction accuracy.
--	---

**II. PROJECT SUMMARY**

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

The main objective of this project is to empirically evaluate the current ACA risk adjustment model's ability to maximize prediction accuracy as well as minimize upcoding bias. Detailed patient level information regarding diagnosis coding, utilization and insurance types are needed to perform the proposed evaluation. Specifically,

1. To investigate the upcoding incentives under ACA risk adjustment models, we will empirically test whether the Risk Adjustment Covered Plans (RACP) have significantly different diagnosis coding compared to the diagnosis coding of other health plans. Moreover, we will conduct the comparison in medical condition level to see which conditions tend to suffer from the upcoding behaviors.
2. To investigate the prediction accuracy of ACA risk adjustment models, we will first test whether there are omitted variables which bias the model predictions. Subsequently, we will test there are individual level heterogeneity in the current models which may encourage adverse selection behaviors among consumers.
3. Based on the empirical findings from the first two questions, we will develop an optimization model to choose the best subset of risk adjusters which simultaneously maximizes model prediction accuracy and minimizes upcoding bias.

The Center for Health Information and Analysis (CHIA) data provides detailed patient level information to study the evaluation and design of risk adjustment models. The Medical Claims (MC) file provides the necessary diagnosis coding data; the Member Eligibility (ME) file contains the necessary demography data to calibrate risk adjustment models; the Product (PR) file helps us to identify incentives of upcodings from insurance types; the Provider (PV) file is needed to test for potential adverse selection behaviors among consumers.

**III. FILES REQUESTED**

Please indicate which MA APCD file(s) you are requesting, the year(s) of data requested, and your justification for requesting each file. Please refer to the MA APCD Release 4.0 Documentation Guides for details of the file contents.

MA ALL PAYER CLAIMS DATABASE FILES	Year(s) Of Data Requested Current Yrs. Available
<input checked="" type="checkbox"/> Medical Claims	<input checked="" type="checkbox"/> 2010 <input checked="" type="checkbox"/> 2011 <input checked="" type="checkbox"/> 2012 <input checked="" type="checkbox"/> 2013 <input checked="" type="checkbox"/> 2014
<input checked="" type="checkbox"/> Medical Claims	<b>Please provide justification for requesting Medical Claims file:</b>  The Medical Claims (MC) file provides the necessary diagnosis coding data
<input type="checkbox"/> Pharmacy Claims	<b>Please provide justification for requesting Pharmacy Claims file:</b>

<input type="checkbox"/> Dental Claims	Please provide justification for requesting Dental Claims file:
<input checked="" type="checkbox"/> Member Eligibility	Please provide justification for requesting Member Eligibility file: The Member Eligibility (ME) file contains the necessary demography data to calibrate risk adjustment models
<input type="checkbox"/> Provider (encrypted NPI) Standard or <input checked="" type="checkbox"/> Provider* (unencrypted NPI)	Please provide justification for requesting Provider file: The Provider (PV) file is needed to test for potential adverse selection behaviors among consumers. *Please provide justificaiton for requesting unencrpted NPI (if requested). Refer to specifics in your methodology: We need to distinguish high quality providers from the low quality providers in order to identify potential adverse selection behaviors among consumers.
<input checked="" type="checkbox"/> Product	Please provide justification for requesting Product file: The Product (PR) file helps us to identify incentives of upcodings from insurance types

**IV. GEOGRAPHIC DETAIL**

Please choose one of the following geographic options for MA residents:

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

**V. DATE DETAIL**

Please choose one option from the following options for dates:

<input type="checkbox"/> Year (YYYY) (Standard)	<input checked="" type="checkbox"/> Month (YYYYMM) ***	<input type="checkbox"/> Day (YYYYMMDD) ***
---	--	---

[for selected data elements only]

**\*\*\* If requested, please provide justification for requesting Month or Day. Refer to specifics in your methodology:**

We need to perform event study to access the implications of certain ACA policy change. This type of studies require at least data at the monthly level to perform.

**VI. FEE INFORMATION**

Please consult the fee schedules for MA APCD data, available at [http://chiamass.gov/regulations/#957\\_5](http://chiamass.gov/regulations/#957_5), and select from the following options:

**APCD Applicants Only**

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

Are you requesting a fee waiver?

- Yes
- No

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

**VII. MEDICAID DATA [APCD Only]**

Please indicate here whether you are seeking Medicaid Data:

- Yes
- No

Federal law (42 USC 1396a(a)7) restricts the use of individually identifiable data of Medicaid recipients to uses that are directly connected with the administration of the Medicaid program. If you are requesting Medicaid data from Level 2 or above, please describe in detail why your use of the data meets this requirement. Applications requesting Medicaid data will be forwarded to MassHealth for a determination as to whether the proposed use of the data is directly connected to the administration of the Medicaid program. MassHealth may impose additional requirements on applicants for Medicaid data as necessary to ensure compliance with federal laws and regulations regarding Medicaid.

**VIII. PURPOSE AND INTENDED USE**

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

The current ACA risk adjustment models deserve careful testing and validation. The performance of risk adjustment models can significantly affect the sustainability and consumer welfare of ACA health exchange. Specifically, if the model indeed provides incentives for upcoding behaviors, the market may unravel because the honest health plans have to pay for health plans who upcode their patients. As such, the honest plans will be driven out of the market because of the financial burdens. On the other hand, if the risk adjustment model does not provide accurate enough individual level medical expense predictions, the market can also unravel because those high quality health plans and hospitals will adversely select sicker consumers and ultimately be driven out of the market due to the high cost incurred.

Therefore, through studying the risk adjustment models, this project can help to monitor and improve the performances of ACA individual health plan market place, and thus make better health services more accessible to the general public.

2. **Attach** a brief (1-2 pages) description of your research methodology. (This description will not be posted on the internet.)
3. Has your project received approval from your organization’s Institutional Review Board (IRB)? Please note that CHIA will not review your application until IRB documentation has been received (if applicable).
  - Yes, and a copy of the approval letter is attached to this application.
  - No, the IRB will review the project on \_\_\_\_\_.
  - No, this project is not subject to IRB review.
  - No, my organization does not have an IRB.

#### IX. APPLICANT QUALIFICATIONS

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

Turgay Ayer (PI) is an Assistant Professor in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. He also has a courtesy appointment at the Winship Cancer Institute at Emory University. His research interests include development of applied mathematical and computational models for health policy and medical decision making problems. Dr. Ayer has been working and publishing in healthcare area for the past 10 years since beginning of his Ph.D., and is experienced with developing theoretical [Ayer et al., 2012, 2015a] and empirical models [Ayer et al., 2010, 2015b, Rust et al., 2016]. He has a strong collaboration history with the researchers from the medical world, and his past research on various healthcare problems has been published in top tier journals in the respective areas, such as Operations Research, Management Science, Annals of Internal Medicine, JAMA Oncology, and Journal of Clinical Oncology [Ayer et al., 2012, 2015a, b, Goldstein et al., 2015a, b, He et al., 2015].

Zhaowei She (Co-PI) is a Ph.D. student in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology, working with Dr. Ayer. Mr. She has a formal training in Economics and is interested in studying the HIX via game-theoretical and empirical analyses. The proposed work will be part of Mr. She's doctoral thesis work.

2. Attach résumés or curricula vitae of the applicant/principal investigator, key contributors, and of all individuals who will have access to the data. (These attachments will not be posted on the internet.)

**X. DATA LINKAGE AND FURTHER DATA ABSTRACTION**

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

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

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

2. If yes, will the CHIA Data be linked or merged to other individual patient level data (e.g. disease registries, death data), individual provider level data (e.g., American Medical Association Physician Masterfile) , facility level (e.g., American Hospital Association data) or with aggregate data (e.g., Census data)? [check all that apply]

- Individual Patient Level Data

What is the purpose of the linkage:

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

- Individual Provider Level Data

What is the purpose of the linkage:

We want to link individual providers to health plans in the insurance market. The purpose of this linkage is to be able to analyze the distribution of patients among different health plans.

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

The health plans data will be extracted from public accessible website from each insurance companies.

Individual Facility Level Data

What is the purpose of the linkage:

We want to link individual facilities to health plans in the insurance market. The purpose of this linkage is to be able to analyze the distribution of patients among different health plans.

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

The health plans data will be extracted from public accessible website from each insurance companies.

Aggregate Data

What is the purpose of the linkage:

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

3. If yes, for each proposed linkage above, please describe your method or selected algorithm (e.g., deterministic or probabilistic) for linking each dataset. If you intend to develop a unique algorithm, please describe how it will link each dataset .

We will link facilities and providers to the health plans they participate deterministically.

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

Since we only plan to link the identifiers of providers and facilities to their health plans, we will not and cannot identify individual patients.

5. If yes, and the data mentioned above is not in the public domain, please attach a letter of agreement or other appropriate documentation on restrictions of use from the data owner corroborating that they agree to have you initiate linkage of their data with CHIA data and include the data owner's website.

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

1. Describe your plans to publish or otherwise disclose CHIA Data, or any data derived or extracted from such data, in any paper, report, website, statistical tabulation, seminar, conference, or other setting.

The research findings will be widely disseminated to health policy and management communities. The PIs have a track record of attending both health policy and management science conferences, publishing in leading journals of these two different disciplines, and disseminating their research findings to both communities. For this specific project, in addition to leading journals in their own area such as Management Science, the team targets Health Affairs journal for publication. The team also plans to present their findings at the annual AcademyHealth conference, which is typically attended by health policy researchers and policy-makers.

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

Our results will mainly be available through academic journals. These journals may charge a certain amount of access fee.

3. Will you use the data for consulting purposes?

Yes  
 No

4. Will you be selling standard report products using the data?

Yes  
 No

5. Will you be selling a software product using the data?

Yes  
 No

6. Will you be reselling the data?

Yes  
 No

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



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

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

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

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

8. Will the agent/contractor have access to the data at a location other than your location, your off-site server and/or your database?

- Yes
- No

If yes, please provide information about the agent/contractor’s data management practices, policies and procedures in your Data Management Plan.

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

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

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

**XIII. ASSURANCES**

Applicants requesting and receiving data from CHIA pursuant to 957 CMR 5.00 ("Data Recipients") will be provided with data following the execution of a data use agreement that requires the Data Recipient to adhere to processes and procedures aimed at preventing unauthorized access, disclosure or use of data, as detailed in the DUA and the applicant's CHIA-approved Data Management Plan.

Data Recipients are further subject to the requirements and restrictions contained in applicable state and federal laws protecting privacy and data security, and will be required to adopt and implement policies and procedures designed to protect CHIA data in a manner consistent with the federal Health Insurance Portability and Accountability Act of 1996 (HIPAA).

By my signature below, I attest to: (1) the accuracy of the information provided herein; (2) my organization's ability to meet CHIA's minimum data security requirements; and (3) my authority to bind the organization seeking CHIA data for the purposes described herein.

Signature:	By: <i>D. Duane Hutchison</i>
Printed Name:	Duane Hutchison
Title	Executive Director, Office of Sponsored Programs
Original Data Request Submission Date:	
Dates Data Request Revised:	

REVIEWED BY GT LEGAL AFFAIRS: *RA*

## References

- Ayer, T., Alagoz, O., Chhatwal, J., Shavlik, J. W., Kahn, C. E., and Burnside, E. S. (2010). Breast cancer risk estimation with artificial neural networks revisited. *Cancer*, 116(14):3310–3321.
- Ayer, T., Alagoz, O., and Stout, N. (2012). A pomdp approach to personalize mammography screening decisions. *Operations Research*, 60(5):1019–1034.
- Ayer, T., Alagoz, O., Stout, N., and Burnside, E. (2015a). Heterogeneity in women’s adherence and its role in optimal cancer screening policies. *Management Science*, *in press*.
- Ayer, T., Ayvaci, M., Karaca, Z., and Vlachy, J. (2015b). The impact of health information exchanges on emergency department length of stay. *POMS*, *under second round of review*.
- Goldstein, D. A., Ahmad, B. B., Chen, Q., Ayer, T., Howard, D. H., Lipscomb, J., El-Rayes, B. F., and Flowers, C. R. (2015a). Cost-effectiveness analysis of regorafenib for metastatic colorectal cancer. *Journal of Clinical Oncology*, pages 3727–3732.
- Goldstein, D. A., Chen, Q., Ayer, T., Howard, D. H., Lipscomb, J., Ramalingam, S. S., Khuri, F. R., and Flowers, C. R. (2015b). Nectinmab in metastatic squamous cell lung cancer: establishing a value-based cost. *JAMA oncology*, 1(9):1293–1300.
- He, T., Li, K., Roberts, M. S., Spaulding, A. C., Ayer, T., Grefenstette, J. J., and Chhatwal, J. (2015). Prevention of hepatitis c by screening and treatment in us prisons. *Annals of internal medicine*, *in press*.
- Rust, G., Zhang, S., Yu, Z., Caplan, L., Jain, S., Ayer, T., McRoy, L., and Levine, R. (2016). Counties eliminating racial disparities in colorectal cancer mortality. *Cancer*. *Forthcoming*.