Risk Adjustment Dataset Documentation

Grace Guan, Adviser: Mark Braverman

October 9, 2018

1. Part I: BASIC INFORMATION

1.1. Contact

E: Grace Guan XXXX@cs.princeton.edu and Mark Braverman XXXXXX@cs.princeton.edu replacing X's with gg17 and mbraverm respectively.

1.2. Title

Affordable Care Act Insurer Risk Adjustment Data

1.3. Brief Description

For each year, we provide a combination of all "Insurer Report" Excel files and "Summary Report" PDF files. The "Insurer Report" Excel files include companies' basic information and insurer financial data; the "Summary Report" PDF files include accurate risk adjustment and reinsurance values.

1.4. Dates of Data Collection

2014, 2015, and 2016 Affordable Care Act Benefit Years.

2. Part II: DESCRIPTION OF DATA

2.1. Data Source

The publicly-available insurer data used was downloaded from the CMS website. The 2014 and 2015 benefit year data was downloaded in late 2017, and the 2016 benefit year data was downloaded in early 2018; the 2014 and 2015 data has not changed since its downloading in late 2017. The datasets were created for 2014, 2015, and 2016. 3,472 companies were represented in the dataset in 2014; 3,070 in 2015; and 2,623 in 2016. There were 1,563 columns scraped for 2014; 1,589 for 2015, and 1,593 for 2016.

The Insurer Report data was downloaded from https://www.cms.gov/apps/mlr/mlr-search.aspx. The Summary Report data was downloaded from https://www.cms.gov/CCIIO/Programs-and-Initiatives/Premium-Stabilization-Programs/index.html.

2.2. Data Attributes

The data includes:
1. "Insurer Report" "Company Information" Sheet - including Company Name, HIOS ID, Nonprofit Status, State
2. "Insurer Report" "Summary of Data" Sheet - including Premiums, Claims, Taxes, Healthcare Quality Improvement Expenses, Non-Claims Costs, Income from Uninsured Plans' Fees, Amount Insured (Member Months, Number of Covered Lives, etc.), and Net Investment Income

3. "Insurer Report" "Premiums and Claims" Sheet - including more detailed Premiums and Claims
4. "Summary Report" Risk Adjustment Data - including Individual Group Reinsurance and Risk Transfers, and Small Group Risk Transfers

2.3. Data Format

All files are in the CSV format. Each row represents a company operating in a state. Each column represents a quantity, such as risk adjustment in the individual group, or member months in the small group market. The "Summary Report" and "Insurer Report" files were merged using an outer join on "HIOS ID," the 5-digit Health Insurance Oversight System Identification number for each company.

The names for columns corresponding to data from the "Summary Report" PDF Files are the names as they appear on the file: "HIOS ID," "HIOS INPUTTED INSURANCE COMPANY NAME," "STATE," "REINSURANCE PAYMENT AMOUNT (OR NOT ELIGIBLE)," "HHS RISK ADJUSTMENT TRANSFER AMOUNT (INDIVIDUAL MARKET, INCLUDING CATAS-TROPHIC)," "HHS RISK ADJUSTMENT TRANSFERS AMOUNT (SMALL GROUP MAR-KET)." These comprise the first six columns in our dataset.

The names for columns corresponding to data from the "Insurer Report" Excel Files are con-catenated row and column names from that file without spaces. For example, if the "Insurer Report" row was "1.1 Direct premium written" and column was "2 Health Insurance INDIVIDUAL Total as of 3/31/17", then the column name in our dataset would be "1.1Directpremiumwrit-ten2HealthInsuranceINDIVIDUALTotalasof3/31/17". These comprise the remaining columns in the dataset.

2.4. Data Limitations

This dataset is limited by the number of companies that decided to report their estimated risk adjustment values. While the original PDF data sums to zero for every state, when matched with the Excel data, the actual transfer payments do not sum to zero.