Creation and Analysis of a Medical Loss Ratio Dataset

Grace Guan
Adviser: Professor Mark Braverman

Background

- **Reinsurance** = payment to plans w/ higher-cost enrollees
- **Risk Adjustment** = redistribution of money from plans with lower-cost enrollees to plans with higher-cost enrollees through “transfer payments”
- **Medical Loss Ratio (MLR)** = benchmark for insurers to provide value to enrollees
    - If an insurer uses 80 cents per premium dollar to pay for its customer’s medical claims, the MLR will be 80%; Gov’t sets MLR minimums

Motivation and Goal

- **Original Project:** A Big Data Examination of the Accuracy of Risk Adjustment under the Affordable Care Act (ACA)
- **Original Goal:** Backwards engineer the plan liability risk score, risk adjustment transfer payment formula, and actuarial value calculation
- **Motivation:** Understand how transfer payment formula works

Motivation and Goal

- No data! --> **New Project:** Creation and Analysis of a Medical Loss Ratio (MLR) Dataset
- **New Goal:** Create a clean, new dataset from all of the data online and do some preliminary analysis on it
- **Motivation:** Data has never been cleaned/analyzed before; analyzing this new data may provide insights into how to make risk adjustment more efficient

Significance

- Online “Public Use Files” Excel files are incomplete
- Online “Summary Report” PDF files are hard to parse
- Online “Insurer Report” Excel files are split between >20,000 files, one for each company, and only have the expected (not actual) values of risk adjustment
- There is no easy way to analyze all of this data!

Data ETL

“Insurer Report” Excel Files
Contains insurer-reported estimates of risk adjustment data (10,000+ fields per sheet)

“Summary Report” PDF File
Contains accurate risk adjustment data

“Public Use” PDF File
Supposed to contain all data from “Insurer Report”

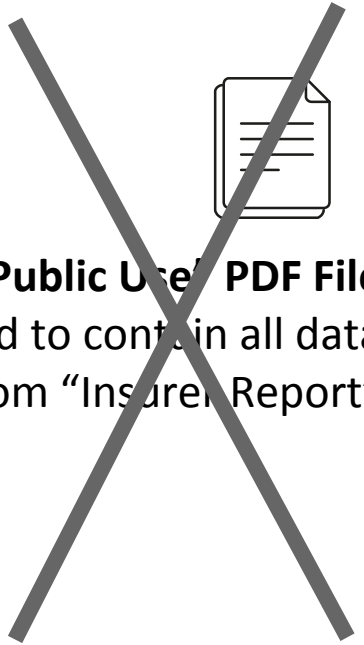
Approach: Data ETL

- Input: >20,000 “Insurer Report” Excel files, scraped from online by Jeremie Lumbroso
- Take 5260 fields of expected values from each Excel file and add as a row into our CSV
- Combine with “Summary Report” PDF file
- Output: 1 CSV file for each year containing risk adjustment data with both expected and actual values for all companies

Data ETL

“Insurer Report” Excel Files
Contains insurer-reported estimates of risk adjustment data (10,000+ fields per sheet)

“Summary Report” PDF File
Contains accurate risk adjustment data



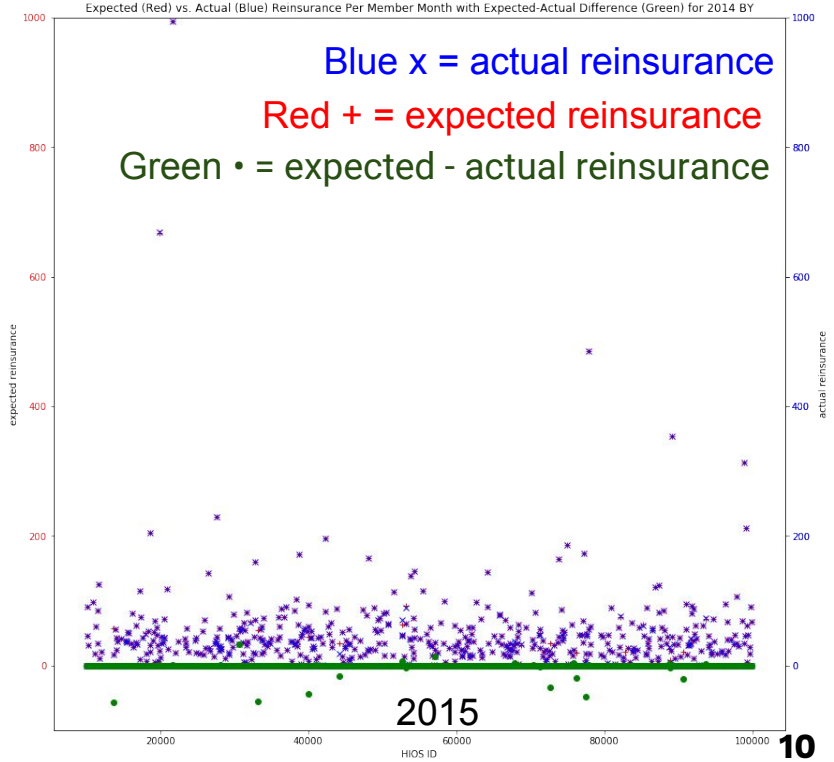
“Public Use” PDF File
Supposed to contain all data from “Insurer Report”

Our output file, one per year, easier to analyze

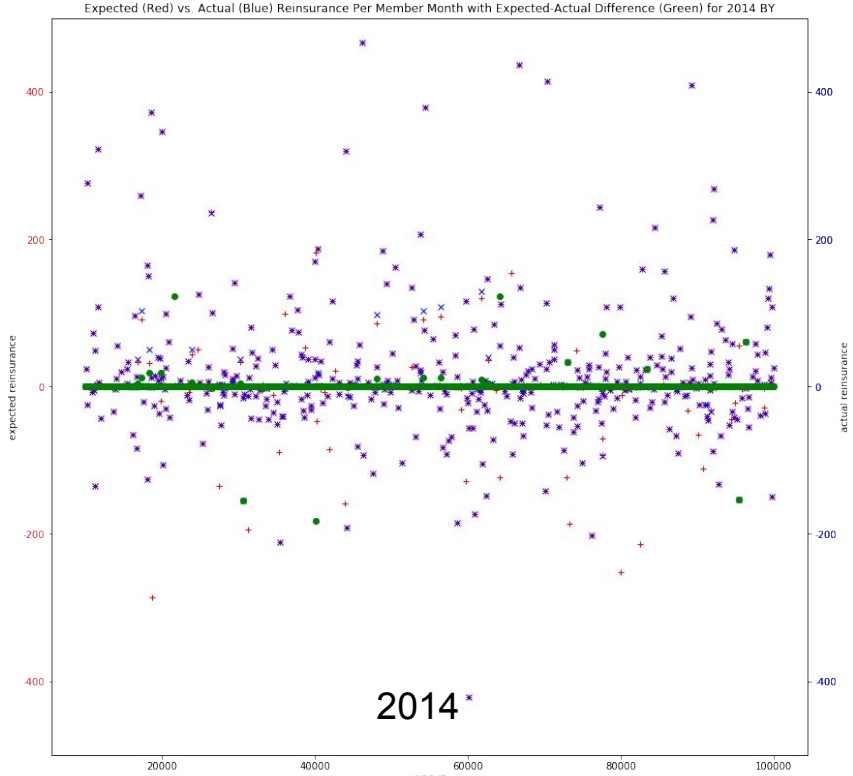
Data Analysis

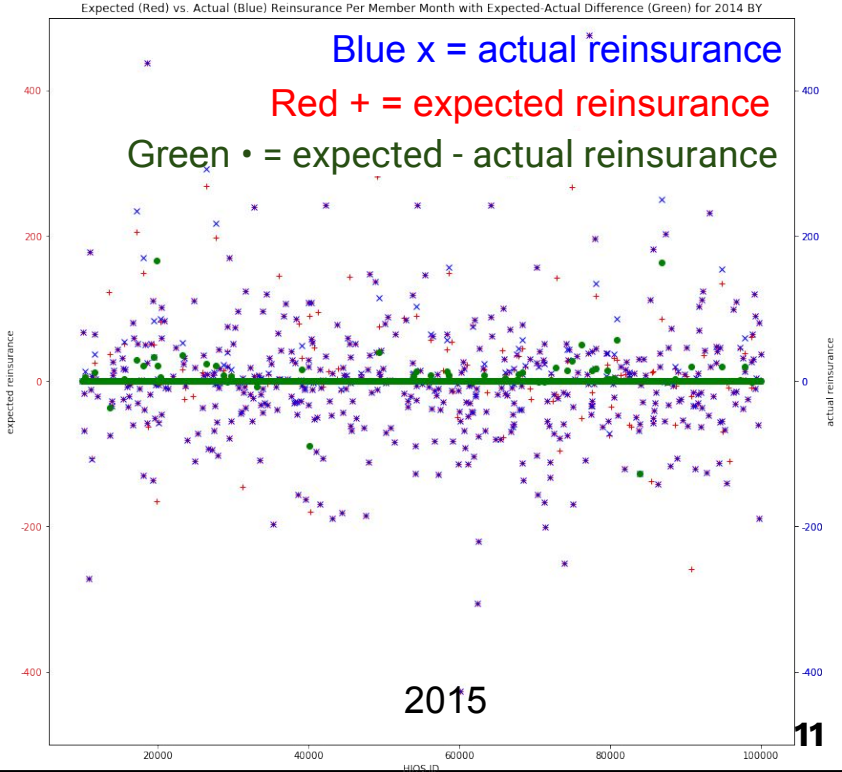
- Year to Year (YTY) per member month (# enrollees * # months) analysis for reinsurance, individual risk adjustment, small group risk adjustment
- Mean/median/etc. basic statistical analysis
- Correlation between total loss & change in risk adjustment

YTY analysis: reinsurance PMM

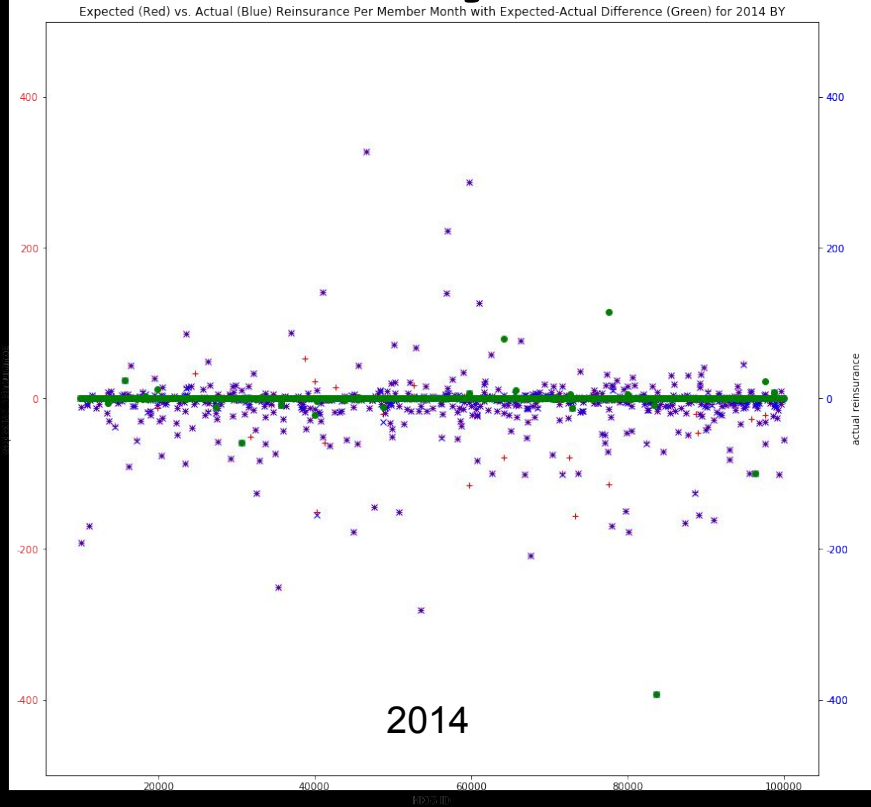


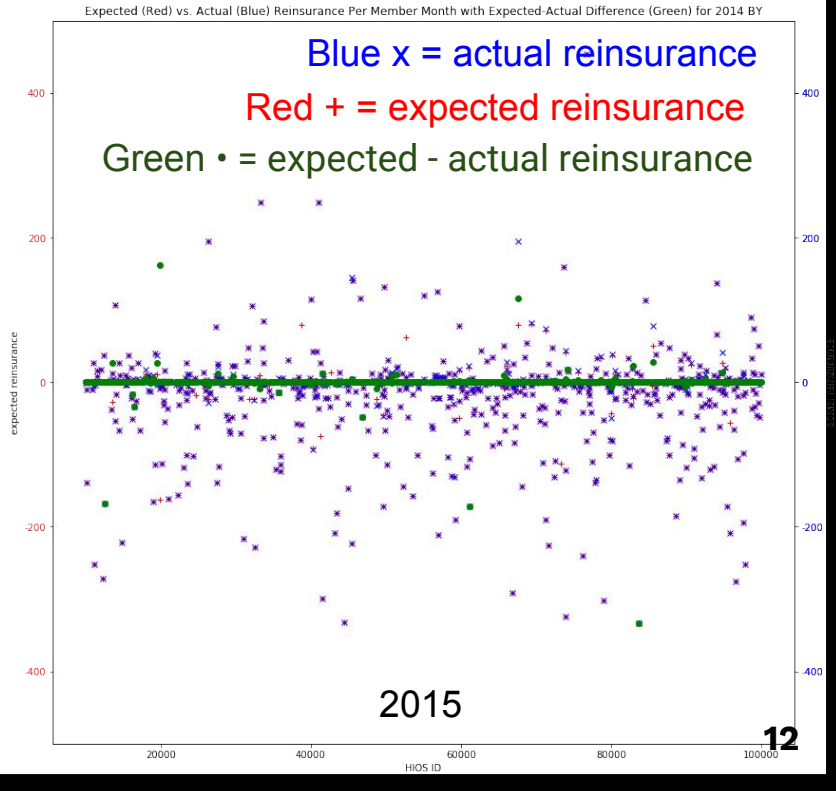
YTY analysis: indiv. RA PMM





YTY analysis: small group RA PMM





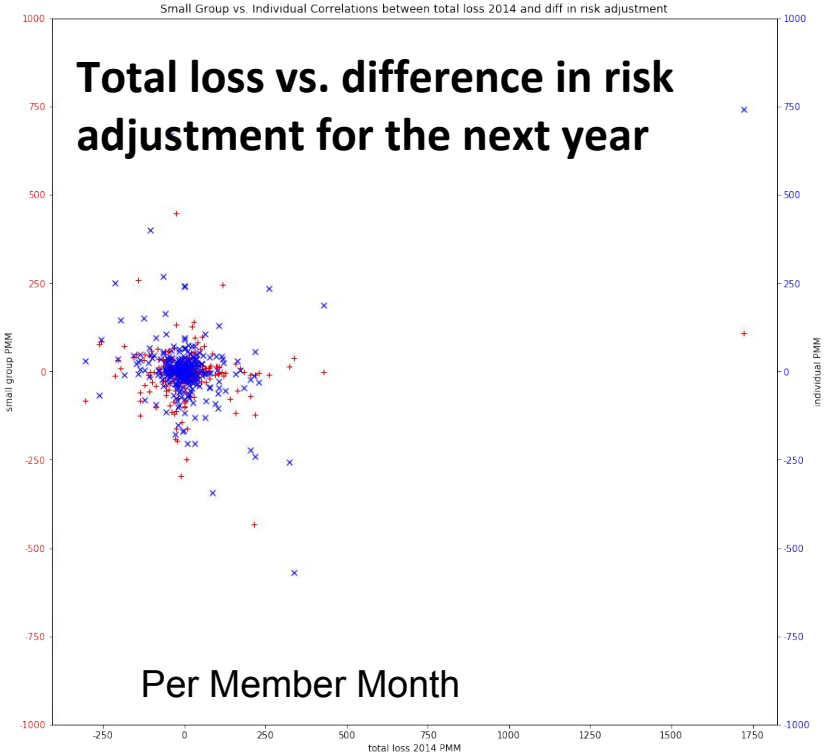
Expected Minus Actual, PMM	Mean 2014	Mean 2015	Std Dev 2014	Std Dev 2015
Reinsurance	0.48	-0.17	19.31	3.23
Individual Risk Adjustment	0.03	1.26	9.18	21.14
Small Group Risk Adjustment	-0.42	-0.37	14.72	16.88

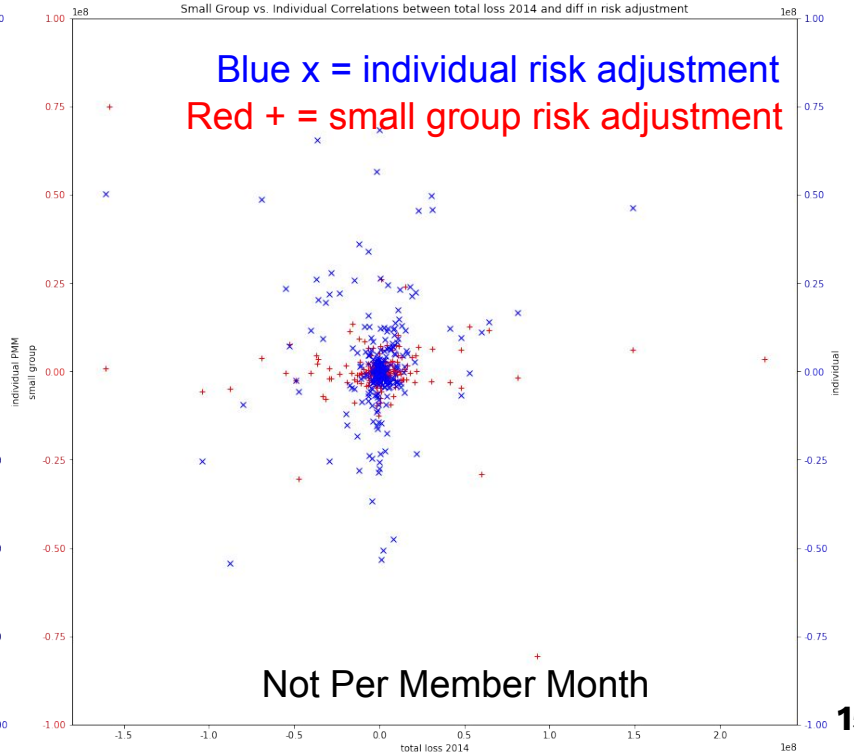
Is Total Loss in 2014 Correlated with Change in Risk Adjustment, 2015 - 2014?

- Basically no correlation! Not what we expected

	Individual RA	Small Group RA	Individual PMM RA	Small Group PMM RA
Correlation (r) w/ Total Loss	-0.091	-0.013	0.193	0.024

Can you see a correlation?





Conclusion

- Government “Public Use” data is relatively inaccessible
- Risk Adjustment can be made more efficient
- Tangible Result: Dataset CSV files on GitHub!
- And some analysis...

- GitHub: http://tiny.cc/riskadjustment
- Thanks to Prof. Braverman and Jeremie
- Questions?