

Retro-Corrected Enrollment Data Comes to the GHRI Data Warehouse

Authors: Roy Pardee, JD, MA¹ & Luesa Jordan, BA¹

Affiliations: ¹ Group Health Research Institute



Introduction

- ▶ Historically, GHRI enrollment data consists of monthly extracts from Membership & Billing.
 - Everyone enrolled on the day of the extract is credited with enrollment for the month.
 - Any post-hoc changes to people's enrollment status, snapshots are lost.
- ▶ This is good for being able to reproduce official monthly enrollment reports after the fact.
- ▶ But bad for exactly describing the sets of people enrolled in a given month very accurately.

Enter Retro

- ▶ In order to address this problem, EIM created a new content area on the Enterprise Data Warehouse: Retro Enrollment.
- ▶ This area makes post-hoc corrections over a sliding six-month window.
 - So each month, the enrollment figures for the previous five months are restated based on current knowledge.
- ▶ When they happen, corrections are typically to:
 - coverage start/stop dates,
 - group memberships, and
 - Primary Care Clinic/Physician elections.
- ▶ Unfortunately, the retro data on EDW has some shortcomings—in particular:
 - It doesn't start until 2002.
 - It is structured *very* differently from traditional M&B extract data, making it very difficult to combine.
- For these reasons (and due to competing priorities) GHRI has not integrated retro into our local data infrastructure.

How Different Is Retro?

In order to assess how different the two types of enrollment data are, we identified the group of people who showed as enrolled at least one month in 2010 in either file (N = 714,902). We then compared how well the two data sources agreed on:

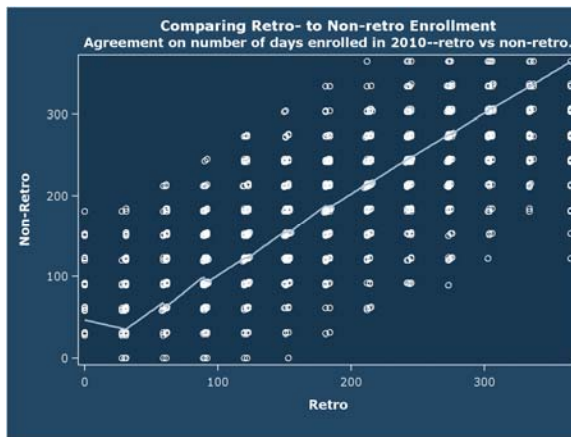
- ▶ Who was/was not enrolled at all in 2010.
- ▶ The number of days in 2010 that they were enrolled

Table 1: Cohort Agreement

Frequency Percent Row Pct Col Pct	Has NON-retro enrollment in 2010?		
	No	Yes	Total
Has retro enrollment in 2010?			
No	0 0.00 0.00	6,701 0.94 100.00	6,701 0.94
Yes	1,129 0.16 100.00	707,072 98.90 99.84	708,201 99.06 99.06
Total	1,129 0.16	713,773 99.84	714,902 100.00

- ▶ Agreement was generally strong—in 98.9% of cases, the two sources agreed on the fact of 2010 enrollment.
- ▶ However as you can see from the retro-no/nonretro-yes cell highlighted in blue above, almost one person in a hundred were not in fact enrolled in 2010 as the uncorrected data would lead you to believe.
- ▶ Conversely, a smaller number of people who were in fact enrolled would be excluded using the non-retro data.

Figure 1: Agreement in duration of enrollment



- ▶ This scatter plot shows the number of enrolled days in 2010 for everyone in our cohort, as given by retro and non-retro data.
- ▶ You can see that because of the monthly granularity of our enrollment data (something retro does *not* improve upon) the points only take a relatively few values. That's what produces the polka-dot pattern—there is a lot of overplotting.
- ▶ The line on the graph shows the LOWESS smoother (smoothing parameter = 5%) to help show where the weight of the observations lie.

The New File: Retro MBHist

- ▶ The new file available on the GHRI data warehouse is a knitting together of the M&B snapshot data we are used to using + reshaped data from the retro enrollment data area on the EDW.
 - Periods prior to 2002 come from MBHist.
 - Periods from 2002 to present come from retro.
- ▶ The structure of the new file matches MBHist.
 - There are some caveats to this—specifically, not every M&B variable made it into the retro version. Please see the upcoming Data Insider article for details.
- ▶ The extent-in-time of the new file matches MBHist.
- ▶ So programmers should in many cases be able to swap Retro MBHist in to most any program that used the old MBHist.
- ▶ In addition to being a good replacement for old MBHist, the retro version features some new MyGroupHealth usage variables that give added benefit:
 - ▶ Registered_mygh_user_flag
 - ▶ Verified_mygh_user_flag

Future Directions

- ▶ As this data is further vetted through use and analysis, we expect to begin sourcing VDW Enrollment from it rather than MBHist.
- ▶ We are also looking into using a new source of EIM data—a daily census of members—to try and increase the resolution of our enrollment data.
- ▶ The big thing on the horizon is of course the retirement of the M&B system and the beginning of Ika Systems.