MERATIVE MARKETSCAN® RESEARCH DATABASES

Commercial Insurance Weights User Guide

Data Year 2022

Merative
100 Phoenix Drive
Ann Arbor, Michigan 48108

The file should be cited as follows:
SOURCE: Merative™ MarketScan® Research Databases.

Contents

Background 1

Definition of commercial insurance weights and strata 2

Overview of commercial insurance weight calculations 3

- Application of weights 3

Selection of a MarketScan enrollment sample 4

- Sample selection criteria 4
- Use of annual or detailed enrollment tables for sampling 5
- Example of weight construction 6

Sample programs and tools 8

- Structure of the weight key 8
- Available data formats 10

Appendix: Analysis of annual and monthly enrollment data 11

- Introduction 11
- Weight computation using annual and detailed enrollment 11
- Differences in estimates based on annual and detailed enrollment 13
- Comparison of projected costs: annual, monthly, and adjusted annual weights 15
- Conclusion 17

Background

The Merative™ MarketScan® Commercial Claims and Encounters (CCAE) Database and Medicare Supplemental and Coordination of Benefits (MDCR) Database contain data from individuals who have employer-sponsored insurance (ESI) as either primary coverage or Medicare supplemental coverage. To project this population to the national population of individuals with ESI, the MarketScan Commercial Insurance Weights were constructed using the Public Use Microdata Sample (PUMS) of the American Community Survey (ACS) conducted by the U.S. Census Bureau.[1] [2] These estimates are designed to weight individuals in the MarketScan databases to reflect the national ESI population by demographic stratum; health care use varies appreciably across demographic strata.

The MarketScan Commercial Insurance Weights currently project to the total U.S. ESI population. This group represents a large market share of health care services, drugs, and medical devices.

[1] U.S. Census Bureau. American Community Survey (ACS).

[2] In prior years, national projections were derived from the Agency for Healthcare Research and Quality (AHRQ) Medical Expenditure Panel Survey (MEPS).

Definition of commercial insurance weights and strata

To construct the commercial insurance weights, ACS respondents first are selected as those indicating that they have ESI at the point in time when the survey is completed. Specifically, survey respondents are asked whether they currently have "Insurance through a current or former employer or union (of this person or another family member)." Each respondent with ESI then is classified using combinations of demographic variables ("strata" or "demographic cells") that account for substantial differences in utilization and expenditures. These variables are as follows:

- → Census Division
  New England, Middle Atlantic, East North Central, West North Central South Atlantic, East South Central, West South Central, Mountain, Pacific
- → Age Group
  0 years, 1–4 years, 5–9 years, 10–14 years, 15–17 years, 18–24 years, 25–29 years, 30–34 years, 35–39 years, 40–44 years, 45–49 years, 50–54 years, 55–59 years, 60–64 years, 65–74 years, 75+ years
- → Sex
  male, female
- → Relationship to the insurance policy holder
  policy holder, spouse/dependent

ACS respondents then are weighted (using ACS population weights supplied with the PUMS) and summed to estimate the U.S. ESI population in each stratum.[3]

If the number of MarketScan individuals in each demographic cell is aggregated for the year (several methods for which are described in a subsequent section), the commercial insurance weight then can be calculated as the ratio of the ACS national ESI estimate in a stratum to the number of MarketScan individuals in the same stratum.

$$\text{Commercial Insurance Weight} = \frac{\text{ACS-estimated ESI population}}{\text{MarketScan-observed convenience sample}}$$

[3] The ACS releases PUMS files capturing 1-year, 3-year, and 5-year population estimates; we provide only 1-year population estimates.

Overview of commercial insurance weight calculations

Application of weights

The commercial insurance weight computations are applicable to the enrolled CCAE and MDCR databases. In all, 416 strata are used to construct the weights for the CCAE enrollees, and 24 strata are used for the MDCR enrollees.

Creation and use of commercial insurance weights with MarketScan typically involve the following steps:

1. Summarize MarketScan enrollment counts by the stratum used for the commercial insurance weights. Strata are identified by the variable "MSWGTKEY" on the MarketScan enrollment tables.
2. Look up the ACS ESI population estimate in the lookup tables for each stratum represented in the MarketScan enrollment summary from step 1.
3. Compute a commercial insurance weight for each stratum as the ratio of the ACS ESI population estimate to the MarketScan enrollment count.
4. Apply the stratum-specific commercial insurance weight to each enrollee in the enrollment table.

Users can apply the weights to enrollment and utilization measures at the enrollee × time period level. Although weights are calculated on the basis of annual ACS-to-MarketScan ratios, they can be applied to more granular time dimensions, such as month or quarter. The weight category for each enrollee is obtained from the MarketScan annual or detailed enrollment file. The appropriate ACS ESI population estimate is obtained from a lookup table using the MarketScan Commercial Insurance Weight Link variable ("MSWGTKEY") as the lookup key.

Selection of a MarketScan enrollment sample

Sample selection criteria

Because the user calculates the commercial insurance weights on the basis of a MarketScan enrollment sample, the selection of that sample is a key consideration. Whatever sample is selected will generate weights that project enrollees to the national ESI population corresponding to the stratum represented in the enrollment sample. If the objective of using the weights is to calculate national estimates of metrics (enrollment, spending, etc.), then the most common enrollment sample selection criterion is MarketScan data contributor status (employer, health plan).[4]

We emphasize this point because it is quite possible to generate weights that do not function as desired. For instance, suppose we select a MarketScan enrollment sample for a specific state and then construct commercial insurance weights from that enrollment sample. Use of such weights will produce totals or averages projecting MarketScan enrollment to the census division in which the state is located, quite likely not the original intent. Although it is possible to project use or expenditure figures from a single state to the entire U.S. ESI population, one must assume that for the outcome being projected, the average per enrollee value in the state being analyzed is representative of the average per enrollee value across the entire U.S. ESI population. It is up to the analyst to judge the appropriateness of such an assumption.

Many studies involve a subset of the MarketScan sample, such as those who are continuously enrolled, have prescription drug coverage, or have a specific chronic condition. We recommend that analysts first create commercial insurance weights using an enrollment sample representative of the entire ESI population, then use appropriate statistical software to analyze the subset of data of immediate interest. Sample programs described below illustrate how this can be accomplished for the subset of enrollees with diabetes.

---

[4] In addition, any MarketScan enrollee with missing information on census division, age, sex, or policy holder status will not be assigned a stratum or weight, and thus not projected. Note that weights are not available for the under age 65 MDCR population.

Use of annual or detailed enrollment tables for sampling

MarketScan enrollment information is supplied in two forms: (1) a detailed, monthly table of demographic and coverage information for each enrollee and (2) an annual enrollment summary table. The annual enrollment table is derived from information on the detailed enrollment table. Two variables used in the commercial insurance weight stratum—age and census division—may change from month to month in the detailed table as people age or relocate. By convention, the annual table assigns age on the basis of the individual's age on his or her first enrolled date of the year (often January 1), whereas census division is derived from the modal value of state of residence on the detailed table.

Although an individual in MarketScan may change stratum within the calendar year, ACS ESI population estimates are available only on an annual basis. Therefore, users of the MarketScan Commercial Insurance Weights can select among several methods of aggregating MarketScan data for the purpose of calculating commercial insurance weights:

1. Annual weight
   Use the annual enrollment table to compute weights, on the basis of the **stratum** assigned to the enrollee in that table ("MSWGTKEY"). This means that all enrollment and outcome variables will be projected to the national ESI population on the basis of the assignment of census division, age group, sex, and policy holder status found in the annual enrollment table.
2. Monthly weight
   If the analyst is interested in conducting analyses on a monthly scale, the detailed enrollment table can be used to compute weights on the basis of the **stratum and month** assigned to the enrollee in that table. This option exploits the accuracy of the MarketScan detailed enrollment table but requires a uniformity assumption for the ACS ESI population estimates. Also, claims data will need to be weighted using information on the month of service for the claim as well as the stratum for the enrollee in the same month.
3. Adjusted annual weight
   A third option is to create an annual weight from the detailed enrollment table. To do this, summarize the enrollment count by stratum and month from the detailed enrollment table and then average the enrollee counts over the 12 months for each stratum.

We have explored the impact of using different weighting approaches on total spending estimates. The results are presented in the Appendix, along with formulas that define the weighting approaches just described. Overall, we found that the largest differences among the three weight calculation options on enrollment and spending estimates were in the newborn

(age 0 years) age group; census division differences in estimates were much smaller in magnitude. The adjusted annual weights appear to produce estimates closer to those obtained from the monthly weights.

Example of weight construction

The following example describes how the weight should be generated.

MarketScan example population: 25–29-year-old females from the Pacific region who are the policyholder would fall into the stratum for weight key 90721 as shown below.

Character	Value
Character 1: Census Division:	9 = Pacific (West region)
Characters 2–3: Age Group:	07 = 25–29
Character 4: Sex:	2 = Female
Character 5: Policy Holder Status:	1 = Policy Holder

When compared with the number of people in the national population based on this weight key, the following would result:

Group	Description	Value
A	People in the nation with employer-sponsored insurance using the American Community Survey	403,583
B	Enrollees in the full Commercial Claims and Encounters Database (created from frequency generating a distribution from the annual enrollment file for the full population)	79,893
C	Number of people in this stratum with Disease X (example)	5,000

Group	Description	Value
A divided by B	Calculated weight	5.0515

Therefore, all those in stratum 90721 will be given the calculated weight (5.0515), and that number then can be used to weight MarketScan up to the national ESI population.

It is important to note that the weight is based on the full denominator population and not on any subset (e.g., Group C).

Sample programs and tools

Structure of the weight key

The MarketScan weight key ("MSWGTKEY") is a five-character variable, with the following structure:

Character 1: Census Division

Census Division
1 = New England (Northeast region)
2 = Middle Atlantic (Northeast region)
3 = East North Central (Midwest region)
4 = West North Central (Midwest region)
5 = South Atlantic (South region)
6 = East South Central (South region)
7 = West South Central (South Region)
8 = Mountain (West region)
9 = Pacific (West region)
Blank = Unknown

Characters 2–3: Age Group

Age Group, Years
01 = 0
02 = 1–4
03 = 5–9
04 = 10–14
05 = 15–17
06 = 18–24
07 = 25–29
08 = 30–34
09 = 35–39
10 = 40–44
11 = 45–49
12 = 50–54
13 = 55–59
14 = 60–64
15 = 65–69
16 = 75+
Blank = Unknown

Character 4: Sex

Sex
1 = Male
2 = Female

Character 5: Policy Holder Status

Policy Holder Status
1 = Policy Holder
0 = Not Policy Holder

For example, the weight key value "10921" corresponds to an enrollee residing in the Northeast census division who is between the ages of 35 and 39 years, female, and a policy holder.

Available data formats

The MarketScan National Weights are provided in the following formats. Users can select the format that works best for their environment. The contents between the two options are identical.

1. **CCAEWYYV.sas7bdat, MDCRWYYV.sas7bdat.** SAS datasets used to obtain the ACS ESI population estimate for a given MSWGTKEY value in year "YY" in the MarketScan CCAE and MDCR datasets respectively, with year and version "YYV"
2. **CCAEWYYV.csv, MDCRWYYV.csv.** CSV files corresponding to CCAEWYYV.sas7bdat and MDCRWYYV.sas7bdat

Appendix: Analysis of annual and monthly enrollment data

Introduction

The MarketScan commercial insurance weights are computed as the ratio of the American Community Survey (ACS) employer-sponsored insurance (ESI) population to the MarketScan enrollee volumes, for the same census division, age group, sex, and policy holder status. The ACS supplies only annual estimates of population, based on continuous collection of data throughout the survey year. There is no information on the ACS to identify the month or quarter in which data are collected.

MarketScan is released with two enrollment tables: (1) an annual table (the "A" table) and (2) a detailed enrollment table (the "T" table). The enrollee characteristics in the annual table are created from the detailed enrollment table, which contains information on enrollee and coverage characteristics by month. In particular, enrollee age and state of residence are determined for each month present in the detailed enrollment table.

When the annual table is created, age is set equal to the enrollee's age on the first day of enrollment during the year (as of DTSTART on the first enrollment detail record) and state is a modal value (the state value from the enrollment detail records having the most enrollment days during the year). Thus, calculations using a single age group or geographic location from the annual enrollment table may not be entirely accurate. Enrollees may cross an age group boundary during the year or move to a different census division.

An initial question is How different are the enrollment counts obtained by using the annual table versus the detailed table? A secondary question is How different are national estimates of enrollment and health outcomes (e.g., costs) using the two enrollment tables?

Weight computation using annual and detailed enrollment

The person-level annual weight is computed as follows: $W_a(s) = E(s)/e_a(s)$, where $E(s)$ represents the ACS population estimate for stratum s and $e_a(s)$ is the MarketScan enrollee count for stratum s derived from the annual enrollment table in a given year. Note that the assignment of stratum s in the annual table may disagree with one or more age group or location assignments in the detailed table depending on whether an enrollee ages or relocates out of the stratum with which he or she first was associated during the year.

When computing the annual weight, the numerator and denominator quantities are point-in-time estimates. That is, they reflect the enrollment status and stratum at a given point in time. For the ACS, this is the point during the year when the respondent is interviewed. In the case of the MarketScan annual enrollment, it is the first enrollment date during the calendar year of enrollment. In both cases, a respondent or enrollee will not necessarily be covered for the entire year or remain in the same stratum for the entire year.

One possible approach for using the monthly MarketScan enrollment data for weight computation is to assume that the annual ACS stratum-specific ESI population estimate is constant from month to month. In this case, an approximate monthly weight can be computed as $W_m^{*}(s,m) = E(s)/e_m(s,m)$, where $E(s)$ represents the annual ACS population estimate for stratum s and $e_m(s,m)$ is the MarketScan enrollee count for stratum s and month m. This approach facilitates analysis on a monthly scale and allows for analysis on an annual scale by aggregating or averaging over the monthly estimates.

When using the weights $W_m^{*}(s,m)$, it is necessary to look up the stratum assignment by month for each enrollee from the detailed enrollment table, compute the MarketScan weight, and then associate that weight with MarketScan claims information for the same enrollee and month. This computation is more complex than simply associating an enrollee with an annual weight or adjusted annual weight for all services in a given year, but it will yield the most accurate estimates.

Another approach, if annual estimates are desired, is to adjust the annual weight by correcting the MarketScan stratum-specific enrollment counts for changes in age group or location during the year. One way to calculate such an adjusted annual weight is by counting the enrollee months spent in each stratum, using the detailed table and then computing an average annual enrollee count for each stratum: $W_a^{*}(s) = E(s)/e_a^{*}(s)$, where $e_a^{*}(s) = (e_m(s,.)\div 12)$, where "." denotes summation over a subscript. This approach is less accurate than the monthly approach that constructs $W_m^{*}(s,m)$ but more accurate than using unadjusted annual weights.

Differences in estimates based on annual and detailed enrollment

How different are the weights computed using the annual and detailed enrollment tables? Because the weights rely on the enrollee counts by stratum, it is informative to compare annual versus monthly enrollment counts by age group and census division, the two nonstatic factors in the stratification scheme.

The following analysis uses the MarketScan Commercial Claims and Encounters (CCAE) datasets for 2014. We selected only those enrollees with a data contribution source of employer. Also, any enrollee with missing data on any of the stratum variables (census division, age group, sex, and policy holder status) was excluded.

Figure 1 compares the age distribution of enrollees from the annual and detailed enrollment tables. On a relative basis, the largest discrepancy is in the 0-year age group: 1.9 percent for the annual table and 1.0 percent for the detailed table. It is easy to visualize the reason for the discrepancy: on average, newborns spend one-half year in the age group. This is accounted for in the detailed enrollment table but not the annual table. Other age groups (25–29 years, 30–34 years) exhibit greater enrollee inflows on age group boundaries that are detected in the detailed enrollment table; the opposite is true for some of the older age groups (50–54 years, 55–59 years, 60–64 years).

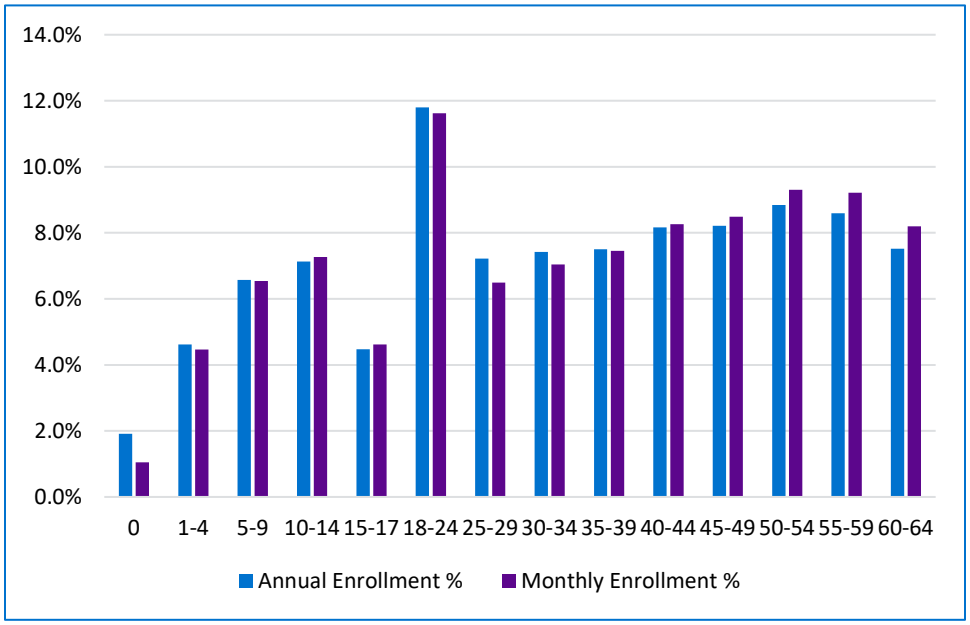

Figure 1. Distribution of Enrollees by Age Group: Adjusted Annual and Detailed Enrollment Tables, 2014

Figure 2 compares enrollee census division residence for the annual and detailed enrollment tables. The differences between the two sources are much smaller than those for the age distribution, as would be expected.

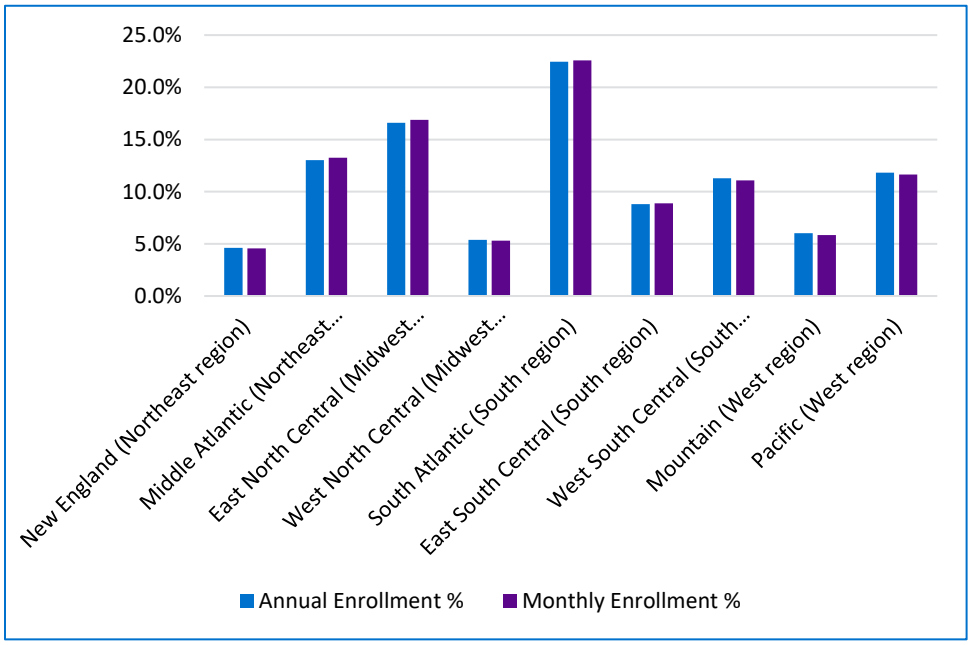

Figure 2. Distribution of Enrollees by Census Division: Adjusted Annual and Detailed Enrollment Tables, 2014

Comparison of projected costs: annual, monthly, and adjusted annual weights

Figures 3 and 4 provide insight into differences in enrollment counts using the annual and monthly enrollment tables. How much do these differences matter in practice, when we are estimating national enrollment counts or outcomes such as costs?

Figure 3 contains estimated mean total payments by age group for 2014, using annual, monthly, and adjusted annual weights. The bars labeled "Annual Enrollment Weights" are estimated average payments computed using weights computed from the annual enrollment table, namely, W_a (s) as defined above. The bars labeled "Monthly Enrollment Weights" represent estimates obtained by using the weights W_m^* (s,m). Finally, the bars labeled "Adjusted Annual Weight" are estimates that use the "adjusted" annual weights W_a^* (s).

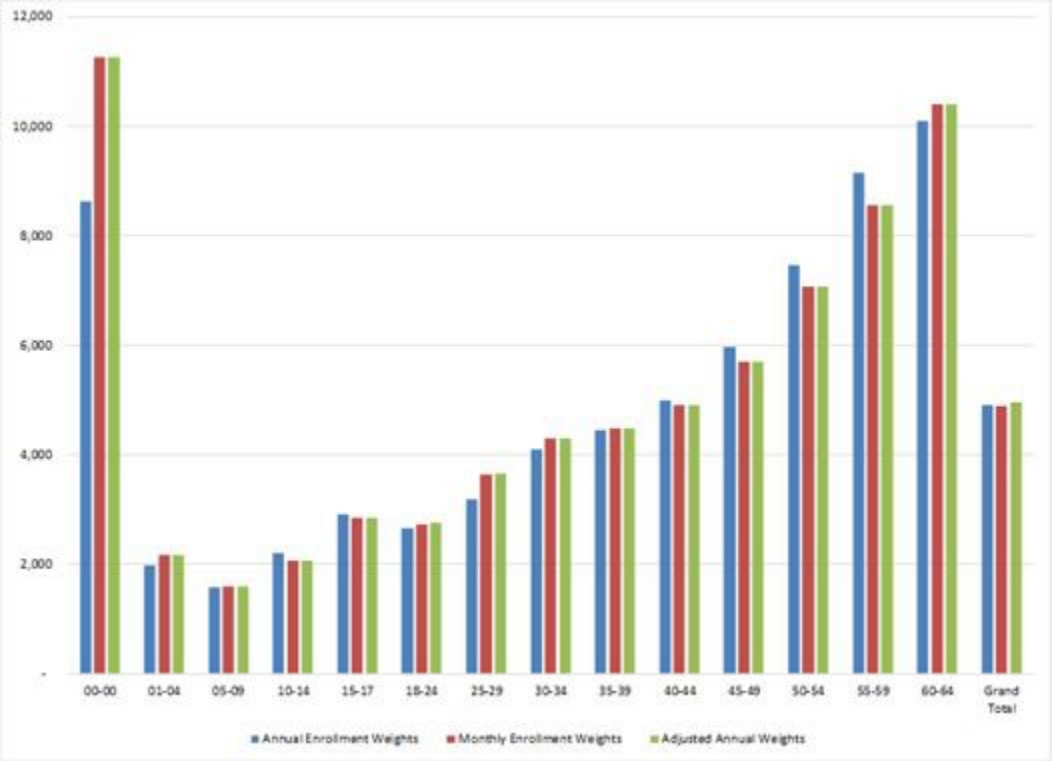

Figure 3. Average Total Payment by Age Group: Annual and Monthly Weights, 2014

Figure 4 is a similar chart for census division of residence. The largest discrepancies in mean payment are found between the annual weighted estimates and the two that use monthly enrollment and claims detail. The discrepancies are larger for certain age groups, most notably age group 0 years. Smaller discrepancies are found for age groups 45–49 years, 50–54 years, and 55–59 years. Mean payment discrepancies are much smaller for census division of residence.

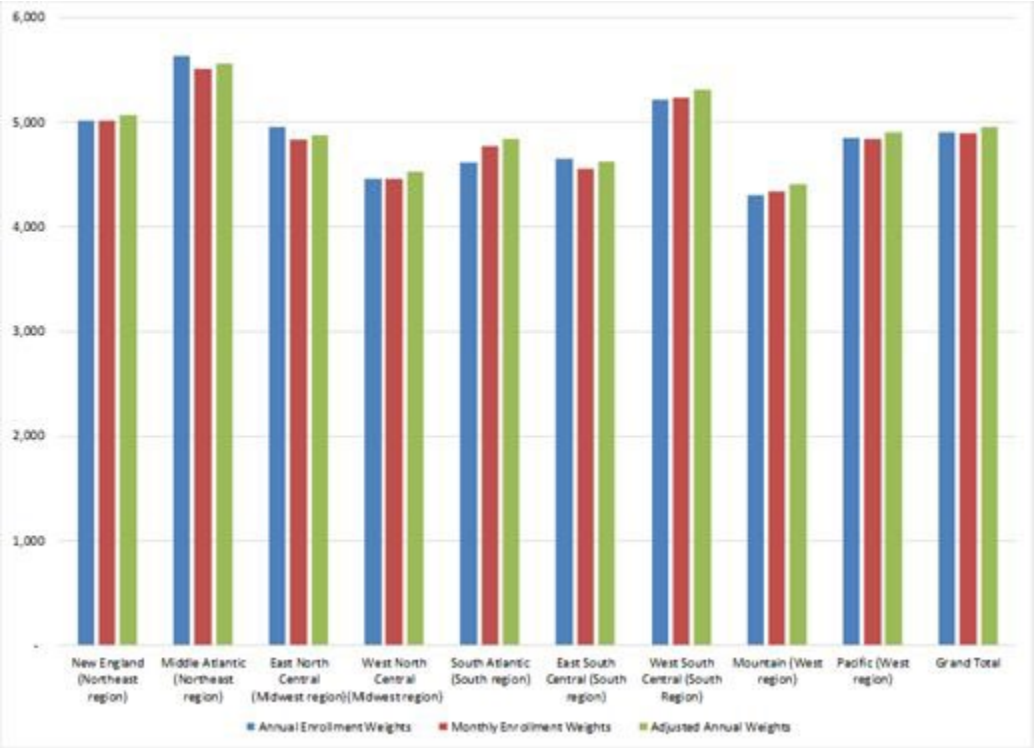

Figure 4. Average Total Payment by Census Division: Annual and Monthly Weights, 2014

Conclusion

Use of the MarketScan monthly enrollment tables to construct commercial insurance weights for estimation is logically the most accurate method, but it introduces programming and analytic complexity. From this preliminary investigation, with aggregated strata, calculation and use of adjusted annual weights appear to yield estimates similar to those produced using monthly enrollment data with little additional programming complexity. With the exception of newborns, estimates derived from the unadjusted annual weights also are similar at the aggregated levels studied.

Merative
100 Phoenix Drive
Ann Arbor, Michigan 48108

Produced in the United States of America
December 2023

This document is current as of the initial date of publication and may be changed by Merative at any time. Not all offerings are available in every country in which Merative operates.

The information in this document is provided “as is” without any warranty, express or implied, including without any warranties of merchantability, fitness for a particular purpose and any warranty or condition of non-infringement. Merative products are warranted according to the terms and conditions of the agreements under which they are provided.

The client is responsible for ensuring compliance with all applicable laws and regulations applicable to it. Merative does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation. The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on the specific configurations and operating conditions. It is the user’s responsibility to evaluate and verify the operation of any other products or programs with Merative product and programs.

Statement of Good Security Practices: IT system security involves protecting systems and information through prevention, detection and response to improper access from within and outside your enterprise. Improper access can result in information being altered, destroyed, misappropriated or misused or can result in damage to or misuse of your systems, including for use in attacks on others. No IT system or product should be considered completely secure and no single product, service or security measure can be completely effective in preventing improper use or access. Merative systems, products and services are designed to be part of a lawful, comprehensive security approach, which will necessarily involve additional operational procedures, and may require other systems, products or services to be most effective. Merative does not warrant that any systems, product or services are immune from, or will make your enterprise immune from, the malicious or illegal conduct of any party.