Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

Journal of Evaluation in Clinical Practice ISSN 1365-2753

Estimating measurement error when annualizing health care costs

Ariel Linden DrPH[1,2] and Steven J. Samuels PhD[3]

[1]President, Linden Consulting Group, Ann Arbor, Michigan, USA
[2]Adjunct Associate Professor, Department of Health Policy & Management, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA
[3]Adjunct Associate Professor, Department of Epidermiology & Biostatistics, School of Public Health, State University of New York at Albany, Albany, New York, USA

Keywords
absolute percentage error, annualized costs, measurement error, per member per month, per member per year

Correspondence
Dr Ariel Linden
Linden Consulting Group
LLC, 1301 North Bay Drive
Ann Arbor, MI 48103
USA
E-mail: alinden@lindenconsulting.org

Accepted for publication: 8 June 2012

doi:10.1111/j.1365-2753.2012.01885.x

Abstract

Objective Health insurers routinely annualize members' health care costs for reporting, predicting high cost cases and evaluating health management programmes. Annualization is the practice of extrapolating to a yearly cost from less than a year of data. In this paper, we systematically estimate the measurement error inherent in this approach.
Study design The paper uses a retrospective observational study using longitudinal claims data from three types of insured populations: Medicare managed care, public employees and a self-insured employer.
Methods The unit of analysis was a block 'year' consisting of 12 consecutive months of cost data for any individual member. These blocks were constructed recursively allowing use of all available data that an individual could contribute. We tested the accuracy of the annualized costs by calculating the absolute error (AE) representing the difference, in dollars, between the actual annual costs and the predicted annual costs, and the absolute percentage error (APE) which is the absolute error divided by the actual 12-month costs.
Results Under the best case scenario (when 11 months of data were used to annualize costs), the mean AE ranged from approximately $2700 for the Medicare population to about $400 for the two working-aged populations; and the mean APE ranged from 9.6% to 11.0% in the three populations. Accuracy diminished systematically with fewer months of available data.
Conclusions Due to the largely unpredictable nature of monthly costs, annualization can produce substantial measurement error. Given the importance of cost metrics for decision making, we offer several alternative approaches that insurers should consider to improve measurement accuracy.

Introduction

Health insurers routinely annualize members' health care costs when reporting medical claims experience at the population level, predicting which members are likely to incur high costs, and evaluating the cost-effectiveness of health management interventions. Annualization is the practice of extrapolating to a yearly cost from less than a year of data. Typically, if monthly costs are known for less than a year, the mean monthly cost is multiplied by 12 to generate the annualized value. Annualization can be highly problematic, however, because of the significant month-to-month variability in health care costs [1]. For example, a person who incurred $10 000 of medical costs in 1 month of enrolment will have a predicted annualized cost of $120 000 ($10 000 × 12 months). If the costs in that 1 month were due to a hospitalization, a known rare event, it is unlikely that 11 additional admissions would occur in the remainder of the year.

Inaccuracy introduced by annualized cost estimates can have serious implications. For example, individuals incorrectly categorized as high cost would be invited to participate in interventions intended to reduce medical costs, while patients incorrectly categorized as low cost would be inappropriately excluded. Insurance premium rates based on partial data might be set excessively high, causing employers or individuals to forego health insurance. Finally, as the US health care system considers innovative approaches to incentivizing cost-effective health care (via accountable care organizations, medical home models, etc.), inaccurate estimation of patient costs could lead to either under- or over-payment.

Despite these significant implications, limited previous work is available to quantify the degree of error introduced by annualization. To address this question, we estimate the error with data from three populations. We hope that that the results will inform organizations that rely on annualized metrics for operational decisions and research.

Table 1 Example of the data structure and measures constructed for use in the current analyses

ID	Month	A Actual monthly costs	B Cumulative monthly costs	C Cumulative mean monthly costs	D Annualized costs (PMPY)*	E Actual annual costs	F Absolute error†	G Absolute percent error‡
1	1	49.64	49.64	49.64	595.68	2118.54	1522.86	71.88
1	2	77.59	127.23	63.62	763.38	2118.54	1355.16	63.97
1	3	0.00	127.23	42.41	508.92	2118.54	1609.62	75.98
1	4	0.00	127.23	31.81	381.69	2118.54	1736.85	81.98
1	5	187.93	315.16	63.03	756.38	2118.54	1362.16	64.30
1	6	207.27	522.43	87.07	1044.86	2118.54	1073.68	50.68
1	7	49.64	572.07	81.07	980.69	2118.54	1137.85	53.71
1	8	249.83	821.90	102.74	1232.85	2118.54	885.69	41.81
1	9	0.00	821.90	91.32	1095.87	2118.54	1022.67	48.27
1	10	0.00	821.90	82.19	986.28	2118.54	1132.26	53.45
1	11	49.77	871.67	79.24	950.91	2118.54	1167.63	55.11
1	12	1246.87	2118.54	176.55	2118.54	2118.54	0.00	0.00
Mean							1167.20	55.09

*Per-member-per-year (PMPY) = Column C × 12.
†Absolute error = Column E – Column D.
‡Absolute percentage error = [(Column E – Column D)/Column E].

The paper is organized as follows. In the Methods section, we briefly describe the three datasets, define measurement error and explain the methodology we employ to derive it. In the Results section, we present the estimates of annualization errors for individuals in the study from the three populations. In the Discussion section, we summarize our findings, discuss the implications for health insurers, and briefly describe existing methods that are more suitable for incomplete longitudinal data. In the Conclusion Section, we share some final thoughts.

Methods

Populations and data

We chose medical cost data from three types of insured populations. The first population consisted of 17 700 public employees enrolled in a managed care plan for a consecutive 36-month period between 2004 and 2010. The second population consisted of 7868 employees of a self-insured employer who were eligible for health insurance benefits for a consecutive 30-month period between 2005 and 2009. The third population consisted of 8586 chronically ill beneficiaries in a Medicare managed care plan who had between 12 and 36 months of consecutive enrolment between 2005 and 2009.

The datasets included costs for all inpatient and outpatient services provided to each plan member, including hospitalizations, surgeries, laboratory, imaging, pharmacy and rehabilitation services. The costs represent the actual amount paid by the insurer on claims submitted by providers net of patient co-pays and coinsurance.

Dataset structure and measures

The unit of analysis was a block 'year' consisting of 12 consecutive months for any individual member. *Actual* annual costs were calculated by summing the costs over the 12 months in the block. The *Predicted* annualized costs based on k months of data (k = 1 to 11) for member's block year were calculated by multiplying the mean cost through month k by 12. Thus, for each block, we create 11 predicted annual costs, based on costs from month 1, from months 1 and 2, months 1 through 3 and so on, with the final prediction based on the annual monthly cost of month 1 through month 11.

We utilize two measures to test the accuracy of the 11 predicted annual costs for an individual. The *absolute error* (AE) represents the difference, in dollars, between the actual annual costs and the predicted annual costs. In the case of perfect accuracy (the actual and predicted are equal) the absolute error is zero. The *absolute percentage error* (APE) is the AE divided by the actual 12-month cost [2]. To derive summary scores at the population level, we calculate the mean AE and the mean APE across all the 12-month blocks in the population.

Table 1 illustrates the construction of the AEs and APEs for one 12-month block from a single individual. As an example, we describe the values derived for month 8. The total actual paid cost in that month (column A) was $249.83. The cumulative monthly cost up to and including month 8 (column B) was $821.90. The cumulative mean cost for month 8 (column C) was $102.74, which equals the value in column B divided by 8 (the cumulative number of months up to that point). The predicted annual cost (column D) was $1232.85, calculated by multiplying the cumulative mean value after 8 months of enrolment (column C) by 12 ($102.74 × 12). The total of actual costs for this individual over the 12-month period was $2118.50 (column E). Therefore, the AE is $885.69 – the absolute difference between the actual annual value and the estimated annual value (column E – column D) for that month. The APE is calculated as [(column E – column D)/column e] = [($2118.54 – $1232.85)/$2118.54] and reflects a 41.81% difference between actual and predicted annual costs using 8 months of data for this individual. The last row of the table indicates that this individual had a mean AE of $1167.20 and mean APE of 55.1% for this 12-month block.

We considered several approaches to define 12-month blocks for the analysis. For example, we could have analysed the first 12

Table 2 Demographic and cost characteristics of the three study populations

	Medicare chronically ill	Commercial employer	Public employer
n	8586	7868	17 700
Percent female	51	53	51
Mean age (standard deviation)	75 (11)	47 (11)	37 (19)
Total months of data	249 116	236 040	637 200
Percent months of data with cost >$0	90	50	39
Monthly costs			
Mean	2293	298	292
25th percentile	125	0	0
50th percentile	385	0	0
75th percentile	1128	159	114

Table 3 Mean absolute percentage error (mean APE) and bootstrapped confidence intervals for predicted annualized costs, by month, for the three populations

	Medicare chronically ill ($n = 8586$; 154 670 years)		Commercial employer ($n = 7868$; 149 492 years)		Public employer ($n = 17\,700$; 442 500 years)	
Month	Mean APE (%)	95% CI	Mean APE (%)	95% CI	Mean APE (%)	95% CI
1	98.7	(97.9, 99.3)	104.0	(102.4, 104.8)	117.6	(116.8, 118.5)
2	79.4	(78.6, 79.8)	84.8	(83.6, 85.7)	95.9	(95.3, 96.6)
3	67.1	(66.8,67.9)	71.9	(71.0, 72.5)	81.6	(81.0, 81.9)
4	57.8	(57.3, 58.5)	61.9	(61.3, 63.2)	70.4	(69.8, 70.8)
5	49.8	(49.4, 50.3)	53.3	(52.4, 54.0)	60.7	(60.3, 61.1)
6	42.5	(42.1, 42.8)	45.5	(44.9, 45.7)	51.8	(51.5, 52.2)
7	36.0	(35.6, 36.3)	38.2	(38.0, 38.5)	43.7	(43.3, 43.8)
8	29.6	(29.2, 29.8)	31.3	(31.0, 31.7)	35.7	(35.4, 35.9)
9	23.3	(23.1, 23.5)	24.4	(24.1, 24.7)	27.7	(27.6, 27.9)
10	16.7	(16.6, 16.9)	17.4	(17.2, 17.6)	19.7	(19.6, 19.8)
11	9.6	(9.5, 9.7)	9.8	(9.7, 9.9)	11.0	(11.0, 11.1)

Mean APE is calculated as the absolute (actual annual costs – predicted annualized costs) / actual annual costs, averaged across all individuals.

months of data for an individual, or we could have selected a 12-month block at random. Instead, we decided to use all available information by extracting all the different 12-month blocks that an individual could contribute. Suppose, for example, that the individual in Table 1 had been followed for 13 months. In this case, the individual would have contributed two different years to the analysis, namely months 1–12 and 2–13.

After constructing a dataset of all such blocks and aligning them to start at month 1, we calculated the AE and APE in predicted annual values, as shown in Table 1. We then averaged these errors, for a given month, across the population of blocks to estimate the mean absolute error and mean absolute percentage errors. To provide valid confidence intervals, we bootstrapped the estimation procedure with 1000 replicates [3,4]. Each bootstrap 'draw' consisted of all the blocks for a given month, contributed by a single individual. This ensures that the confidence intervals reflect variation between individuals. We report bias corrected estimates. All analyses were conducted in Stata, Version 11 (StataCorp, College Station, TX, USA).

Results

Table 2 provides demographic and cost characteristics for the three study populations. As expected, the chronically ill Medicare population was older than the two working-aged populations, had substantially higher mean monthly costs and a considerably higher percentage of months with costs greater than $0. This last point explains why the 25th and 50th percentile of costs in the two working aged populations was $0.

Table 3 reports mean APE values and bootstrapped confidence intervals for predicted annualized costs, by month, for the three populations. For the Medicare chronically ill, commercial employer and public employer populations, the number of blocks included was 154 670, 149 492 and 442 500, respectively. Estimates based on a single month of data produced predicted annual costs with very high mean APEs of 99%, 104% and 117% respectively. These mean APEs decreased with each month of data added. However, even when 11 months of data were used to annualize costs, the mean APE ranged from 9.6% to 11.0%.

Table 4 reports mean absolute errors and bootstrapped confidence intervals for predicted annualized costs, by month, for the three populations. As shown, the results are consistent with those of Table 3. When only 1 month of data are used to annualize costs, the difference between actual and predicted is approximately $27 000 for the Medicare chronically ill population and about $4000 for the two working-aged populations. When 11 months of data are used to annualize costs, the mean absolute error is approximately $2700 for the Medicare chronically ill population

Table 4 Mean absolute error (mean AE) and bootstrapped confidence intervals for predicted annualized costs, by month, for three populations

	Medicare chronically ill (n = 8586; 154 670 years)		Commercial employer (n = 7868; 149 492 years)		Public employer (n = 17 700; 442 500 years)	
Month	Mean AE ($)	95% CI	Mean AE ($)	95% CI	Mean AE ($)	95% CI
1	26 804	(26 003, 27 331)	4129	(3924, 4341)	4174	(3965, 4310)
2	22 242	(21 761, 22 870)	3513	(3245, 3726)	3563	(3407, 3701)
3	19 304	(18 931, 19 679)	3036	(2930, 3223)	3096	(2981, 3202)
4	16 770	(16 507, 17 328)	2669	(2476, 2824)	2723	(2623, 2811)
5	14 569	(14 318, 15 074)	2315	(2214, 2520)	2385	(2262, 2478)
6	12 535	(12 336, 12 988)	1990	(1915, 2105)	2037	(1970, 2142)
7	10 637	(10 432, 10 815)	1679	(1585, 1764)	1719	(1665, 1813)
8	8 740	(8 432, 8 944)	1372	(1284, 1441)	1393	(1340, 1458)
9	6 877	(6 690, 7 026)	1046	(1005, 1109)	1065	(1027, 1121)
10	4 881	(4 777, 5 008)	735	(689, 782)	740	(715, 764)
11	2 747	(2 674, 2 771)	398	(371, 417)	398	(381, 418)

Mean AE is calculated as the absolute (actual annual costs – predicted annualized costs), averaged across all individuals.

and about $400 for the two working-aged populations. To illustrate the magnitude of error for the Medicare chronically ill population under the best case scenario (i.e. when 11 months of data are used to annualize costs), we multiply the MAE at 11 months ($2747) by the population size (8586) and get an estimated error amounting to $23 585 742. Obviously, much larger discrepancies will be found when fewer months of available data are used to annualize costs.

Discussion

Our results indicate that annualized costs have substantial error, regardless of whether costs are estimated in healthy working-aged populations or a chronically ill Medicare population. While reliance on only a single month of data naturally resulted in a huge degree of error, even 11 months of data resulted in meaningful error. More broadly, this study underscores the unpredictable nature of health care costs, whether the data are from an older chronically ill population or a younger working-aged population.

Given that the approach to annualizing costs used in our analysis relies solely on accrued past costs for estimation, it is fair to question how much the model could be improved by including other predictors. In a recent study comparing the accuracy of several commercial predictive modelling applications (which include basic demographic data and other data element to supplement past medical claims costs), the best performing model achieved an R^2 of only 32% and a mean APE of 75.2% [5]. Given that health insurers are increasingly relying on such applications to identify individuals at risk for high medical costs in order to provide targeted interventions, these findings suggest that even the best available models include a meaningful level of error.

Another approach that is commonly used when there are missing data is to examine per-member-per-month (PMPM) costs based on data from all available months. However, as with the annualization approach, the PMPM method also introduces error by ignoring temporal variability. To illustrate this flaw, Fig. 1 depicts the cost trajectories of three individuals from the Medicare chronically ill population who had the same average monthly cost ($1000) but differed in their lengths of enrolment. The first individual was enrolled for 3 months and had costs somewhat evenly

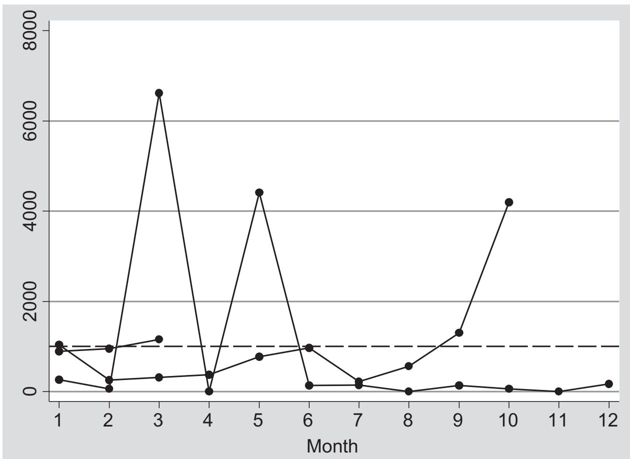

Figure 1 Actual cost patterns for three individuals with equal mean per-member-per-month cost of $1000, but different lengths of enrolment.

distributed over that period. The second individual was enrolled for 10 months, and had a spike in costs in months 9 and 10. The third individual was enrolled for 12 months, and had two spikes in cost (months 3 and 5) but then had little or no cost for the remainder of the year. Thus, the PMPM approach assumes that the individual who accrues an average of $1000 in costs in 3 months is the same as the individual who accrues the equal average costs over the course of a year.

As we have demonstrated in this paper, naïve approaches used to overcoming the limitations of missing data introduce significant measurement error. In the annualization approach, the error is introduced from extrapolation, because it assumes that *within individuals*, costs are consistent over time. In the PMPM approach, the error is introduced by assuming comparability *between individuals* without regard for differing number of months for which data are available. Thus, neither approach represents an accurate method to dealing with missing cost data.

Given the large month-over-month variability in medical claims costs, the most obvious approach to mitigating measurement error is to ensure that comparisons are made for all individuals who have the same length of enrolment. This would eliminate the need to extrapolate beyond the range of data for any individual, and it would improve comparability between individuals (at least on the temporal component) when comparing PMPM costs. Thus, if an annual per-member-per-year rate is desired, only those members who have a complete year of claims data could be included in the analysis. Of course, this approach limits the number of individuals with sufficient data. Therefore, one must balance the trade-off between accuracy and sample size.

One approach to consider for forecasting future costs is time series analysis (TSA) [2,6]. This regression-based technique relies on past observations to predict future behaviour of the observed outcome variable (i.e. costs) without attempting to measure independent relationships that influence it. This approach generally requires many past observations to produce accurate predictions, and so its utility may be limited if the period under study is relatively short.

Linden *et al*. [7] introduced a regression-based approach to emulate the key characteristics of TSA and its ability to adjust for seasonality, and survival analysis and its ability to adjust for the impact of length of programme enrolment. Similarly, Etzioni *et al*. [8] proposed a sophisticated survival analysis-based approach to predict medical costs when data are missing due to attrition or censoring.

Another potential technique worth exploring to forecast future costs when there are missing months of data is recursive regression [9]. In this approach, a regression model would be estimated predicting annual costs using current and lagged covariates. These covariates can include past costs in addition to possible characteristics that may indicate the direction of future costs. What makes this approach 'recursive' is that the starting point is fixed and each new monthly observation is then added. As new observations are added, theoretically the model should become more accurate. However, given the instability in month-over-month medical costs, it is questionable how well this model will perform.

There are several other statistical approaches for handling longitudinal data with non-uniform lengths of enrolment, such as general estimating equations, random effects, fixed effects or mixed effects models (Rabe-Hesketh & Skrondal [10] and Fitzmaurice *et al*. [11] have a comprehensive discussion on these models). Longitudinal models differ from standard regression-type models in that they account for within-individual patterns of change over time in addition to the between-individual patterns of changes estimated in standard models. These techniques are particularly suited for evaluation of interventions, whereas the models described above are more suited for forecasting.

We took several steps to ensure the robustness of our results. First, we used large datasets from three different populations. Additionally, by operating on 'block years' we utilized all the information in the data. This methodology served to neutralize the effects of seasonality and any other temporally related biases. Finally, we ensured that the confidence intervals were correctly estimated by using the non-parametric bootstrap technique.

Nonetheless, there are potential limitations. For example, individuals with longer observation times received more weight in our analysis than individual with shorter times. We believe it was better to utilize all the data from each person, but the possibility remains that average annual cost, as well as total cost, is related to the number of months with coverage. Another limitation is that we did not study how the annualization method performs when the goal is to predict costs for groups of people as opposed to individuals. For groups, averaging costs over individuals should reduce the magnitude of prediction errors.

Conclusion

Our analyses demonstrate that annualization, the approach commonly used by health insurers to predict annual member costs, can produce substantial measurement error. Even when 11 months of data are used to annualize costs, the difference between actual and predicted total costs is approximately 10%. These results have important implications for health insurers who routinely use such data for reporting medical claims experience, rating populations, predicting which members are likely to incur high costs and evaluating the cost effectiveness of health management interventions. There are statistical approaches specifically designed to account for length of enrolment in the estimation process. And while such models may require a trained statistician, the improvement in accuracy will likely translate into better decision making regarding allocation of resources to improve the quality and reduce the costs of care.

References

1. Diehr, P., Yanez, D., Ash, A., Hornbrook, M. & Lin, D. Y. (1999) Methods for analyzing health care utilization and costs. *Annual Review in Public Health*, 20, 125–144.
2. Linden, A., Adams, J. & Roberts, N. (2003) Evaluating disease management program effectiveness: an introduction to time series analysis. *Disease Management*, 6 (4), 243–255.
3. Efron, B. & Tibshirani, R. (1993) An Introduction to the Bootstrap. New York: Chapman & Hall.
4. Linden, A., Adams, J. & Roberts, N. (2005) Evaluating disease management program effectiveness: an introduction to the bootstrap technique. *Disease Management & Health Outcomes*, 13 (3), 159–167.
5. Winkelman, R. & Mehmud, S. (2007) A Comparative Analysis of Claims-based Tools for Health Risk Assessment. Schaumburg, IL: Society of Actuaries.
6. Makridakis, S. G., Wheelwright, S. C. & Hyndman, R. J. (1998) Forecasting: Methods and Applications, 3rd edn. New York: John Wiley and Sons.
7. Linden, A., Adams, J. & Roberts, N. (2004) Evaluating disease management program effectiveness adjusting for enrollment (tenure) and seasonality. *Research in Healthcare Financial Management*, 9 (1), 57–68.
8. Etzioni, R. D., Feuer, E. J., Sullivan, S. D., Lin, D., Hu, C. & Ramsey, S. D. (1999) On the use of survival analysis techniques to estimate medical care costs. *Journal of Health Economics*, 18 (3), 365–380.
9. Kmenta, J. (1997) Elements of Econometrics, 2nd edn. Ann Arbor, MI: University of Michigan Press.
10. Rabe-Hesketh, S. & Skrondal, A. (2008) Multilevel and Longitudinal Modeling Using Stata, 2nd edn. College Station, TX: Stata Press.
11. Fitzmaurice, G. M., Laird, N. M. & Ware, J. H. (2004) Applied Longitudinal Analysis. Hoboken, NJ: John Wiley & Sons, Inc.

