Paper SGF1341-0215

Random vs. Fixed Effects: Which Technique More Effectively Addresses Selection Bias in Observational Studies?

Jessica Navratil-Strawn MS MBA, Optum; Stephen Hartley BS, Optum; Ronald J. Ozminkowski PhD, Optum

ABSTRACT

Retrospective case-control designs are frequently used to evaluate healthcare programs when it is not feasible to randomly assign members to participate or not in those programs. Without randomization, estimates of program impact are more susceptible to selection bias because those who participate in an intervention may be different than those who do not, in ways that are difficult to address. This paper will discuss how multinomial propensity score weighing and random effects techniques can be used to reduce the impact selection bias has on observational study outcomes. All results shown are drawn from a return on investment (ROI) analysis using a participant (cases) versus non-participant (control) study design for a fitness reimbursement program (FRP) aiming to reduce healthcare expenditures among participating members.

Results for the fixed effects model showed members experienced between $6 (low moderate – 4 to 8 gym visits per month) to $20 (high – 12 or more gym visits per month) lower per participant per month (PPPM) post period expenditures in comparison to the lowest participant group (members visiting the gym less than 4 times per month). Using the random effects model, members experienced between $4 (low moderate – 4 to 8 gym visits per month) to $19 (high – 12 or more gym visits per month) lower PPPM post period expenditures in comparison to the lowest participant group (members visiting the gym less than 4 times per month). Comparing these results, we found the random effects results were consistent with the fixed effects model, suggesting the within subject correlation does not appear to be biasing the post-period expenditure estimates for the FRP.

INTRODUCTION

Randomized Control Trials (RCTs) use the most rigorous form of experimental design because they effectively remove case-mix differences by randomly assigning subjects to a treatment or control group before estimating program impact. Even though RCTs are considered the industry gold-standard, utilization of this design is not practical when healthcare programs allow members to decide for themselves whether to participate or not in the program interventions. Retrospective case-control designs are frequently used to estimate program impact when it is not feasible to randomly assign members to participant or non-participant status. Without randomization, observational studies are more susceptible to selection bias when there are underlying and unmeasured differences between the two study groups of interest. Without properly adjusting for selection bias the measured outcome estimates will likely be biased with unknown magnitude and direction, and may therefore lead to a misinterpretation of participant effects (Curtis, 2007).

Researchers frequently use a statistical technique known as inverse propensity weighting (IPW) to reduce the bias caused by lack of study group randomization. Using a parametric model such as logistic regression, the first step of IPW is to determine the propensity score or the probability (on a scale of 0 to 1) of being in the participant group conditional on observed covariates (Curtis, 2007). This predicted probability is then used to create a weighting variable for subsequent statistical analyses. The value of the weight for each participant group member is defined as 1.0 divided by his or her predicted probability of participation, while the non-participant group member's IPW is defined as1.0 divided by one minus his or her predicted probability of participation. After obtainment of these weights, an examination of the standardized differences of the weighted and unweighted covariates between the participant and non-participant groups should be assessed prior to estimating program outcomes. If covariates are balanced, the participant effect can be estimated as the difference between the average outcomes of the two groups. To test if participation bias is adequately controlled for, some analysts also like to conduct additional sensitivity analyses looking at the degree of influence unmeasured covariates have on program

participation and outcomes using residual inclusion (not discussed in this paper) or random effects methods (Terza et al, 2008; Nyman, 2009).

This paper discusses how to develop a multinomial inverse propensity weight and evaluates the impact unmeasured covariates had on healthcare expenditure outcomes using a random effects statistical design in comparison to a fixed effects model. Results included in this paper evaluate the impact varying levels of participation had on post period healthcare expenditures.

DESCRIPTION OF STUDY

Optum FRP is a prevention program that reimburses participants $20 every month in which they visit a participating fitness center at least 12 times. The goal of service is to reduce sedentary lifestyles by offering incentives for gym participation because regular physical activity has been associated with better health, improved functioning, increased quality of life, and reduced healthcare expenditures (Nguyen, 2008).

A retrospective, multinomial cohort study applying inverse propensity score weighting was used to evaluate the effectiveness of varying levels of FRP participation had on downstream healthcare expenditures for small employer group customers (e.g. 2 to 99 employees). Eligible FRP study group members (N=8,723) were divided into four participation frequency groups based on the average number of gym visits as follows:

- Low Participation: Under 4 visits (non-participant group)
- Low Moderate Participation: Between 4 and 8 visits (participant group)
- High Moderate Participation: Between 8 and 12 visits (participant group)
- High Participation: 12 or more visits (participant group)

Impact on healthcare expenditures was estimated by comparing the difference in post PPPM expenditures between those with higher levels of participation (high, high moderate & low moderate) versus lower level participants after controlling for various demographic, healthcare supply, health status and healthcare utilization characteristics (e.g. age, gender, income, pre-intervention period healthcare costs, etc.). In addition, a sensitivity analysis using a random effects model was conducted to assess the impact participation bias or unmeasured covariates had on study outcomes.

MULTINOMIAL INVERSE PROPENSITY SCORE WEIGHTING

DEVELOPING PROPENSITY SCORE

Using the parametric model of logistic regression, the first step in building out a multinomial propensity weight is to determine the propensity score. The propensity score (on a scale of 0 to 1) is the probability of each subject being in the participant group conditional on observed covariates prior to participant status assignment. The basic syntax used is as follows:

```
proc logistic data= <Analysis_Data> <desc>;
class <Class Covariates>;
model <Group_Level>(ref = 'n') = <Covariates> / link = <glogit>;
output out= <Propensity_Scores> predprobs = <i>;
run;
```

The PROC LOGISTIC statement is interpreted as follows:

- <Analysis_Data> is the subject-level dataset that includes a multinomial categorical variable for participant level assignment along with the baseline variables that the propensity score model requires.
- <Class Covariates> the CLASS statement contains all of the categorical covariates to be used in propensity score assignment.

- <Group_Level> is the multinomial variable indicating which participant group level the subject is a member of and the ref='n' defines the reference group used for comparison.
- <Covariates> are all the variables being used in the propensity score model to generate the propensity score. These are measurable variables that occurred prior to participant assignment and should be theoretically related to the decision to participate in the intervention of interest or at least shown empirically to influence that decision in previous analyses.
- <Propensity_Scores> is the output dataset containing the original <Analysis_Data> variables along with the new propensity score value output for each subject.
- <i> outputs individual probabilities generated from the propensity scores with a value between 0 and 1.
- <Desc> is needed if there is order to dependent variable.
- <glogit> fits a multinomial model.

DEVELOPING MULTINOMIAL INVERSE PROPENSITY WEIGHT

This predicted probability is then used to create a weighting variable for subsequent statistical analyses. Inverse propensity weighting uses the propensity score to weight the subject in the outcome variable modeling. The value of the weight for each participant group level for a multinomial variable is defined as 1.0 divided by his or her predicted probability of being in the respective participation group (Crowson, 2013). The inverse propensity weight is assigned in SAS® as follows:

```
%let levels = <Multinomial_Levels>; /*Number of dependant variable levels*/
%macro IPWeight;
    data <Propensity_Weight> (drop=_from_ _into_);
    set <Propensity_Scores>;
    %do i = 0 %to &levels.-1;
          if <Group_Level> = &i. then Progpswt = (1/IP_&i.);
    %end;
   run;
%mend IPWeight;
%IPWeight;
```

- <Multinomial_Levels> defines the number of levels for the categorical dependent variable.
- <Propensity_Weight> is the output data set containing the inverse propensity weight <progpswt> of the propensity score and is used for outcome assessment.

WEIGHT ADJUSTED OUTCOMES – FIXED EFFECTS

The inverse propensity weight is used as a case weight when estimating the impact of the intervention on the outcome(s) of interest. Since the measured variables used to construct the model occur prior to engagement in the FRP, propensity score weighting helps adjust for unbalanced baseline covariates between participant and non-participant (Faries, 2010). Application of the inverse propensity weighting using Proc Genmod is shown as follows:

```
title "FIXED EFFECTS";
  proc genmod data= <Propensity_Weight>;
     class <Group_Level>;
           model <Dependent_Var> = <Group_Level> / dist=<Dist> link=<Link>;
           lsmeans <Group_Level> / <pdiff> <cl>;
           weight <Progpswt>;
```

```
run;
```

- <Dist> and <Link> options model can be used to model a non-normal distribution and are used when outcome data is skewed. These options can be adjusted to another distribution depending on the data. If you specify no distribution and no link function, then the GENMOD procedure defaults to the normal distribution with the identity link function.
- <Dependent_Var> is the outcome variable being assessed after adjusting for the effect of participation <Group_Level> and case mix adjustment via the inverse propensity weight <Progpswt>.
- <Progpswt> is inverse propensity weight determined from the proc logistic code above.
- <pdiff> requests that *p*-values for differences of the LS Means to be produced.
- <cl> requests that confidence limits for predicted values be displayed (see the OBSTATS option).

Fixed effects (FE) explore the relationship between predictor and outcome variables and attempt to remove the net influence predictor variables have on the outcome variable by assuming the effect size is the same across observations and the only reason effect size may vary is due to sampling error (Nyman, 2009). FE adjusts only for the unmeasured variables that do not change over time but does not control for confounding factors that may vary over time (Cameron, 2005).

The PPPM expenditure results for FRP (after adjusting for pre-period spend and other demographic predictors) for the fixed effects model design are listed in Table 1.

TABLE ONE – FIXED EFFECTS HEALTHCARE EXPENDITURES BY PARTICIPANT LEVEL

Under 4 (Low)	4 to 8 (Low Moderate)	8 to 12 (High Moderate)	12 or More (High)	Difference (Low Mod - Low)	Difference (High Mod - Low)	Difference (High - Low)
$239*	$233	$223	$219	$ (6.14)	$ (16.40)	$ (20.01)

* p-value<.1, ** p-value<.01, *** p-value<.001

For the fixed effects model, members experienced between $6 (low moderate – 4 to 8 gym visits per month) to $20 (high – 12 or more gym visits per month) lower PPPM post period expenditures in comparison to the lowest participant group (members visiting the gym less than 4 times per month). These results indicate a significant difference between the highest participant group relative to the lowest participant group using a 90 percent confidence interval.

These results show positive effects on healthcare expenditures as the frequency of gym visits increases, but it is a good quality check to assess what degree of influence unmeasured covariates have on outcomes using a random effects methods (Nyman, 2009). Validating random effects (RE) helps determine if the fixed effects estimates are biased due to lack of controlling for unmeasured variables.

RANDOM EFFECTS

FE models adjust only for the unmeasured variables that do not change over time, making it hard to generalize beyond the sample included in the study. RE supports this extrapolation because it attempts to control for unmeasured confounders that may vary over time. Another key difference between RE and FE is an assumption about whether the unmeasured variables are correlated with the independent variables included in the regression model. FE assumes they are, while RE assumes they are not, so it is good practice to conduct sensitivity analysis to test the impact these assumptions have on outcomes.

In repeated measures and longitudinal studies, observations can be clustered within a subject. When this is the case, the observations and residuals may be correlated with each other and may lead to biased outcomes. Literature varies on how much variance is introduced by using fixed instead of random effects.

A common approach is to evaluate both fixed and random effects models using a Hausman test. This test is designed to indicate the discrepancies in parameter estimates between the two techniques. However,

according to Clark, “the Hausman test is neither a necessary nor a sufficient metric for deciding between fixed and random effects”. Instead, they suggest focusing on an examination of the correlation matrices between the covariate and subject effects and the degree of within-subject variation between the predictor variables and the outcome of interest (Clark, 2012).

Code designed to model fixed and random effects simultaneously (e.g. referred to as a mixed model) and test total variance between subjects can be applied in SAS Proc HPMIXED by using a random statement and specifying the preferred variance structure in the following syntax:

```
Title2 "RANDOM EFFECTS";
  Proc HPMixed Data=<Propensity_Weight>;
      Class <Group_Level> <Time> <Group_Level*Time>;
      Model <Dependent_Var> = <Group_Level> <Time>
            <Group_Level*Time>/solution
      Random INT / Subject = <Member_ID> Type = <Corr>;
      Lsmeans <Group_Level> <Time> <Group_Level*Time>;
      Weight <Progpswt>;

RUN;
QUIT;
```

- <Member_ID> identifies the subjects in your mixed model.
- <Corr> specifies the type of covariance structure to model for G side effects. If type is not specified than the default is Variance Components (VC) is used and this specifies a heterogeneous variance model as shown in the figure below.

Variance components VC (default)

$$\begin{bmatrix} \sigma_A^2 & 0 & 0 & 0 \\ 0 & \sigma_A^2 & 0 & 0 \\ 0 & 0 & \sigma_B^2 & 0 \\ 0 & 0 & 0 & \sigma_B^2 \end{bmatrix}$$

The FRP PPPM expenditures were modeled using a random effects model design with a VC covariance structure. Due to the presence of only one random effect, having only two observations per subject and needing to process random effects for a large data set, the default covariance structure of VC was selected. This analysis also did not require the use of a positive definitive matrix or homogeneous variances like other correlations structures enforce. Finally, this structure is preferable when working with large datasets as it uses considerably less processing time than more complicated correlation structures (Kiernan, 2012). To test the impact of the covariance structure, additional testing of alternative correlation structures was done using type equals Compound Symmetry (CS), Unstructured (UN), and First-Order Autoregressive (AR(1)). All of the least squares (LS) Means estimates and standard errors were consistent between the covariance structures and did not impact final outcomes. Results for based on random effects using the VC covariance structure are listed in Table 2.

TABLE TWO: RANDOM EFFECTS HEALTHCARE EXPENDITURES BY PARTICIPANT LEVEL (VC COVARIANCE STRUCTURE)

Under 4 (Low)	4 to 8 (Low Moderate)	8 to 12 (High Moderate)	12 or More (High)	Difference (Low Mod - Low)	Difference (High Mod - Low)	Difference (High - Low)
$252	$248	$249	$233	$ (3.95)	$ (3.33)	$ (18.82)

* p-value<.1, ** p-value<.01, *** p-value<.001

Using the random effects model and VC covariance structure, members experienced between $4 (low moderate – 4 to 8 gym visits per month) to $19 (high – 12 or more gym visits per month) lower PPPM post period expenditures in comparison to the lowest participant group (members visiting the gym less than 4 times per month). The findings do not show significant differences between the highest participant groups relative to the lowest. These results are also consistent with the fixed effects model, suggesting the within subject correlation is not biasing the post-period expenditure estimates.

MODEL SELECTION

The models showed consistent findings (fixed effects and random effects using the VC correlation matrix). However, due to the slight differences in results, the covariance parameter estimate should be evaluated to determine which model more accurately estimates the outcomes.

Based on the HP Mixed SAS output, the covariance parameter estimate can be used to calculate the ratio of the between-cluster variance to the total variance, which is known as the Intraclass Correlation Coefficient (ICC). The ICC provides the proportion of the total variance in Y that is accounted for by the clustering (e.g. variable specified as the subject in the random statement) and is interpreted as the correlation among observations within the same cluster (Grace-Martin, Singer, 2008). For the FRP, the ICC would measure the correlation within the same member relative to the total variance. The ICC formula is:

$$\text{Corr}\,(Y_{ij,}\, Y_{ij'}\,) = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}$$

The numerator is the sum of the total variance in Y (σ_o^2) that is not explained by X (σ^2). The lower the ICC score the less likely that unmeasured random effects would be influencing the outcome estimates. The covariance parameter estimate used to calculate the ICC for the FRP study is shown in Table 3 below.

TABLE THREE: RANDOM EFFECTS COVARIANCE PARAMETER ESTIMATES ICC CALCULATION

Correlation Type	Covariance Parmeter	Subject	Estimate	ICC
VC	Intercept (σ_o^2)	INDV_ID	19826	5%
	Residual (σ^2)		341084	

Based on the similarity in the outcome estimates between the models and the small value of ICC, this indicates usage of the fixed effects model over the random effects (Grace-Martin).

CONCLUSION

Randomized controls trials are not feasible when members can elect to participate or not participate in programs. As a result, case control observation studies using inverse propensity weighting can help to adjust for participation bias, resulting in more accurate estimates of program effect. However, aside from using inverse propensity weighting to adjust for bias, it is also important to evaluate the impact clustered estimates may have on study outcomes by evaluating a random effects model. As shown in this paper, Proc HPMixed is one technique that can be used to control for variation within subjects. Through an evaluation of the means and the ICC test one can evaluate whether or not a random or fixed effects model produces less biased estimations of the overall effect. Using FRP as an example, it was apparent that fixed effects modeling was appropriate because the means were similar between the tests and the ICC test was small after adjusting for random effects.

Finally, there are other methods to address selection bias related to both measurable and unmeasurable outcome predictors. An explication of these methods is beyond the scope of this paper. So far we have not seen full discussions in the literature of how these other methods compare to fixed effects and random effects approaches or discussions or whether it makes sense to combine these other approaches

with fixed or random effects modeling. Nevertheless, interested readers can turn to papers by Hainmueller (2011) for an entropy weighting approach to making apples to apples comparisons based on measurable factors, and to Terza et al (2008) for more information about how to reduce selection bias due to unmeasurable variables that influence outcomes of interest.

REFERENCES

Cameron AC, Trivedi PK. (2005). Microneconometrics: Methods and applications. New York, NY: Cambridge University Press, 2005, p 701.

Clark, T., Linzer, D. (2012). Should I use fixed or random effects? Available at: <polmeth.wustl.edu/media/Paper/ClarkLinzerREFEMar2012>. Accessed January 15, 2015.

Crowson, C., Schenck, L., Green, A., Atkinson, E., Therneau, T. (2013). The basics of propensity scoring and marginal structural models. Available at: http://www.mayo.edu/research/documents/biostat-84-pdf/doc-20024406. Accessed on January 27, 2015.

Curtis, L., Hammill, B., Eisenstein, E., Kramer, J., Anstrom, K. (2007). Using inverse probability-weighted estimators in comparative effectiveness analyses with observational database. *Journal of Medical Care, 45(10): S103-S107.*

Faries, D., Leon, A., Haro, J., Obenchain, R. (2010). Analysis of observational healthcare data using SAS. Cary, NC: SAS Institute, Inc.

Grace-Martin, K. The intraclass correlation coefficient in mixed models. Available at:

<http://www.theanalysisfactor.com/the-intraclass-correlation-coefficient-in-mixed-models>. Accessed on January 16, 2015.

Hainmueller, J. (2011). Entropy balancing for casual effects: A multivariate reweighting method to produce balance samples in observational studies. *Political Analysis, 20(1): 25-46.*

Kiernan,K.,Tao, J., Gibbs, P. (2012). Tips and strategies for mixed modeling with SAS/STAT procedures.

SAS Global Forum 2012: Paper 332-2012.

Nyman, J., Barleen, N., Dowd, B. (2009). A return-on-investment analysis of the health promotion program at the University of Minnesota. *Journal of Occupational and Environmental Medicine, 51(1):54-65.*

Nguyen, H., et al. (2008). Managed-Medicare health club benefit and reduced health care costs among older adults. *Journal of Preventing Chronic Disease, 5(1).*

Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23 (4): 323-355.*

Terza, R., A. Basu, and P. Rathouz. (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling, *Journal of Health Economics, 27 (3), 531-543.*

ACKNOWLEDGMENTS

The methods and examples presented here are derived from Optum's Fitness Reimbursement Program product support, conducted by the Consumer Solutions Group Healthcare Analytics team. The authors would like to thank Optum for its continued support of knowledge sharing in the healthcare industry.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jessica Navratil-Strawn
Optum
763-797-2489
jessica.navratil-strawn@optum.com