Supplemental digital content for Kashner TM, Henley SS, Golden RM, et al. Studying the effects of ACGME duty hours limits on resident satisfaction: Results from VA Learners' Perceptions Survey. *Academic Medicine* 2010;85:1130-1139.

MATHEMATICAL APPENDIX

"ROBUST DIFFERENCING VARIABLE TECHNIQUE"

This Appendix contains a description of the Robust Differencing Variable Technique (RDV), a statistical method used to estimate ACGME mandatory duty-hour limits effects on resident satisfaction while simultaneously controlling for covariate- and trend-biases. Extending traditional difference-in-differences approaches (DD)[1, 2] and accepted methodologies,[3-7] RDV was necessary to compute effect sizes on these data for two reasons. First, information classifying respondents into "effect" and "no-effect" control settings was missing for pre-limits periods. That is, the ACGME duty-hour limits effect question was not asked during pre-limits periods for the 2001-2003 LPS surveys. Second, the final estimating model needed to adjust for covariate-biases may be misspecified.[3, 6–11] Using the results from Golden et al.,[11] RDV estimators and statistical tests can be shown to be asymptotically unbiased in the presence of MNAR missing data and model misspecification. ACGME duty-hour limits effects may be properly inferred from model estimates provided the impact of setting on respondent satisfaction ratings is a log-linear, time-invariant, function.

Modeling Assumptions

<u>Notation.</u>

The random variable Y is a binary random variable that takes on values such that:

$$Y = \begin{cases} 0 & if \quad respondent\ is\ "not\ satisfied" \\ 1 & if \quad respondent\ is\ "satisfied" \end{cases}$$

For the binary *period* covariate d_1, respondents were administered the survey in either pre- (d_1 = 0) or post- (d_1 = 1) duty-hour limits periods. The ACGME duty-hour limits question was asked only during post-limits periods (d_1 = 1). For the binary *setting* covariate d_2, the respondent reported on the LPS survey that ACGME duty-hour limits had "no effect" (d_2 =0) or "effect" (d_2 =1) on their clinical training environment. The variables

$$d_1 = \begin{cases} 0 & \textit{if} \quad \text{pre-mandatory limits period} \\ 1 & \textit{if} \quad \text{post-mandatory limits period} \end{cases}$$

and

$$d_2 = \begin{cases} 0 & \textit{if} \quad \text{no effect setting} \\ 1 & \textit{if} \quad \text{effect setting} \end{cases}$$

are always included in all models. In some cases, k additional covariates $\mathbf{x} = \left[x_1, ..., x_k\right]$, are also included to improve predictive performance.

<u>Data Generating Process Assumptions.</u>

Let the $k+3$-dimensional vector $\boldsymbol{o}_i \equiv \left(y^i, d_1^i, d_2^i, \mathbf{x}^i\right)$ denote the ith observation, $i = 1,\ldots,n$. The $k+3$-dimensional binary vector $\boldsymbol{h}_i$ will be used to specify which elements in $\boldsymbol{o}_i$ are observable by setting the jth element of $\boldsymbol{h}_i$ equal to zero when the jth element of $\boldsymbol{o}_i$ is not observable; and setting the remaining elements of $\boldsymbol{h}_i$ equal to the value of one, $i = 1,\ldots,n$. It is assumed that $\left(\boldsymbol{o}_1, \boldsymbol{h}_1\right), ..., \left(\boldsymbol{o}_n, \boldsymbol{h}_n\right)$ is a realization of a sequence of n independent and identically distributed random variables. It is additionally assumed that y^i and d_1^i are always observable.

Researcher's Probability Model of the Complete Data.

(A1) Let $\boldsymbol{\beta} \equiv \left[\beta_{1,0} \quad \beta_{2,0} \quad \left(\beta_{1,2} + \beta_{2,1}\right) \quad \left(\boldsymbol{\beta}_x\right)^T \quad \beta_0 \right]^T$ be a *k*+*4*-dimensional column vector where

$\boldsymbol{\beta}_x$ is a *k*-dimensional column vector. Let $p\left(Y=1 \mid d_1, d_2, \mathbf{x}; \boldsymbol{\beta}\right)$ be defined such that:

$$\log\left(\frac{p\left(Y=1 \mid d_1, d_2, \mathbf{x}; \boldsymbol{\beta}\right)}{p\left(Y=0 \mid d_1, d_2, \mathbf{x}; \boldsymbol{\beta}\right)}\right) = \beta_1 d_1 + \beta_2 d_2 + \mathbf{x}^T \boldsymbol{\beta}_x + \beta_0$$

$$\beta_1 \equiv \beta_{1,2} d_2 + \beta_{1,0}$$
$$\beta_2 \equiv \beta_{2,1} d_1 + \beta_{2,0}$$

By using the definitions of β_1 and β_2 we obtain:

$$\log\left(\frac{p\left(Y=1 \mid d_1, d_2, \mathbf{x}; \boldsymbol{\beta}\right)}{p\left(Y=0 \mid d_1, d_2, \mathbf{x}; \boldsymbol{\beta}\right)}\right) = \left(\beta_{1,2} d_2 + \beta_{1,0}\right) d_1 + \left(\beta_{2,1} d_1 + \beta_{2,0}\right) d_2 + \mathbf{x}^T \boldsymbol{\beta}_x + \beta_0$$

$$= \beta_{1,0} d_1 + \beta_{2,0} d_2 + \left(\beta_{1,2} + \beta_{2,1}\right) d_1 d_2 + \mathbf{x}^T \boldsymbol{\beta}_x + \beta_0$$
$$= \boldsymbol{\beta}^T \left[d_1 \quad d_2 \quad d_1 d_2 \quad \mathbf{x}^T \quad 1 \right]^T.$$

Assumption A1 states that the researcher is modeling the data generating process as a logistic regression model[12] with dependent binary variable *Y* and covariates $d_1, d_2, d_1 d_2$, and *k*-dimensional covariate vector **x** when no data is missing. Also note that, ignoring the experimental context and considering the above expression from a purely formal perspective, the interaction term $\beta_{1,2}$ specifies how the impact of D_2 is influenced by D_1 while the interaction term $\beta_{2,1}$ specifies how the impact of D_1 is influenced by D_2.

When no data are missing, it is not necessary for the researcher to specify the joint distribution of the covariates. For the more general case, however, when maximum likelihood estimation in the presence of general types of data decimation mechanisms is desired, it is necessary that the researcher model the joint distribution of the covariates that are not fully observable. Let $x_{j,miss}$ denote the value of the *j*th covariate that contains missing data. Such a covariate will be referred to as a *partially observable covariate*. Ibrahim et al.[13-15] have proposed to model the covariate distribution of the partially observable covariates as a product of one-dimensional parametric conditional distributions so that:

$$p_0\left(d_2,\mathbf{x}\right)=p_0\left(d_2\right)p_0\left(x_1\right)\prod_{j=2}^{k}p_o\left(x_j\mid x_{j-1},...,x_1\right).$$

In addition, make the stronger assumption that the joint distribution, $p_0\left(d_2,\mathbf{x}\right)$, of the partially observable covariates may be expressed as:

(A2) Let $p_0\left(d_2,\mathbf{x}_{miss}\right)=p_0\left(d_2\right)\prod_{k}p_o\left(x_{k,miss}\right)$.

Assumption A2 states that the additional partially observable covariates in **x** will only be included in the model if they provide a source of information that is not redundant with the information source d_2 (i.e., $p_0\left(\mathbf{x}_{miss}\mid d_2\right)=p\left(\mathbf{x}_{miss}\right)$). In addition, A2 states that the *j*th partially observable covariate was added to the model only if it provided a source of information that was not redundant with the previous *j-1* partially observable covariates included in the model (i.e., $p_0\left(x_j\mid x_{j-1},...,x_1\right)=p\left(x_j\right)$).

It is important to emphasize that while the covariate modeling distribution A2 may not be completely satisfied in practice, our empirical investigations have shown that this choice of covariate prior resulted in the development of missing data probability models that did not evidence any signs of model misspecification. Moreover, if A2 does not hold, Golden et al.[11] provide explicit regularity conditions on the researcher's complete data model that ensures the asymptotic consistency of all estimators and statistical test results based upon the missing data probability model.

<u>Researcher's Model of the Decimation Mechanism Assumed to be "Ignorable."</u>

(A3) Assume Y, D_1, and some subset (possibly an empty subset) of the covariates X are observable.

(A4) Let $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, d_2^i, \mathbf{x}^i\right) = p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$ where $\mathbf{x}_{obs}^i$ denotes the covariates which are observable for the ith data record, $i = 1,\ldots,n$.

Assumption A4 states that that the researcher's model of the missing data has the ignorability property as defined by Golden et al.[11] (see also Little and Rubin[16] for a review). Such a property is highly desirable since estimators and statistical tests derived from an ignorable missing data model will not be biased by different forms of the resulting *data decimation mechanism model* $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$. Thus, because of the ignorability assumption, it is not necessary to provide a more specific specification of $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$.

(A5) Assume $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$ satisfies the constraint that whenever $d_1^i = 0$ that the value of d_2^i is not observable.

Assumption A5 shows how the decimation mechanism $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$ is used to represent two fundamentally distinct types of "missingness". First, we have missingness since the questionnaire in the pre-program phase (i.e., the case where $d_1 = 0$) differed from the post-program questionnaire by not including the "duty-hour limits" question that is the basis for determining the distribution of D_2. Second, we have missingness in the post-program phase (i.e., the case where $d_1 = 1$) when the question about "duty-hour limits" does in fact exist because it is possible that the distribution of D_2 may not be observable due to various factors (e.g., participants chose to not answer that question and so on). Both of these two types of missingness may be simultaneously modeled using the decimation mechanism $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$ since this mechanism is functionally dependent upon the observed value of $d_1^i, i = 1, \ldots, n.$ Indeed, if the decimation mechanism $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$ were not dependent on d_1^i (i.e., $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right) = p\left(\boldsymbol{h}^i \mid y^i, \mathbf{x}_{obs}^i\right)$) and given A5, it follows that the variable d_2^i must be eliminated from the model.

Note that if the binary variable $d_1^i = 0$ and the binary variable d_2^i is not observable, then the interaction term $d_1^i d_2^i = 0$ and is observable. To see this, note that (without any loss in generality) it may be assumed that the data generating process generates a complete data record and then subsequently decimates the complete data record. In the situation where the complete data record has $d_1^i = 0$ it is always the case that $d_1^i d_2^i = 0$. However, in the case where $d_1^i = 1$ and d_2^i is not observable, then the interaction term $d_1^i d_2^i$ must be defined as not observable since the value of $d_1^i d_2^i$ cannot be logically inferred without observing the value of d_2^i.

Semantic Interpretation of Interaction Term for the Complete Data Case.

Let $\rho\left(d_1, d_2 \mid \mathbf{x}; \boldsymbol{\beta}\right) \equiv \dfrac{p\left(y=1 \mid d_1, d_2, \mathbf{x}; \boldsymbol{\beta}\right)}{p\left(y=0 \mid d_1, d_2, \mathbf{x}; \boldsymbol{\beta}\right)}$.

Let $r\left(d_1, d_2 \mid \mathbf{x}; \boldsymbol{\beta}\right) \equiv \log\left(\rho\left(d_1, d_2 \mid \mathbf{x}; \boldsymbol{\beta}\right)\right) = \begin{bmatrix} d_1 & d_2 & d_1 d_2 & \mathbf{x}^T & 1 \end{bmatrix} \boldsymbol{\beta}$.

Following standard methods (see Page 11 of Section V in Mullahy[2]), in the special case where no data is missing it follows that the "ratio of ratios" measures the impact of duty-hour limits on the dependent variable while controlling for the effects of time-trends and other covariates. In particular, the Ratio of Ratios (ROR) formula is defined as:

$$ROR \equiv \frac{\rho\left(d_1 = 1, d_2 = 1 \mid \mathbf{x}; \boldsymbol{\beta}\right) / \rho\left(d_1 = 0, d_2 = 1 \mid \mathbf{x}; \boldsymbol{\beta}\right)}{\rho\left(d_1 = 1, d_2 = 0 \mid \mathbf{x}; \boldsymbol{\beta}\right) / \rho\left(d_1 = 0, d_2 = 0 \mid \mathbf{x}; \boldsymbol{\beta}\right)}.$$

The log ROR may be rewritten as:

$$\log ROR = \left(r\left(d_1 = 1, d_2 = 1 \mid \mathbf{x}; \boldsymbol{\beta}\right) - r\left(d_1 = 0, d_2 = 1 \mid \mathbf{x}; \boldsymbol{\beta}\right)\right) - \left(r\left(d_1 = 1, d_2 = 0 \mid \mathbf{x}; \boldsymbol{\beta}\right) - r\left(d_1 = 0, d_2 = 0 \mid \mathbf{x}; \boldsymbol{\beta}\right)\right) =$$

$$\left(\begin{bmatrix} \beta_{1,0} \\ \beta_{2,0} \\ \beta_{1,2} + \beta_{2,1} \\ \boldsymbol{\beta}_x \\ \beta_0 \end{bmatrix}^T \begin{bmatrix} 1 \\ 1 \\ 1 \\ \mathbf{x} \\ 1 \end{bmatrix} - \begin{bmatrix} \beta_{1,0} \\ \beta_{2,0} \\ \beta_{1,2} + \beta_{2,1} \\ \boldsymbol{\beta}_x \\ \beta_0 \end{bmatrix}^T \begin{bmatrix} 0 \\ 1 \\ 0 \\ \mathbf{x} \\ 1 \end{bmatrix} \right) - \left(\begin{bmatrix} \beta_{1,0} \\ \beta_{2,0} \\ \beta_{1,2} + \beta_{2,1} \\ \boldsymbol{\beta}_x \\ \beta_0 \end{bmatrix}^T \begin{bmatrix} 1 \\ 0 \\ 0 \\ \mathbf{x} \\ 1 \end{bmatrix} - \begin{bmatrix} \beta_{1,0} \\ \beta_{2,0} \\ \beta_{1,2} + \beta_{2,1} \\ \boldsymbol{\beta}_x \\ \beta_0 \end{bmatrix}^T \begin{bmatrix} 0 \\ 0 \\ 0 \\ \mathbf{x} \\ 1 \end{bmatrix} \right) = \beta_{1,2} + \beta_{2,1}.$$

However, for the experimental context considered here, the presence or absence of the duty limit effect D_2 does not influence how the program implementation indicator factor D_1 impacts respondent satisfaction rating implying that: $\beta_{2,1} = 0$. Given the *identifiability assumption* that $\beta_{2,1} = 0$, it follows from the above analysis of the ROR that:

$$ROR = \exp\left(\beta_{1,2} + \beta_{2,1}\right) = \exp\left(\beta_{1,2} + 0\right) = \exp\left(\beta_{1,2}\right)$$

and we may write:

$$\begin{aligned}\boldsymbol{\beta} \equiv \begin{bmatrix} \beta_{1,0} & \beta_{2,0} & (\beta_{1,2}+\beta_{2,1}) & \boldsymbol{\beta}_x & \beta_0 \end{bmatrix}^T &= \begin{bmatrix} \beta_{1,0} & \beta_{2,0} & (\beta_{1,2}+0) & \boldsymbol{\beta}_x & \beta_0 \end{bmatrix}^T \\ &= \begin{bmatrix} \beta_{1,0} & \beta_{2,0} & \beta_{1,2} & \boldsymbol{\beta}_x & \beta_0 \end{bmatrix}^T\end{aligned}.$$

Thus, the interaction term coefficient has the semantic interpretation of measuring how the program implementation factor D_1 influences the impact of the duty limit effect D_2 on respondent satisfaction rating.

Missing Data Theory Results.

As described by Golden et al.,[11] the maximum likelihood estimate $\hat{\boldsymbol{\beta}}_n$ of a possibly misspecified missing data model with an ignorable decimation mechanism consistent with assumptions A1-A5 may be computed using the negative log-likelihood:

$$l_n(\boldsymbol{\beta}) \equiv -n^{-1}\sum_{i=1}^{n}\left(\sum_{d_2^i,\mathbf{x}_{miss}^i} \log p\left(y^i \mid d_1^i, d_2^i, \mathbf{x}_{obs}^i, \mathbf{x}_{miss}^i; \boldsymbol{\beta}\right) p_0\left(d_2^i, \mathbf{x}_{miss}^i\right)\right)$$

by setting: $\hat{\boldsymbol{\beta}}_n = \arg\min l_n\left(\hat{\boldsymbol{\beta}}_n\right)$. We refer to $\hat{\boldsymbol{\beta}}_n$ as the RDV maximum likelihood estimate and $\hat{l}_n \equiv l_n\left(\hat{\boldsymbol{\beta}}_n\right)$ as the RDV negative log-likelihood.

Moreover, the missing data theory of Golden et al.[11] formally establishes that the RDV maximum likelihood estimate $\hat{\boldsymbol{\beta}}_n$ is an asymptotically consistent estimator with an asymptotic Gaussian distribution even if a model satisfying A1-A5 is misspecified and even if the missing data generating process is of the most general type (i.e., the data generating process is type MNAR). Furthermore, the methods of Golden et al.[11] were used to derive new asymptotically consistent RDV odds ratio estimators and new asymptotically consistent RDV statistical tests which are valid in the presence of both model misspecification and MNAR statistical environments.

References

1. Bertrand M, Duflo E, Mullainathan S. How much should we trust differences-in-differences estimates? Quarterly Journal of Economics 2004;119:249-275.

2. Mullahy J. Interaction effects and difference-in-difference estimation in loglinear models. Technical working paper #245. Cambridge MA: National Bureau of Economic Research, 1999.

3. White H. Estimation, Inference, and Specification Analyses. New York: Cambridge University Press, 1994.

4. Robins JM, Wang N. Inference for imputation estimators. Biometrika. 2000;87:113-124.

5. Qin J, Zhang B. Empirical-likelihood-based difference-in-differences estimators. Journal of the Royal Statistical Society B. 2008;70 Pt 2:329-349.

6. Golden RM. Making correct statistical inferences using a wrong probability model. Journal of Mathematical Psychology 1995;39:3-20.

7. Golden RM. Mathematical Methods for Neural Network Analysis and Design. Cambridge MA: MIT Press, 1996.

8. Golden RM. Statistical tests for comparing possibly misspecified and nonnested models. Journal of Mathematical Psychology 2000;44:153-170.

9. Golden RM. Discrepancy risk model selection test theory for comparing possibly misspecified or nonnested models. Psychometrika 2003;68:165-332.

10. White H. Maximum likelihood estimation of misspecified models. Econometrica 1982;50:1-25.

11. Golden RM, Henley SS, White H, Kashner TM. Maximum likelihood estimation for misspecified models within missing data: Theory. Manuscript, 2010.

12. Hosmer DW, Lemeshow S. Applied Logistic Regression, Second Edition. New York: John Wiley & Sons, 2000.

13. Ibrahim JG, Chen H, Lipsitz SR, Herring AH. Missing-data methods for generalized linear models: A comparative review. Journal of the American Statistical Association 2005;100(469):332-346.

14. Lipsitz SR, Ibrahim JG. A conditional model for incomplete covariates in parametric regression models. Biometrika 1996;83(4):916-922.

15. Ibrahim JG, Lipsitz SR, Chen MH. Missing covariates in generalized linear models when the missing data mechanism is nonignorable. Journal of the Royal Statistical Society Series B 1999;61:173-190.

16. Little RJA and Rubin DB. Statistical Analysis with Missing Data. 2nd Edition. New York: Wiley and Sons, 2002.