VERMONT LEGAL AID, INC.

OFFICE OF THE HEALTH CARE ADVOCATE

264 NORTH WINOOSKI AVE.
BURLINGTON, VERMONT 05401
(800) 917-7787 (TOLL FREE HOTLINE)
(802) 863-7152 (FAX)

OFFICES:

BURLINGTON
RUTLAND
ST. JOHNSBURY

OFFICES:

MONTPELIER
SPRINGFIELD

August 13, 2020

Kevin Mullin, Chair
Green Mountain Care Board
144 State Street
Montpelier, VT 05602

Dear Chair Mullin:

Please forward the questions below to the Northwestern Medical Center in advance of the hearings, pursuant to 18 V.S.A. §9456(d)(3).

Feel free to contact us at hcapolicystaff@vtlegalaid.org with any questions.

Thank you,

s\ Mike Fisher, Chief Health Care Advocate

s\ Kaili Kuiper, Staff Attorney

s\ Eric Schultheis, Staff Attorney

s\ Julia Shaw, Health Care Policy Analyst

Contact information:

Email: hcapolicystaff@vtlegalaid.org
Phone: (802) 383-2211 (Julia Shaw)

The Office of the Health Care Advocate (HCA) has questions about the multivariate linear regression models and associated analysis found in Northwestern Medical Center's (NMC) 2021 FY2021 budget narrative submitted to the Green Mountain Care Board (GMCB). The models and associated analysis can be found on pages 9 through 10 of NMC's budget narrative and also in Appendix V of the same document. Because NMC has not provided key statistics related to the intermediate and final regression models, we are unable to assess NMC's hypotheses. It is our assumption that NMC has the information we are requesting readily available or can easily produce it using a basic statistical software package. Additionally, we seek to clarify some of the conclusions NMC draws from its models.

We would also like to note that the HCA and other actors have attempted to fit various statistical models to Vermont hospital data. Based on our experience of fitting multivariate linear regression and time-series statistical models (fixed effect and hierarchal linear models) to Vermont hospital data, we believe that fitting statistical models to Vermont hospital data yields highly unstable models of limited use due to the small number of observations.[1] While we hope that NMC has overcome the issues that the HCA encountered when fitting statistical models to Vermont hospital data, our experience with similar models informs our evaluation of NMC's methodology.

HCA Follow-Up Questions on NMC's Regression Analysis:

This set of questions relates to the data used in and specification of the regression models:

1. It appears that the data used to fit both models are from a single cross-section.
   a. Is this correct?
   b. Please specify the length of the period the data is from (i.e. day, month, year, etc.).
2. Please list the hospitals, other than UVMMC and NMC, that the data are from.
3. Please provide the estimator each model (i.e. OLS, maximum likelihood, etc.).
4. Please provide the formula of each model.
   a. Based on our reading of NMC's budget narrative, we believe that you fitted the two regression models listed below. Is this correct?
      - $Y_1 = a + \beta X_1 + \beta X_2 + \beta X_3 + \beta X_4 + \varepsilon$
         - Where $Y_1 = $ *Total Operating Revenue*, $a = $ *intercept*, $X_1 = $ *Admissions*, $X_2 = $ *Operating Room Cases*, $X_3 = $ *Physician Practice Visits*, and $X_4 = $ *Emergency Room Visits*.
      - $Y_2 = a + \beta X_1 + \beta X_2 + \beta X_3 + \beta X_4 + \varepsilon$
         - Where $Y_2 = $ *Total Operating Expense*, $a = $ *intercept*, $X_1 = $ *Admissions*, $X_2 = $ *Operating Room Cases*, $X_3 = $ *Physician Practice Visits*, and $X_4 = $ *Emergency Room Visits*.
5. Please provide the null hypothesis(es) that you tested and the basis for your acceptance or rejection of this hypothesis(es).
6. Please provide the data used to fit the regression models.

---

[1] Neither the HCA nor other actors found these regression models to be sufficiently stable or predictive to be useful for internal or external purposes.

These questions relate to model stability and the explanatory power of the regression models:

7. To our knowledge, the consensus opinion is that a sample size of 10 is too small to fit a linear regression with one, let alone three, independent variables. Please provide the basis for your decision that this was not the case for NMC.
8. For each model presented in the budget narrative, please provide the model information listed below:
    o R^2
    o SS_{Model}
    o SS_{Error}
    o df_{Error}
9. For each model in the budget narrative, please provide the statistics listed below for a baseline control model (a model with only one independent variable)[2]:
    o   Intercept
    o   Coefficient of the independent variable
    o   P-value of the independent variable
    o R^2
    o SS_{Model}
    o SS_{Error}
    o df_{Error}
10. Did you conduct any analysis of whether the explanatory power of including an independent variable was worth the loss in degrees of freedom? For instance, did you conduct any general linear hypothesis testing? If you did conduct such testing, please provide the results of the tests.

This question relates to the assumptions associated with linear regression (linearity, homoscedasticity, independence, and normality).

11. NMC addressed normality in its budget narrative, but not the three other assumptions associated with linear regression. Do the models satisfy the three other assumptions associated with linear regression (linearity, homoscedasticity, and independence)?
    a. Linear regression assumes that the relationship X and the mean of Y is linear.
        i. For each model, please provide scatter plots of the independent variables on the dependent variable.
    b. Linear regression assumes that the observations are independent of each other.
        i. Conceptually, it appears that volume statistics between hospitals could be dependent on each other. For instance, admission and operating volumes could be partially dependent on utilization and/or the mix of available services in neighboring hospitals. Please discuss what led you to conclude that the observations are sufficiently independent of each other.
    c. Linear regression assumes that the variance of residual is the same for any value of the dependent variable.
        i. For each regression model, please provide a scatterplot of the regression standardized residual versus the regression standardized predicted value.

---

[2] For instance, if the model presented in the budget narrative was $Y_1 = a + \beta X_1 + \beta X_2 + \beta X_3 + \beta X_4 + \varepsilon$ then a baseline control model would be $Y_1 = a + \beta X_1 + \varepsilon$.

These questions relate to concerns about possible collinearity of independent variables. Collinearity of independent variables is a substantial problem if present. Conceptually, it appears that some of the independent variables could be dependent on each other. For instance, the number of admissions at a given hospital likely depends on emergency room volume at that hospital. Similarly, the number of operating room cases at a given hospital could depend on admissions.

12. Please provide the bivariate correlation between each of the independent variables.
13. Please provide the variance inflation factor (VIF) for each independent variable coefficient in each of the regression models.
14. Please provide a condition index for each regression model.

These questions relate to the limitations of point-in-time linear regression models to make causal inferences or inferences related to change. Linear regression is normally used to explore correlations between the dependent variable and the primary predictor variable controlling for other variables. Unlike some statistical models, multivariate linear regression does not support causal inference. Additionally, non-time series regression models cannot speak to change over time. However, NMC states that the regression models "support the supposition that limiting net patient revenue will force hospitals to limit expenses." Further, NMC states that the regression models show that "the historical and cumulative implementation of rate regulation has created a disparity between hospitals" and that hospitals "who have been successful in achieving higher rates, either through higher rate allowances or by having higher rates in 2013 when enforcement began, have been able to support higher than expected costs." The HCA makes no substantive judgement on the veracity of these claims at this time; however, we assert that using a point-in-time linear regression to suggest proof of these claims is inappropriate and bordering on misleading.

15. Please explain how the regression models show that limiting NPR will "force" hospitals to take any action. Again, our concern is not whether or not this statement is true but how the regression models could support it.
16. Please explain how the regression models, which appear to be cross-sectional, show how "historical or cumulative implementation of rate regulation has created a disparity between hospitals." Again, our concern is not the veracity of this statement but how the regression models could speak to either historical change or causal forces (i.e. "create a disparity").