

WORKING P A P E R

The Effect of HMOs on the Inpatient Utilization of Medicare Beneficiaries

Technical Appendix

NASREEN DHANANI, JUNE F. O'LEARY,
EMMETT KEELER, ANIL BAMEZAI, GLENN MELNICK

WR-138

February 2004

This Working Paper is the technical appendix to an article published in a scientific journal. It has been subject to the journal's usual peer review process. **RAND**® is a registered trademark.

TECHNICAL APPENDIX

“The Effect of HMOs on the Inpatient Utilization of Medicare Beneficiaries.”

In this Appendix, we explain and justify our analytic approach to estimating the pure managed care effect of Medicare HMOs.

We used a “before and after with a comparison group” design to evaluate the impact of Medicare HMOs on hospital utilization.¹ To take advantage of our enormous data set with up to five years of inpatient data (1991-1995) on millions of Medicare beneficiaries in which the HMO people spent considerable time in and out of an HMO, we were somewhat restrictive in selecting a clean HMO sample. (See Study Design Overview and Sample Selection portions of paper).

After estimating the reduction in utilization due to being in an HMO in our selected sample, we assume that this managed care effect is proportionally the same for everyone, both in and out of this sample. We then calculate the selection effect indirectly as the difference between FFS use and what we hypothesize the people in HMOs would use if they were in FFS.

Other assumptions and choices made:

The unit of analysis is the person/year. We descriptively evaluated utilization and deaths by quarter, and even by month in smaller samples, and found no cyclical patterns or short-run differences in these variables around the time of changes from FFS to HMO or changes from HMO to FFS. Therefore, we defined the “in HMO” variable for each year by the ratio of the months people are in an HMO over the months that people are alive in that year.

Using the person/year in a two-part model means that the first part is whether a person has any hospital days in the year, and the second part is how many days, conditional on having one or more. This specification is a slight departure from standard admission/length of stay

models. The focus on days means that admissions with zero length of stay are excluded. Such admissions were rare in these data (0.2% of people in HMOs had an admission but 0 total days each year as did 0.14% in FFS).

Assume each person/year is an independent observation. Originally we had planned to use panel data methods to exploit the switch from FFS to HMO in our selected analytic sample (xtgee in Stata²). After some preliminary diagnostics, we estimated general linear models (with log link and gamma family errors, as suggested by the patterns in residuals discussed below, which is the same as a one-part exponential regression). These models fit reasonably well despite the large number of zeros. The runs showed that the correlation of residuals of days across years is very low (for random coefficients models it was 0.075 between years). Correlation of health care utilization over time is low in general, but this correlation is particularly low because we are only modeling inpatient use and we are controlling for current and future year death. Ignoring this correlation reduces the precision of our estimates, but does not lead to bias in estimated means, so instead of panel data methods we used simpler models that treat each person/year as a separate observation. The before and after with the always in FFS comparison group aspect of the data is captured using cohort indicator variables that represent switching to HMO and staying or always remaining in FFS, which average the behavior and other unobserved characteristics of individuals in these cohorts.

The correlation of the up to five years of data for each individual does affect the estimated error in estimates, so we estimated robust standard errors to control for clustering of residuals within people.³ The robust confidence intervals of the variable coefficients are up to 30% wider for the logistic regression of any use and 20% wider for the regression of log days

given any use compared to the unadjusted confidence intervals. Increases are greatest for variables that are constant over time, and strong predictors of use (e.g. disabled >65); for most constant variables, cluster corrections are around 15% and for variables that change from year to year such as “in HMO” or die they are almost the same as unadjusted confidence intervals.

We tested the impact of assuming independence of years indirectly by evaluating the sensitivity of the results from the first part of the model using conditional logit, a method that does not assume independence. We performed conditional logit regressions of the probability that a beneficiary had any days in different years with the varying HMO membership variables and the varying death variables. All constant variables drop out of conditional logit, which studies the impact of changes in predictor variables over time on people who sometimes have hospitalizations and sometimes do not. The method predicts for these people the years they have events and is the equivalent of a fixed effects model for logistic regression. The estimated effect of being in an HMO on years with use in these conditional logistic models was identical (odds ratio =1.01) to results from our standard logistic regression when each year was considered independent.

Two-Part Model. We chose the two-part model for hospital days based on theory and statistics. In theory, the decision to hospitalize is often a separate decision from the extent of use (i.e., length of stay) once hospitalized. While the provider and patient may know an expected length of stay, the patient’s ultimate condition and the practice pattern of the physician as well as the system of care (i.e., HMO or FFS) will impact the actual length of stay. Many managed care organizations separate the management of inpatient services into pre-admission certification (whether to hospitalize or not) and concurrent review (length of stay and discharge disposition)

highlighting the two-step nature of hospitalization decisions. Statistically, more than 80% of the sample per year uses zero days and among those who are hospitalized, length of stay has a long right tail. Both these reasons support the two-part model. We did the diagnostic tests recommended for specifying a model in Manning and Mullahy.⁴ We found the two-part model, with the second part a log transformation, fit the data very well. First the variance of raw scale residuals is quadratic with predictions on the raw scale suggesting gamma and log transformation. Also the log scale residuals are homoscedastic with length of stay, close to normal (skewness = 0.08, kurtosis = 2.77), and the smearing factor is almost exactly $\exp(\sigma^2/2)$.

Because of concern about the undue influence of possible data errors or unusual cases, we looked at the distribution of log residuals from the length of stay regression. They were not quite normal, with skewness = .16. However, we tested winsorizing days greater than 91 days (the top 1/4% of annual days) to 91 days, finding it only changed the effect of the HMO membership coefficients in at most the 4th significant figure. Because we defined year by month of admission, and considered length of stay to be the days since admission, a few people had lengths of stay over 365 days in a particular year and these were excluded from our sample.

Retransformation. After studying log (days) in the second part of the two-part model it is necessary to retransform the results back to days. Effects on total days for each variable were calculated by multiplying three parts: the effect on the probability of a year with at least one day, the mean effect on days conditional on any days, and the smearing factor. To calculate the effect of a variable on the probability of a year with at least one day, take for example, “continuous HMO.” First, we recycled the entire sample to calculate the predicted probability p_1 of a year with at least one day if all people were always in the HMO, and the probability p_0 if all people in the sample were in FFS. The impact $p_1/p_0 = 0.855$. This result is closer to 1 than the odds ratio

of 0.78 in Table 3a in the paper because some people are quite likely to have a hospitalization, based on other variables such as death, so changing one value for a predictor does not reduce their chance of hospitalization much.

If the residuals in the log scale of days truly have a normal distribution then mean days are the exponential (exp) of mean log days multiplied by the retransformation factor, which is $\exp(\sigma^2/2)$ for σ the standard deviation of the residuals. An alternative nonparametric retransformation factor is called the smearing estimate, which is the average of $\exp(e_i)$ for the residuals e_i of the regression on the log scale.^{5,6} We report results using the smearing factor to retransform the data, but the median difference between the two retransformation factors on all our contrasts was less than 1%.

Prediction standard errors in retransformed total days. Effects on total days are obtained by multiplying three effects, those from the logistic regression of any use, the $\exp(\text{coefficient of log days})$, and the smearing factor. The standard deviation of the estimated logistic coefficients is almost exactly represented by the first order Taylor series expression (the proof of this is given below), and therefore we can simply evaluate the impact of small changes in the coefficients on predicted probabilities to estimate the factor needed to transform the standard deviation of the coefficient in the original equation to the standard deviation of the impact on population probabilities. For example, a listed standard error of 0.0098 on the logit is 0.0075 in terms of increased log (years with a hospitalization). Because residuals were close to log normal and the smearing factor was so close to the log normal factor of $\exp(\sigma^2/2)$, we assumed the error in the smearing estimate was equal to the error in $\exp(\sigma^2/2)$. Assuming log normality, we can use the standard result that the variance in the estimated variance of σ^2 is

$2\sigma^4/(N-p)$ to compute that error. We assume that errors in the three factors that are multiplied together to estimate effects on total days are independent, so we can add the variance of the log of those factors to get the variance of the log of the estimated effects on total days. Residuals and predictions for $\log(\text{days})$ are independent by construction.

Proof that the first order Taylor series approximation to the confidence intervals is a good assumption for the logistic reduction in years with a hospitalization.

The log of the ratio of probability (with $x=1$)/probability (with $x=0$) can be written out as a function f of β where β is the estimated coefficient of x , the variable of interest. Let K_i represent the rest of the index for person I , and let g be the inverse logit function, $g(x) = (1+\exp(-x))^{-1}$.

Then $f(\beta) = \log [\Sigma g(\beta+K_i) / \Sigma g(K_i)] = \log[\Sigma g(\beta+K_i)] - \log[\Sigma g(K_i)]$. We are interested in $f(\beta + d)$ for small changes d .

The Taylor series is $f(\beta + d) = f(\beta) + d f' + d^2 f''/2$.

The second term of f does not depend on β , so drops out of the derivative. Also we will suppress the argument $\beta+K_i$ in what follows.

Now $f' = \Sigma g' / \Sigma g$ where $g' = \exp(-x) g^2$,

and $f'' = \Sigma g'' / \Sigma g - \Sigma g' \Sigma g' / (\Sigma g)^2$, where $g'' = g' (x) (2g \exp(-x) - 1)$.

Now $h = g \exp(-x) = (1+\exp(x))^{-1}$ is between 0 and 1 so $g' = gh < g$, and the second term of f'' is less than 1. When h is small, $g'' = g'(2h-1)$ can be negative but in absolute value it is always less than g' and hence less than g . So the first part of f'' also has absolute value less than 1, and f'' has absolute value less than 2. Empirically, the absolute value of f'' is largest for small values of x , where g' is close to g , and the expression is negative. At $\beta = -0.287$, which is the smallest it gets, we compute that $f'' = (0.072/0.142) - [(0.111)^2/(0.142)^2] = -0.104$.

The largest standard error in estimating any coefficient is 0.018 for “die in that year”, so 2 standard errors = 0.036. So the second term of the Taylor series with $d = 2$ standard errors satisfies $|f''d^2/2| < 0.104(0.036)^2/2 = 0.00007$, which is negligible, so we will use the first order Taylor series values in the calculation of standard error of f .

REFERENCES TO THE TECHNICAL APPENDIX

1. Cook TD, Campbell DT. Quasi-experimentation: Design & analysis issues for field settings. Chicago,IL: Rand McNally College Pub. Co.; 1979.
2. StataCorp. *Stata Statistical Software: Release 6.0*. College Station, TX: Stata Corporation; 1999.
3. Huber PJ. The behavior of maximum likelihood estimates under non-standard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA: University of California Press; 1967:221-233.
4. Manning WG, Mullahy J. Estimating Log Models: To Transform or Not to Transform? *Journal of Health Economics*. 2001;20(4):461-494.
5. Duan N, Manning WG, Jr., Newhouse JP. A comparison of alternative models for the demand for medical care. *Journal of Business & Economic Statistics*. 1983;1(2):115-126.
6. Duan N. Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association*. 1983;78(383):605-610.