

A collection of statistical mistakes

E.P. Martens, Statisticor, 2015

Remarks on the paper "**Similar patient outcomes yet different hospital costs between flowable hemostatic agents**", Guy David, Sangtaeck Limb, Candace Gunnarssonc, Richard Kochariand and Sanjoy Royin, Journal of Medical Economics Volume 18, Issue 9, 2015.

1. Introduction

The main objective of the paper is to assess the difference in total costs when either using SURGIFLO or FLOSEAL (both with Thrombin) during surgeries in order to stop bleeding. Because this is not a randomized, but an *observational study*, there is the need to adjust for possible *confounding factors*. In this paper adjustment has been made for

- month and year,
- used product quantity,
- patients characteristics,
- hospital characteristics and
- surgical characteristics.

Using data from more 16,000 patients, they performed three types of models:

- linear regression models to assess the difference in costs between SURGIFLO and FLOSEAL,
- linear regression models to assess the difference in costs between SURGIFLO and FLOSEAL, but including hospital, which they call 'hospital fixed-effectsmodels'
- linear regression (?) models to assess the difference in surgical complications.

Next to the effect of SUGIFLO (yes/no), the authors show the effects of surgery duration and its interaction with SURGIFLO/FLOSEAL.

In the next paragraphs we will comment on this research from a statistical and methodological point of view.

2. Inclusion of used quantity (ml) as a confounder

The authors present as most important cost models those with number 7 and 14 (table 4), in which adjustment has been made for all confounders mentioned in the paper. From model 3 resp. 10 onwards, the *product quantity* is

included as a confounder. It seems that this is the total quantity used for the surgery, which means for SURGIFLO the number of units multiplied by 8 ml and for FLOSEAL the number of small packages multiplied by 5 ml plus the number of large packages multiplied by 10 ml.

The main comment on including this variable as a confounder in the analysis, is that this is not correct. The reason is quite simple: the main objective of this study is to explain *whether there is a difference in total cost between both products* when used in clinical practice. When including product quantity (ml) in the analysis, one tries to explain the difference in *cost per ml* between both products, i.e. trying to explain the difference in unit price. That is not the objective of the study and is not interesting at all. When one product needs more units (because of less efficiency or other reasons), the analysis should *not* adjust for that.

When looking at the differences between model 2 and model 3 in table 4 (where quantity used was added), a large change in coefficients can be seen. The coefficient for SURGIFLO changes from -236 to -54.0, for surgery hours from 33.9 to 1.96 and for the interaction from -1.84 to -22.3. Of course, the percentage of explained variance (R^2) also changes from 22.3% to 78.4%, but changes the interpretation of the model completely; a model that tries to explain the total cost to a model trying to explain the total costs, given the quantity.

Then, in models 3-7 and models 10-14, adjustment has been made for all other confounders, but hardly any change emerges in coefficients. This is not surprising, because *given the quantity used* patient, surgery and hospital characteristics does not have any impact any more. Whether it has an impact on total costs, can not be assessed with these analyses.

3. Models

There were two sets of models applied on costs, as they call it 'ordinary least squares cross-sectional' and a 'hospital-level fixed-effects'. In the first it is clear that a linear regression has been used, but it is unclear what was used in the second. Their description is unclear:

"...; specification; and the second specification added hospital-level fixed effects. The fixedeffects model assists in controlling for unobserved heterogeneity when this heterogeneity is constant over time and correlated with the choice of flowable agent within hospitals. This constant is removed from the data through differencing. These variations in the model specifications were included to assess effect of flowable hemostat agent use on costs and clinical outcomes across hospitals and within hospitals."

A quick search for some text strings on the internet led directly to Wikipedia "http://en.wikipedia.org/wiki/Fixed_effects_model", in which the next description was found:

"... assist in controlling for unobserved heterogeneity when this heterogeneity is constant over time and correlated with independent variables. This constant can be removed from the data through differencing, ..."

It is clear that this text has been directly copied from this internet source. Although that piece of text in that source isn't clear either, the context is quite different because there it concerns a *longitudinal model*. The model used in this paper has *no* longitudinal component at all. Therefore, this

description is not correct and also the formula given in the section Statistical Analyses can not apply to these data. The subscript 't' that is used in this formula only applies when measurements are taken over time on the same measurement units (here patients), which is not the case. This gives the feeling that the authors put in a formula without knowing how to read and explain it. They didn't even explain the terms ϕ_t , μ_i and ϵ_{jikt} .

4. Dependent observations

Possibly, in the second set of models only hospital is included in the model as a fixed-effect. Why a fixed-effect has been chosen is unclear. What should be addressed here, is that observations within hospitals probably are correlated because the use of SURGIFLO and FLOSEAL is a choice made on hospital level (may be to be confirmed). For the first set of linear regression models this means that the assumption of independent observations is not fulfilled, leading to possibly wrong conclusions on the factors of interest.

To address this, one should take into account a 'hospital effect', which can better be incorporated as a random effect, because it is not the objective to assess effects of specific hospitals.

5. Wrong statements on coefficients

The main conclusion is taken from model 7 and model 14 in table 4. Although it is quantitatively not a major difference, it is amazing to see that the authors don't know how to interpret coefficients in a regression model.

In the abstract:

"Regression models showed a reduction in cost of \$65 associated with use of SURGIFLO with Thrombin and an additional \$21 reduction in hospital cost for each additional hour of surgery."

In the text:

"These results were also confirmed by the cross-sectional model (Table 4) which showed that reduction in costs for utilizing SURGIFLO with Thrombin compared to FLOSEAL was estimated at \$65, with additional estimated reduction of \$21 for each additional hour of surgery."

and:

"The results suggest that in addition to a gap of approximately \$300, a baseline favoring SURGIFLO, every additional hour of surgery was associated with an additional reduction in hospital costs of approximately \$25.7. For example, an eight hour surgery with SURGIFLO with Thrombin would reduce hospital costs by approximately \$500 compared to an eight hour surgery with FLOSEAL performed at the same hospital, and controlling for surgery, patients and other characteristics."

In order to calculate the difference in costs between the two products one should take into account all three coefficients, not only the coefficient for SURGIFLO and the interaction with duration, but *also the coefficient for duration itself*. That means that 'an additional hour' does not cost \$21.3, but $\$21.26 - \$1.518 = \$19,7$. For their 'hospital fixed-effects regression', this should be \$22.9 instead of \$25.7.

6. Results 'hospital fixed-effect' model

The unadjusted difference in costs between the two products is $\$349.8 - \$222.66 = \$127.14$, favoring SURGIFLO. When an average of 3.85 hours of

surgery is filled in the regression model 1 (which also includes year and month), one comes close to this difference (= **\$115**).

In the hospital fixed-effect model 8, the only difference with model 1 is that hospital is included (still patient level, same dependent, same number of patients, no adjustment except for year and month not reported). When in this model the average duration of 3.85 is filled in the model, the difference between the two products turns out to be **\$224**, which is much higher than \$115 as found in model 1. This difference I really don't understand.

When adjustment has been done for all factors (model 14), this difference has grown to **\$387**. That means that adjustment leads to an additional difference of **\$163**. When this is compared to the effect of adjustment in the first set of models (model 1 and 7), this is much higher; there it is only **\$25**. In conclusion: the meaning of this second set of models is completely unclear to me.

7. Skewness and residuals

It has been explained that a linear regression has been performed on costs. In general, the distribution of costs is highly skewed to the right and I suppose that this is also the case for these analyses. Because no information is given on outliers or residuals, it can not be assessed whether this has an impact on the results. It is an omission that no information is available.

8. Robustness check

The authors state that they performed 'robustness checks': "Robustness checks were performed by varying surgery length 5-to-8 hours versus surgeries lasting 9-to-12 hours and by creating interaction terms for hemostatic matrices by surgery time categories in hours (1-to-4, 5-to-8, 9-to-12)."

It is not clear what they have done. They 'varied surgery length', but this is an observed characteristic/variable in the data set that has variation between individuals, with mean=3.85 and sd=2.00. That means it is a (very) right skewed distribution, probably extending to 9 or 10 hours. 'Varying surgery hours' seems as if some simulation has been performed, but this is not mentioned at all. May be they only filled in 5-8 hours in the model to see what will be the estimated difference between products. In that case, it is strange to choose these categories while most of the data is located around 3-4 hours. A similar remark can be made for their example when explaining the results: '... an eight hour surgery with SURGIFLO ...'. Why to choose 8 hours when the mean is around 3-4?

9. Models on clinical outcomes

The clinical outcomes are dichotomous variables. As I do not see any other explanation on used models, I assume that here also OLS has been used. This is not suitable for dichotomous dependent variables.