

“Before and After” Models in Observational Research Using Random Slopes and Intercepts

David J. Pasta, ICON Clinical Research, San Francisco, CA

ABSTRACT

In observational data analyses, it is often useful to use patients as their own controls by comparing their outcomes “before” and “after” some signal event, such as the initiation of a new therapy. It may be useful to have a control group that does not have the event but instead is evaluated before and after some arbitrary point in time, such as their birthday. In this context, the change over time is a continuous outcome that can be modeled as a (possibly discontinuous) line, with the same or different slope before and after the event. Mixed models can be used to estimate random slopes and intercepts and compare patients between groups. A specific example published in a peer-reviewed journal is presented.

SCIENTIFIC BACKGROUND

The complex model used as our example arose in connection with a study of cystic fibrosis (Konstan et al., 2011). Cystic fibrosis (CF) is a hereditary disease that leads to long term decline in lung function. One measure of lung function is FEV₁, the Forced Expiratory Volume in 1 second, which is obtained during a pulmonary function test (PFT). Because the volume of air that can be expelled in 1 second varies considerably based on the size of the lung, it is common in CF to calculate a “% predicted” measure that is based on the sex, age, height, and race/ethnicity of the patient. This relates the absolute FEV₁ to the expected (mean) value based on the patient's characteristics, expressed as a percentage. It is common practice to track FEV₁ % predicted over time to document patients' lung function decline. There are some disadvantages to modeling the % predicted values (rather than z-scores or other alternative measures), but this approach models the measures most commonly used clinically.

Consider choosing an arbitrary point in time and assessing the lung function of a cystic fibrosis patient, as measured by FEV₁ % predicted, before and after that index time. At first thought, it may seem that there is no reason to believe the rate of decline would be any different before and after the arbitrary index time. However, previous research has shown that high lung function is an independent risk factor for decline. Therefore, patients with higher than average lung function are expected to experience a steeper than average decline going forward and patients with lower lung function are expected to experience a less steep decline going forward. Furthermore, it stands to reason that patients with relatively high lung function at that index time are likely to have had more gradual prior decline than patients with relatively low lung function. (This is a sort of regression to the mean effect looking backwards in time.) These two factors combine to produce the expectation that patients with relatively high lung function at the index time are likely to show a change from mild decline to steeper decline, whereas those with relatively low lung function are likely to show a change from steep decline to milder decline. Thus, the null hypothesis of no change in average decline before and after an arbitrary index time may need to be adjusted depending on the measured lung function at that index time.

THE STATISTICAL MODEL

In the statistical model, we wanted to quantify the average rates of decline in FEV₁ % predicted before and after an index time, separately by age and treatment group. For untreated (comparator) patients, an index pulmonary function test was defined as the PFT closest (within 30 days) to the first encounter within one year following the eighth or subsequent even numbered birthday. (Even numbered birthdays were used to avoid having overlapping pre-index periods.) For patients treated with dornase alfa (the treatment under study), the beginning of treatment was used rather than the birthday. The pre-index and post-index periods – each 2 years in duration – were each required to have ≥ 1 encounter and ≥ 3 FEV₁ values spanning at least six months to estimate the slope of FEV₁. Comparator patients were included for as many sets of pre-index and post-index periods as they had available data.

For every patient and index value, separate regression lines were fit during each of the two-year pre-index and post-index periods. The index PFT was excluded from both the pre-index and post-index periods to minimize issues associated with regression to the mean. The regression lines were fit using

PROC MIXED in SAS® with four random effects: intercept (at the index PFT) and slope before the index event, and change in intercept and change in slope after the index event.

THE RESULTS

Figure 1 shows the average linear trends by age and treatment group, and Figure 2 shows the annual rates of decline “before” and “after” the index event by age and treatment group.

Figure 1: Estimated linear trends by age group (8-17 or 18+ years) and treatment group (Dornase alfa or Comparator)

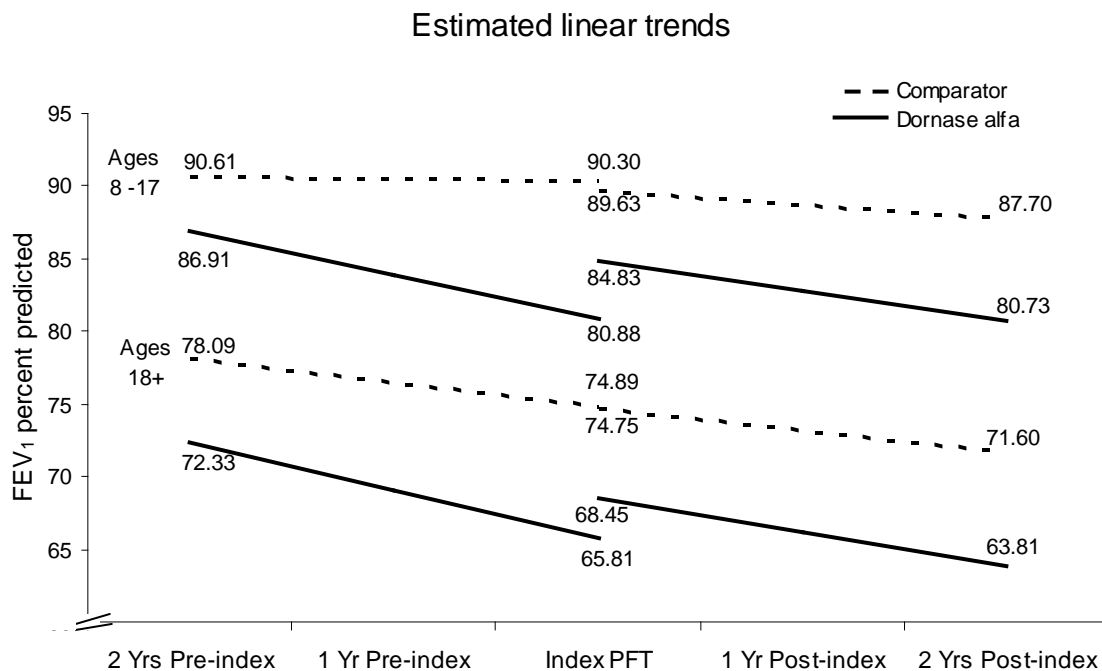
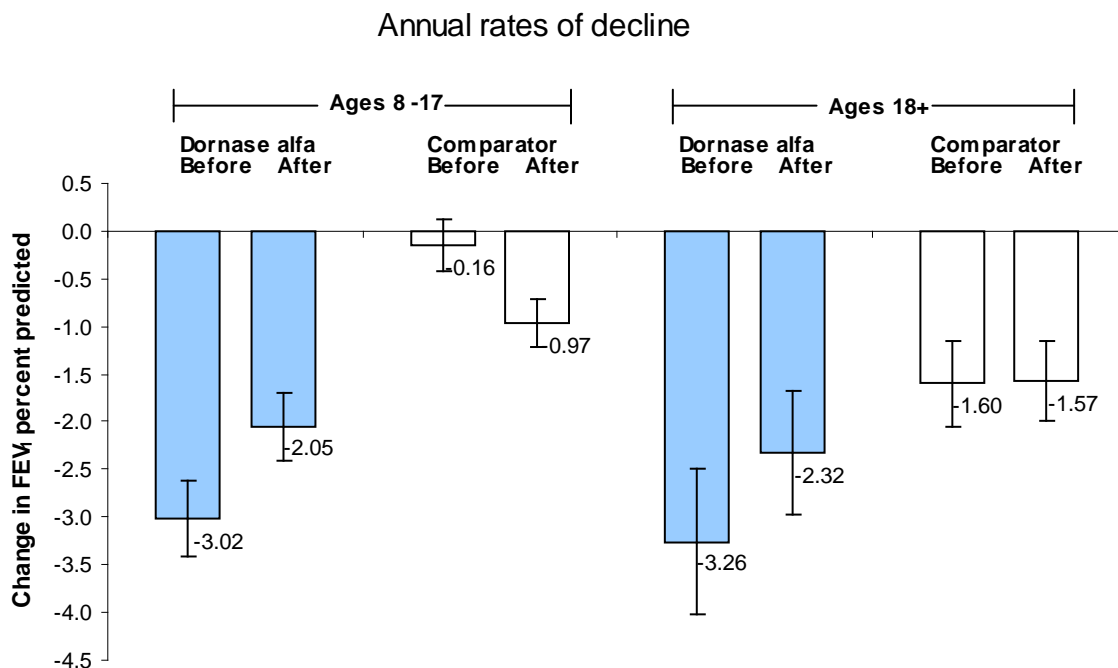


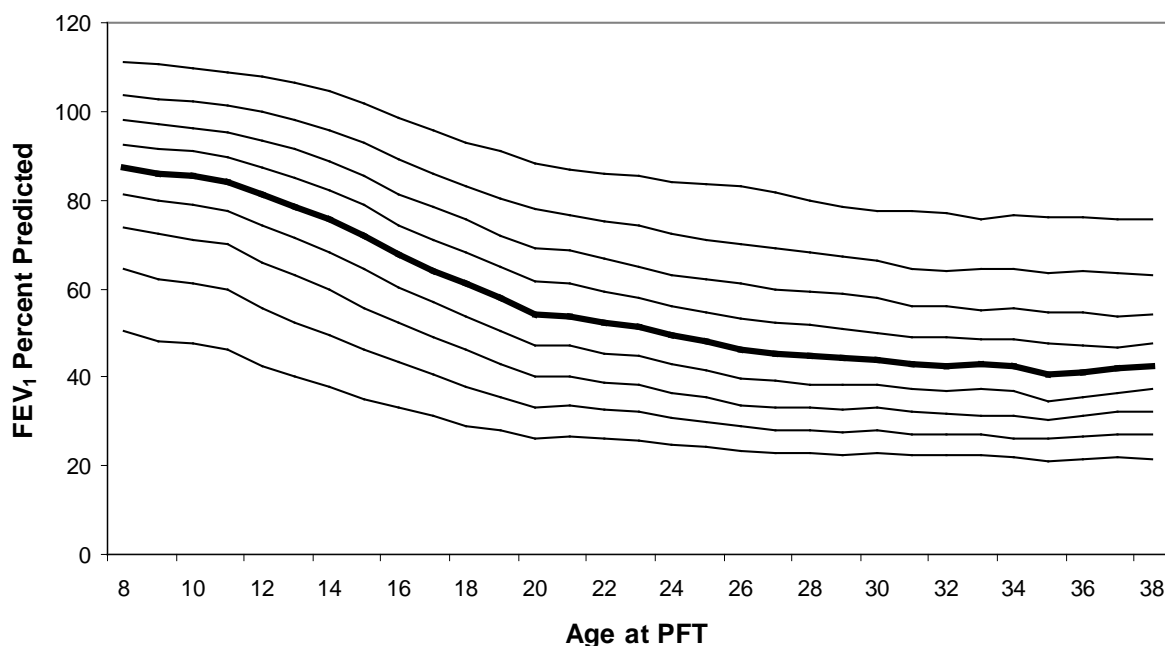
Figure 2: Annual rates of decline before and after the index event by age group (8-17 or 18+ years) and treatment group (Dornase alfa or Comparator)



A REFINED STATISTICAL MODEL

These results were interesting and encouraging that dornase alfa had a desirable effect of lung function decline, but there were very substantial differences in the baseline severity between the dornase alfa and comparator groups. Could we adjust for severity in the models? When we tried characterizing severity using FEV₁ % predicted values, we ran into the difficulty that there were few younger patients in the most severe categories and few older patients in the least severe categories. To provide for a more balanced distribution across categories by age, we characterized lung function relative to other CF patients at every age from 8 to 38 years using all PFTs in the dataset to establish age-specific deciles of FEV₁ % predicted. Figure 3 shows the deciles of lung function by age.

Figure 3: Deciles of FEV₁ Percent Predicted by Age

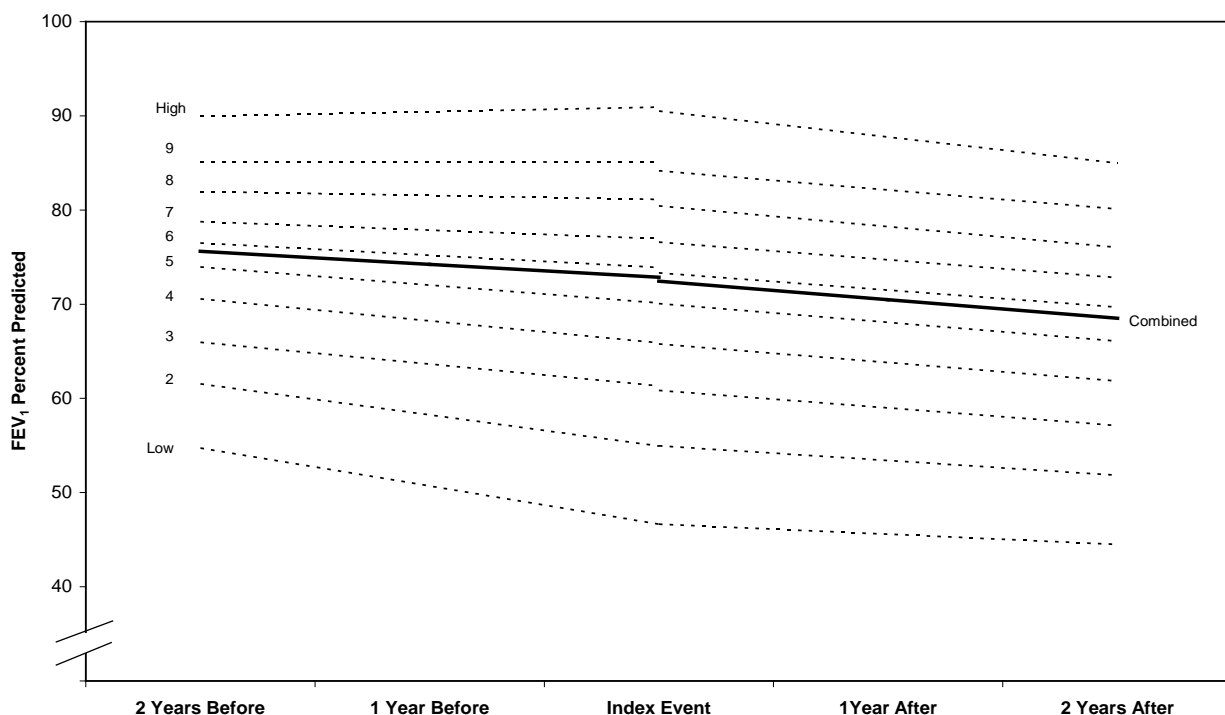


The refined model calculated separate regression lines for pre- and post-index for each patient and index value and then estimated average values by treatment group (dornase alfa or comparator) and by severity decile (1 to 10) based on the FEV₁ % predicted at the index PFT.

Figure 4 shows the average pre-index and post-index fitted lines by decile for the comparator group; Table 1 provides the details.

In addition to estimating the average lines by decile, an overall estimate was obtained by combining the deciles using equal weighting (each decile counted equally) and the observed distribution (each decile counted according to the number of patients represented). These two ways of combining the deciles differ because the number of patients with available data varied by decile; the figure presents the version based on the observed distribution.

Figure 4: Pre- and Post-Index Slopes and Increment at Index Event by FEV₁ Decile for Comparator Patients



The results show the anticipated "bowing." The middle deciles have similar slope pre- and post-index with little change in intercept. For the lower deciles, the pre-index slopes are fairly steep compared to the post-index slopes, which are fairly flat. The opposite is the case for the higher deciles, where the pre-index slopes are fairly flat and the post-index slopes are fairly steep. The differences in estimated intercept are an indication that the straight lines do not adequately fit what is presumably a curved trajectory. Although it is reasonable to approximate the rate of change over short times using a straight line, fitting straight lines to up to two years of data may be more problematic if the shape of the curve is not exactly quadratic. The more curved the true underlying trends, the more likely there is to be an observed difference in intercept when straight lines are fit in the two time periods.

Table 1 includes many different values, some with standard errors and some with *P* values, most of which came from ESTIMATE statements.

Table 1: Details of Pre- and Post-Index Slopes and Increment at Index Event by FEV₁ Decile and Combined for Comparator Patients

Decile	N	Pre-index Slope (SE)	Post-index Slope (SE)	Slope Difference (SE)	P Difference	Post-index Increase (SE)	P Increase	Pre-index Start	Pre-index Stop	Post-index Start	Post-index Stop
Combined (observed)	32355	-1.38 (0.05)	-1.98 (0.04)	-0.60 (0.06)	<.001	-0.40 (0.06)	<.001	75.61	72.85	72.45	68.50
Combined (uniform)	32355	-1.60 (0.05)	-1.92 (0.04)	-0.31 (0.06)	<.001	-0.36 (0.06)	<.001	73.91	70.70	70.33	66.50
1	2155	-4.05 (0.16)	-1.07 (0.16)	2.97 (0.23)	<.001	0.00 (0.21)	1.00	54.74	46.65	46.65	44.50
2	2511	-3.29 (0.15)	-1.56 (0.15)	1.73 (0.21)	<.001	0.01 (0.20)	0.95	61.53	54.95	54.96	51.84
3	2874	-2.33 (0.15)	-1.86 (0.14)	0.47 (0.20)	0.021	-0.47 (0.19)	0.014	65.97	61.32	60.85	57.12
4	3091	-2.35 (0.14)	-1.97 (0.13)	0.38 (0.20)	0.057	-0.10 (0.19)	0.59	70.57	65.88	65.78	61.85
5	3292	-1.87 (0.14)	-2.00 (0.13)	-0.13 (0.20)	0.52	-0.14 (0.19)	0.46	73.95	70.21	70.07	66.08
6	3428	-1.30 (0.14)	-1.82 (0.13)	-0.52 (0.19)	0.007	-0.58 (0.19)	0.002	76.49	73.90	73.32	69.68
7	3639	-0.90 (0.14)	-1.90 (0.13)	-1.00 (0.19)	<.001	-0.36 (0.19)	0.059	78.76	76.96	76.60	72.81
8	3768	-0.43 (0.14)	-2.20 (0.13)	-1.77 (0.19)	<.001	-0.69 (0.19)	<.001	81.97	81.12	80.43	76.03
9	3782	-0.02 (0.14)	-2.03 (0.13)	-2.01 (0.19)	<.001	-0.92 (0.19)	<.001	85.13	85.08	84.16	80.10
10	3815	0.49 (0.14)	-2.77 (0.14)	-3.26 (0.20)	<.001	-0.40 (0.20)	0.040	89.94	90.92	90.51	84.98

THE STRATIFIED PIECEWISE LINEAR MODEL AND SOME ESTIMATE STATEMENTS

The model stratifies patients into ten deciles (based on their lung function at an index event) and fits two separate lines for each patient, one before the index event and one after the index event. Shown below is the model and the first group of ESTIMATE statements. The model includes four variables that define the two lines. The intercept estimates the value at time 0 (the index event) using the pre-index data. The variable t represents time and ranges from -2 to +2. The variable tafter is 1 for the time after index and 0 before; it represents the change in intercept from pre- to post-index. The variable $t0$ equals $\max(t, 0)$ and therefore represents the change in slope between the pre-index and post-index periods. This is a convenient parameterization for piecewise linear models because it provides a direct test of the change in intercept and the change in slope, both of which are likely to be of interest. For more on parameterization of piecewise linear models, see Pasta 2005.

The model includes the decile variable (which takes on values 1 to 10 to represent the 10 deciles of severity) alone and interacted with t , $t0$, and tafter . This causes the MIXED procedure to calculate, for each decile, an estimated average value for the intercept, t , $t0$, and tafter .

Technical Note: Because the same patient could contribute more than one index event, the mixed model allows for within-patient correlation through the use of two RANDOM statements. At the level of the index event, all four parameters of the lines were treated as random effects with an unstructured covariance matrix parameterized as $\text{fa0}(4)$ to ensure it is positive semi-definite. At the level of the patient, the slope and intercept of the pre-index line were treated as random effects with unstructured covariance; we found empirically that the patient-level variances associated with the change in intercept and the change in slope were near zero, so we set them to zero to avoid numerical instabilities. This implies that there is a correlation within patient for the overall slope and intercept but not for the change values.

```

proc mixed data = anal01 noclprint;
  class patid patid_age decile;
  model fevlpct = decile
              decile*t
              decile*t0
              decile*tafter / solution ddfm=bw;
  random intercept t / sub=patid type=fa0(2) g gcorr;
  random intercept t t0 tafter / sub=patid_age(patid) type=fa0(4) g gcorr;
  *** Estimates for slope BEFORE index event ***;
  estimate 'Before:1'  decile*t    1 0 0 0 0 0 0 0 0 0 0;
  estimate 'Before:2'  decile*t    0 1 0 0 0 0 0 0 0 0 0;
  estimate 'Before:3'  decile*t    0 0 1 0 0 0 0 0 0 0 0;
  estimate 'Before:4'  decile*t    0 0 0 1 0 0 0 0 0 0 0;
  estimate 'Before:5'  decile*t    0 0 0 0 1 0 0 0 0 0 0;
  estimate 'Before:6'  decile*t    0 0 0 0 0 1 0 0 0 0 0;
  estimate 'Before:7'  decile*t    0 0 0 0 0 0 1 0 0 0 0;
  estimate 'Before:8'  decile*t    0 0 0 0 0 0 0 1 0 0 0;
  estimate 'Before:9'  decile*t    0 0 0 0 0 0 0 0 1 0 0;
  estimate 'Before:10' decile*t    0 0 0 0 0 0 0 0 0 1 0;
  estimate 'Before:U'  decile*t    1 1 1 1 1 1 1 1 1 1 1 / divisor=10;
  estimate 'Before:O'  decile*t    0.0666048524 0.0776077886 0.0888270746
                                0.0955339206 0.1017462525 0.1059496214
                                0.1124710246 0.1164580436 0.1168907433
                                0.1179106784 ;

  [MORE . . . ]

```

These estimate statements start out pretty easy. To get the estimated slope before the index event for each decile, we just need to pick out the interaction of the decile and the time variable for that decile. Note that the 0s after the 1 are not strictly necessary; I like to include them to make clear what is going on. The decile*t effect represents 10 parameters, so I like to have all 10 coefficients appear.

Is it possible to get these same estimates simply by specifying "solution" on the model statement? Indeed it is. Later values will not be possible to get that way and this is a good check on the coding.

The next two ESTIMATE statements obtain an overall average "before" slope. The one labeled U uses a uniform distribution across the deciles – each decile is given a weight of 0.1. This could be expressed as a coefficient of 0.1 for each or, as is done here, by specifying a 1 for each coefficient and a divisor of 10. The divisor approach is especially convenient for avoiding long fractions when there are, say, 7 or 13 categories. The one labeled O uses the observed distribution across the deciles. The extent to which some deciles are over- or under-represented here is a reflection of the actual data available. There are times when the uniform approach makes the most sense and there are times when the observed approach makes the most sense. This choice corresponds to the default in LSMEANS (the uniform approach) or the version you get when you specify OBSMARGINS (the observed approach). Note that the observed version requires calculating to many decimal places to get good accuracy and making sure the coefficients sum exactly to 1.0 (if they do not, because of rounding, they need to be adjusted so that they do). For other ways to specify these coefficients, including get SAS to do much of the work, see Pasta (2010).

MORE ESTIMATE STATEMENTS

In addition to the pre-index slope, we want to look at the post-index slope and various other values derived from the four values (intercept, t, t0, and tafter). Here are more ESTIMATE statements.

```

*** Estimates for slope AFTER index event ***;
estimate 'After:1'  decile*t    1 0 0 0 0 0 0 0 0 0 0
                   decile*t0   1 0 0 0 0 0 0 0 0 0 0;
estimate 'After:2'  decile*t    0 1 0 0 0 0 0 0 0 0 0
                   decile*t0   0 1 0 0 0 0 0 0 0 0 0;
estimate 'After:3'  decile*t    0 0 1 0 0 0 0 0 0 0 0
                   decile*t0   0 0 1 0 0 0 0 0 0 0 0;
estimate 'After:4'  decile*t    0 0 0 1 0 0 0 0 0 0 0

```

```

estimate 'After:5' decile*t0 0 0 0 1 0 0 0 0 0 0;
decile*t 0 0 0 0 1 0 0 0 0 0
decile*t0 0 0 0 0 1 0 0 0 0 0;
estimate 'After:6' decile*t 0 0 0 0 0 1 0 0 0 0
decile*t0 0 0 0 0 0 1 0 0 0 0;
estimate 'After:7' decile*t 0 0 0 0 0 0 1 0 0 0
decile*t0 0 0 0 0 0 0 1 0 0 0;
estimate 'After:8' decile*t 0 0 0 0 0 0 0 1 0 0
decile*t0 0 0 0 0 0 0 0 1 0 0;
estimate 'After:9' decile*t 0 0 0 0 0 0 0 0 1 0
decile*t0 0 0 0 0 0 0 0 0 1 0;
estimate 'After:10' decile*t 0 0 0 0 0 0 0 0 0 1
decile*t0 0 0 0 0 0 0 0 0 0 1;
estimate 'After:U' decile*t 1 1 1 1 1 1 1 1 1 1
decile*t0 1 1 1 1 1 1 1 1 1 / divisor=10;
estimate 'After:O' decile*t 0.0666048524 0.0776077886 0.0888270746
0.0955339206 0.1017462525 0.1059496214
0.1124710246 0.1164580436 0.1168907433
0.1179106784
decile*t0 0.0666048524 0.0776077886 0.0888270746
0.0955339206 0.1017462525 0.1059496214
0.1124710246 0.1164580436 0.1168907433
0.1179106784 ;

```

*** Estimates for DIFFERENCE in slope between before and after index event ***;

```

estimate 'Diff:1' decile*t0 1 0 0 0 0 0 0 0 0 0;
estimate 'Diff:2' decile*t0 0 1 0 0 0 0 0 0 0 0;
estimate 'Diff:3' decile*t0 0 0 1 0 0 0 0 0 0 0;
estimate 'Diff:4' decile*t0 0 0 0 1 0 0 0 0 0 0;
estimate 'Diff:5' decile*t0 0 0 0 0 1 0 0 0 0 0;
estimate 'Diff:6' decile*t0 0 0 0 0 0 1 0 0 0 0;
estimate 'Diff:7' decile*t0 0 0 0 0 0 0 1 0 0 0;
estimate 'Diff:8' decile*t0 0 0 0 0 0 0 0 1 0 0;
estimate 'Diff:9' decile*t0 0 0 0 0 0 0 0 0 1 0;
estimate 'Diff:10' decile*t0 0 0 0 0 0 0 0 0 0 1;
estimate 'Diff:U' decile*t0 1 1 1 1 1 1 1 1 1 1 / divisor=10;
estimate 'Diff:O' decile*t0 0.0666048524 0.0776077886 0.0888270746
0.0955339206 0.1017462525 0.1059496214
0.1124710246 0.1164580436 0.1168907433
0.1179106784 ;

```

*** Estimates of INCREASE AFTER the index event ***;

```

estimate 'IncAfter:1' decile*tafter 1 0 0 0 0 0 0 0 0 0;
estimate 'IncAfter:2' decile*tafter 0 1 0 0 0 0 0 0 0 0;
estimate 'IncAfter:3' decile*tafter 0 0 1 0 0 0 0 0 0 0;
estimate 'IncAfter:4' decile*tafter 0 0 0 1 0 0 0 0 0 0;
estimate 'IncAfter:5' decile*tafter 0 0 0 0 1 0 0 0 0 0;
estimate 'IncAfter:6' decile*tafter 0 0 0 0 0 1 0 0 0 0;
estimate 'IncAfter:7' decile*tafter 0 0 0 0 0 0 1 0 0 0;
estimate 'IncAfter:8' decile*tafter 0 0 0 0 0 0 0 1 0 0;
estimate 'IncAfter:9' decile*tafter 0 0 0 0 0 0 0 0 1 0;
estimate 'IncAfter:10' decile*tafter 0 0 0 0 0 0 0 0 0 1;
estimate 'IncAfter:U' decile*tafter 1 1 1 1 1 1 1 1 1 1 / divisor=10;
estimate 'IncAfter:O' decile*tafter 0.0666048524 0.0776077886 0.0888270746
0.0955339206 0.1017462525 0.1059496214
0.1124710246 0.1164580436 0.1168907433
0.1179106784 ;

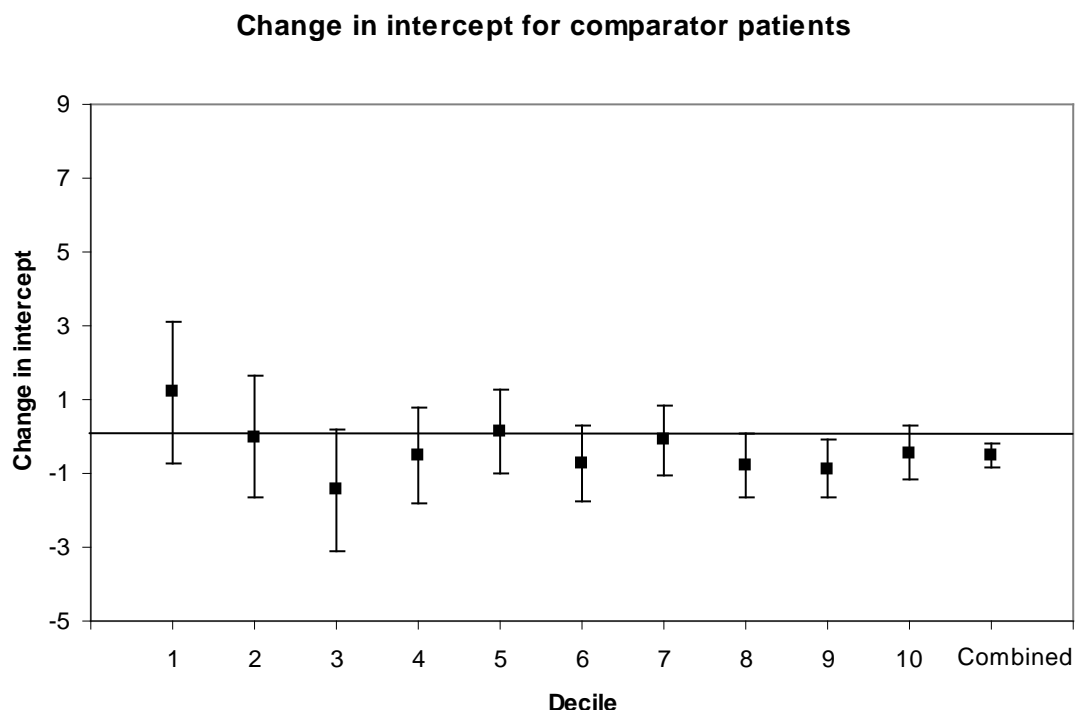
```

To get the slope after the index event, it is necessary to include the coefficient of t and of t_0 for the corresponding decile. Note that there is *not* a semicolon between the two lines; you are creating a single estimate using coefficients from two different terms in the model.

GRAPHICAL DISPLAY OF THE RESULTS

There are graphs that can go with these analyses. They are not necessarily familiar or easy to understand at first, but once you understand them they provide a compact visualization of the results.

Figure 5: Change in Intercept for Comparator Patients by FEV₁ Decile



Note: Error bars represent 95% CI.

Figure 5 shows the change in intercept (the discontinuity in the lines at the index event) for the comparator patients. These should be about zero, and indeed 9 of the 10 deciles have confidence intervals that are at or cross the zero line. However, the overall combined estimate is a little less than zero. This is an indication that the two fitted straight lines do not meet exactly, on average. This is an indication that the underlying curvature may not be well approximated by a quadratic curve. If the underlying curve were quadratic, two straight lines fitted in this fashion would, on average, meet.

Figure 6 shows the **difference** in the change in intercept for the dornase alfa group compared to the corresponding comparator group. This adjusts the change in intercept for dornase alfa patients for any change in the comparator group in that decile. For 9 of the 10 deciles, the difference is statistically significant (the confidence interval does not cross zero). The combined result shows about a 4-point greater change in intercept for dornase alfa patients than comparator patients. This is a very substantial effect. As is typical with observational studies, the effect is smaller than observed in controlled clinical trials.

Figure 6: Difference in Change in Slope, Dornase Alfa versus Comparator by FEV₁ Decile

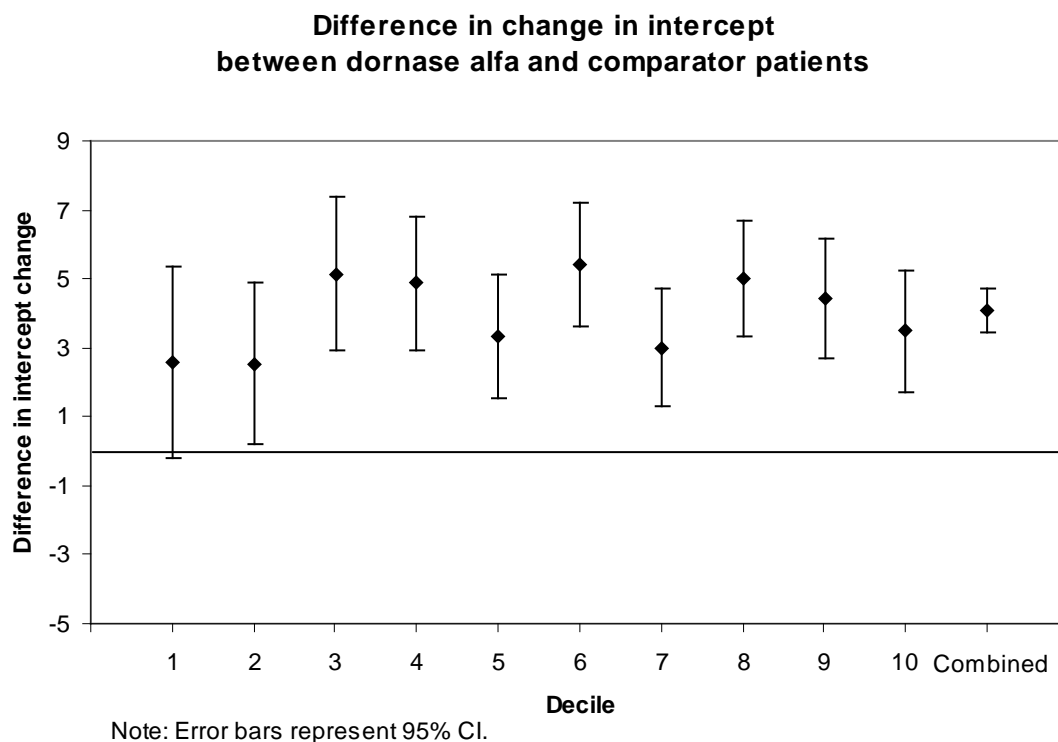


Figure 7: Change in Slope for Comparator Patients by FEV₁ Decile

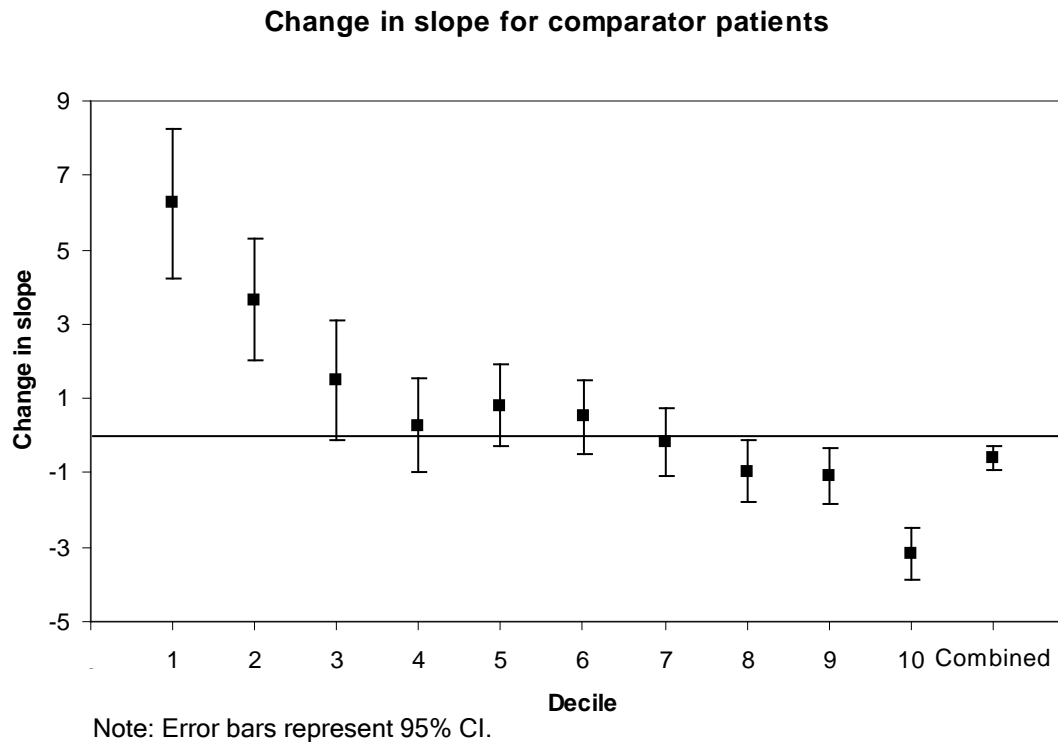
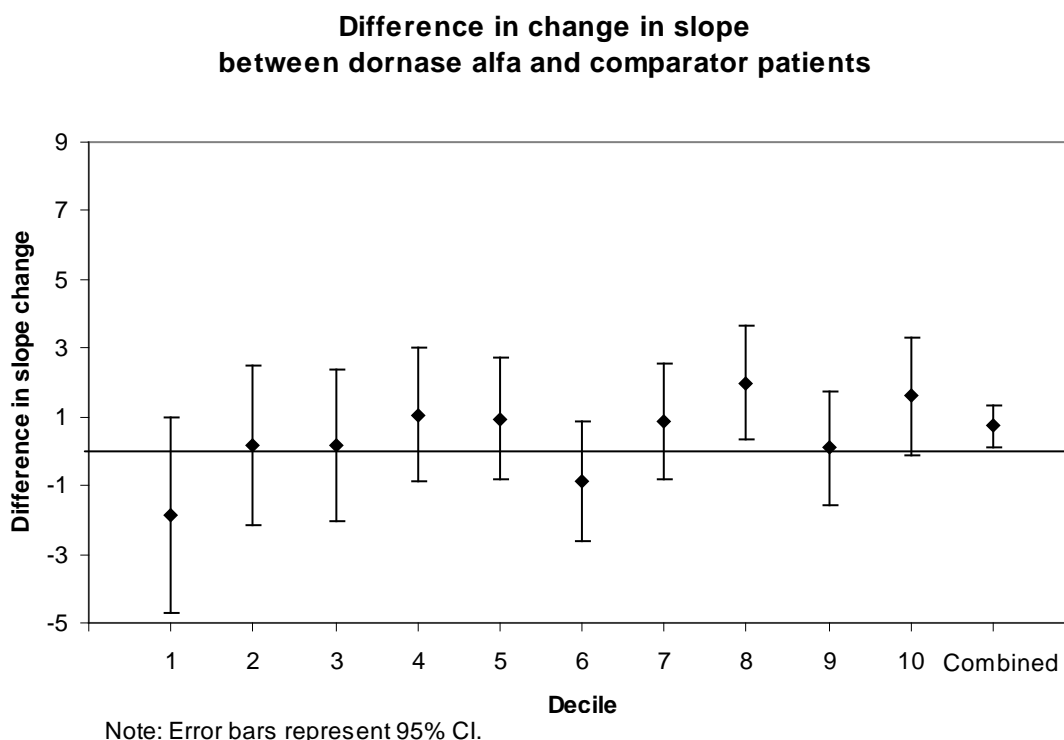


Figure 7 shows the change in slope for comparator patients by decile and shows a clear trend that reflects the “bowing” we saw in Figure 4. At the lowest deciles, the change in slope is positive (curved upwards) and at the highest deciles, the change in slope is negative (curved downwards). In the middle,

the change in slope is not materially different from zero; the lines are approximately straight. In Figure 8, we show the difference in the change in slope for dornase alfa patients compared to the change in slope for comparator patients. This is the “bottom line” graph, that shows the slope change is consistently more positive for the dornase alfa patients than the comparator patients (in 8 of 10 deciles and overall). The effect – about 1 point per year – may look small but it has potentially profound implications over the course of a patient’s life.

Figure 8: Difference in Change in Slope, Dornase Alfa versus Comparator by FEV₁ Decile

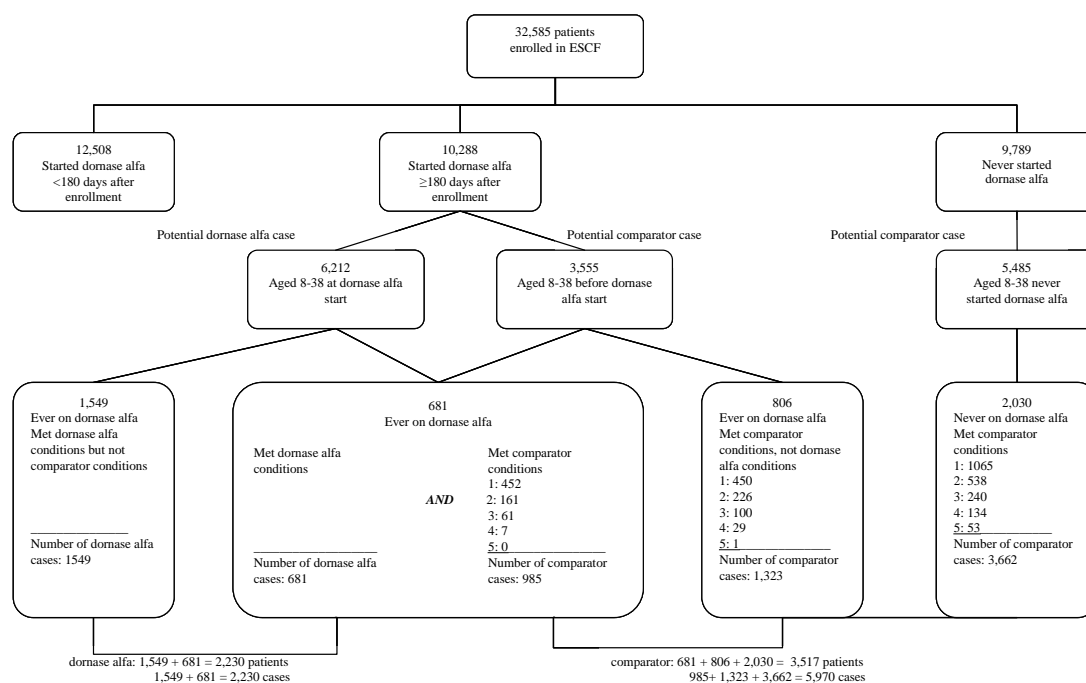


CONSTRUCTING COHORTS

It is not easy to construct cohorts in a study like this. It turns out that most CF patients will eventually be on dornase alfa. Some researchers suggest keeping groups “pure” by omitting patients from the comparator group if they later take dornase alfa, but this approach is seriously biased. One way to think about it is that you should not “penalize” a potential comparator patient for having additional data collected later. If, based on the information collected up to a certain point, the patient is eligible to be a comparator patient then they should be included in that cohort. Furthermore, patients can reasonably contribute multiple time periods to the analysis. This “reuse” of patients violates the assumption of independence and makes some researchers very nervous. (In practice, there is frequently correlation among patients treated at the same site that we do not address, but the correlation within a patient across different times some years apart creates much greater concerns). One approach that avoids including patients more than once would be to select just one of the eligible times in some systematic way, but that introduces bias depending on whether you take the first, last, or most central of the available times. The approach of selecting one of those times at random reduces statistical power unnecessarily. It raises the question why you would sample from a population when you have the population information available for analysis.

Accordingly, we included all the patients we could in as many ways as we could, while avoiding overlapping time periods. The same period was allowed to be used as both the “after” for one time point and the “before” for another time point, but the “before” time periods did not overlap and the “after” time periods did not overlap. A schematic of the cohort construction is rather complex (Figure 9).

Figure 9: Schematic of the Cohort Construction Process



CONCLUSION

Statistical models that evaluate an outcome “before” and “after” an intervention are common in many fields. In observational research, there is a particular challenge finding suitable comparison groups. There is also much to be gained by careful model parameterization, an important skill that is rarely taught explicitly. Once a model is parameterized, the construction of ESTIMATE and CONTRAST statements to test the desired hypotheses can be confusing and is usually tedious. New tools for constructing the desired hypothesis tests are available in SAS, but an understanding of the fundamentals is always desirable. Graphical displays of complex results can sometimes make the conclusions clearer. When constructing cohorts of treated and comparison patients, try hard to avoid biasing the cohorts or losing statistical power unnecessarily.

REFERENCES

Konstan, Michael W., Wagener, Jeffrey S., Pasta, David J., Millar, Stefanie J., Jacobs, Joan R., Yegin, Ashley, Morgan, Wayne J., “Clinical use of dornase alfa is associated with a slower rate of FEV₁ decline in cystic fibrosis,” *Pediatr Pulmonol.* 2011; 46:545-553

Pasta, David J. (2005), “Parameterizing models to test the hypotheses you want: coding indicator variables and modified continuous variables,” *Proceedings of the Thirtieth Annual SAS Users Group International Conference*, 212-30 <http://www2.sas.com/proceedings/sugi30/212-30.pdf>

Pasta, David J. (2010), “Practicalities of using ESTIMATE and CONTRAST statements,” *Proceedings of the SAS Global Forum 2010*, 269-280 <http://support.sas.com/resources/papers/proceedings10/269-280.pdf>

ACKNOWLEDGEMENT

My thanks go to all the coauthors of the published manuscript. This project has been joint with my colleague Stefanie Millar, who wrote code and text describing the model as well as producing the figures and table. Much of the material in this paper previously appeared in Pasta (2010).

CONTACT INFORMATION

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

David J. Pasta, Vice President
Medical Affairs Statistical Analysis
ICON Clinical Research
456 Montgomery Street, Suite 2200
San Francisco, CA 94104
+1.415.371.2111
david.pasta@iconplc.com
www.iconplc.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.