

Analysis of fracture and NTX data

J Martin Bland
Professor of Health Statistics
University of York

The questions

Two questions are addressed in this report.

The first question addressed is: Do the data generated by Aubrey Blumsohn's work support Procter & Gamble's conclusion that there is a plateau effect in the relationship of bone fractures with changes in the marker NTX when using Risedronate? The plateau is said to occur at and below a 30% fall in NTX.

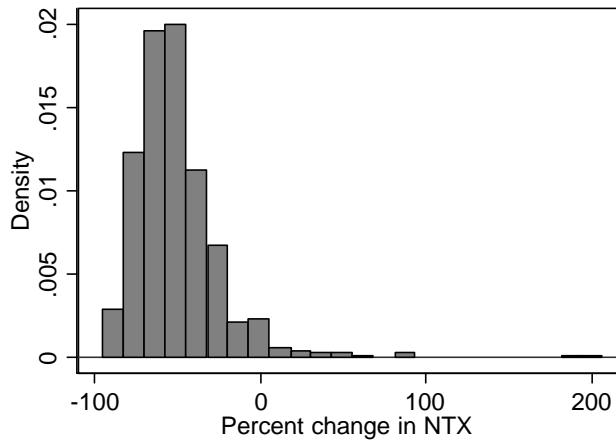
The second question is: Do the data support the conclusions made in Professor Eastell's paper published in *Journal of Bone & Mineral Research*. The paper in question is Eastell, R, Barton I, Hannon R, Chines A, Garnero P, Delmas P. (2003) Relationship of early changes in bone resorption to the reduction in fracture risk with risedronate. *Journal of Bone and Mineral Research* **18:6**, 1051-1056. The conclusion in question is that 'The relationships between vertebral fracture risk and changes from baseline in CTX and NTX were not linear ($p < 0.05$). There was little further improvement in fracture benefit below a decrease of 55–60% for CTX and 35–40% for NTX.'

The data

The data were supplied to me by Vivienne Parry as an Excel file. Only four variables are used in the analysis:

- The study, HIP or VERT
- The % change in baseline at 3 to 6 months in NTX
- Whether a fracture was observed within 3 years
- The time to fracture or time of follow-up

The distribution of % change in NTX is shown in the following histogram:



This shows a highly positively skew distribution with only a few positive increases in NTX and some extremely high outliers.

Deleted: negatively

A table of fracture and time to event (months) follows:

Time to event	Fracture		Total
	No	Yes	
3	1	0	1
6	4	1	5
9	3	0	3
12	43	11	54
15	6	3	9
18	8	1	9
21	5	0	5
24	51	13	64
27	2	1	3
30	2	0	2
33	6	1	7
36	362	16	378
Total	493	47	540

This shows that there were 47 fractures out of 540 subjects for whom data were available and that not all subjects were observed for the full 36 months. Fractures could be observed only on X-ray and although these were at intervals of three months this was clearly not done for all subjects, as there are clear peaks at 12, 24, and 36 months, both in fractures and people who were no longer followed up.

The numbers in the two studies were:

Study	Fracture		Total
	No	Yes	
HIP	213	11	224
VERT	285	36	321
Total	498	47	545

I understand that VERT was subdivided into VERT-MN and VERT-NA, but we do not have this information.

Statistical methods

Because subjects were observed for varying lengths of time, we must take this into account in the analysis. People observed for 12 months are clearly less likely to experience a fracture than those who were observed for 36 months. The appropriate method for doing this is survival analysis, where the survival time is the time to a fracture. Two statistical methods will be used: Kaplan Meier survival curves and Cox proportional hazards regression

Kaplan Meier survival curves are plots of the estimated proportion of subjects who have not yet experienced a fracture against the time of follow-up. This takes into account that some subjects were not followed for the full 36 months. We must assume that subjects who are not followed for the full 3 years do not differ in their risk of fracture from those subjects who are.

Cox proportional hazards regression will be done to estimate the effect of change in NTX. This uses a model where anything which alters the risk of a fracture does so in a constant ratio. For example, if the HIP trial patients had twice the risk of a fracture as the VERT trial patients, we would assume that this doubling of risk continued throughout the period of follow-up. This ratio of risk is called the hazard ratio. For a quantitative variable such as % change in NTX, we estimate the hazard ratio per unit difference in the variable, e.g. per percentage point difference in % change in NTX. To be valid, Cox regression should have at least 10 events per variable included as a predictor of survival, preferably 20. Hence we should not have more than four predicting variables, at most.

We do not know the exact times of fractures, nor do we know at what intervals each patient was followed up. Our event will be the detection of a fracture rather than the fracture itself.

As there are two different studies here, any differences between the patient populations from which trial subjects were drawn may produce different risks of fracture in the two trial groups. I shall therefore check for a trial effect and include it in any subsequent analyses should one be present.

Analyses were done using Stata version 8.

Analysis strategy

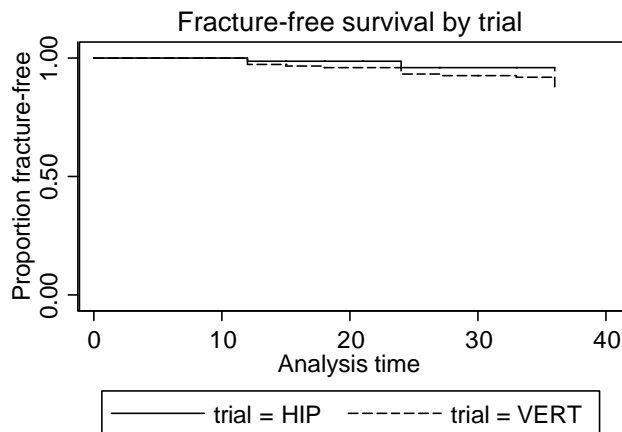
I have taken two approaches to the analysis. The first is to test for the presence of a 'plateau' effect, whereby reductions in NTX greater than 30% have the same effect as a reduction of 30%. The problem with doing this is that the plateau hypothesis may have been generated in part from the data with which we test it, giving it a greater chance than we expect of producing a spurious plateau.

For this reason, I have also carried out an analysis from the standpoint of ignorance and attempted to model the relationship between time to fracture and % change in NTX.

The primary analysis will include all the available data. However, because there are a few subjects for whom there appears to have been a treatment failure, some with a very large increase in NTX, it will be repeated excluding those subjects for whom NTX increased.

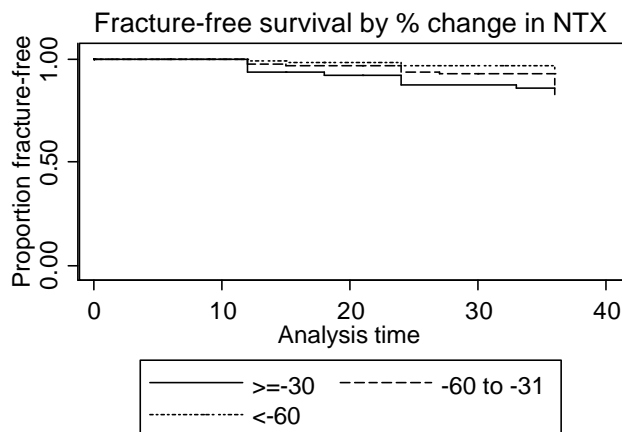
Preliminary analyses and survival plots

The Kaplan Meier survival curves for the two trials were as follows:



The survival without fracture appeared to be better for the HIP patients than for the VERT patients. This was confirmed by Cox regression, which estimated the hazard ratio as 2.33. This means that the risk of fracture at any time for the VERT group was 2.33 times that for the HIP group. This was statistically significant ($P = 0.01$) which means that it was unlikely that we could get a ratio as different from 1.0 in a sample as is 2.33 if there were really no difference between the trial populations. In other words, we have good evidence that the trial populations were different. (Effects with $P < 0.05$ are usually taken as providing some evidence for an effect, and smaller P values mean stronger evidence.) The 95% confidence interval was 1.18 to 4.57, meaning that we estimate the hazard ratio between the two trial populations, as opposed to the trial samples, to be between 1.18 and 4.57.

To show graphically the relationship between fracture-free survival and % change in NTX, the change was classified into three groups $> -30\%$, i.e. above the putative plateau, -31% to -60% , and below -60% . If there is a plateau at values more negative than -30% , the second and third groups should not differ in their fracture-free survival:



This is not the case and survival appears best in the $<-60\%$ group, those with the largest fall in NTX.

As a simple linear effect, % change in NTX was estimated to have a hazard ratio 1.0070, $P = 0.02$. The hazard ratio seems very small because it is the hazard ratio for a difference of one percentage point in % change in NTX. It means that the risk of fracture at any moment is multiplied by 1.007 for each increase of one percentage point in the % change. An increase here means becoming more positive, moving from -100% , the maximum possible fall in NTX, upwards toward 0% and on into the positive area. The P value indicates that there is evidence that this is a real effect. The question now becomes whether the linear model is a good fit to the data.

Results of test for a plateau effect

To test the plateau at changes more negative than -30% , the % change in NTX was split into two variables. The first, above the plateau, was equal to the % change unless the % change was less than -30% (i.e. more negative), when it was set equal to -30% . The second variable, on the plateau, was equal to the % change + 30 unless the % change was greater than -30% (i.e. in the positive direction), when it was set equal to 0. The effect of this is that the sum of these two variables is the % change. We include both of them in our Cox regression model, along with the trial variable. If there were a plateau, we would expect changes more negative than -30% to have no effect. We therefore test whether the second variable, on the plateau, has any effect upon fracture-free survival. In fact, the hazard ratio was 1.019, $P=0.05$, for the percentage change less than minus 30% , i.e. on the putative plateau, giving some rather weak evidence that there is an effect in the region of the plateau. For the percentage change on the positive side of the plateau the hazard ratio was smaller, 1.0032, and was not statistically significant, $P = 0.5$, so there was no evidence for any effect on the positive side of minus 30% in this analysis. This does *not* mean that there no effect. It means that this analysis failed to demonstrate one. However, these results are the opposite of what we would expect if there were a plateau at % change in NTX more negative than -30% .

Deleted: survival

We can test the hypothesis that the plateau model fits the data better than does a simple linear model. We do this by a chi-squared test, using the difference between the chi-squared statistics for the two models. The difference is $12.29 - 10.54 = 1.75$ with 1 degree of freedom. This gives $P = 0.2$, so there is no evidence that a plateau model fits better than a linear model.

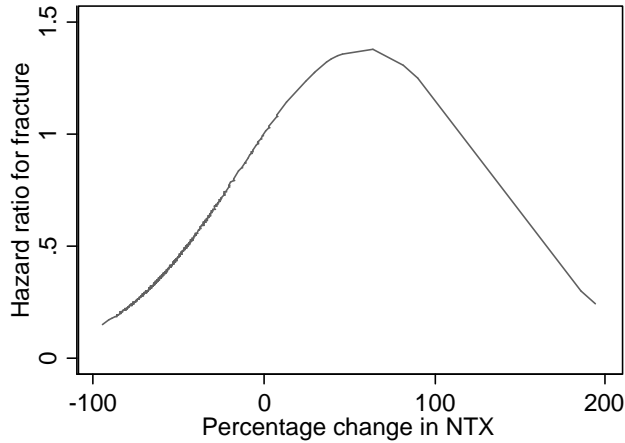
If we repeat all this excluding subjects with positive changes, the results are similar, although everything ceases to be statistically significant. The hazard ratio was 1.013, $P=0.3$, for the percentage change on the putative plateau and for the percentage change on the positive side of the plateau the hazard ratio was smaller, 1.0027, $P = 0.9$. For the chi-squared test, the difference is $11.14 - 11.06 = 0.08$ with 1 degree of freedom, $P = 0.8$, so there is no evidence that a plateau model fits better than a linear model.

As it is not clear at which point the plateau is thought to begin, I carried out the analysis, using all subjects, with a change point at -40% rather than -30% . The hazard ratio was 1.025, $P=0.05$, for the percentage change on the putative plateau and for the percentage change on the positive side of the plateau the hazard ratio was smaller, 1.0033, $P = 0.5$. For the chi-squared test, the difference is $12.84 - 10.54 = 2.30$ with 1 degree of freedom, $P = 0.3$, so there is no evidence that a plateau model fits better than a linear model.

Results of the open analysis

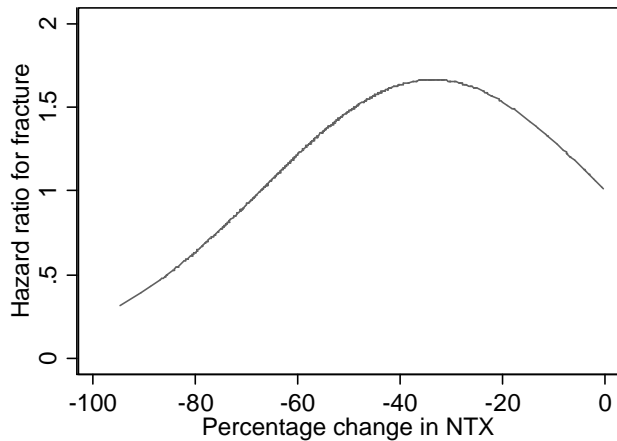
The first step is to test whether a linear fit is adequate. To do this we include a non-linear term, which we get by squaring the % change. We then use a chi-squared test to see whether the model including the squared term fits the model better than the model with only the linear term. This gives $15.73 - 10.54 = 5.19$ with 1 degree of freedom. This gives $P = 0.02$, so there is evidence that a non-linear model fits better than a linear model.

The question now becomes what non-linear model best predicts fracture-free survival. The numbers of fractures are too small to answer this question properly, but I have plotted the hazard ratio against the value of % change in NTX:



(The ratio is set to 1.00 when the NTX change is zero.) This looks very odd, as the hazard rises then plunges down as the percentage change in NTX becomes large and positive. This is clearly nonsense and arises because there are just two outlying observations with very high positive % change, where the subjects did not have fractures. For this analysis it may well be better to exclude them.

Repeating the analysis with only those subjects with negative changes shows $12.42 - 11.06 = 1.36$, $P=0.2$, gives no evidence of non-linearity and the following plot:



Again this looks very unsatisfactory. I think that general curve fitting is not possible with so few fractures in the data.

I also attempted a more powerful curve -fitting technique using fractional polynomials, but the curves produced were very similar to these .

About all we can say about these curves is that they do not suggest a plateau.

Comment

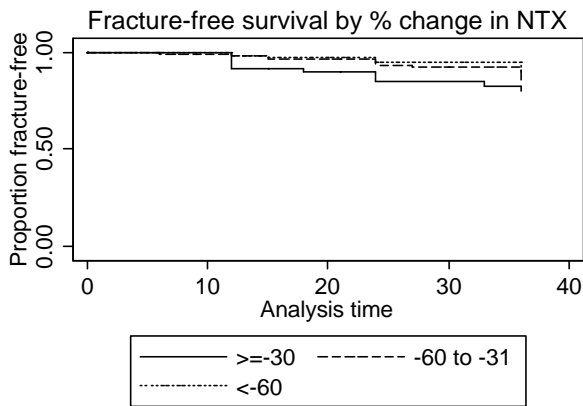
Analyses of these data are unreliable because of the small number of fractures. These studies were designed as placebo-controlled trials and, as there is little doubt that bisphosphonates reduce the risk of bone fracture in susceptible people, I suspect that they were of adequate size to compare the active treatment to no treatment. I do not think that they are of adequate size to give a reliable estimate of the shape of the NTX change response curve.

I could find no evidence for a plateau effect beginning around minus 30% change in NTX

Do the data support the conclusions made by Eastell (2003)?

The conclusion in question is that ‘The relationships between vertebral fracture risk and changes from baseline in CTX and NTX were not linear ($p < 0.05$). There was little further improvement in fracture benefit below a decrease of 55–60% for CTX and 35–40% for NTX.’ I have only looked at NTX in this analysis. I have used all the data from two trials. Clearly, this analysis does not support the conclusions in that paper. I have repeated the analysis using only the data from the VERT trial. I have only the data for 321 patients on Risedronate. I have no way to tell whether these are the same data as used by Eastell *et al.* (2003). I think that they are, however, because Eastell *et al.* report 36 Risedronate patients to have experienced fractures at three years, of whom 9 had fractures at one year, exactly as in these data.

I have repeated my analysis for the VERT data alone. The results are similar to those for the full data. However, the survival plot by % change in NTX certainly makes a plateau look more plausible:



The two lines for the plateau region are more similar than they are for the whole data. (The corollary to this is that they are more different in the HIP data.)

For the test of the plateau effect, the hazard ratio was 1.013, $P=0.2$, for the percentage change on the putative plateau and for the percentage change on the positive side of the plateau the

hazard ratio was smaller, 1.001, $P = 0.9$. As noted above, this should be larger if there were a plateau. For the chi-squared test, the difference is $1.71 - 0.98 = 0.73$ with 1 degree of freedom, $P = 0.4$, so there is no evidence that a plateau model fits better than a linear model.

With a change point at -40% rather than -30%, the hazard ratio was 1.016, $P=0.3$, for the percentage change on the putative plateau and for the percentage change on the positive side of the plateau the hazard ratio was smaller, 1.0016, $P = 0.8$. For the chi-squared test, the difference is $1.71 - 0.98 = 0.73$ with 1 degree of freedom, $P = 0.4$, so there is no evidence that a plateau model fits better than a linear model.

We should remember that Eastell *et al.* had the placebo group data, in which the subjects had 64 fractures and 33 of these were in the first year. This would give them far more power than either my VERT trial analysis or my analysis for both trials.

In the methods section of the paper, Eastell *et al.* wrote:

‘Cox regression was used to explore the relationship between fracture incidence and selected baseline measures. Univariate models were constructed to examine the simple relation between baseline measures and fracture incidence, ignoring all the other measures. A multiple regression model was consequently constructed, comprising baseline measures that were statistically significantly associated ($p < 0.05$) with fracture incidence.

‘To visualize the association between fracture incidence and early changes in bone turnover makers, the probability of sustaining a fracture was plotted against the 3- to 6-month bone turnover maker (sic) data. Empirical displays of the incidence were constructed using a smoothing curve. Because these displays were not model-dependent, no confidence intervals were constructed. Cox regression polynomial models were formed to compare the fit of the data when using linear, quadratic, and cubic functions. These models were statistically compared using the likelihood ratio χ^2 test.’

Although it appears that they have used the same statistical approach as I have, the key sentences are the first two of the second paragraph. I have no idea what they mean. What were these probabilities? How was this empirical curve constructed? No reference is given for this method. How were the data from the Risedronate and placebo groups combined? We cannot tell from the paper. The plateau is stated on the basis of the empirical lines on Figure 1, but there is no statistical modelling or testing of it presented. The Cox models mentioned at the end of the second paragraph do not appear to be presented, although these may have given rise to the test for linearity reported in paragraph 1 of page 1054.

As this plateau result is presented in the abstract of the paper, and so is presented as an important finding of the paper, I think the basis on which the conclusion is drawn should have been described more fully.

Conflicts of interest

I do not think that I have any conflict of interest, but I will describe here all my contact with pharmaceutical companies over the past five years.

The most relevant was a review which I conducted of papers reporting a trial comparing two bisphosphonates, Zoledronic Acid and Pamidronate, for Mayne Pharma, for which I was paid. Zoledronic Acid is manufactured by Novartis, Pamidronate is a generic drug which Mayne manufacture. I also attended a meeting on this paid for by Mayne. It is possible that I met someone involved in this dispute at that meeting, and I certainly met a French professor there who may well be the one mentioned by Aubrey Blumsohn.

I have analysed, for a fee, a small cancer vaccine trial for Onyvox and a vaccine safety study for Chiron. I have given two paid lectures, on the design and analysis of measurement studies, for Novartis.

I have carried out no other paid work for a pharmaceutical company. I was joint holder of a grant for a large study of death related to asthma, funded by the National Asthma Campaign but using an ear-marked grant from Glaxo.

I do not recall any personal contact with either Procter & Gamble or Merck. I have analysed data from and advised clinicians on innumerable projects and it is quite possible that one of these was funded by one of these companies, I wouldn't know.

Output of the Stata analysis

```
-----
log: C:\projyork\radio4\anal.log
log type: text
opened on: 6 Oct 2005, 15:34:37

. do "C:\TEMP\STD01000000.tmp"

. cd \projyork\radio4
C:\projyork\radio4

.
. clear

.
. set scheme slmono

.
. use data_from_object

.
. * this command sets up the survival analysis
. stset ntimevs, failure( ncensor)

      failure event:  ncensor != 0 & ncensor < .
obs. time interval:  (0, ntimevs]
exit on or before:   failure

-----
      545 total obs.
      5 event time missing (ntimevs>=.)          PROBABLE ERROR
-----
      540 obs. remaining, representing
      47 failures in single record/single failure data
      16626 total analysis time at risk, at risk from t =      0
              earliest observed entry t =      0
              last observed exit t =      36

.
. * graphs
.
. histogram pchntx, xtitle(Percent change in NTX) yline(0) scale(1.5)
saving(histpch, re
> place)
(bin=23, start=-94.838997, width=12.545739)
(file histpch.gph saved)

. graph export histpch.emf, replace
(file C:\projyork\radio4\histpch.emf written in Enhanced Metafile format)

.
. egen trial = group(study)

. lab def triallb 1 "HIP" 2 "VERT"

. lab val trial triallb

.
. gen pchntxgp = 1 if pchntx !=.

. replace pchntxgp = 2 if pchntx <-30
(463 real changes made)

. replace pchntxgp = 3 if pchntx <-60
(212 real changes made)

. lab def pchlb 1 ">=-30" 2 "-60 to -31" 3 "<-60"
```

Formatted: Spanish
Spain-Traditional Sort

```

. lab val pchntxgp pchlb

.
. sts graph, by(trial) xtitle("Analysis time") ylabel(0 (.5) 1, nogrid)
yttitle(Proportio
> n fracture-free) title(Fracture-free survival by trial, size(medium)) scale(1.5)
savin
> g(trial, replace)

        failure _d: ncensor
        analysis time _t: ntimevs
(file trial.gph saved)

. graph export trial.emf, replace
(file C:\projyork\radio4\trial.emf written in Enhanced Metafile format)

.
. sts graph, by(pchntxgp) xtitle("Analysis time") yttitle(Proportion fracture-free)
ylab
> el(0 (.5) 1, nogrid) title(Fracture-free survival by % change in NTX,
size(medium)) le
> gend(order(1 2 3) label(1 ">=-30") label(2 "-60 to -31") label(3 "<-60"))
scale(1.5)
> saving(pchntx, replace)

        failure _d: ncensor
        analysis time _t: ntimevs
(file pchntx.gph saved)

. graph export pchntx.emf, replace
(file C:\projyork\radio4\pchntx.emf written in Enhanced Metafile format)

.
.
. * basic data
. tab ntimevs ncensor

```

ntimevs	ncensor		Total
	0	1	
3	1	0	1
6	4	1	5
9	3	0	3
12	43	11	54
15	6	3	9
18	8	1	9
21	5	0	5
24	51	13	64
27	2	1	3
30	2	0	2
33	6	1	7
36	362	16	378
Total	493	47	540

```

. tab trial ncensor

```

group(study)	ncensor		Total
	0	1	
HIP	213	11	224
VERT	285	36	321
Total	498	47	545

```

.
. * simple linear model
.
. stcox trial

      failure _d: ncensor
      analysis time _t: ntimevs

Iteration 0:  log likelihood = -286.37608
Iteration 1:  log likelihood = -282.99969
Iteration 2:  log likelihood = -282.9593
Iteration 3:  log likelihood = -282.95927
Refining estimates:
Iteration 0:  log likelihood = -282.95927

Cox regression -- Breslow method for ties

No. of subjects =          540          Number of obs =          540
No. of failures =           47
Time at risk   =        16626

Log likelihood = -282.95927          LR chi2(1) =          6.83
          Prob > chi2 =          0.0089

-----+-----
      _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      trial |   2.325883   .8013453    2.45  0.014   1.183914   4.569362
-----+-----

. stcox trial pchntx

      failure _d: ncensor
      analysis time _t: ntimevs

Iteration 0:  log likelihood = -286.37608
Iteration 1:  log likelihood = -283.1412
Iteration 2:  log likelihood = -281.27568
Iteration 3:  log likelihood = -281.10697
Iteration 4:  log likelihood = -281.10431
Iteration 5:  log likelihood = -281.10431
Refining estimates:
Iteration 0:  log likelihood = -281.10431

Cox regression -- Breslow method for ties

No. of subjects =          540          Number of obs =          540
No. of failures =           47
Time at risk   =        16626

Log likelihood = -281.10431          LR chi2(2) =          10.54
          Prob > chi2 =          0.0051

-----+-----
      _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      trial |   2.201561   .762554    2.28  0.023   1.116607   4.340713
      pchntx |   1.006963   .0030802   2.27  0.023   1.000944   1.013018
-----+-----

.
. * threshold model
. * % change greater than -30
. gen pchg30 = pchntx

. replace pchg30 = -30 if pchntx<-30
(463 real changes made)

. * % change less than -30, i.e. more negative.  Variable is % change plus 30.
. gen pchlt30 = pchntx + 30

```

```

. replace pchltm30 = 0 if pchntx>=-30
(82 real changes made)

.
. * NB pchntx = pchltm30 + pchgtm30
. gen x = pchltm30 + pchgtm30 - pchntx

. sum x

-----+-----
Variable |      Obs      Mean   Std. Dev.   Min       Max
-----+-----
x       |      545         0         0         0         0

.
. stcox trial pchltm30 pchgtm30

      failure _d:  ncensor
      analysis time _t:  ntimevs

Iteration 0:  log likelihood = -286.37608
Iteration 1:  log likelihood = -281.04017
Iteration 2:  log likelihood = -280.29868
Iteration 3:  log likelihood = -280.23118
Iteration 4:  log likelihood = -280.23002
Iteration 5:  log likelihood = -280.23002
Refining estimates:
Iteration 0:  log likelihood = -280.23002

Cox regression -- Breslow method for ties

No. of subjects =          540          Number of obs =          540
No. of failures =           47
Time at risk   =       16626
Log likelihood =   -280.23002          LR chi2(3) =          12.29
                                          Prob > chi2 =          0.0064

-----+-----
      _t | Haz. Ratio   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
      trial |  2.144378   .742322    2.20  0.028   1.088027   4.226324
  pchltm30 |  1.018938   .0097729   1.96  0.050   .9999623   1.038274
  pchgtm30 |  1.003154   .0049499   0.64  0.523   .993499   1.012902
-----+-----

.
. * % change greater than -40
. gen pchgtm40 = pchntx

. replace pchgtm40 = -40 if pchntx<-40
(413 real changes made)

. * % change less than -40, i.e. more negative. Variable is % change plus 40.
. gen pchltm40 = pchntx + 40

. replace pchltm40 = 0 if pchntx>=-40
(132 real changes made)

.
. * NB pchntx = pchltm40 + pchgtm40
. gen z = pchltm40 + pchgtm40 - pchntx

. sum z

-----+-----
Variable |      Obs      Mean   Std. Dev.   Min       Max
-----+-----
z       |      545         0         0         0         0

```

```
. stcox trial pchltm40 pchgtm40
```

```
failure _d: ncensor  
analysis time _t: ntimevs
```

```
Iteration 0: log likelihood = -286.37608  
Iteration 1: log likelihood = -280.64341  
Iteration 2: log likelihood = -280.0025  
Iteration 3: log likelihood = -279.95652  
Iteration 4: log likelihood = -279.95601  
Iteration 5: log likelihood = -279.95601  
Refining estimates:  
Iteration 0: log likelihood = -279.95601
```

Cox regression -- Breslow method for ties

```
No. of subjects = 540 Number of obs = 540  
No. of failures = 47  
Time at risk = 16626  
Log likelihood = -279.95601 LR chi2(3) = 12.84  
Prob > chi2 = 0.0050
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
trial	2.123969	.7355291	2.18	0.030	1.077402	4.187152
pchltm40	1.02458	.0125165	1.99	0.047	1.00034	1.049408
pchgtm40	1.003259	.0046761	0.70	0.485	.9941354	1.012465

```
. * modelling  
. * quadratic model  
. sum ntimevs
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ntimevs	540	30.78889	8.808405	3	36

```
. gen pch2 = (pchntx+50)^2
```

```
. stcox trial pchntx pch2, nohr
```

```
failure _d: ncensor  
analysis time _t: ntimevs
```

```
Iteration 0: log likelihood = -286.37608  
Iteration 1: log likelihood = -278.59516  
Iteration 2: log likelihood = -278.50932  
Iteration 3: log likelihood = -278.50914  
Refining estimates:  
Iteration 0: log likelihood = -278.50914
```

Cox regression -- Breslow method for ties

```
No. of subjects = 540 Number of obs = 540  
No. of failures = 47  
Time at risk = 16626  
Log likelihood = -278.50914 LR chi2(3) = 15.73  
Prob > chi2 = 0.0013
```

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
trial	.7647699	.3453055	2.21	0.027	.0879835	1.441556
pchntx	.0205689	.006906	2.98	0.003	.0070333	.0341045

```

-----
      pch2 | -.0000951  .0000551  -1.73  0.084  -.0002031  .0000128
-----

```

```

.
. * now do it with a true zero for the plot
. gen pch02 = pchntx^2

```

```

. stcox pchntx pch02 trial , nohr

```

```

      failure _d:  ncensor
      analysis time _t:  ntimevs

```

```

Iteration 0:  log likelihood = -286.37608
Iteration 1:  log likelihood = -278.59516
Iteration 2:  log likelihood = -278.50932
Iteration 3:  log likelihood = -278.50914
Refining estimates:
Iteration 0:  log likelihood = -278.50914

```

```

Cox regression -- Breslow method for ties

```

```

No. of subjects =          540                Number of obs =          540
No. of failures =           47
Time at risk   =          16626
Log likelihood = -278.50914
LR chi2(3)     =          15.73
Prob > chi2    =          0.0013

```

```

-----
      _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      pchntx |   .0110545   .0048025     2.30   0.021   .0016418   .0204673
      pch02 |  -.0000951   .0000551    -1.73   0.084  -.0002031   .0000128
      trial |   .7647699   .3453055     2.21   0.027   .0879835   1.441556
-----

```

```

. gen hr02 = exp(pchntx * .0110545 - pch02 * 0.0000951)

```

```

. graph twoway line hr02 pchntx, sort scale(1.5) xtitle(Percentage change in NTX)
ytitl
> e(Hazard ratio for fracture) saving(hr, replace)
(file hr.gph saved)

```

```

. graph export hr.emf, replace
(file C:\projyork\radio4\hr.emf written in Enhanced Metafile format)

```

```

. * Now repeat it all for negative changes only

```

```

. drop if pchntx>0
(20 observations deleted)

```

```

. sts graph, by(trial) xtitle("Analysis time") ylabel(0 (.5) 1, nogrid)
ytitl(Proportio
> n fracture-free) title(Fracture-free survival by % change in NTX, size(medium))
scale
> (1.5) saving(trial2, replace)

```

```

      failure _d:  ncensor
      analysis time _t:  ntimevs
(file trial2.gph saved)

```

```

. graph export trial2.emf, replace
(file C:\projyork\radio4\trial2.emf written in Enhanced Metafile format)

```

```
. sts graph, by(pchntxgp) xtitle("Analysis time") ytitle(Proportion fracture-free)
ylab
> el(0 (.5) 1, nogrid) title(Fracture-free survival by % change in NTX,
size(medium)) 1
> egen(order(1 2 3) label(1 ">=-30") label(2 "-60 to -31") label(3 "<-60"))
scale(1.5) s
> aving(pchntx2, replace)
```

```
failure _d: ncensor
analysis time _t: ntimevs
(file pchntx2.gph saved)
```

```
. graph export pchntx2.emf, replace
(file C:\projyork\radio4\pchntx2.emf written in Enhanced Metafile format)
```

```
. stcox trial pchntx
```

```
failure _d: ncensor
analysis time _t: ntimevs
```

```
Iteration 0: log likelihood = -248.34712
Iteration 1: log likelihood = -242.91824
Iteration 2: log likelihood = -242.81702
Iteration 3: log likelihood = -242.8168
Refining estimates:
Iteration 0: log likelihood = -242.8168
```

Cox regression -- Breslow method for ties

```
No. of subjects =          520          Number of obs =          520
No. of failures =           41
Time at risk    =          16065
Log likelihood  =       -242.8168          LR chi2(2)      =          11.06
                                          Prob > chi2     =          0.0040
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
trial	2.865336	1.132106	2.66	0.008	1.320877 6.215683
pchntx	1.01071	.0080423	1.34	0.181	.9950701 1.026597

```
. stcox trial pchltm30 pchgtm30
```

```
failure _d: ncensor
analysis time _t: ntimevs
```

```
Iteration 0: log likelihood = -248.34712
Iteration 1: log likelihood = -242.88495
Iteration 2: log likelihood = -242.77784
Iteration 3: log likelihood = -242.77761
Refining estimates:
Iteration 0: log likelihood = -242.77761
```

Cox regression -- Breslow method for ties

```
No. of subjects =          520          Number of obs =          520
No. of failures =           41
Time at risk    =          16065
Log likelihood  =       -242.77761          LR chi2(3)      =          11.14
                                          Prob > chi2     =          0.0110
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
trial	2.846287	1.126564	2.64	0.008	1.310304 6.182801


```

    pchltm30 | 1.012841 .0111301 1.16 0.246 .9912597 1.034892
    pchgtm30 | 1.002662 .0303139 0.09 0.930 .9449744 1.063872
-----

```

```

. stcox trial pchntx pch2, nohr

```

```

    failure _d:  ncensor
    analysis time _t:  ntimevsv

```

```

Iteration 0:  log likelihood = -248.34712
Iteration 1:  log likelihood = -242.27275
Iteration 2:  log likelihood = -242.08688
Iteration 3:  log likelihood = -242.08636
Refining estimates:
Iteration 0:  log likelihood = -242.08636

```

```

Cox regression -- Breslow method for ties

```

```

No. of subjects =          520                Number of obs   =          520
No. of failures =           41                LR chi2(3)         =         12.52
Time at risk    =        16065                Prob > chi2        =          0.0058
Log likelihood   =   -242.08636

```

```

-----
      _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
    trial |  1.018392   .3956575     2.57  0.010   .2429177   1.793867
    pchntx |  .0146026   .0094878     1.54  0.124  -.0039931   .0331984
    pch2   | -.0004483   .0003929    -1.14  0.254  -.0012183   .0003217
-----

```

```

. stcox trial pchntx pch02, nohr

```

```

    failure _d:  ncensor
    analysis time _t:  ntimevsv

```

```

Iteration 0:  log likelihood = -248.34712
Iteration 1:  log likelihood = -242.27275
Iteration 2:  log likelihood = -242.08688
Iteration 3:  log likelihood = -242.08636
Refining estimates:
Iteration 0:  log likelihood = -242.08636

```

```

Cox regression -- Breslow method for ties

```

```

No. of subjects =          520                Number of obs   =          520
No. of failures =           41                LR chi2(3)         =         12.52
Time at risk    =        16065                Prob > chi2        =          0.0058
Log likelihood   =   -242.08636

```

```

-----
      _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
    trial |  1.018392   .3956575     2.57  0.010   .2429177   1.793867
    pchntx | -.0302253   .0373462    -0.81  0.418  -.1034225   .042972
    pch02  | -.0004483   .0003929    -1.14  0.254  -.0012183   .0003217
-----

```

```

. gen hr022 = exp(-pchntx * 0.0302253 - pch02 *0.0004483)

```

```

. graph twoway line hr022 pchntx, sort scale(1.5) xtitle(Percentage change in NTX)
ytit1
> e(Hazard ratio for fracture) saving(hr2, replace)
(file hr2.gph saved)

. graph export hr2.emf, replace
(file C:\projyork\radio4\hr2.emf written in Enhanced Metafile format)

.
.
end of do-file

. log close
  log: C:\projyork\radio4\ana1.log
  log type: text
  closed on: 6 Oct 2005, 15:46:34
-----

  log: C:\projyork\radio4\ana2.log
  log type: text
  opened on: 6 Oct 2005, 16:32:49

. do "C:\TEMP\STD01000000.tmp"

. * analysis for VERT only
.
. cd \projyork\radio4
C:\projyork\radio4

.
. clear

.
. set scheme slmono

.
. use data_from_object

.
. * vert trial data only
. egen trial = group(study)

. drop if trial == 1
(224 observations deleted)

.
. * this command sets up the survival analysis
. stset ntimevs, failure( ncensor)

      failure event: ncensor != 0 & ncensor < .
obs. time interval: (0, ntimevs]
exit on or before: failure

-----
      321 total obs.
       0 exclusions
-----

      321 obs. remaining, representing
       36 failures in single record/single failure data
 9762 total analysis time at risk, at risk from t =          0
                                     earliest observed entry t =          0
                                     last observed exit t =          36

.
. * graphs
.

```

```

. histogram pchntx, xtitle(Percent change in NTX) yline(0) scale(1.5)
saving(histpchv, r
> eplace)
(bin=17, start=-88.692001, width=16.612059)
(note: file histpchv.gph not found)
(file histpchv.gph saved)

. graph export histpchv.emf, replace
(file C:\projyork\radio4\histpchv.emf written in Enhanced Metafile format)

.
. gen pchntxgp = 1 if pchntx !=.

. replace pchntxgp = 2 if pchntx <-30
(270 real changes made)

. replace pchntxgp = 3 if pchntx <-60
(111 real changes made)

. lab def pchlb 1 ">=-30" 2 "-60 to -31" 3 "<-60"

. lab val pchntxgp pchlb

.
. sts graph, by(pchntxgp) xtitle("Analysis time") ytitle(Proportion fracture-free)
ylab
> el(0 (.5) 1, nogrid) title(Fracture-free survival by % change in NTX,
size(medium)) le
> gend(order(1 2 3) label(1 ">=-30") label(2 "-60 to -31") label(3 "<-60"))
scale(1.5)
> saving(pchntxv, replace)

failure _d: ncensor
analysis time _t: ntimevs
(file pchntxv.gph saved)

. graph export pchntxv.emf, replace
(file C:\projyork\radio4\pchntxv.emf written in Enhanced Metafile format)

.
. * basic data
. tab ntimevs ncensor

```

ntimevs	ncensor		Total
	0	1	
3	1	0	1
6	2	1	3
9	2	0	2
12	25	8	33
15	5	3	8
18	3	1	4
21	4	0	4
24	33	8	41
27	0	1	1
30	2	0	2
33	3	1	4
36	205	13	218
Total	285	36	321

```

.
. * simple linear model
.
. stcox pchntx

failure _d: ncensor

```

```

analysis time _t: ntimevs

Iteration 0:  log likelihood = -199.81294
Iteration 1:  log likelihood = -199.41891
Iteration 2:  log likelihood = -199.32565
Iteration 3:  log likelihood = -199.32426
Iteration 4:  log likelihood = -199.32426
Refining estimates:
Iteration 0:  log likelihood = -199.32426

Cox regression -- Breslow method for ties

No. of subjects =          321          Number of obs =          321
No. of failures =           36
Time at risk   =          9762
Log likelihood =  -199.32426          LR chi2(1) =          0.98
                                          Prob > chi2 =          0.3229

```

```

-----
      _t | Haz. Ratio   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
    pchntx |  1.004387   .0039973    1.10  0.271   .9965829   1.012252
-----+-----

```

```

.
. * threshold model
. * % change greater than -30
. gen pchgtm30 = pchntx

. replace pchgtm30 = -30 if pchntx<-30
(270 real changes made)

. * % change less than -30, i.e. more negative. Variable is % change plus 30.
. gen pchltm30 = pchntx + 30

. replace pchltm30 = 0 if pchntx>=-30
(51 real changes made)

.
. * NB pchntx = pchltm30 + pchgtm30
. gen x = pchltm30 + pchgtm30 - pchntx

```

```

. sum x

Variable |      Obs      Mean   Std. Dev.   Min   Max
-----+-----
      x |      321         0         0         0     0

```

```

. stcox pchltm30 pchgtm30

      failure _d: ncensor
analysis time _t: ntimevs

Iteration 0:  log likelihood = -199.81294
Iteration 1:  log likelihood = -198.96563
Iteration 2:  log likelihood = -198.95916
Iteration 3:  log likelihood = -198.95913
Refining estimates:
Iteration 0:  log likelihood = -198.95913

Cox regression -- Breslow method for ties

No. of subjects =          321          Number of obs =          321
No. of failures =           36
Time at risk   =          9762
Log likelihood =  -198.95913          LR chi2(2) =          1.71
                                          Prob > chi2 =          0.4258

```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
pchltm30	1.013131	.0111709	1.18	0.237	.9914711	1.035264
pchgtm30	1.001013	.0065355	0.16	0.877	.9882857	1.013905

```

.
. * % change greater than -40
. gen pchgtm40 = pchntx

. replace pchgtm40 = -40 if pchntx<-40
(234 real changes made)

. * % change less than -40, i.e. more negative. Variable is % change plus 40.
. gen pchltm40 = pchntx + 40

. replace pchltm40 = 0 if pchntx>=-40
(87 real changes made)

.
. * NB pchntx = pchltm40 + pchgtm40
. gen z = pchltm40 + pchgtm40 - pchntx

. sum z

```

Variable	Obs	Mean	Std. Dev.	Min	Max
z	321	0	0	0	0

```

. stcox pchltm40 pchgtm40

```

```

      failure _d:  ncensor
      analysis time _t:  ntimevs

```

```

Iteration 0:  log likelihood = -199.81294
Iteration 1:  log likelihood = -198.9721
Iteration 2:  log likelihood = -198.95982
Iteration 3:  log likelihood = -198.95977
Iteration 4:  log likelihood = -198.95977
Refining estimates:
Iteration 0:  log likelihood = -198.95977

```

Cox regression -- Breslow method for ties

```

No. of subjects =          321                Number of obs   =          321
No. of failures =           36
Time at risk   =          9762
Log likelihood = -198.95977
LR chi2(2)     =           1.71
Prob > chi2    =           0.4261

```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
pchltm40	1.015503	.013959	1.12	0.263	.9885088	1.043234
pchgtm40	1.001574	.0059461	0.26	0.791	.989987	1.013296

```

.
. * modelling
. * quadratic model
. sum ntimevs

```

Variable	Obs	Mean	Std. Dev.	Min	Max
ntimevs	321	30.41121	9.0218	3	36

```

. gen pch2 = (pchntx+50)^2
.
. stcox pchntx pch2, nohr
      failure _d:  ncensor
      analysis time _t:  ntimevs

Iteration 0:  log likelihood = -199.81294
Iteration 1:  log likelihood = -198.13064
Iteration 2:  log likelihood = -198.1143
Iteration 3:  log likelihood = -198.11405
Refining estimates:
Iteration 0:  log likelihood = -198.11405

Cox regression -- Breslow method for ties

No. of subjects =          321          Number of obs =          321
No. of failures =           36
Time at risk   =          9762
Log likelihood = -198.11405
LR chi2(2)     =           3.40
Prob > chi2    =           0.1829

```

```

-----+-----
      _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
    pchntx |   .0150371   .0081315     1.85   0.064   - .0009004   .0309746
    pch2   |  -.000008   .0000674    -1.19   0.235   - .000212   .0000521
-----+-----

```

```

.
. * now do it with a true zero for the plot
. gen pch02 = pchntx^2
.
. stcox pchntx pch02, nohr
      failure _d:  ncensor
      analysis time _t:  ntimevs

Iteration 0:  log likelihood = -199.81294
Iteration 1:  log likelihood = -198.13064
Iteration 2:  log likelihood = -198.1143
Iteration 3:  log likelihood = -198.11405
Iteration 4:  log likelihood = -198.11405
Refining estimates:
Iteration 0:  log likelihood = -198.11405

Cox regression -- Breslow method for ties

No. of subjects =          321          Number of obs =          321
No. of failures =           36
Time at risk   =          9762
Log likelihood = -198.11405
LR chi2(2)     =           3.40
Prob > chi2    =           0.1829

```

```

-----+-----
      _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
    pchntx |   .0070417   .0061673     1.14   0.254   - .005046   .0191295
    pch02  |  -.000008   .0000674    -1.19   0.235   - .000212   .0000521
-----+-----

```

```

. gen hr02 = exp(pchntx * 0.0070417 - pch02 * 0.00008)
.

```

```
. graph twoway line hr02 pchntx, sort scale(1.5) xtitle(Percentage change in NTX)
ytit1
> e(Hazard ratio for fracture) saving(hrv, replace)
(file hrv.gph saved)

. graph export hrv.emf, replace
(file C:\projyork\radio4\hrv.emf written in Enhanced Metafile format)

.
.
end of do-file

. log close
   log: C:\projyork\radio4\ana2.log
   log type: text
closed on: 6 Oct 2005, 16:33:14
-----
```

Further analysis of fracture and NTX data, 7 October 2005

J Martin Bland
Professor of Health Statistics
University of York

The question

In this report I attempt to answer the question as to what Eastell *et al.* did to produce Figure 1 in Eastell, R, Barton I, Hannon R, Chines A, Garnero P, Delmas P. (2003) Relationship of early changes in bone resorption to the reduction in fracture risk with risedronate. *Journal of Bone and Mineral Research* **18:6**, 1051-1056.

The data

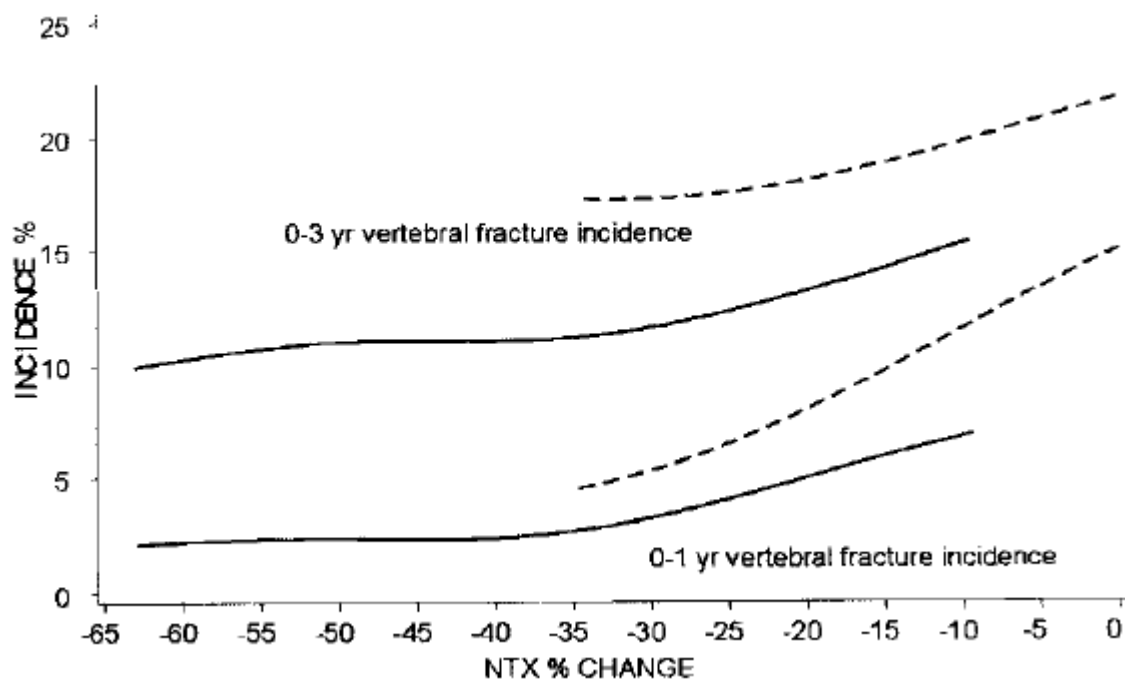
The data were supplied to me by Vivienne Parry as an Excel file. Only four variables are used in the analysis:

- The study, HIP or VERT
- The % change in baseline at 3 to 6 months in NTX
- Whether a fracture was observed within 3 years
- The time to fracture or time of follow-up

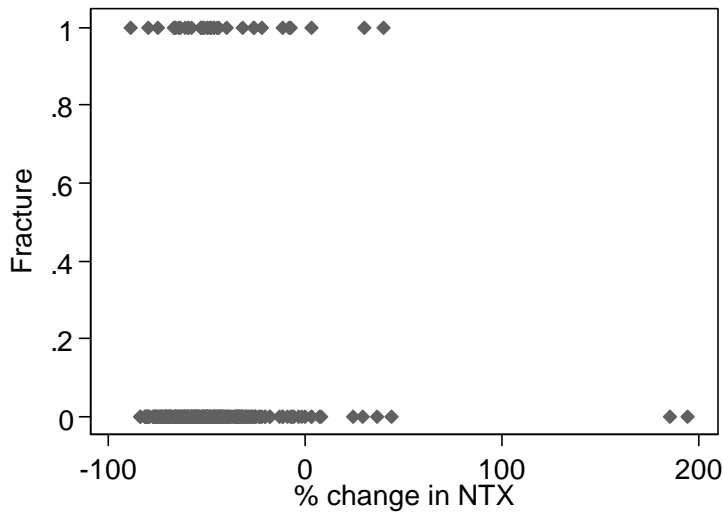
For this analysis I dropped the HIP trial data.

Analysis

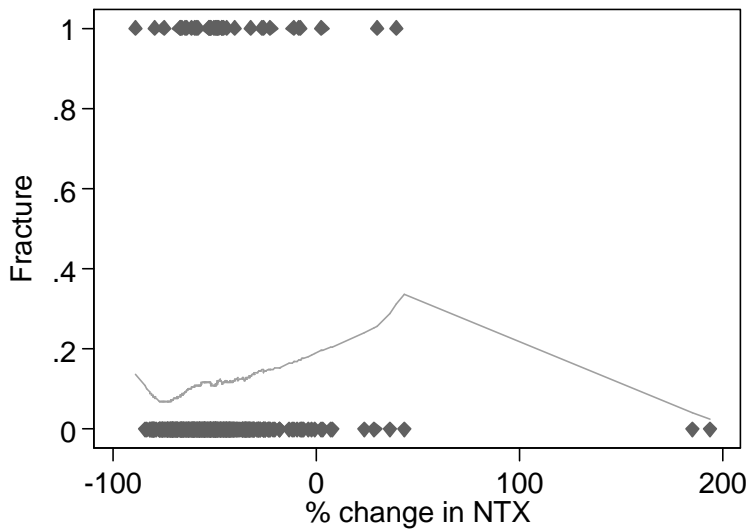
Figure 1 in the paper shows a plot of fracture incidence per cent for 0-3 years and 0-1 year against % change in NTX, as shown below.



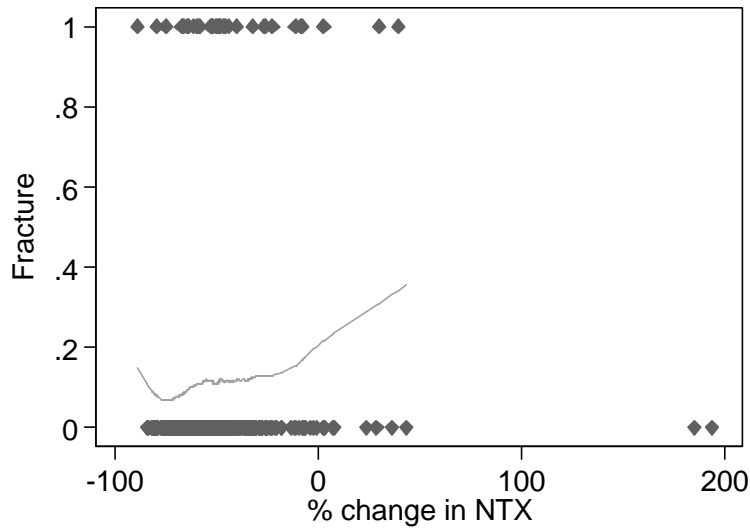
Only the solid lines concern us as these are for the risedronate group. All the other analysis in the paper used survival analysis but this did not. What did they do? I think that they started with a plot of fracture (1 = fracture, 0 = no fracture) against % change in NTX:



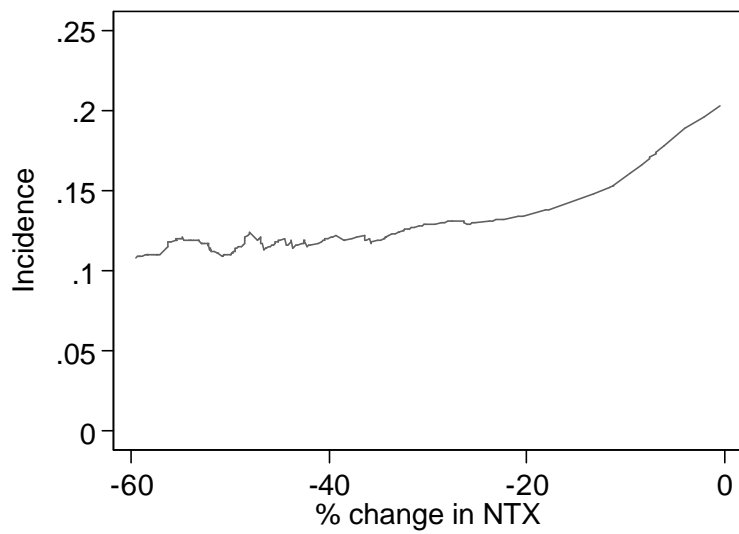
This completely ignores the time of follow-up. They then fitted a curve to this using some sort of smoothing algorithm such as LOWESS (locally weighted smoothing for scatter plots). I used the lowess command in Stata. This gives:



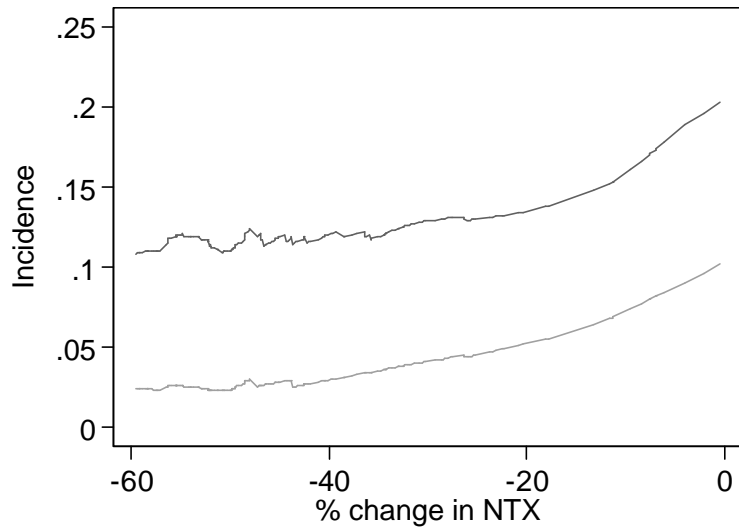
This is not attractive, as the extreme outliers have produced an unconvincing plunge to zero incidence at % change in NTX = 200. So we omit these two outliers:



This looks much better. Now we get rid of the fracture variable points at zero and one, because they look very odd. At the same time, we cut off the curve at the top and bottom. We don't like the upward sweep at the extreme left, but if we just cut that off we have a downturn at about -70 , so we cut it off above that:



We now relabel the vertical axis as “Incidence”, because it represents an estimate of the proportion of subjects who had fracture. Now we have a plateau. We can do it for 12 months, too:



If we were to smooth these out a bit more, we would get something quite like Figure 1 of Eastell *et al.*

Comments

This is pure speculation on my part. If this is what they did, it has no analytical value in my opinion. Cox regression is the correct analysis, but the number of fractures is too small for any but the simplest modelling, as I did in my first analysis.

The authors themselves used Cox regression for everything else in the paper.