Linear meta-analysis results from nonlinear data

This example is given in the context of logistic regression. Consider 4 studies with 9 BP categories and let

$$n_{sb} = \text{ the number of subjects in study } s \text{ and BP category } b$$

Let the n_{sb} have the values

```
10 10 10 10  0  0  0  0  0
 0 10 10 10 10 10  0  0  0
 0  0  0 10 10 10 10 10  0
 0  0  0  0  0 10 10 10 10
```

Note that the BP distributions shift to the right as one moves down from one study to the next.

Let

$$r_{sb} = \text{ the observed risk for study } s \text{ and BP category } b$$

Not all of these are defined because some of the n_{sb} are zero. Let

$$\ell_{sb} = \log\left(\frac{r_{sb}}{1 - r_{sb}}\right)$$

This is the observed logit, that is, the logit of the observed risk r_{sb}.

The data are defined by giving the following values to the ℓ_{sb}:

```
-4 -4 -3 -2
   -4 -4 -4 -3 -2
         -4 -4 -4 -3 -2
               -4 -4 -4 -3
```

These are plotted in Figure 1. A slight displacement is used to help identify the 4 studies. Note that the logit profile for each study is very nonlinear and that all the profiles differ by more than an additive constant. The profiles are actually horizontal splines of the form considered by Port et.al. (2000).

Consider a model of the form

$$\lambda_{sb} = \alpha_s + \beta_b$$

for the population logits λ_{sb}. This is a proper meta-analysis model because it is additive. The common logit profile for the 4 studies is given by

$$\beta_1. \cdots, \beta_9$$

Logistic regression gives the maximum likelihood estimate

```
-3.41 -2.98 -2.37 -1.52 -1.19 -0.33  0.05  0.84  1.30
```

for this profile which is plotted in Figure 2.

This plot is well approximated by a straight line in spite of the fact that the data for each study is distinctly nonlinear. The problem of course is that the data do not satisify the standard meta-analysis assumptions. Because of this it would be a large mistake to conclude from the linearity of Figure 2 that the studies themselves have a linear logit-risk profile.

Reference

Port S, Demer L, Jennrich R, Walter D, Garfinkel A. Systolic blood pressure and mortality. *Lancet* 2000; **355**:175-80.

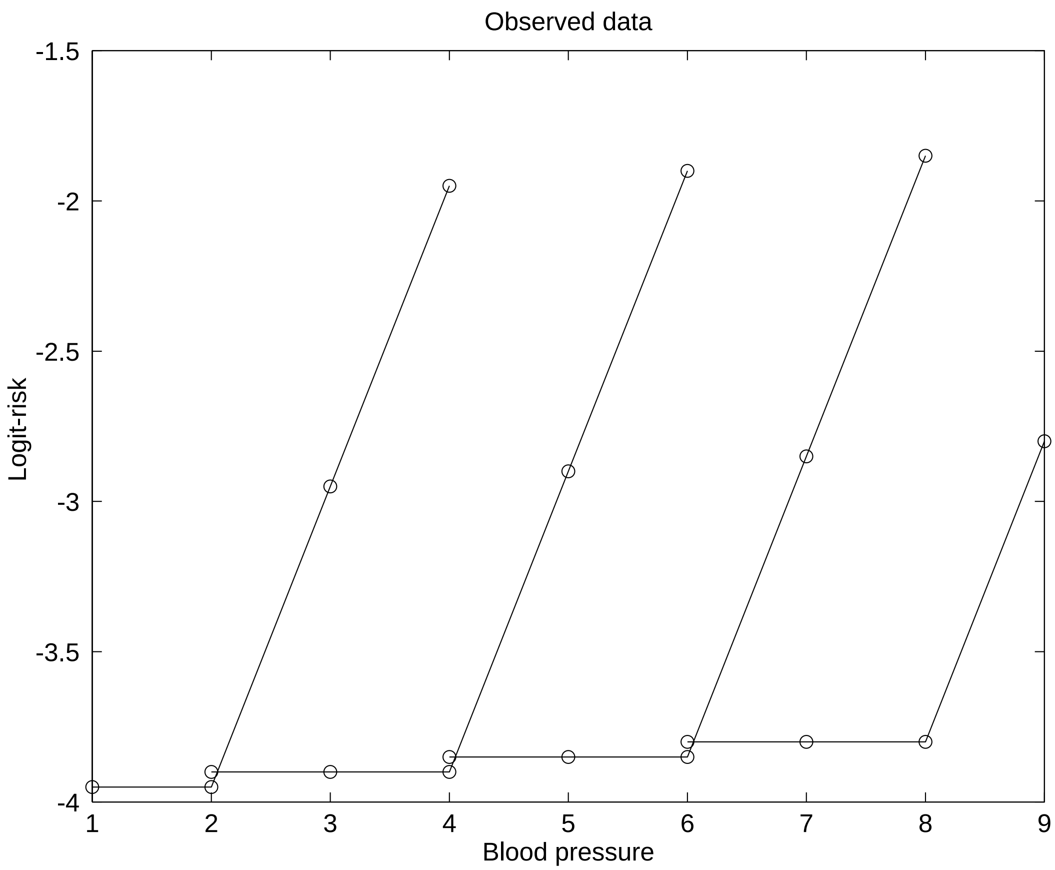

Figure 1: The observed data are denoted by 'o's. They are logits of the observed risks plotted against blood pressure. The four broken lines connect the data from each of the four studies. The data are displaced slightly to separate the studies for viewing.

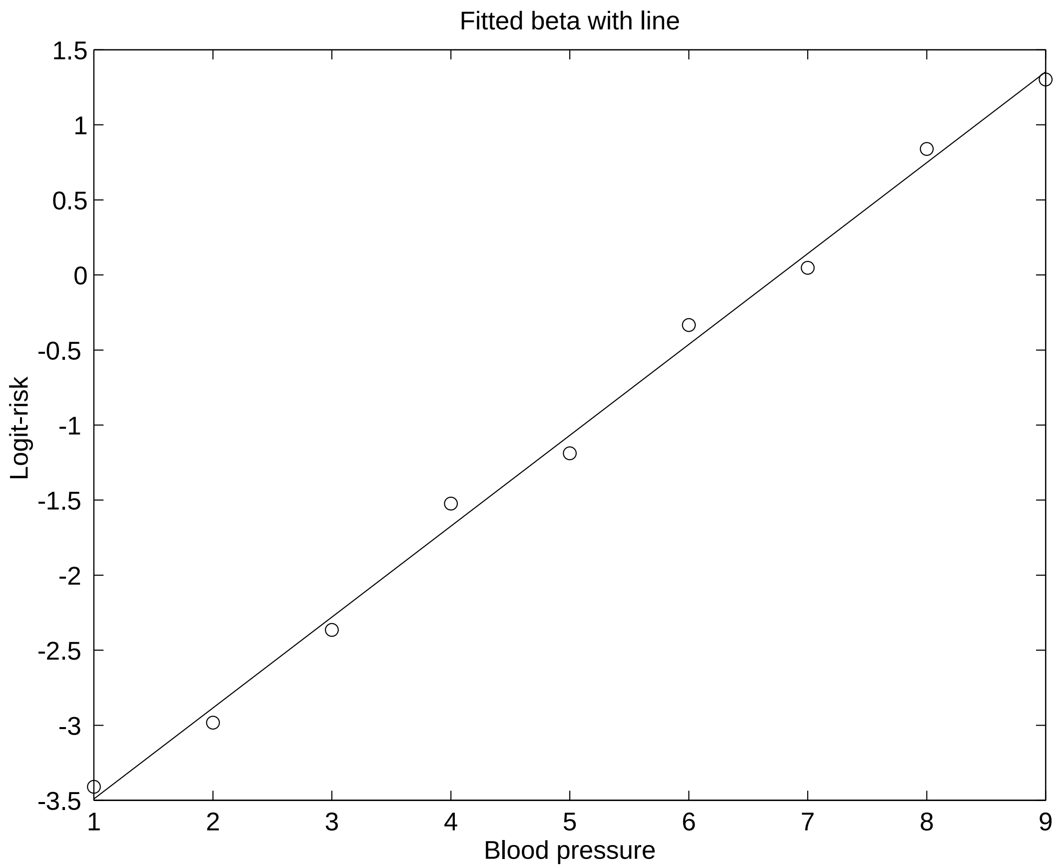

Figure 2: The 'o's denote the logistic regression estimate of the logit-risk profile from the meta-analysis. The line is added to show it is nearly linear.

Technical note

The author used a logistic regression routine that accepted the observed logits ℓ_{sb} as input. One may wish to check the results given using a logistic regression routine that uses deaths y_{sb} and numbers at risk n_{sb}. Then $r_{sb} = y_{sb}/n_{sb}$ and one can find y_{sb} that give the ℓ_{sb} used above. The y_{sb}, however, will not be integers. If his logit regression routine does not object to this he may proceed. If it does he might proceed by replacing the nonzero n_{sb} by 1000. That is replace each 10 by 1000. Then use the death counts y_{sb} given by:

18	18	47	119					
	18	18	18	47	119			
			18	18	18	47	119	
					18	18	18	47

This will give a close approximation to the observed logits and results given above.