Dieldrin in Human Milk

Arden Miller

Executive Summary

The probability a donor's milk contains above 0.009 parts per million Dieldrin is much higher if the donor lives in a house that has been treated with Dieldrin within the past three years than not. It is also higher for donors that live in new suburbs than old suburbs and increases as the donor's age increases.

The Dieldrin Data

An investigation into the presence of the pesticide dieldrin in human milk was carried out in Western Australia. Dieldrin is used to control termites in Western Australia and new houses, by law, must be treated for termites. There is concern that dieldrin may be present in the milk of nursing mothers who are exposed to it. Samples of milk from 43 donors were tested to see they contained above 0.009 parts per million of the pesticide Dieldrin. The donor's age, whether the donor lived in an old or new suburb, and whether the donor lived in a house that had been treated with Dieldrin within the last 3 years were also recorded.

For this data set 12 of the donors lived in a new suburb and 31 lived in an old suburb. Twenty-four of the donors lived in houses that had been treated for dieldrin within the previous three years. The ages of the donors ranged from 21 to 37.

A Logistic Regression Model

A logistic regression model was used to model the probability that the donor's milk contained (more than 0.009 ppm) Dieldrin, $\hat{\pi}$, as a function of the three other variables. The fitted model for this data is:

$$\hat{\pi} = \frac{\exp(-8.973 + 2.597I_t + 2.141I_s + 0.208\text{age})}{1 + \exp(-8.973 + 2.597I_t + 2.141I_s + 0.208\text{age})}$$

where I_t is 1 if the donor lives in a house that has been treated with Dieldrin and 0 otherwise, I_s is 1 if the donor lives in a new suburb and 0 otherwise, and age is the donor's age in years. This model indicates that the probability a donor's milk contains Dieldrin is higher if the donor lives in a treated house than not. It is also higher for donors that live in new suburbs and increases as the donor's age increases.

To quantify how the presence of Dieldrin is affected by each of these factors, it is easier to discuss the odds of Dieldrin occurring in the donor's milk rather than the probability. The odds are simply defined as the probability that Dieldrin occurs divided by the probability it doesn't occur. Thus an odds of 5 indicates that a donor's milk is five times more likely to contain Dieldrin than to not contain Dieldrin whereas an odds of 1 means the donor's milk is equally likely to contain or not contain Dieldrin.

Consider two donors who are the same age and live in the same type of suburb (old or new). One of these lives in a treated house and the other in an untreated house. Our model predicts that the odds of the donor's milk containing dieldrin are 13.5 times higher for the donor who lives in the treated house than the donor who lives in the untreated house. This is an estimated value and so the true value will be somewhat different. Taking the uncertainty in the fitted model into account, we can be reasonably certain (95% confident) that true multiplicative factor is between 2 and 89.

Now compare two donors from different types of suburb but who are the same age and who both live in the same type of house (treated or untreated). The odds for the donor from the new suburb will be 8.5 times the odds for the donor from the old suburb. Taking model uncertainty into account the true factor should be between 1.4 times and 52.7 times.

Finally consider 2 donors who live in the same type of suburb and the same type of house but are 1 year apart in ages. The odds for the older donor are estimated to be 1.23 times those for the younger donor. Taking model uncertainty into account the true factor should be between 0.98 times and 1.55 times.

Estimated Probabilities

Figure 1 contains a plot of the estimated probabilities that a donor's milk contains Dieldrin versus the age of the donor for each of the four combinations of suburb type and house type. The estimated probabilities increase with age and range from approximately 0.01 for young donors (21 years old) who live in untreated houses in old suburbs to approximately 0.90 for older donors (> 35) who live in treated houses in new suburbs. The curve for donors that live in treated houses in new suburbs is clearly the highest and the curve for donors that live in untreated houses in old suburbs is the lowest. The curve for a donors that live in untreated houses in new suburbs is very close to the curve for donors who live in treated houses in old suburbs.

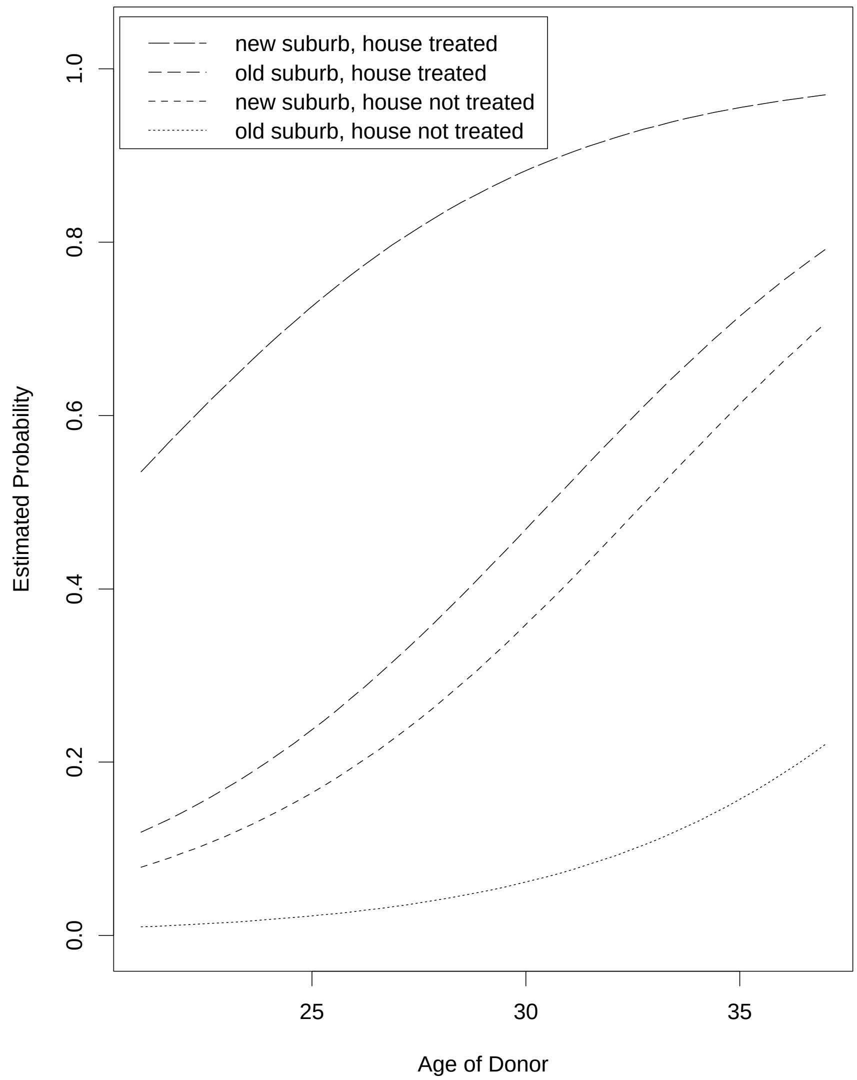

Figure 1: Probability that Donor's Milk contains Dieldrin versus Age

The curves in Figure 1 are subject to estimation error. To illustrate the amount of uncertainty in these estimates, intervals of the plausible values for π were calculated for each of the four combinations of house type and suburb type assuming the donor was 30 years old. We can be fairly certain (95 % confident) that the true probabilities are within these intervals.

House	Suburb	Estimate	Interval
untreated	old	0.06	0.04 to 0.09
untreated	new	0.36	0.26 to 0.47
treated	old	0.47	0.36 to 0.58
treated	new	0.88	0.83 to 0.92

Statistical Appendix

I found that the logistic regression model that simply contains the three main effects was suitable.

```
Coefficients:
                 Value Std. Error   t value
(Intercept) -8.9727497  3.8532397 -2.328625
    treated  2.5969714  0.9683143  2.681951
     suburb  2.1406872  0.9308404  2.299736
        age  0.2083864  0.1172264  1.777640

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 56.76518 on 42 degrees of freedom

Residual Deviance: 41.08711 on 39 degrees of freedom
```

The only term that may not be needed in the model is age. I used a χ^2 test to see if "age" was needed given that "treated" and "suburb" were in the model. The P-value for this test was 0.053 which gives some evidence that "age" is needed.

I used 95% confidence intervals for each of the coefficients, $\hat{\beta} \pm 1.96\text{se}(\hat{\beta})$, to represent the range of plausible values. Taking the exponentials of these gives the multiplicative factors for the odds.

The 95% confidence intervals for π presented in the table were obtained by creating 95% confidence intervals for logit π and then applying the logistic function to get the interval for π.