Assessing Sensitivity to Multiple Factors in Calculating Attributable Risks

Sue TAYLOR and Richard TWEEDIE

March 28, 1996

Abstract

In many calculations of attributable risk, there are a number of parameters (excess risks, exposure rates, and so on) that are not known accurately, and one should take into account the variability in these parameters in determining a range of "plausible values" of the attributable number (AN) of incidences due to the exposure. This is often done through simple tabulation of extreme combinations of the factors, but this can give a spurious picture of the real dimensions of the attributable risk. In this paper we describe graphical displays leading to more accurate evaluations of the AN range using multiple box plots, and a multiple regression approach to deciding on the most sensitive parameters for display in this way. The methods are illustrated by giving calculations of the AN's for lung cancer and for heart disease attributed to exposure to active and passive smoking. These expand analyses previously given by the United States EPA, the Australian NH&MRC and the United States OSHA.

KEYWORDS Attributable risk, excess risk, ischaemic heart disease, smoking, ETS, passive smoking, lung cancer, spousal smoking

*Department of Preventive Medicine and Biometrics, University of Colorado Health Sciences Center, 4200 E. Ninth Ave, Denver CO 80262 and Department of Statistics, Colorado State University, Fort Collins, CO 80523
1 Introduction

When a particular factor is shown to cause an elevated relative risk (RR) of a disease in a population, it is common to calculate the attributable number (AN) of incidences of the disease that can be ascribed to the existence of the factor: that is, the number of incidences that would not have occurred but for the exposure of some part of the population to the factor.

The concepts used in calculating such AN's are often simple, as in Section 2 (Breslow and Day, 1980), and can be based on no more than the parameters RR and

\[ p = \text{Proportion of population exposed to the cause} \]

\[ N = \text{Total number of incidences of the disease in the population}. \]  \hspace{2cm} (1)

Even in such a straightforward case there will usually be uncertainty in all three of the parameters RR, p, N, however, and this will flow through to the accuracy with which we know AN. An obvious method of considering the effect of such uncertainty in the input parameters is to decide on a sensible range of variation for each parameter and then to describe the corresponding range of the output AN under the combinations of parameters chosen, usually by giving the extreme values to be taken by the output. This is done, for example, by the EPA (EPA Review, 1992, Section 6.3.5) or Gross (1995), in assessing attributable risk for lung cancer associated with exposure to environmental tobacco smoke (ETS), a rather more complex situation that we will return to in Section 4 below.

However, simple tabulation of such extreme values can give a somewhat spurious view of the real range of plausible values, since the mutual extremes only occur once in the many combinations of parameter levels; similarly, looking at the univariate parameters and their ranges in combination only with one selected value of each of the other parameters can also give a picture which is difficult to interpret (EPA Review, 1992, Table 6-6).

In this paper, we suggest a new graphical approach which enables a more appropriate impression of the sensitivity of the estimate of AN to uncertainty in the input parameters. In this approach we deal with the multivariate problem by displaying a gallery of univariate box plots (UBP's) that show the range of outcomes, given separately for each level of any parameter judged to have a significant impact on the variability in AN.

For more complicated models, where there are rather more input parameters, we also address the associated question of deciding on the parameters for which one should display UBP's by using a multiple regression-based approach.

This is illustrated in Section 3 and Section 4, where we provide detailed analyses, based on recent age-dependent models given by the the Australian National Health and Medical Research Council (NH&MRC) of the association of environmental tobacco smoke with the incidence of ischaemic heart disease (IHD) and lung cancer. This is in principle more accurate than in the Environmental Protection Agency (EPA) Report (1992) or the Occupational Safety and Health Administration (OSHA) Report (1994). Prior to this, in Section 2, we first illustrate the graphical method in a more straightforward example of the association between IHD and active smoking: here the parameters involved are only RR, p and N, and the depiction of uncertainty through UBP's is clearcut.
In what follows, we will assume that the associations we use for illustration have been established as causal. It has been argued elsewhere that this is not well-established for some of these examples (Lee, 1992; Gross, 1995; Tweedie et al., 1996; Merrilees et al., 1996), but here we will not address these issues. Of course (and this is true of every such association), if the association is not causal, then the AR calculations are totally irrelevant and should not be carried out.

2 Univariate displays: smoking and IHD

2.1 The simple AR model

In this section we consider the association between IHD and active smoking in the United States. We will not differentiate in this first example between males and females, nor between the attributable risks in different age groups: we add this level of complication in the next sections.

The simplest model for attributable risk begins with the parameters $RR, p, N$ in (1). From these we have that the population attributable risk $PAR$ is given by

$$PAR = \frac{p(RR - 1)}{1 + p(RR - 1)}$$

and the attributable number $AN$ of incidences is

$$AN = N \times PAR.$$  

Clearly then $AN$ increases as any of $RR, p, N$ increases. To assess the sensitivity in this case we need reasonable ranges for each of these three input parameters (based on data rather than on mere hypothetical variation (Tweedie et al., 1994)). We now consider possible sources of such data.

2.2 Data for values of $RR$, $p$ and $N$

We first consider values of $N$. In the United States the number of IHD deaths in 1990 is approximately $N = 490,000$ based on ICD9 codes 410-414 (World Health Organisation, 1994). Although the value of $N$ is often taken as fixed in PAR calculations, this is clearly not as true as one might like. It might be appropriate to account, for example, for possible errors in death certificates and the like, which can have inaccuracies of at least 10% (Garfinkel et al., 1985). Moreover, the values over different time-periods (or from different sources) seem to oscillate by at least this much, and we find the 1985 value of $N$ being quoted at 536,800 (Wells, 1994).

In the sensitivity analysis below we will therefore take $N$ as having a range of 440,000-540,000, in five steps of 25,000.

We next consider the proportion of the population exposed to active smoking, which we denote as $p_c$. Different authors give widely differing values for $p_c$. Rates vary from as low as $p_c = 25 - 30\%$ (Steenland, 1992; NH&MRC Working Party, 1995), based on the proportion of currently active smokers in the US, Australia and similar societies, to as high as $p_c = 60\%$, based on an average proportion of male and female "ever smokers" (Wells, 1994).
This is a wide range, and the relevant exposure proportion to use in assessing the risk for IHD depends on the mechanism which is thought to be causal. This is far from clear, since there are only fairly limited scientific studies which try to explain the pathways by which smoking might cause IHD. On the one hand, the association between cigarette smoking might be caused by recent exposures: this could follow from the experimental data in Howard et al. (1994) or Celermajer et al. (1996) where it is shown that tobacco smoke may be related to essentially instant but reversible hardening of the vascular system, or the data in Zhu et al. (1993) showing that tobacco smoke may be related to similarly rapid and reversible diminution of anti-clotting activities related to platelets. In this case the proportion of current smokers is relevant. If on the other hand the association is caused by some longterm mechanism, related say to carcinogenic activity, then the proportion of ever smokers may be relevant.

Here we will consider the effect of varying $p_c$ over the range 25%-60% in steps of 5% (i.e at 8 levels).

Thirdly we consider the relative risk for incidence of ischaemic heart disease (IHD) associated with active smoking. This is often quoted as $RR_c = 1.7$ (Wells, 1994; Steenland, 1992; NH&MRC Working Party, 1993), based on the 1983 US Surgeon's Report (United States Department of Health and Human Services, 1983). To assess the range of values that might be reasonable for this relationship, we consider English et al. (1995, Section 4.6.32), where we find data from almost 20 recent studies. They conclude that a pooled estimate for the $RR_c$ for those 65 years of age or over is 1.66 (95% CI 1.59-1.74), whilst for those under 65, they give a pooled estimate for the $RR_c$ of 3.06 (95% CI 3.00-3.13).

We will defer consideration of different rates for different age groups to the next two sections, since it is easier to see the principles we are describing without this complication. Consideration of the details of the studies in English et al. (1995) show that for a sensitivity analysis we might plausibly use a range for $RR_c$ of 1.3-2.5. (While there are studies (English et al., 1995) for which $RR_c$ for those under 65 is in the higher range around 3.06, there are many fewer actual deaths in the lower age ranges (World Health Organisation, 1994), so that $AN$ is less sensitive to these values, which justifies ignoring them in this first model.)

Here then we will illustrate the graphical methods by considering the effect of varying $RR_c$ over the range 1.3-2.5 in steps of 0.2 (i.e at 7 levels).

### 2.3 Results of the sensitivity analysis

The values of the attributable number of deaths $AN$ calculated from such a sensitivity analysis, if presented in tabular form, would give 280 different numbers to try to assimilate. In Table 2.1, to indicate this difficulty, we present the values just for the midrange of $N = 490,000$.

Over the full range of all three parameters, the minimum value of $AN$ is 30,700 and the maximum is 255,800. This 8-fold variation makes interpretation difficult, even though the ranges of parameters may seem well justified from the different data sources. However, the detailed pattern of the values of $AN$ is not easy to see, even for just the one value of $N$ in Table 2.1.

The graphical method we introduce will, in this case, display three graphs, one for each of the parameters. In each graph the outcomes are broken down by values of the single parameter being considered, as in each graph in Figure 2.1. The outcomes, given as a set of UBPs (using SPSS
format in this case, with outliers and extremes omitted), enable one to see, not just the ends of the range, but an indication of where the bulk of the values lie as input parameters vary.

Table 2.1 Sensitivity analysis with $N = 490,000$

<table>
<thead>
<tr>
<th>$RR_c$</th>
<th>1.3</th>
<th>1.5</th>
<th>1.7</th>
<th>1.9</th>
<th>2.1</th>
<th>2.3</th>
<th>2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_c$</td>
<td>0.25</td>
<td>34186</td>
<td>54444</td>
<td>72979</td>
<td>90000</td>
<td>105686</td>
<td>120189</td>
</tr>
<tr>
<td></td>
<td>0.30</td>
<td>40459</td>
<td>63913</td>
<td>85041</td>
<td>104173</td>
<td>121579</td>
<td>137482</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
<td>46561</td>
<td>72979</td>
<td>96426</td>
<td>117376</td>
<td>136209</td>
<td>153230</td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>52500</td>
<td>81667</td>
<td>107188</td>
<td>129706</td>
<td>149722</td>
<td>167632</td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>58282</td>
<td>90000</td>
<td>117376</td>
<td>141246</td>
<td>162241</td>
<td>180852</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>63913</td>
<td>98000</td>
<td>127037</td>
<td>152069</td>
<td>173871</td>
<td>193030</td>
</tr>
<tr>
<td></td>
<td>0.55</td>
<td>69399</td>
<td>105686</td>
<td>136209</td>
<td>162241</td>
<td>184704</td>
<td>204286</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>74746</td>
<td>113077</td>
<td>144930</td>
<td>171818</td>
<td>194819</td>
<td>214719</td>
</tr>
</tbody>
</table>

On each graph we also give a summary boxplot of the whole range of outcomes, for comparison with the UBPs broken down into levels of the factor. Note immediately that:

(a) The influence of the relative risk $RR_c$ is very strong, but the approximate 5-fold increase in the excess risk $RR_c - 1$ only causes the medians of the values of $AN$ to vary from approximately 55,250 to 188,000;

(b) The influence of the population proportion $p_c$ is also strong, but is close to linear as one would expect. Although we more than double $p_c$, the median values of $AN$ slightly less than double (90,000 to 171,800) over the range, with the interquartile ranges of the boxplots following suit;

(c) As the “summary” box-plots show, the number of attributable deaths in the overall interquartile range is 85,500–167,600, with a median of 126,900: obviously neither the overall maximum nor the minimum is as plausible as values in this range, and even though this “plausible range” is still wide it is much less so than the full range of the $AN$.

FIGURE 2.1 NEAR HERE

2.4 Weighting by probabilities

There is an interesting probabilistic variation on this approach that one might adopt. In some cases it might be that we have prior distributions on the input parameters involved in the sensitivity analysis: for example, if we used, not just a range of $RR_c$ based on different studies, but the pooled mean and CI of a meta-analysis, then we might want to weight the various values used in the sensitivity by their relative probability, and this would lead to a (rather degenerate form of) posterior distribution that could be depicted by UBPs.
Figure 2.1
Attributable Numbers (AN): Association of IHD with active smoking
Sensitivity to variation in N, p, and RRc.
We have not done this here because we do not have, in general, enough information to support
distributional assumptions. In effect, then, we are weighting the values chosen in a uniform or
uninformative manner, since we do not think some are necessarily much more likely than others:
this is the situation, for example, with the values chosen for $p_c$ or $N$.

The use of the more probabilistic approach when data exist to support the choice of informative
priors would give an even better picture of the real range of attributable risks and the credence to
be placed on them, and UBPs would again provide a suitable method of display for them.

3 Identifying key parameters in complex models: ETS and IHD

3.1 The NH&MRC model

In this section we again consider IHD but we use as our example the attributable risk from exposure
to environmental tobacco smoke (ETS) through spousal smoking, and to do this we move to a model
described in the recent Australian NH&MRC Draft Report on the Health Effect of Passive Smoking
(1995). Here we apply the model to the US context: we recognise that some of the point estimates
of US parameters might be different in detail, but since our aim is to consider variation in sensitivity
analyses, the Australian health experience is similar enough to that in the United States for these
to serve as a useful starting point.

In the NH&MRC model the source of ETS is taken solely to be spousal smoking. This is in contrast
to the recent report of the United States EPA (1992), who developed estimates for AN for lung
cancer that included an evaluation of "background risk": we will consider this in Section 4.

The NH&MRC model adds extra complexity in a number of other ways:

(i) it assesses the attributable risk from current and ex-smoking as well as considering ETS, so
that we have several more parameters to consider;

(ii) it takes into account different parameter values in different age groups, and as we will see this
is important for heart disease; and

(ii) it treats males and females separately, and differently.

We will consider the question of identifying sensitive parameters in this more complex models in
Section 3.3.

The parameters needed for this model are, for each sex and age group:

\[
\begin{align*}
N &= \text{number of deaths in the population} \\
 p_c &= \text{proportion of current smokers in the population} \\
 RR_c &= \text{RR for current smokers relative to never-smokers} \\
 p_x &= \text{proportion of ex-smokers in the population} \\
 RR_x &= \text{RR for ex-smokers relative to never-smokers} \\
 p_s &= \text{proportion of never-smokers exposed to spousal ETS} \\
 RR_s &= \text{RR for never-smokers exposed to spousal ETS relative to never-smokers not so exposed}
\end{align*}
\]
Using these parameters, the model used by the NH&MRC, based on the papers by Taylor (1993b; 1993a), is given:

\[
PAR = \frac{p_n p_x (RR_x - 1)}{[p_n (RR_n - 1)] + 1} + [p_x (RR_x - 1)] + 1
\]

where \( p_n = 1 - p_x - p_c \) is the proportion of never-smokers in the population. The attributable numbers are then summed over the age groups to give separate male and female totals.

### 3.2 Choosing Parameter Values

**Exposure rates: \( p_c, p_x, p_s \)**

The values of \( p_c \) and \( p_x \) given in Table 3.1 are taken from Table A2.1 and A2.2 in the NH&MRC Draft Report (1995). These enable us to consider the model in Section 2 for separate age groups and separately also for current and ex-smokers. Because of the complexity of the rest of the model we will take these as fixed for the remainder of the paper although clearly they could be varied if one felt they were not reliable.

**Table 3.1: Percentage (\( p_s \)) of never-smokers with spouse who smokes*, percentages of current (\( p_c \)) and ex-smokers (\( p_x \)*), and deaths from IHD in the population (\( N \))**

<table>
<thead>
<tr>
<th>Age Group</th>
<th>( p_s )</th>
<th>( p_s )</th>
<th>( p_c )</th>
<th>( p_c )</th>
<th>( p_x )</th>
<th>( p_x )</th>
<th>( N )</th>
<th>( N )</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 – 34</td>
<td>9.3</td>
<td>18.0</td>
<td>36.8</td>
<td>29.9</td>
<td>22.2</td>
<td>20.4</td>
<td>407</td>
<td>125</td>
</tr>
<tr>
<td>35 – 39</td>
<td>11.1</td>
<td>17.8</td>
<td>36.8</td>
<td>26.1</td>
<td>25.6</td>
<td>18.1</td>
<td>2634</td>
<td>603</td>
</tr>
<tr>
<td>40 – 44</td>
<td>7.0</td>
<td>19.9</td>
<td>31.3</td>
<td>23.8</td>
<td>29.4</td>
<td>17.5</td>
<td>2635</td>
<td>603</td>
</tr>
<tr>
<td>45 – 49</td>
<td>9.2</td>
<td>20.8</td>
<td>33.3</td>
<td>24.4</td>
<td>31.9</td>
<td>20.6</td>
<td>7571</td>
<td>2157</td>
</tr>
<tr>
<td>50 – 54</td>
<td>11.1</td>
<td>15.0</td>
<td>30.5</td>
<td>22.1</td>
<td>33.7</td>
<td>16.6</td>
<td>7572</td>
<td>2157</td>
</tr>
<tr>
<td>55 – 59</td>
<td>10.2</td>
<td>14.4</td>
<td>29.5</td>
<td>20.7</td>
<td>41.2</td>
<td>17.9</td>
<td>18687</td>
<td>7553</td>
</tr>
<tr>
<td>60 – 64</td>
<td>8.0</td>
<td>15.3</td>
<td>26.9</td>
<td>17.9</td>
<td>45.8</td>
<td>18.4</td>
<td>18688</td>
<td>7553</td>
</tr>
<tr>
<td>65 – 69</td>
<td>4.1</td>
<td>7.8</td>
<td>22.2</td>
<td>14.0</td>
<td>53.4</td>
<td>21.7</td>
<td>35522</td>
<td>21047</td>
</tr>
<tr>
<td>70 – 74</td>
<td>6.9</td>
<td>3.0</td>
<td>16.3</td>
<td>13.5</td>
<td>54.1</td>
<td>21.0</td>
<td>35522</td>
<td>21048</td>
</tr>
<tr>
<td>75 – 79</td>
<td>4.5</td>
<td>2.0</td>
<td>11.7</td>
<td>9.8</td>
<td>55.9</td>
<td>18.6</td>
<td>61412</td>
<td>86761</td>
</tr>
</tbody>
</table>


In order to carry out calculation of \( AN \) in this context it is necessary to have data on the propensity of non-smoking women (or men) to marry partners who smoke, through the parameter \( p_s \). One of the difficulties in this area has been in obtaining estimates of this parameter. The NH&MRC Report (1995, Appendix A) provides detailed data which enable this to be estimated. These are based on an Australian National Health Survey carried out in 1989-1990, which is far more extensive.
than the small and rather fragile data-sets used to estimate \( p_s \) in previous studies, and even allow age-dependent models to be set up.

We shall use the values in Table A2.8 of the NH&MRC Report (1995), given in Table 3.1, as a basis for our sensitivity analyses. It is clear that a break in level of exposure \( p_s \) occurs near age 65. It is cumbersome to vary parameters in each age group, and the parameter values we use will be the same within each of the two age groups (under 65 and over 65), but different between them. The ranges we use are given in Table 3.2, and pick up the spread in specific age groups in Table 3.1.

**Table 3.2: Parameter ranges for \( p_s, RR_c, RR_x, RR_s \)**

For females under 65

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>( RR_x )</td>
<td>1.00</td>
<td>1.25</td>
<td>1.50</td>
</tr>
<tr>
<td>( RR_c )</td>
<td>1.50</td>
<td>2.50</td>
<td>3.50</td>
</tr>
<tr>
<td>( RR_s )</td>
<td>1.10</td>
<td>1.20</td>
<td>1.30</td>
</tr>
<tr>
<td>( p_s )</td>
<td>0.12</td>
<td>0.15</td>
<td>0.18</td>
</tr>
</tbody>
</table>

For females over 65

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>( RR_x )</td>
<td>1.00</td>
<td>1.10</td>
<td>1.20</td>
</tr>
<tr>
<td>( RR_c )</td>
<td>1.50</td>
<td>2.00</td>
<td>2.50</td>
</tr>
<tr>
<td>( RR_s )</td>
<td>1.10</td>
<td>1.20</td>
<td>1.30</td>
</tr>
<tr>
<td>( p_s )</td>
<td>0.02</td>
<td>0.035</td>
<td>0.05</td>
</tr>
</tbody>
</table>

For males under 65:

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>( RR_x )</td>
<td>1.00</td>
<td>1.25</td>
<td>1.50</td>
</tr>
<tr>
<td>( RR_c )</td>
<td>1.50</td>
<td>2.50</td>
<td>3.50</td>
</tr>
<tr>
<td>( RR_s )</td>
<td>1.10</td>
<td>1.20</td>
<td>1.30</td>
</tr>
<tr>
<td>( p_s )</td>
<td>0.06</td>
<td>0.075</td>
<td>0.09</td>
</tr>
</tbody>
</table>

For males over 65:

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>( RR_x )</td>
<td>0.80</td>
<td>0.90</td>
<td>1.00</td>
</tr>
<tr>
<td>( RR_c )</td>
<td>1.50</td>
<td>2.00</td>
<td>2.50</td>
</tr>
<tr>
<td>( RR_s )</td>
<td>1.10</td>
<td>1.20</td>
<td>1.30</td>
</tr>
<tr>
<td>( p_s )</td>
<td>0.02</td>
<td>0.035</td>
<td>0.05</td>
</tr>
</tbody>
</table>

**Choice of relative risks**

We also need to choose the range of relative risks for active smokers, ex-smokers and non-smokers exposed to ETS through spousal smoking. The ranges we use are also given in Table 3.2 and are based on

(i) varying the values for \( RR_x \) and \( RR_c \) for males and females and in each of the two age groups (over and under 65), using the studies in English *et al.* (1995) as a guide;

(ii) varying the values of \( RR_s \), with the lower values chosen as discussed in the analysis elsewhere (Tweedie and Mengersen, 1992; Tweedie *et al.*, 1996); the higher values are consistent with the higher values in the NH&MRC Draft Report (1996) and Gross (1993) although we feel
that these are probably excessive given the US range of about 1.04-1.35 considered reasonable by the EPA (EPA Review, 1992, Table 6-6). (Note that Gross (1995) claims (p. 409) to vary over 1.0-1.2 for this parameter but in Tables I and II he appears to use 1.0-1.5; and that in discussion of that paper, Hanley (1995) gives some reasons why the lower value of 1.0 used by Gross is inappropriate)

**Numbers of population deaths N**

The values used for the number of deaths in each age group are also given in Table 3.1. These were calculated by a simple interpolation into the appropriate age groups from recent data (World Health Organisation, 1994, Table D-1). This is somewhat simplistic, and clearly the values of these parameters could be varied as in Section 2, although the ranges would change linearly. Thus if, as we discussed in Section 2, there is a 10% possible variation either way in these estimates over different years and from sampling errors and the like, then one can factor this in simply, and we do not pursue this here.

### 3.3 Parameter variation and the regression approach

The approach above could lead to a sensitivity analysis over 9 different parameters in each age range, but we vary only 4 of them here, and these are not all varied over every age range.

After summing over the age-groups this leads to a total of 625 estimates for $AN$ for each gender. The second step in our suggested analysis is that one should treat these values as variables dependent on the input parameters and should carry out a standard multiple regression to decide on which of the parameters are most crucially affecting the output range. We have done that here and the results are as in Table 3.3.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Females</th>
<th>Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RR_s$</td>
<td>54.9</td>
<td>54.1</td>
</tr>
<tr>
<td>$p_s$</td>
<td>89.2</td>
<td>85.2</td>
</tr>
<tr>
<td>$RR_c$</td>
<td>91.7</td>
<td>89.9</td>
</tr>
<tr>
<td>$RR_x$</td>
<td>91.9</td>
<td>91.1</td>
</tr>
</tbody>
</table>

Taking logs of the outcomes improved the fits of the models [as measured by $R^2$] without altering the order of entry of variables into the models.

We see from this analysis that for this model the main effects come from variation in the parameters $p_s, RR_s$, and that the level of variation we allow in the active smoking relative risks $RR_c, RR_x$ has limited impact; and so in the next section we show only the effects of $p_s, RR_s$ in the UBP displays.
3.4 Results of the analysis

Table 3.4 shows the descriptive statistics for both females and males over the ranges chosen. For females, the results are illustrated in Figures 3.1-3.2, where we give the UBPs related to the main parameters $p_s, RR_s, RR_c, RR_e$ identified above. (The results for males are similar and we do not repeat them here.) The “Low”, “Medium” and “High” values in the Figures indicate that we have chosen simultaneously the low, medium and high values in both age groups in Table 3.2.

Table 3.4: Summary statistics on sensitivity analyses

<table>
<thead>
<tr>
<th>Sex</th>
<th>Median Percentile</th>
<th>25th Percentile</th>
<th>75th Percentile</th>
<th>IQ</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>1320</td>
<td>739</td>
<td>2139</td>
<td>1400</td>
<td>204</td>
<td>4496</td>
</tr>
<tr>
<td>Males</td>
<td>630</td>
<td>357</td>
<td>1010</td>
<td>653</td>
<td>92</td>
<td>2447</td>
</tr>
</tbody>
</table>

Note that the lack of impact of variation in $RR_c, RR_e$ as described in the regression analysis is obvious: we give the graphs here purely to illustrate how little difference these parameter changes do make across the range chosen.

FIGURES 3.1-3.2 NEAR HERE

Summing the median values over males and females gives a total $AN = 1950$. This should be contrasted with the values of around 40,000-60,000 (Steenland, 1992; Wells, 1994; Glantz and Parmley, 1995). There are a number of reasons for this:

(i) We consider only the effect of spousal exposure, whereas Wells (1994), following the EPA Report (1992), allows for an effect of so-called “background exposure”: we discuss a model for this in the next section, and note it can make a serious difference to the value estimated for $AN$;

(ii) We consider only the effect on never-smokers, whereas Wells (1994) also adds in an additive effect of exposure to ETS on ex-smokers in particular;

(iii) The values of $p_s$ used by the NH&MRC, which we have adopted as being appropriate for IHD, are very much lower than the value of 60% used by the EPA and Wells; as we again find in the next section, higher values of $p_s$ are crucial in increasing the estimates of $AN$. The values we give in Table 3.1 fit well to a model of relatively rapid causation, as discussed in Section 2 in choosing $p_s$, and this is presumably the basis on which they are chosen by the NH&MRC: choosing the higher values on the right hand end of Figure 3.2 for this parameter does lead to estimates more in tune with those of Steenland (1992).

Each of the factors above can increase the value of $AN$ by a factor of 2-4, as is easily checked; and if all three of these factors are taken into account with parameters being chosen at rather high values as in the EPA Report (1992), it is not hard to increase the estimates of $AN$ by an overall factor of 20.

All of these points illustrate, as a sensitivity analysis should, the crucial importance of confirming the correct values of these parameters and the groups to which the model should be applied.
Figure 3.1
Attributable Numbers (AN): Association of IHD with exposure to spousal ETS in females
Sensitivity to variation in $RR_s$ and $p_s$
Figure 3.2
Attributable Numbers (AN): Association of IHD with exposure to spousal ETS in females
Sensitivity to variation in $RR_c$ and $RR_x$
4  Adding other factors: ETS and lung cancer

4.1  The EPA Model

In this section we consider models for the association of environmental tobacco smoke and lung cancer rather than heart disease.

In its estimation of AN in this context, the EPA (1992) introduces a further factor to the model in Section 3. They assume that every person in the population is exposed to a background level of ETS, and that the estimated \( RR_s \) in the “exposed” subgroup who have a smoking spouse is due to this further exposure of ETS from the spouse.

Let us use the notation in (4) to denote, in this section, relative risks for lung cancer rather than heart disease. In order to implement this model involving background exposure, the EPA introduces the further parameter

\[
Z = \frac{\text{level of ETS in “exposed” population}}{\text{level of ETS in “unexposed background” population}},
\]

and then needs to assume that the ratio of excess risks from each source is also proportional to \( Z \).

If we assume this model to be valid, then it is possible to calculate \( RR_{0b} \), the “true background” RR relative to a hypothetical totally unexposed population and \( RR_{0s} \), the “true spousal ETS” RR also relative to a hypothetical totally unexposed population: we have that the relative risks satisfy

\[
[RR_{0s} - 1]/[RR_{0b} - 1] = Z,
\]

and since

\[
RR_s = RR_{0s}/RR_{0b}
\]

we get

\[
RR_s = (1 + [RR_{0s} - 1]Z)/RR_{0b};
\]

hence (9) and (8) can be solved for \( RR_{0b} \) and \( RR_{0s} \). This approach is due to Wald et al. (1986), who originally used a value of \( Z = 3 \), based on cotinine marker studies. This value of \( Z \) was used also in the first EPA Draft Report (1990), but values of \( Z \) around 1.75-2.0 were used in the final EPA Report (1992), based on different studies. These assumptions involve rather detailed dependence on a linear dose-response relationship: see Gross (1995) for a detailed evaluation of this point. Gross (1995) also points out that the final attributable risk calculations are very sensitive to the choice of this parameter.

Our analysis below confirms this sensitivity, and it will be seen that the UBP displays are valuable in an overall evaluation of the variability of AN based on different choices of \( Z \).

The EPA approach also adjusts the relative risk for ever-smokers to the same hypothetical completely unexposed population. The ever-smokers are not broken into the two current and ex-smoking categories we used in Section 3: instead a relative risk of ever-smoking is used, which is the weighted average

\[
RR_e = \frac{p_e}{p_c + p_e} RR_c + \frac{p_e}{p_c + p_e} RR_e
\]
If we denote the relative risk adjusted for background in this context by $RR_{0e}$, then some manipulation of the relationships yields

$$RR_{0e} = RR_e[p_sRR_{0s} + (1 - p_s)RR_{0b}];$$

$$RR_{0s} = (Z - 1)/(Z/RR_s - 1).$$

From these relationships the PAR in each “exposure” group can be calculated. Details can be found in the EPA Report (1992) and in equations 7-15 of Gross (1995). Note that these references contain a further step allocating some risk to ex-smokers for spousal and background exposure to ETS, and we have not tried to add this extra complication here.

4.2 Choice of parameters

Values of $N, RR_c, RR_x, RR_s$

We extend the EPA model here by using values that are relevant to each age-group, as in Section 3, and summing over these to get totals for males and for females: within each age group the principles are the same. We need the numbers of deaths in each age group again: these are given in Table 4.1 based again on interpolating known WHO values (World Health Organisation, 1994, Table D-1), and again we assume they are known accurately even though as in Section 2 we could allow for these to vary also in a sensitivity analysis.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>$N$ Females</th>
<th>$N$ Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 – 34</td>
<td>56</td>
<td>96</td>
</tr>
<tr>
<td>35 – 39</td>
<td>494</td>
<td>788</td>
</tr>
<tr>
<td>40 – 44</td>
<td>494</td>
<td>788</td>
</tr>
<tr>
<td>45 – 49</td>
<td>2214</td>
<td>3653</td>
</tr>
<tr>
<td>50 – 54</td>
<td>2215</td>
<td>3653</td>
</tr>
<tr>
<td>55 – 59</td>
<td>5860</td>
<td>11097</td>
</tr>
<tr>
<td>60 – 64</td>
<td>5861</td>
<td>11097</td>
</tr>
<tr>
<td>65 – 69</td>
<td>9004</td>
<td>17015</td>
</tr>
<tr>
<td>70 – 74</td>
<td>9004</td>
<td>17016</td>
</tr>
<tr>
<td>75 – 79</td>
<td>7462</td>
<td>12885</td>
</tr>
<tr>
<td>Total</td>
<td>42664</td>
<td>78088</td>
</tr>
</tbody>
</table>


The values of $RR_c, RR_x$ and $RR_s$ that we have used in this lung cancer model are given in Table 4.2. The values for the former are again based on a current range of values (English *et al.*, 1995), and the spousal exposure risk is typical of that discussed in the EPA Report (1992) and, for example, Tweedie and Mengersen (1992).
Table 4.2: Parameter Ranges (Step Lengths) Used in EPA Model for Lung Cancer

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$RR_a$</th>
<th>$RR_x$</th>
<th>$RR_e$</th>
<th>$Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>7 - 15(2)</td>
<td>3 - 7(1)</td>
<td>1.05 - 1.35(0.05)</td>
<td>1.5 - 3.5(0.5)</td>
</tr>
<tr>
<td>Males</td>
<td>9 - 17(2)</td>
<td>4 - 8(1)</td>
<td>1.05 - 1.35(0.05)</td>
<td>1.5 - 3.5(0.5)</td>
</tr>
</tbody>
</table>

Models A and B: different choices of $p_s$

The remaining crucial decision here is the choice of $p_s$: as we saw in Section 3 this can have considerable impact on the results.

In assessing lung cancer risk, the NH&MRC (NH&MRC Working Party, 1995) used the same values of $p_s$ as in Table 3.2, and in Model A we have carried out the analysis below using these values: that is, for the under 65 age-group we take $p_s$ as ranging from 0.12 to 0.24 in steps of 0.03 for females, with males ranging similarly over 0.06 to 0.12 in steps of 0.015; and for the 65 and over age-group we take $p_s$ as ranging from 0.02 to 0.08 in steps of 0.015 for both males and females.

It is also possible to argue that, given the longer time for cancer to develop, an “ever-married” proportion might be more suitable. Tables A2.6 and A2.7 of the NH&MRC Draft Report (1995) give data from the same data-source and give values of around 55% for females married to smoking husbands (21% to current smokers, and 34% to ex-smokers), and around 27% for males married to smoking wives (12% to current smokers, and 15% to ex-smokers). These values appear consistent with those of the 1986 NRC Report in the US (NRC Committee on Passive Smoking, 1986). They all appear much lower than those cited, with very little attribution, in the EPA Report (1992, p. 6-15): here we find 60% being used for both males and for females. Gross (1995) uses the range 17%-60% in his sensitivity analysis. Based on all of these data, and bearing in mind the demonstrated sensitivity of the results in Section 3 to this parameter, we analyse a second model (Model B) with all parameters as in Table 4.1 and with $p_s$ varying in each age group identically across the values 20%-60% (in 5 steps of width 10%) for females, and across the values 10%-30% (in 5 steps of width 5%) for males.

Output for spousal and background exposure

We also need to differentiate the output for those with “background” exposure only and those with “background plus spousal” exposure. We use $AN_b$ to denote the former and $AN_{bs}$ to denote the latter.

Thus in both Model A and Model B, for each of these outcome types we have now 4375 combinations of parameters to consider for each sex after summing over the age groups, and we carry out the calculations on $AN$ and evaluate and display its sensitivity to these values as in the previous section.

4.3 Results of Regression Analyses

For each of these models we give the results of the regression analyses in Table 4.3, where we tabulate the cumulative percentage of the variance explained for the females only as we add in more parameters. (The male results are similar and we do not display them.)
In Model A, for those spousally exposed, the critical parameters are $RR_s, p_s$ and $Z$; and over the range we consider $Z$ has considerably less effect than the other two variables, somewhat in contrast to the claims in Gross (1995). For those who just have “background” exposure, $RR_s$ and $Z$ are critical and the other parameters are not of importance.

This is essentially the same in Model B, with the exception that for those with background exposure only, the values of $AN_b$ do vary somewhat over the very wide range allowed for $p_s$.

There is definite non-linearity in this model over this range: taking logs of the $AN$ values results in a distinctly better fit to the model, but does little to the order of the parameters chosen, except to indicate that for $AN_b$ the parameter $RR_s$ may be influencing the outcome more does $Z$.

<table>
<thead>
<tr>
<th>Attributable Number</th>
<th>Parameter</th>
<th>Model A</th>
<th>Parameter</th>
<th>Model B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$AN_{bs}$</td>
<td>$RR_s$</td>
<td>46.6</td>
<td>$RR_s$</td>
<td>43.2</td>
</tr>
<tr>
<td></td>
<td>$p_s$</td>
<td>63.9</td>
<td>$p_s$</td>
<td>64.0</td>
</tr>
<tr>
<td></td>
<td>$Z$</td>
<td>78.4</td>
<td>$Z$</td>
<td>78.7</td>
</tr>
<tr>
<td></td>
<td>$RR_c$</td>
<td>82.3</td>
<td>$RR_c$</td>
<td>82.0</td>
</tr>
<tr>
<td></td>
<td>$RR_x$</td>
<td>83.6</td>
<td>$RR_x$</td>
<td>83.6</td>
</tr>
<tr>
<td>$AN_b$</td>
<td>$Z$</td>
<td>41.4</td>
<td>$Z$</td>
<td>36.8</td>
</tr>
<tr>
<td></td>
<td>$RR_s$</td>
<td>72.1</td>
<td>$RR_s$</td>
<td>62.0</td>
</tr>
<tr>
<td></td>
<td>$RR_c$</td>
<td>74.2</td>
<td>$p_s$</td>
<td>69.9</td>
</tr>
<tr>
<td></td>
<td>$RR_x$</td>
<td>75.2</td>
<td>$RR_c$</td>
<td>71.8</td>
</tr>
<tr>
<td></td>
<td>$p_s$</td>
<td>75.4</td>
<td>$RR_x$</td>
<td>72.7</td>
</tr>
</tbody>
</table>

### 4.4 Sensitivity Plots and Summary Statistics

Based on this we show in Figures 4.1-4.4 the sensitivity of the outcomes for Models A and B for each of the two outcome measures, and for females only.

**Figures 4.1-4.4 Near Here**

Table 4.4 indicates the simple summaries of the outcomes for males and females. Under Model A, as developed by the NH&MRC, in the spousally exposed group we get a median of 244 female and 79 male attributed deaths; in the unexposed groups we get 987 and 468 respectively. This gives a plausible total of around 1500 attributable deaths in never-smokers. If we follow Model B with its much higher exposure rates, then adding the medians gives around 2200 attributable deaths in never-smokers.

From Figures 4.1-4.4 we see that these results respond in a non-linear way to the choice of $Z$ even though this is not the most sensitive parameter identified by the (linear) regression analysis. It is very obvious that $AN$ increases very much if the lowest value of $Z$ is used; and although it is not clear that the claim of Gross (1995) (that a value of $Z = 4.5$ is preferable to the value of 1.75 given by the EPA) can be substantiated (Hanley, 1995), it is certainly clear that the choice of this value makes a key impact on the values of $AN$ that are given.
Figure 4.1
Attributable Numbers for Model A [females]
Sensitivity of $AN_{05}$ to variation in $RR_a$, $p_a$ and $Z$
Figure 4.2
Attributable Numbers for Model A [females]
Sensitivity of ANb to variation in RR, p, and Z
Figure 4.3
Attributable Numbers for Model B [females]
Sensitivity of $AN_{bo}$ to variation in $RR_{b}$, $p_{s}$ and $Z$

![Graph showing sensitivity of attributable numbers to variation in $RR_{b}$, $p_{s}$, and $Z$.]
Figure 4.4
Attributable Numbers for Model B [females]
Sensitivity of $AN_b$ to variation in $RR_a$, $p_s$ and $Z$
Table 4.4: Summary statistics on sensitivity analyses

<table>
<thead>
<tr>
<th>Model</th>
<th>Attributable Number</th>
<th>Median</th>
<th>25th Percentile</th>
<th>75th Percentile</th>
<th>IQ Range</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td>$AN_{bs}$</td>
<td>244</td>
<td>137</td>
<td>395</td>
<td>23</td>
<td>1572</td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td>$AN_b$</td>
<td>987</td>
<td>571</td>
<td>1723</td>
<td>110</td>
<td>7744</td>
<td></td>
</tr>
<tr>
<td>Model B</td>
<td>$AN_{bs}$</td>
<td>982</td>
<td>551</td>
<td>1582</td>
<td>89</td>
<td>6192</td>
<td></td>
</tr>
<tr>
<td>Model B</td>
<td>$AN_b$</td>
<td>596</td>
<td>331</td>
<td>1072</td>
<td>50</td>
<td>6225</td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td>$AN_{bs}$</td>
<td>79</td>
<td>44</td>
<td>130</td>
<td>7</td>
<td>529</td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td>$AN_b$</td>
<td>468</td>
<td>271</td>
<td>821</td>
<td>52</td>
<td>3652</td>
<td></td>
</tr>
<tr>
<td>Model B</td>
<td>$AN_{bs}$</td>
<td>233</td>
<td>129</td>
<td>385</td>
<td>20</td>
<td>1560</td>
<td></td>
</tr>
<tr>
<td>Model B</td>
<td>$AN_b$</td>
<td>387</td>
<td>223</td>
<td>678</td>
<td>40</td>
<td>3332</td>
<td></td>
</tr>
</tbody>
</table>

We also see that there is very considerable variability in the range of the various attributable numbers. Considerable work is needed to ensure that the input parameters are as accurate as possible.

5 Conclusions

It is trivially obvious that calculations of attributable risks are sensitive to the parameters put into the calculations. The two ideas in this paper (identification of key parameters through regression analysis, and display of their effects through UBPs) can be applied to measure sensitivity in all manner of environmental situations.

The usefulness of the concepts in the examples we have used for illustration is easily seen by comparison with more traditional approaches. In the situation of Section 4, the EPA (EPA Review, 1992, Table 6-6) gave some idea of sensitivity by carrying out univariate assessments of extremes when all other parameters were at their “preferred” fixed values. Gross (1995, Tables I and II) looks at the variation under different parameter combinations but really achieves little more than to give a maximum and a minimum range for $AN$ under these combinations. In both of these it is not at all easy to see the pattern of responses to the different parameters moving in concert.

The parameters we have identified as critical in these examples may seem obvious ones, but our approach gives some valid basis for focussing on them, and not on others which from first principles might seem equally important. For example, Gross (1995) claims that “... very small changes in some key model parameters lead to large changes in the estimates [of $AN$s]”. The regression analysis enables us to see that in fact not all parameters are as important as might be thought: for example, $p_4$ has less impact than one would imagine on $AN_b$ in Section 4.
The UBP displays indicate that under the various combinations a "plausible" (i.e. interquartile) range of AV is rather tighter, and the more extreme values occur from many fewer parameter combinations, than one might initially think. The extreme values can be misleading and focussing on the median and interquartile range can give much more reasonable assessment of the range of outcomes and the weight to be given to them.

Overall, then, the tools proposed in this paper can give a simple but effective approach that enables far better measurement of sensivity than mere tabulation, and assists the researcher to identify key parameters and their influence much more effectively.

Acknowledgements

This research was stimulated by the preparation of an open commentary on the NH&MRC Draft Report (1995) by the second author. The commentary and the work in this paper was partially supported by the Tobacco Institute of Australia. The views in this paper are however entirely independent: they have been developed without consultation with any member or representatives of the tobacco industry, and should not be ascribed to that industry.

References


