A Strategy for Bounding Attributable Risk: a Lung Cancer Example

Minh Ha-Duong[1,2], Elizabeth A. Casman[2], and M. Granger. Morgan[2]

[1] Centre International de Recherche sur l'Environnement et le Développement, Centre National de la Recherche Scientifique
[2]Carnegie Mellon University, Department of Engineering and Public Policy

ABSTRACT

For diseases with more than one risk factor, the sum of probabilistic estimates of the number of cases attributable to each individual factor may exceed the total number of cases observed, especially when uncertainties about exposure and dose-response for some risk factors is high. In this study we outline a method to bound the fraction of lung cancer fatalities *not* attributed to specific well-studied causes. Such information serves as a "reality check" for attributional studies of the minor risk factors, and, as such, complements the traditional risk analysis. With lung cancer as our example, we attribute portions of the observed lung cancer mortality to known causes (such as smoking, residential radon, and asbestos fibers) and describe the uncertainty surrounding those estimates. The interactions among the risk factors are also quantified, to the extent possible. We then infer an upper bound on the residual risk due to "other" causes, using a coherence constraint on the total number of deaths, the maximum uncertainty principle, and the mathematics of imprecise probabilities.

KEY WORDS: bounding analysis, lung cancer risk, imprecise probability

1. INTRODUCTION

The familiar "front-to-back" procedure for calculating disease or mortality risk from exposure to environmental contaminants (which involves estimating toxic releases, modeling environmental and physiological transformations, and employing exposure models and dose-response functions) works best when the relevant science is well developed. However, when there are several risk factors and the uncertainty about some of them is large, such a procedure can lead to estimates for the numbers of cases attributable to the various factors that, summed, exceed the total number of cases actually observed.

Morgan argued that methods of bounding analysis could be used in environmental risk analysis to avoid such absurdities, because, for health endpoints with multiple external causes, the available knowledge constrains the magnitude of the poorly characterized risks.[1] If most risks were known with precision, this would be a simple subtraction problem. However disease risks from environmental causes are often estimated from models or inferred from studies involving limited numbers of subjects and inconsistent notions of controls or have other methodological problems that contribute to the uncertainty of the results. It is common to see even the central tendencies of such risk estimates expressed as ranges, especially when there are competing plausible models. Despite this uncertainty, often there is agreement in the literature concerning the general magnitude of the impacts of the best-studied causes. How to quantify and bound the residual risk, in light of the uncertainty surrounding the relatively well characterized risks, is the subject of this paper.

Using lung cancer mortality as a case study, the goal of our analysis is to generate an upper bound on the mortality attributed to the group of poorly characterized factors. Lung cancer (including cancers of the trachea and bronchus) is the second most common cancer for both men

and women, in terms of incidence (after breast cancer in women and prostate cancer in men), but is the leading cause of cancer death in the United States.[2] Lung cancer accounts for about 30% of all male cancer deaths and 25% of all female cancer deaths (Table I).

Table I. U.S. Lung Cancer Mortality

	Lung Cancer Deaths per 10^5, for year 2000[3]	Estimated Total U. S. Population, July 1, 2000[4]*	Lung Cancer Deaths, 2000	Projected Lung Cancer Deaths, 2003[2]**
All persons	56.5	281,421,906	159,200	157,200
Male	67.2	138,053,563	92,800	88,400
Female	46.3	143,368,343	66,400	68,800

*estimated undercount of 1.17%
**based on NCI SEER data and age adjusted to 2000 population

The death rate among the US civilian noninstitutionalized resident population from malignant neoplasms of the trachea, bronchus, and lung in the year 2000 is estimated to be 56.5 per 100,000 (ICD-10 definition), based on death certificate reporting.[3]

In 2000, the crude lung cancer death rate (not age-adjusted) among men of all ages was 67.2 per 100,000, and among women, 46.3 per 100,000. NCHS estimates that better than 99.3% of all death certificates nation-wide were included in the calculation of these rates.

Some of the major environmental risk factors for lung cancer are shown in Table II. In our method, the observed annual lung cancer mortality is attributed statistically to the major causes of lung cancer (smoking, radon and asbestos). In this paper we use literature values, but this method is well suited for expert elicitation, as well, which is planned as further work. Information about the risks from unspecified causes is then inferred using a coherence constraint on the total number of deaths, and a principle we term maximum uncertainty.

Table II. Examples of Environmental Risk Factors for Lung Cancer

Well Characterized Factors	Less Well Characterized Factors
Cigarette smoking	Occupational exposures
Passive smoking	Arsenic
Indoor radon exposures	Beryllium
Occupational exposure to inhaled asbestos	Chromates
	Chloromethyl ethers
	Diesel exhaust
	Nickel
	Silica
	Soot
	Polycyclic aromatic hydrocarbons (PAHs)
	Ambient air pollution

Our method builds upon the work of Walley.[5] Mathematically, this is an application of Smets' Transferable Belief Model[6], which offers a convenient mathematical framework for implementing the Shafer Theory of Evidence[7] using matrices. We represent our understanding of the impacts of a finite set of risk factors for lung cancer mortality as constraints on a linear

programming problem involving a convex family of probabilities. We then invoke the maximum unspecificity criterion in order to estimate the upper bound for the less well-studied members of the set.

2. MATHEMATICAL THEORY

Let N denote the magnitude of the health end-point, in this case, the total annual number of lung cancer deaths. Let Ω denote the set of all possible causes of lung cancer deaths. For example, $\Omega = \{C,R,A,X\}$ where C means exposure to tobacco smoke primarily from cigarettes (we combine active, passive, and former smoking), R means indoor exposure to radon, A means asbestos, and X is the group of all other more poorly understood environmental factors of interest.

We assume that N is readily observable and therefore known with precision. While this is not strictly true in the case of lung cancer[8], the assumption is not limiting, since the results of the method can be stated in percentage terms and then applied to a range of possible numerical values of N. We also assume exposure to be binary, which is of course not true, but the assumption is consistent with the exposure definitions used in the supporting epidemiological studies. With these two assumptions, each death can be linked to zero or more possible causes in Ω. Synergistic cases, in which more than one cause is involved, are allowed.

Figure 1 shows one way to subdivide N by causes that includes synergistic effects. We denote the number of deaths linked to cause s as $n(s)$, where s is any subset of Ω. In our example we consider four possible causes in Ω, so there could be sixteen (2^4) possible s, but to simplify the analysis and to be consistent with the lung cancer literature, we will consider only the two-factor interactions involving cigarette smoke.
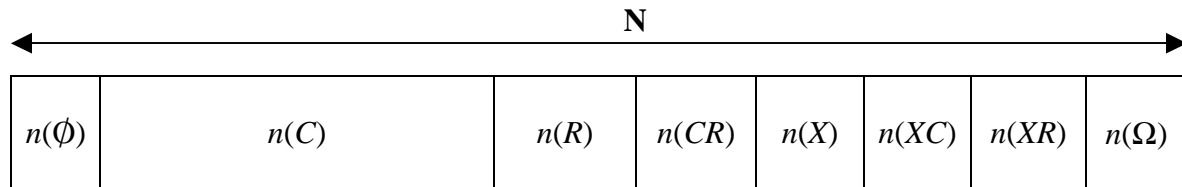

Fig. 1. The basic statistic n, showing only risk factors C, R, and X, for simplicity. N is the total number of lung cancer fatalities. n is the number of fatalities attributable to each risk factor or combination of factors. $n(\emptyset)$ is the background number of lung cancer deaths that would occur absent all the various risk factors. $n(\Omega) = n(CRX)$ and represents the number of cases for which no risk factor can be excluded.

To adopt a more precise and cautious definition, $n(s)$ is the number of cases not exposed to pollutants not in s. This implies that causes not in s are known to be non-contributing to that lung cancer. For deaths in $n(s)$, any cause in s may have caused the lung cancer, but which one is deeply epistemically uncertain and there may have been synergies. The two extreme subsets need more explanation.

The number of lung cancer deaths where all causes of Ω have been positively excluded is $n(\emptyset)$ shown to the left of the bar in Figure 1. Cases that could not be linked to any pollutant in Ω

are considered spontaneous lung cancer. It is important to underline that $n(\emptyset)$ does not have the same status as $n(X)$, which will be deduced as a residual.

On the right, the full set $n(\Omega)$ (a short notation for $n(CRAX)$) corresponds to the situation when no cause of lung cancer has been excluded because either all risk factors have been observed to be present, or there is no information about risk factor exposure.

Direct measurement of the basic statistic n is impossible, since exposure to a pollutant does not necessarily result in a cancer fatality and because retrospectively, lifetime exposures to the various carcinogens can only be roughly estimated. For now, let us assume the values of n are known; a section on attribution of cases to the various causes and the attendant uncertainty follows.

The basic statistic n can be used to bound the number of cases attributable to smoking C as follows, where $\bar{n}(C)$ and $\underline{n}(C)$ denote the upper and lower bounds on $n(C)$, respectively:

- The lower bound is the number of cases attributed only to smoking (we lump both passive and active smoking together). $\underline{n}(C) = n(C)$.
- The upper bound is the number of cases exposed to smoke and possibly other factors. That is, $\bar{n}(C) = n(C) + n(XC) + n(CR) + n(XCR) + n(CA) + n(XCA) + n(CRA) + n(XCRA)$ or: $\bar{n}(C) = \sum_{E} n(E)$ for all E such that $C \in E$

Figure 2 illustrates this definition of the upper and lower bounds of the number of lung cancer deaths attributable to X and C. For clarity the figure is drawn showing only two causes, with $\Omega = \{C,X\}$.
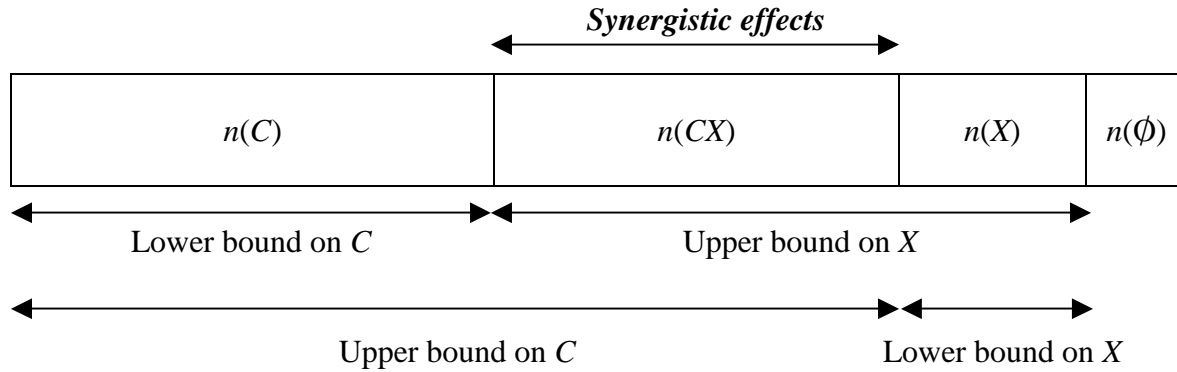

Fig. 2. Upper and lower bounds on the number of lung cancer deaths attributable to C and X. For clarity of illustration, risk factors other than C and X have been excluded from this figure.

In epidemiologists' terms, the attributable fraction of cause C, for example, is the proportion of all cases that could be avoided if exposure to C were eliminated, denoted $af(C)$. The model suggests the following bounds for smoking attributable fraction:

$$\frac{\underline{n}(C)}{N}(1-r_0) \le af(C) \le \frac{\bar{n}(C)}{N}(1-r_0) \tag{1}$$

where r_0 is the background rate of lung cancer mortality. This background rate is the number of lung cancer deaths in the unexposed population divided by the unexposed population. The lower

bound on *af* accounts for the 1 - r_0 share of spontaneous lung cancer cases in those cases exposed to cigarettes. The upper bound attributes all cigarette-exposed deaths to this factor.

Denoting p_C, p_R and p_A as the exposure probabilities of *C,R,A*; and *T* as the total population; assuming independence (meaning that people who smoke are no more or less likely to be exposed to radon or to asbestos):

$$r_0 = \frac{n(\emptyset)}{(1-p_C)(1-p_R)(1-p_A)T} \qquad (2)$$

A geometric interpretation of the space created by the upper and lower bounds on the individual causes in *s* is shown in Fig. 3, using only factors *C*, *R*, and *X*, for simplicity. Denoting π a basic statistic that is interaction-free (meaning that $\pi(C) + \pi(R) + \pi(X) = N - \pi(\emptyset)$) we say π is "compatible" with another basic statistic *n* when it falls within the bounds *n* defines. The set of all interaction-free basic statistics compatible with *n* is denoted $\mathcal{P}n$ in the figure.
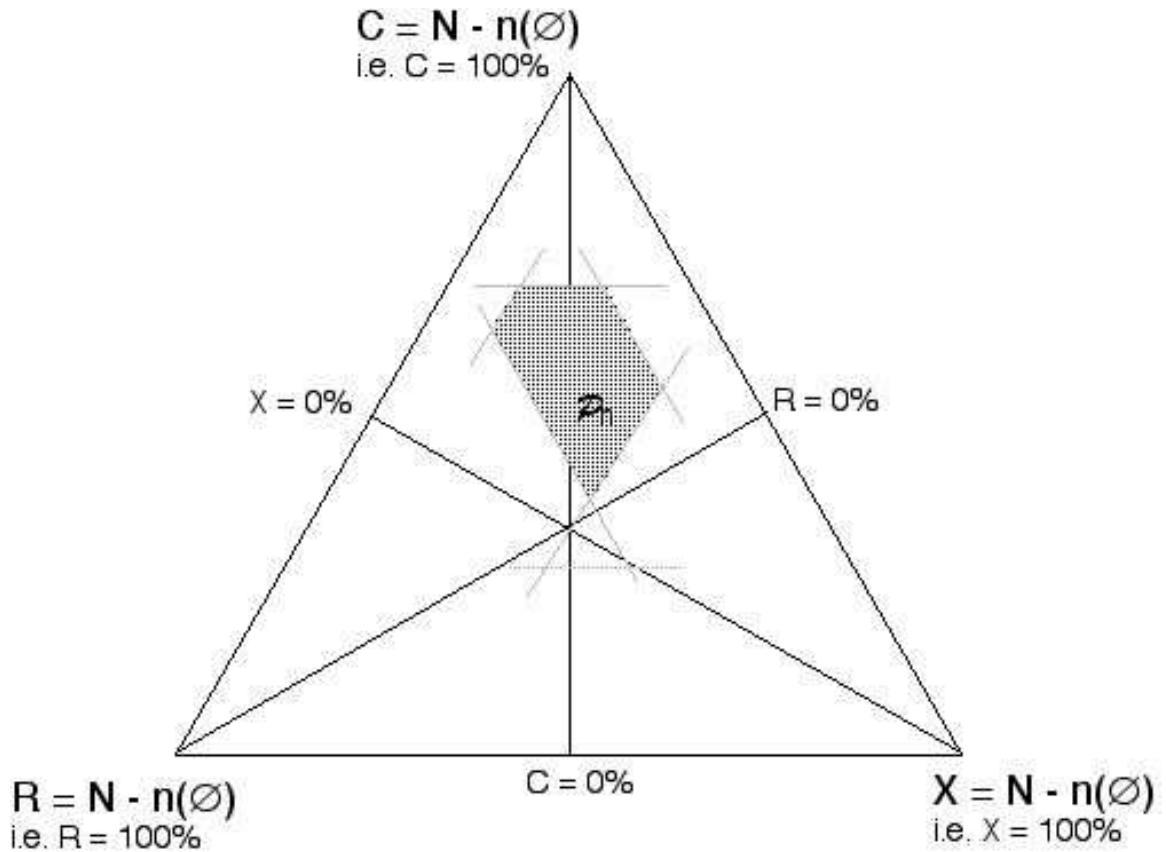

Fig. 3. Graphical representation of the set of all interaction-free π compatible with a specific, *n* defined on *C*, *R*, and *X*.

Consider now the bounds on deaths attributed to multiple synergistic causes. Denote these causes as *s*, a subset of Ω, for example $s = CR$. For the lower bound on the number of deaths attributable to these causes acting jointly, we continue to adopt the number of cases exposed only to these causes, that is:

$$\underline{n}(s) = n(s) \qquad (3)$$

7

and as the upper bound, we continue to adopt the number of cases exposed to *s* and possibly other factors, that is:

$$\bar{n}(s) = \sum_{E, s \subseteq E} n(E) \qquad (4)$$

This $\bar{n}$ corresponds to the commonality function in the Transferable Belief Model[9]. Bounds on the attributable fraction can be computed as in Equation 1.

One measure of the information contained in *n* is its unspecificity. Consider these two (of the three) extreme cases:

- If each death were attributed to exactly one cause, then there would be no uncertainty, and all lower bounds would coincide with their upper counterpart. We would have $n(C) + n(R) + n(A) + n(X) = N - n(\emptyset)$. Note that since *n* is a positive function that sums up to *N*, this implies that $n(s) = 0$ for all the subsets with 2 or more elements.
- If no information on cause of cancer were available, each death would be attributed to the synergy of all factors. Mathematically, this is $n(\Omega) = N$. Note that this constitutes a proper uninformative distribution: it is not the Bayesian uniform prior probability distribution on Ω. Rather, it is the family of all probability distributions that can be defined on Ω.

Unspecificity is a numeric indicator that equals one in the first case, and in the second case equals the number of elements of Ω. It is the expected value of the number of elements of *s* with respect to the probability distribution $m(s) = \dfrac{n(s)}{N}$, that is:

$$U = \frac{n(C) + n(R) + n(A) + n(X) + 2(n(CR) + n(RA) + ...) + 3... + 4n(\Omega)}{N} \qquad (5)$$

In this paper unspecificity is a kind of generalized cardinality that specifies the number of alternatives. The reason for using this word is that when a death is attributed to the synergy of *k* factors, it can be said that the unspecificity of this information is *k*. See Rocha (1997) for an extensive discussion of this concept[10].

A lower unspecificity measure corresponds to better information, so a third extreme case needs discussion: unspecificity is zero when and only when $n(\emptyset) = N$. In this case, all non-spontaneous causes of Ω have been positively excluded. It means that all the substances in Ω are actually safe (with respect to lung cancer). This is the highest level of information achievable, to the point that it makes Ω irrelevant.

Note how this interpretation hinges on the idea of counting missing data with $n(\Omega)$. This is an application of a general principle of maximum uncertainty, also known as Laplace's principle of "raison insuffisante". The principle states that one should select the statistic that is the most unspecific, compatible with existing information. This is the principle that we use in the next section to estimate the bounds on the unknown cause, given information about all others.

Quantitative or comparative statements regarding the likely magnitude of $n(s)$ can be interpreted as linear constraints on *n*. These constraints determine a set *B* of basic statistics. The most unspecific *n* in *B* is chosen according to the maximum uncertainty principle. This amounts to solving a linear program in a space with $2^{|\Omega|}$ dimensions, for which we impose these additional constraints:

- All $n(s)$ are non-negative, summing up to *N*.
- Three-way interactions and higher are not allowed. That is, $n(s) = 0$ if *s* has 3 or more elements.

To ensure that a solution exists, we impose the condition of coherence on the *a priori* attribution:

$$\overline{af}(s_i) + \sum_{j \neq i} \underline{af}(s_j) \leq 1 \leq \underline{af}(s_i) + \sum_{j \neq i} \overline{af}(s_j) \quad \text{for } 1 \leq i \leq |\Omega| \qquad (6)$$

If exposure probabilities are known or assumed, it is possible to also bound the relative risk of s, $rr(s)$. The definition of the relative risk of dying of lung cancer from smoking cigarettes $rr(C)$, for example, is the lung cancer rate associated with exposure to tobacco smoke divided by the lung cancer rate among non-smokers. Given exposure probabilities in the general population, we will assess the bounds on the relative risk for the various pollutants using the following relationship:

$$\overline{af}(s) = \frac{p_s \, (\overline{rr}(s) - 1)}{p_s \, (\overline{rr}(s) - 1) + 1} \quad \text{for any cause } s \text{ in } \Omega \qquad (7)$$

where p_s is the exposure to cause s. The same relationship holds for the lower bound.

This paper's numerical simulations were performed using a *Mathematica* notebook[1]. The code directly implements matrix calculus for belief functions as outlined in Smets (2001)[6]. This is the most straightforward method given that Ω remains small, but it would not scale well to tens of carcinogens, since it involves square matrices with $2^{2|\Omega|}$ elements. For example, 10 pollutants implies storing in memory arrays with 10^6 numbers.

3. *A PRIORI* ATTRIBUTION OF LUNG CANCER RISK

In our illustration Ω, the set of possible causes of lung cancer, consists of *C,* active and passive smoking; *R,* domestic exposure to radon; *A,* occupational exposure to inhaled asbestos; and *X,* the group of all other environmental risk factors.

That tobacco smoking is the single most important cause of lung cancer, has been understood for more than 60 years[11]. For example, a 1983 survey conducted by the World Health Organization of 160 active members of the International Association for the Study of Lung Cancer found that 97% of the respondents agreed that at least 80% of all lung cancer cases were caused by tobacco.[12] About 98% of European and U. S. male lung cancer patients were "ever-smokers" (persons who had smoked more than 100 cigarettes in their lifetime) as were 70% to 90% of the women lung cancer patients.[13] Lung cancer mortality rates are about 23 times higher for current male smokers and 12 times higher for current female smokers than for never smokers.[2] For people having smoked more than 100 cigarettes in their lifetimes, the relative risk of dying of lung cancer is 10 to 15.

The smoking attributable fraction (SAF) of cancers of the lung, trachea, and bronchus (ICD-9 code 162) averaged over the years 1995-1999 was calculated by the CDC using their SAMMEC software[14]. These estimates were derived from relative risks from the American Cancer Society's Cancer Prevention Study II, CPS-II: 1982-1988[15] and current and former cigarette smoking prevalence[16] for two age cohorts: persons aged 35 to 64 years and those aged 65 years, according to this formula:

$$SAF = \frac{p_0 + p_1(rr_1) + p_2(rr_2) - 1}{p_0 + p_1(rr_1) + p_2(rr_2)} \qquad (8)$$

where p_0 is the percentage of persons who have never smoked 100 or more cigarettes, p_1 is the percentage of persons who have smoked 100 or more cigarettes and smoke every day or some days, and p_2 is the percentage of persons who have smoked more than 100 cigarettes but have stopped smoking. The relative risks of lung cancer of these three segments of the population are

---

[1] Available on the web at url http://www.andrew.cmu.edu/user/mduong, or upon request, under the GNU General Public License

rr_0 (which equals 1 by definition), rr_1, and rr_2. Using SAMMEC, the CDC estimated the following annual smoking attributable mortality figures, shown in Table III.

Table III. Smoking-Attributable Mortality (SAM), Cancers of the Trachea, Lung and Bronchus, 1999[17]

		Mortality from Neoplasms of Trachea, Lung, and Bronchus	Lung Cancer Deaths from Secondhand Smoking[18]	Active & Passive Smoking Attributable Fraction
Males	Total deaths	89,337		
	SAM	78,459	1,110	
	percent	87.8%		89.1%
Females	Total deaths	62,613		
	SAM	44,727	1,890	
	percent	71.4%		74.4%
Males and Females	Total deaths	151,950		
	SAM	123,186	3,000	83.0%
	percent	81.1%	2%	

The CDC noted that these figures have some limitations.
- SAM figures were derived from smoking rates in 1999 when the actual smoking attributable deaths were the result of smoking in previous decades, when smoking rates were higher. During periods where smoking prevalence is declining, the attributable fraction methodology will tend to understate the number of deaths caused by smoking. Smoking is currently declining by about 1% a year and between 1993 and 1999, smoking prevalence among adults declined by 25%.
- Relative Risks were adjusted for the effects of age but not for other potential confounders. This may not be a large problem, as CPS-II data showed that education, alcohol, and other confounders had negligible additional impact on SAM estimates for lung cancer[19,20].
- Deaths attributable to cigar and pipe smoking and smokeless tobacco use were not included.
- Although the CPS-II cohort includes over 1.2 million men and women, it is somewhat more white and middle class than the U. S. population as a whole.

Given both the overwhelming evidence of the importance of this risk factor and the potential for underestimation bias in the calculation of its contribution to overall risk, for this calculation we assume the lower bound on attributable risk due to smoking to be 70% and the upper bound, 95%.

Probably the second most important risk factor for lung cancer in the U. S. is radon, which is a gas emitted from the soil, from water and from building materials, that can become concentrated in homes. Once inhaled into human lungs, radioactive decay of radon and its radioactive daughters results in the exposure of lung cells to alpha particles, which can damage DNA by ionization. On the basis of studies of health effects in uranium and other underground miners who worked in radon-rich environments, and backed-up by animal exposure and human epidemiological studies, radioactive radon progeny have been identified as a cause of lung cancer in humans.

The annual number of lung cancers in the U. S. caused by residential radon exposure was estimated in the BIER VI study (Biological Effects of Ionizing Radiation) conducted by the National Research Council[21]. This committee chose to use mathematical models (rather than population studies), because epidemiological case-control studies of lung cancer due to residential radon have not produced clear results. This failure was attributed to a number of reasons. The risk appears to be very small, and the exposures encountered in most homes are low. It is difficult to estimate radon exposures that people have received over their lifetimes. Also, many more lung cancers are caused by smoking than by radon and there is strong evidence for a synergistic interaction between smoking and radon.

The NRC committee constructed two preferred models, both based on information from 11 major studies of lung cancer in underground miners together involving 68,000 men with 2,700 lung cancer fatalities.[22] Since radon levels in the mines were in general many times the levels found in most homes, the assumption was made that risk decreases linearly with exposure and there is no threshold. There were other problems with this approach, as well. Almost all of the miners were men, they inhaled dust and other pollutants in the mines, and most were smokers.

The NRC committee assumed a submultiplicative interaction (more than additive but less than multiplicative) between cigarette smoking and radon exposure, consistent with data from the miner studies. Another important modeling assumption was that the physical and biologic differences between the circumstances of exposures of male miners working underground and of men, women, and children in their homes lead to the same doses at the same exposures. The committee used lung cancer mortality data from 1985-1989 and smoking prevalence (ever and never smoked) estimates for 1993. Based on the National Residential Radon Survey, a lognormal distribution for residential radon concentration with a median of 24.3 Bqm^{-3} (0.67 $pCiL^{-1}$) and a geometric standard deviation of 3.1 was assumed[23]. The attributable risk was calculated for the entire US population and for males and females and ever-smokers and never-smokers (Table IV) as 10% to 15% of the total lung cancer mortality (their central estimates) with a 95% confidence interval of 2% to 21%. The authors of BIER VI could not find a reason to prefer either the Exposure-Age-Concentration Model or the Exposure-Age-Duration Model of the effects of residential exposure to radon on lung cancer so they presented the results of both models. Consistent with their preference, we adopt the upper and lower bound of the Constant Relative Risk Model.

Table IV. The Fraction of all Lung Cancer Deaths Due to Indoor Radon (Attributable Risk) as Estimated in BIER VI [21]

Model*	Central Estimate	95% Confidence Interval
Exposure-Age-Concentration	14%	7-17%
Exposure-Age-Duration	10%	9-25%
Constant Relative Risk**	12%	2-21%

*For central estimate, the NRC prefers the first two models. For confidence interval, they prefer the simple constant relative risk model.
**exposure <0.175 Jhm^{-3}; <50 WLM

In our judgment, the third most important cause of lung cancer is inhaled asbestos fibers. Asbestos refers to a group of 6 naturally occurring fibrous silicate minerals, chrysotile, amosite, crocidolite, tremolite asbestos, anthophyllite asbestos, and actinolite asbestos. These minerals are

made up of long (1 to 100mm), thin fibers that tend to break easily into a dust composed of inhalable particles, a few microns in length. Asbestos is released into the air through the erosion of natural deposits in asbestos-bearing rocks; through the manufacture and wear of clutches and brakes on cars and trucks; from abraded or corroded building materials such as insulation, ceiling and floor tiles, asbestos-cement pipes, and asbestos roofing materials; from building demolition; and from a variety of asbestos-related industries.

Asbestos and all commercial forms of asbestos are considered to be human carcinogens[24]. Occupational exposure to chrysotile, amosite, anthophyllite, and mixtures containing crocidolite has resulted in elevated incidence of lung carcinomas. Mesotheliomas have been associated with occupational exposure to crocidolite, amosite, and chrysotile.[25,26] Asbestos-related medical conditions can occur 15 to 40 years after exposure.

Before the 1970s, an estimated 27 million American workers in high-risk occupations were exposed to asbestos and an unknown number of workers were exposed in other occupations. The U. S. Department of Labor estimates that currently 1.3 million employees could be exposed to asbestos on the job[27]. Workers in the construction industry face the biggest risk of exposure, particularly in the removal of asbestos during renovation or demolition of old buildings and the repair of ships. Employees can also be exposed to fibers during the manufacture of asbestos products (such as textiles, friction products, insulation, and other building materials) and during automotive brake and clutch repair work.

With the passage of the Occupational Safety and Health Act in 1970, increasingly strict asbestos exposure regulations were adopted. The EPA proposed a ban on all uses of asbestos in 1986, but the proposed ban was overturned on appeal[28]. However, some asbestos containing materials were successfully banned, and no new uses in products that had not historically contained asbestos are allowed. A combination of legislation, large class-action suits, and strong public opposition to asbestos resulted in the reduction of asbestos consumption from its peak at 803,000 metric tons in 1973 to 13,100 metric tons in 2001[29,30]. Asbestos use is presently limited primarily to roofing products (61%), gaskets (19%), and automobile clutch, break and transmission components (13%)[31].

In 1982 Nicholson et al. estimated that in the year 2000 there would be 9,700 workplace asbestos related lung cancer deaths per year (about 6% of all lung cancers), based on occupational asbestos exposures from 1940-1979[32]. This is not directly verifiable, because statistics on asbestos-related lung cancer are not collected. However, the Centers for Disease Control and Prevention collect death certificate data on other asbestos related diseases[3] (asbestosis and malignant mesothelioma – the latter condition is 70 to 90% associated with asbestos exposure). In 1999 there were 2,343 reported mesothelioma deaths, and in 2000, 2,384. These figures, combined with the ratio of mesothelioma to lung cancer mortality in asbestos exposed populations, can be used to estimate the number of asbestos-related lung cancers.

In a review of asbestos-related lawsuits from 1991 through 2000, representing over 600,000 individual claimants[2], the RAND Institute for Civil Justice found that 3% of these claims concerned mesothelioma, 7% lung cancer and 89% nonmalignant disease[33] suggesting that the ratio between lung cancer and mesothelioma in asbestos workers is 7 to 3 (2.33 to 1) if

---

[2] A large volume of these claims had been initiated through mass screenings for class action suits. Though overburdening the courts with unimpaired claimants, this process reduces sampling bias for our purposes.

all affected parties sued[3]. Some of the mesotheliomas were not caused by asbestos, some the lung cancers were probably due to factors such as smoking, and there was probably some diagnosis error. However, we use this ratio to calculate that there were <5,560 occupational asbestos related lung cancers in 2000, representing <3.5% of the 159,000 lung cancer deaths of 2000.

This is compatible with other estimates that attribute 5 to 25% of lung cancer mortality to the effects of all occupational carcinogenic exposures combined, coupled with estimates that 10 to 20% of occupational lung cancers are due asbestos exposure (which would attribute 1 to 5% of total lung cancer to asbestos).[35-40] It is also compatible with the year 2000 population attributable lung cancer plus mesothelioma risk for asbestos of 5% calculated for Denmark, Finland, Iceland, Norway, and Sweden[38].

Though there is no definitive statement of the range of attributable risk of lung cancer from asbestos exposure, several methods of estimating this seem to converge on a range of 1% to 5%. For exposure rate we use 10 times the current occupational exposure rate, to account for past exposures.

In summary, based on our review of the literature we have constructed a set of judgments attributing lung cancer deaths among these causes which imply the following constraints: $\overline{af}(C)$ = 0.70, $\underline{af}(C)$ = 0.95, $\overline{af}(R)$ = 0.21, $\underline{af}(R)$ = 0.02, $\overline{af}(A)$ = 0.05, and $\underline{af}(A)$ = 0.01.

Our method also requires an estimate of the background rate of spontaneous lung cancer, r_0 (see Equation 2). Background lung cancer mortality means different things in different studies. For single risk factor studies, it usually means the cancer rate in the population not exposed to that carcinogen. For this analysis, we have defined the background rate as the rate of lung cancer deaths that occur spontaneously, without exposure to any environmental, dietary, occupational, or other carcinogen. The background lung cancer rate in populations not exposed to any carcinogen cannot be directly calculated because exposure history among lung cancer victims is poorly known. Instead, we can make some inferences based on lung cancer rates in populations presumed to have had low carcinogen exposure.

The attempts to define background lung cancer rate for minor carcinogens (anything but smoking) can inform order of magnitude estimates of the true background rate. For instance, the Committee on the Health Risks of Exposure to Radon suggested that lung cancer mortality at the beginning of the 20th century, which was 3 to 4 per 100,000 in the 1930s, or contemporary lung cancer rates among lifetime non-smokers such as Mormon women, about 4 per 100,000[41], could be used to cautiously bound the impact of radon exposure on lung cancer[21].

Summarizing 3 studies published between 1968 and 1982, Koo and Ho reported that the lung cancer rate among non-smoking U.S. females aged 40-74 was 5.8 to 7.5 per 100,000[13]. These rates are not corrected for passive smoking.

Some of the deaths in these cohorts presumably occurred spontaneously (without exposure to radon or environmental tobacco smoke) so we can claim that these rates represent the impact of background occurrence *plus* radon and perhaps other carcinogen exposures.

---

[3] This is comparable to the ratio of lung cancer to mesothelioma deaths among 17,800 asbestos insulation workers in the US and Canada from 1967 to 1977, namely 2.77. See R. Lilis, "Mesothelioma*," in Asbestos Bibliography (Revised)*, (U. S. Department of Health and Human Services, National Institute for Occupational Safety and Health,, Cincinnati, 1997.

If there were no other lung cancer risks than radon exposure in these populations and the relative risk for radon exposure were 1.2 to 1.5, the background "radon free" rate of lung cancer from an observed mortality of 4 per 100,000, would be 3.3 to 3.6 lung cancer deaths per 100,000.[21]

There are 3 major objections to this calculation. In the early part of this century, lung cancer was greatly underdiagnosed, often being mistaken for tuberculosis, in the absence of tissue confirmation. The assumption that radon was the only contributing carcinogen is false. Finally,the relative risk for radon exposure is linked to dose.

Several authors have studied lung cancer in nonsmoking populations. Axelson et al. calculated the background lung cancer rate (defined as lung cancer mortality rate among non-smokers) as

$$r_0 = \frac{r}{p_0 + p_s rr_s + p_P rr_P} \qquad (9)$$

where r_0 is the background rate in the non-smoking population that is also not exposed to passive smoking, r is the total lung cancer mortality among smokers and non-smokers, p_0 is the percent not exposed to smoke twenty years before the time r is measured, p_S and p_P are the proportions of the group who smoke and are exposed to passive smoking, respectively, and rr is the relative risk, with subscripts as for p. This background rate increased over time in several populations, presumably with occupational and other environmental carcinogenic exposure (Table V).

Alavanja et al., observing that 20% of lung cancers in women from Missouri from 1986 to 1991 occurred among non-smokers, studied lung cancer among ex-smoking and lifetime nonsmoking white women from that state. They were able to explain 29% of the lung cancer incidence in this group by non-smoke related variables (saturated fat (22%), previous non-malignant lung disease (10%), occupational exposures to known carcinogens (5%), family history of lung cancer (4%), indoor radon concentrations above 4 pCi/m^3 at current residence (1%)) and 50% of the risk by those plus past smoking history (17%) and environmental tobacco smoke (6%). The percentages in parentheses are individual variable attributable risk whereas the first two percentages refer to groups of variables evaluated together.[43]

Table V. Lung Cancer Mortality Rates Not Attributable to Smoking[42]

Year	Cohort	Rate, deaths per 100,000
1955	Nonsmoking U. S. women	4.9
1965	Nonsmoking U. S. women	5.7
1975	Nonsmoking U. S. women	4.9
1985	Nonsmoking U. S. women	4.5 to 6.13
1978	Nonsmoking Japanese women	3.25
1980-2	Nonsmoking Italian women	6.1 to 8.6
1980-2	Nonsmoking Italian men	10.1 to 10.8

Though it is impossible to calculate a background lung cancer mortality rate for people with no exposure to carcinogens, researchers have calculated mortality rates for groups expected to have low exposures to carcinogens, especially non-smoking groups. These rates are generally 10 deaths per 100,000, with a median around 5 or 6. During the latter part of the last century, probably at least a half of these cancers could be traced to some past carcinogenic exposure if the

patients' exposure histories were known. Accordingly, we estimate the background rate to be 3 per 100,000, or $n(\emptyset) = 0.013N$.

5. CALCULATION OF UPPER AND LOWER BOUNDS ON ATTRIBUTABLE FRACTION AND RELATIVE RISK FOR EACH RISK FACTOR

Table VI shows the implications of the most unspecific imprecise probability distribution compatible with these constraints for bounds of *af* and *rr*.

Table VI. Results of Optimization: Upper and Lower Bounds on Attributable Fractions and Relative Risks of Cancer Risk Factors

	C	R	A	X
Upper bound on *af*	95%	21%	5.0%	3.2%
Lower bound on *af*	70%	2%	1.0%	0.0%
Exposure probability	45%[44]	50%*[21]	5%[27]	5%**
Upper bound on *rr*	43.2	1.53	2.05	1.66
Lower bound on *rr*	6.19	1.04	1.20	1.00

*The fraction of U. S. homes with radon concentrations at or above 25 Bqm^{-1}
**our estimate

This result attributes between 0% and 3% of lung cancer deaths to *X*, the group of unknown environmental pollutants. For the group of known and suspected lung carcinogens other than C, A, and R, the risk analyst concludes that, *if one is confident in the bounds assigned to the well understood risk factors*, the sum of the effects of the other factors accounts for no more than 3% of total lung cancer mortality.

The implication for judging future risk assessments of members of *X* is that, if the assessments project the lung cancer risk to the U. S. population from these pollutants to be in excess of 3% of the annual lung cancer mortality, the assumptions of the models should be re-examined and the upper bound on the resulting estimate constrained.

Two national associations of air quality control offices published a projection of the number of cancers due to exposure to diesel exhaust[45]. They estimated that diesel will be responsible for 125,110 cancers for all metropolitan and non-metropolitan areas of the U. S. (over 70 year lifetimes) for an annual rate of 1,787 cancers. This figure is below 3% (4,716 deaths) of the projected 2003 lung cancer mortality rate of 157,200[2], even without adjusting for the non-lung cancer mortality inherent in their risk estimate, and therefore, to first order, would pass our plausibility test.

The members of *X* are a diverse group and probably do not share a common exposure level. As more is known about the individual members, they would presumably move out of *X* and perhaps the exposure level for *X* will become more tractable. In the absence of some convincing estimate of this exposure, we offer a sensitivity analysis (Table VII). Changing the exposure level does not affect the calculation of the basic statistic; however, it does affect the relative risk (Equation 7).

Table VII. Sensitivity Analysis of Relative Risk of X to Fraction of Population Exposed

p_x	0.005	0.01	0.05	0.10	0.15	0.20	0.25	0.50	0.75
rr(X)	7.64	4.32	1.66	1.33	1.22	1.17	1.13	1.07	1.04

6. CONCLUSION

In this paper we have applied the Smets' Transferable Belief Model to estimate an upper bound on the fraction of lung cancers caused annually by the group of causes for which comprehensive longitudinal studies are lacking. Such a result is interesting from a risk management perspective. For example it might be used to infer the level of control effort these pollutants deserve.

The bounding was accomplished by attributing a portion of the observed cancers to three known causes, smoking, radon, and asbestos, and then deducing information about the residual using maximum unspecificity. The procedure illustrated here can be readily applied to bound risk factors for a variety of other health impacts and environmental pollutants. The critical aspects of this procedure are:

1. Uncertainty in the known causes is explicitly stated, using statements on upper and lower bounds.
2. Synergistic effects in the known causes are part of the framework.
3. Consistency between known causes and poorly understood agents is required.

The bounding methodology presented in this paper is amenable both philosophically and mathematically to coupling with expert elicitation of *a priori* risk attributions[46], a set of such elicitations is planned for the next phase of this research. We also plan to further develop the mathematical treatment to remove the constraint on the number of the interaction terms.

REFERENCES

1. M. G. Morgan, "The neglected art of bounding analysis," *Environmental Science & Technology* **35**, 162A-163A (2001).
2. American Cancer Institute, "Cancer Facts & Figures 2003," American Cancer Society, Inc. Surveillance Research, http://www.cancer.org/docroot/STT/stt_0.asp, *last accessed 2/16/03* (2003).
3. National Center for Health Statistics, in, US Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD, 2002.
4. U. S. Census Bureau, "United States Census 2000," U. S. Department of Commerce Economics and Statistics Administration, http://www.census.gov/prod/2002pubs/c2kprof00-us.pdf, *last accessed 2/16/03* (2002).
5. P. Walley, *Statistical Reasoning with Imprecise Probabilities*, (Chapman and Hall, London, 1991).
6. P. Smets, "Matrix calculus for belief functions," Institut de Recherches Interdisciplinaires et de Développements en Intelligence Artificielle, http://iridia.ulb.ac.be/~psmets/MatrixRepresentation.pdf, *last accessed* (2001).

7.   G. Shafer, *A Mathematical Theory of Evidence*, (Princeton University Press, Princeton, NJ, 1976).

8.   V. E. Archer, J. L. Lyon, "Errors and biases in the diagnosis of cancer of the lung and their influence on risk estimates," *Medical Hypotheses* **54**, 400-407 (2000).

9.   P. Smets, "Belief functions and the transferable belief model," The Imprecise Probabilities Project, http://ippserv.rug.ac.be, *last accessed* (2000).

10. L. M. Rocha, "Relative uncertainty and evidence sets: a constructivist framework," *International Journal of General Systems* **26**, 35-61 (1997).

11. J. Higginson, C. S. Muir, N. Muñoz, *Human Cancer: Epidemiology and Environmental Causes*, (Cambridge University Press, Cambridge (UK), 1992).

12. K. Stanley, H. H. Hansen, "An international survey on attitudes in the control of lung cancer: A collaborative study between WHO and I.A.S.L.C.," *Lung Cancer* **1**, 47-55 (1985).

13. L. C. Koo, J. H. Ho, "Worldwide epidemiological patterns of lung cancer in nonsmokers," *International Journal of Epidemiology* **19**, s14-s23 (1990).

14. Centers for Disease Control and Prevention, "Smoking-attributable mortality, morbidity, and economic costs (SAMMEC): adult SAMMEC and maternal and child health (MCH) SAMMEC software,", http://www.cdc.gov/tobacco/sammec, *last accessed* (2002).

15. M. J. Thun, C. Day-Lally, D. Myers, e. al., "Trends in tobacco smoking and mortality from cigarette use in Cancer Prevention Studies I (1959 through 1965) and II (1982 through 1988)*," in Changes in cigarette-smoking and tobacco control*, (U. S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute, Bethesda, MD, 1997.

16. National Center for Health Statistics, "National Health Interview Survey (NHIS)," Centers for Disease Control and Prevention, http://www.cdc.gov/nchs/about/major/nhis/released200202.htm, *last accessed* (2003).

17. Centers for Disease Control and Prevention, "Annual smoking-attributable mortality, years of potential life lost, and economic costs -- United States, 1995-1999," *Morbidity and Mortality Weekly Report* **51**, 300-303 (2002).

18. Environmental Protection Agency, in, National Center for Environmental Assessment, 1993.

19. M. J. Thun, L. F. Apicella, S. J. Henley, "Smoking vs other risk factors as the cause of smoking-attributable deaths: Confounding in the courtroom.," *Journal of the American Medical Association* **284**, 706-712 (2000).

20. A. M. Malarcher, J. Schulman, L. A. Epstein, et al., "Methodological issues in estimating smoking-attributable mortality in the United States," *American Journal of Epidemiology* **152**, 573-584 (2000).

21. J. M. Samet, D. Brenner, A. L. Brooks, W. H. Ellett, E. S. Gilbert, D. T. Goodhead, E. J. Hall, P. K. Hopke, D. Krewski, J. H. Lubin, R. O. McClellan, P. L. Ziemer, in Biological Effects of Ionizing Radiation, National Research Council, National Academy Press, Washington, D. C., 1999.

22. J. H. Lubin, J. D. J. Boice, C. Edling, R. W. Hornung, G. Howe, E. Kunz, R. A. Kusiak, H. I. Morrison, E. P. Radford, J. M. Samet, M. Tirmarche, A. Woodward, S. X. Yao, D. A. Pierce, in, National Institutes of Health, National Cancer Institute, U. S. Department of Health and Human Services., Washington, D. C., 1994.

23. F. R. Marcinowski, M. Lucas, W. M. Yeager, "National and regional distributions of airborne radon concentrations in U. S. homes," *Health Physics* **66**, 699-706 (1994).

24. U. S. Department of Health and Human Services, "Report on Carcinogens, Tenth Edition," Public Health Service, National Toxicology Program, http://ehp.niehs.nih.gov/roc/toc10.html, *last accessed* (2002).

25. International Agency for Research on Cancer (IARC), in Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, Lyon, France, 1982, p. 292.

26. International Agency for Research on Cancer (IARC), in Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, Lyon, France, 1987, p. 440.

27. U. S. Department of Labor, "Safety and Health Topics: Asbestos," Occupational Safety & Health Administration, http://www.osha-slc.gov/SLTC/asbestos/, *last accessed 3/5/03* (2002).

28. Fifth Circuit Court of Appeals, in, 1989.

29. D. A. Buckingham, R. L. Virta, "Asbestos Statistics 1900-2001," U. S. Geological Survey, minerals.usgs.gov/minerals/pubs/of01-006/asbestos.pdf, *last accessed* (2002).

30. R. L. Virta, "Asbestos," U.S.G.S., http://minerals.usgs.gov/minerals/pubs/commodity/asbestos/070400.pdf, *last accessed 3/5/03* (2000).

31. Agency for Toxic Substances and Disease Registry, in, U. S. Department of Health and Human Services, 2001, p. 146.

32. W. J. Nicholson, G. Perkel, I. J. Selikoff, "Occupational exposure to asbestos: population at risk and projected mortality -- 1980-2030," *American Journal of Industrial Medicine* **3**, 259-311 (1982).

33. S. J. Carroll, D. Hensler, A. Abrahamse, J. Gross, M. White, S. Ashwood, E. Sloss, in, RAND Institute for Civil Justice, Santa Monica, CA, 2002, p. 106.

34. R. Lilis, "Mesothelioma*," in Asbestos Bibliography (Revised)*, (U. S. Department of Health and Human Services, National Institute for Occupational Safety and Health,, Cincinnati, 1997.

35. M. Nurminen, A. Karjalainen, "Epidemiologic estimate of the proportion of fatalities related to occupational factors in Finland," *Scandinavian Journal of Work, Environment, and Health* **27**, 161-213 (2001).

36. O. Axelson, "Alternative for estimating the burden of lung cancer from occupational exposures -- some calculations based on data from Swedish men," *Scandinavian Journal of Work, Environment, and Health* **28**, 58-63 (2002).

37. J. B. Schoenberg, A. Stemhagen, T. J. Mason, J. Patterson, J. Bill, R. Altman, "Occupation and lung cancer risk among New Jersey white males," *Journal of the National Cancer Institute* **79**, 13-21 (1987).

38. L. Dreyer, J. F. Winther, E. Pukkala, A. Andersen, "Avoidable cancers in the Nordic countries; occupation," *Acta Pathologica, Microbiologica, et immunologica Scandinavica. Supplementum* **105**, 68-79 (1997).

39. P. Vineis, L. Simonato, "Proportion of lung and bladder cancers in males resulting from occupation: a systematic approach," *Archives of Environmental Health* **46**, 6-15 (1991).

40. A. Morabia, S. Markowitz, K. Garibaldi, E. L. Wynder, "Lung cancer and occupation: results of a multicentre case-control study," *British Journal of Industrial Medicine* **49**, 721-727 (1992).

41. J. L. Lyon, J. W. Gardner, D. W. West, "Cancer in Utah: risk by religion and place of residence," *J Natl Cancer Inst* **65**, 1063-1071. (1980).

42. O. Axelson, D. L. Davis, F. Forestiere, M. Schneiderman, D. Wagener, "Lung Cancer Not Attributable to Smoking," *Annals of the New York Academy of Sciences* **609**, 165-178 (1990).

43. M. C. R. Alavanja, R. C. Brownson, J. Benichou, C. Swanson, J. D. J. Boice, "Attributable risk of lung cancer in lifetime nonsmokers and long-term ex-smokers (Missouri, United States)," *Cancer Causes and Control* **6**, 209-216 (1995).

44. National Center for Health Statistics, "Percent distribution of smoking status among adults aged 18 years and over, by sex: United States, 2001," Centers for Disease Control and Prevention, Department of Health and Human Services, http://www.cdc.gov/nchs/about/major/nhis/released200207/figures08_1-8_4.htm, *last accessed* (2002).

45. State and Territorial Air Pollution Program Administrators and the Association of Local Air Pollution Control Officials, "Cancer Risk from Diesel Particulate: National and Metropolitan Area Estimates for the United States," http://www.4cleanair.org/comments/cancerriskreport.pdf, *last accessed* (2000).

46. G. C. N. Nadler Lins, F. M. C. Campello de Sousa, "A protocol for the elicitation of prior distributions," Second Symposium on Imprecise Probabilities and Their Applications, Cornell University, 26-29 June 2001, http://ippserv.rug.ac.be/~isipta01/proceedings/009.html, *last accessed* (2001).