Uncertainty of Damage Cost Estimates

May 2003

Ari Rabl, PhD
Ecole des Mines, 60 boul. St.-Michel, 75272 Paris CEDEX 06, FRANCE
ari.rabl@ensmp.fr
http://www-cenerg.ensmp.fr/english/themes/index.html
click on Impacts of Energy Production

Sources of uncertainty

i) **data uncertainty**

e.g. slope of a dose-response function, cost of a day of restricted activity, and deposition velocity of a pollutant;

ii) **model uncertainty**

e.g. assumptions about causal links between a pollutant and a health impact, assumptions about form of a dose-response function (e.g. with or without threshold), and choice of models for atmospheric dispersion and chemistry;

iii) **uncertainty about policy and ethical choices**

e.g. discount rate for intergenerational costs, and value of statistical life;

iv) **uncertainty about the future**

e.g. the potential for reducing crop losses by the development of more resistant species;

v) **idiosyncrasies of the analyst**

e.g. interpretation of ambiguous or incomplete information, and human error.

*The **difficulties** begin with trying to prepare this list: the distinction between these sources is not always clear.*

Appropriate analysis

For data and model uncertainties:

analysis by statistical methods, combining the component uncertainties over the steps of the impact pathway, to obtain formal confidence intervals around a central estimate.

For ethical choice, uncertainty about the future, and subjective choices of the analyst:

sensitivity analysis, indicating how the results depend on these choices and on the scenarios for the future.

For human error:

be careful and guard against overconfidence.

Difficulties

Quantifying the sources of uncertainty in this field is problematic because of a general lack of information.

Usually one has to fall back on subjective judgment, preferably by the experts of the respective disciplines.

The uncertainties due to **strategic choices of the analyst**, e.g. which dose-response functions to include, are difficult to take into account in a formal uncertainty analysis.

$\Rightarrow$ the **comprehensive uncertainties** can be much larger than the ones that have been quantified (uncertainties due to data and parameters).

Uncertainty ≠ variability

Don’t confuse uncertainty and variability of impacts!

Both can cause estimates to change,

but in very different ways and for totally different reasons:

Uncertainty: insufficient knowledge at the present time,

future estimates may be different when we know more.

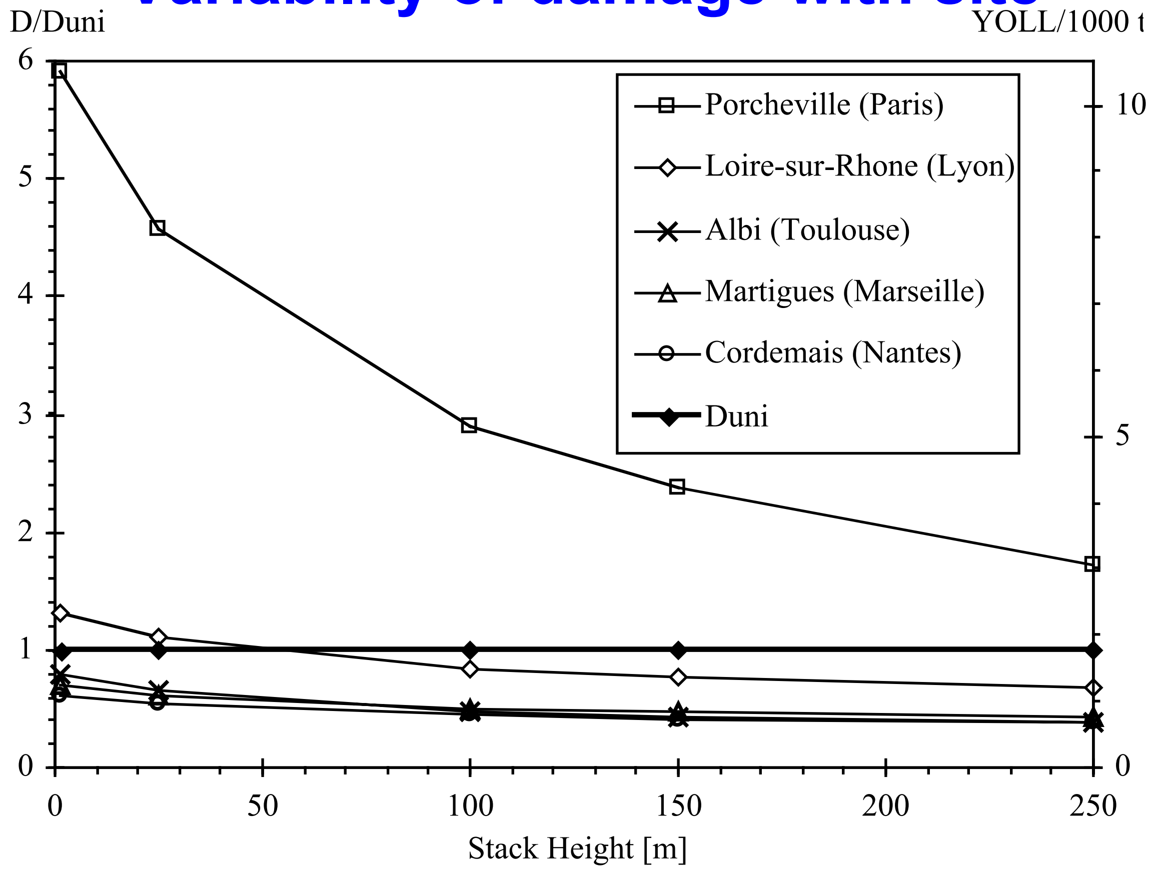
Variability: damage cost can vary with the type of source

(where, ground level or tall stacks, …).

Damage cost per kWh are proportional to the emissions and vary with the technologies used.

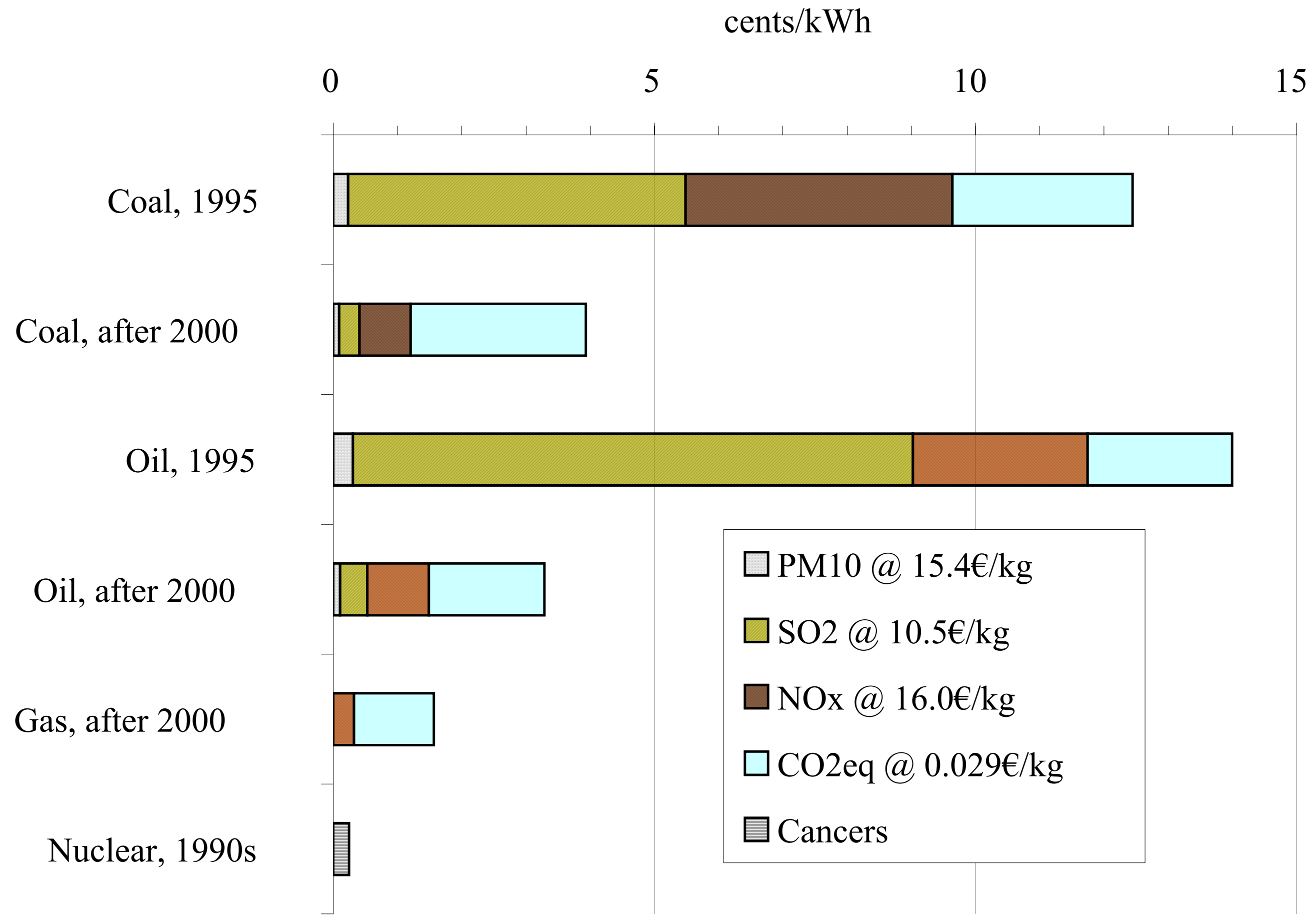
These variations are independent of the uncertainties.

Variability of damage with site



An example of dependence on site and on height of source for a primary pollutant: damage D from SO_2 emissions with linear C-R function, for five sites in France, in units of D_{uni} for uniform world model Eq.10 with ρ = 105 persons/km^2 (the nearest big city, 25 to 50 km away, is indicated in parentheses). The scale on the right indicates YOLL/yr (years of life lost) by acute mortality from a plant with emission 1000 ton/yr. Plume rise for typical power plant conditions is accounted for.

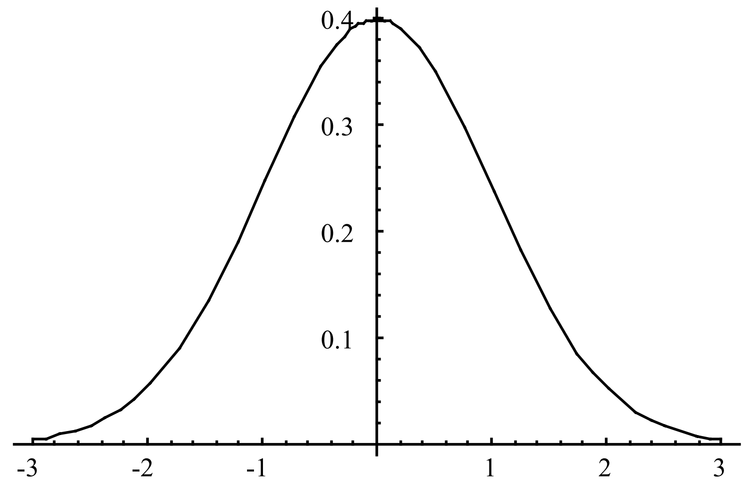
Variability of damage cost, €/kWh in France, due to reduced emissions



Probability distributions of parameter values

The most important characteristics are
the **mean** μ and the **standard deviation** σ

Example: the frequent case of a normal (=gaussian) distribution of $\mu = 0$ and $\sigma = 1$



Combination of errors for a sum of terms

If $y = x_1 + x_2 + \ldots + x_n$
is a sum of **uncorrelated** random variables x_i,
each with mean μ_i and standard deviation σ_i,
the uncertainty distribution of y has mean

$$\boldsymbol{\mu_y = \mu_1 + \mu_2 + \ldots + \mu_n},$$

and standard deviation σ_y given by

$$\boldsymbol{\sigma_y^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}.$$

Central limit theorem of statistics:
In the limit $n \to \infty$, the distribution of y approaches a normal distribution, even if the distributions of the individual x_i are not normal.

In practice the distribution of y is **often close to normal** even for very small n if the individual distributions (especially those with large widths) are not too far from normal.

Combination of errors for a product of terms

If $\mathbf{y = x_1\ x_2\ \ldots\ x_n}$,
the log of y is a sum
$\ln(y) = \ln(x_1) + \ln(x_2) + \ldots + \ln(x_n)$.
Let the x_i be uncorrelated random variables
with probability distributions $p_i(x_i)$.
Define the **geometric mean** $\mu_{\mathbf{gi}}$ of x_i by

$$\ln(\mu_{gi}) = \int p_i(x_i)\ln(x_i)\,dx_i$$

Then the geometric mean μ_{gy} of y is

$$\ln(\mu_{gy}) = \int p_y(x_y)\ln(y)\,dy$$

and it is equal to

$\mathbf{\mu_{gy} = \mu_{g1}\ \mu_{g2}\ \ldots \mu_{gn}}$.

Combination of errors for a product of terms, cont'd

Now define the **geometric standard deviation** $\boldsymbol{\sigma_{gi}}$ of x_i by

$$[\ln(\sigma_{gi})]^2 = \int p_i(x_i)[\ln(x_i) - \ln(\mu_{gi})]^2 \, dx_i$$

Then the geometric standard deviation σ_{gy} of y is given by

$$[\ln(\sigma_{gy})]^2 = [\ln(\sigma_{g1})]^2 + [\ln(\sigma_{g2})]^2 + \ldots + [\ln(\sigma_{gn})]^2 .$$

The central limit theorem implies that for large n the distribution of ln(y) is approximately normal, and in practice this approximation can be quite good even for small n if the distributions of the individual $\ln(x_i)$ are not too far from normal.

A variable whose log has a normal distribution is called **lognormal**. The distribution of a product is often approximately lognormal.

The lognormal distribution

To get the lognormal from the normal distribution $p_n(u) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(u-\mu)^2}{2\sigma^2}\right]$

Change variable u = ln(x). The normalization integral becomes

$$\int_0^\infty \frac{p_n(\ln(x))}{x} dx = 1$$

which allows interpreting the function

$$p_{\log n}(x) = \frac{p_n(\ln(x))}{x} = \frac{1}{\sigma x\sqrt{2\pi}} \exp\left[-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right]$$

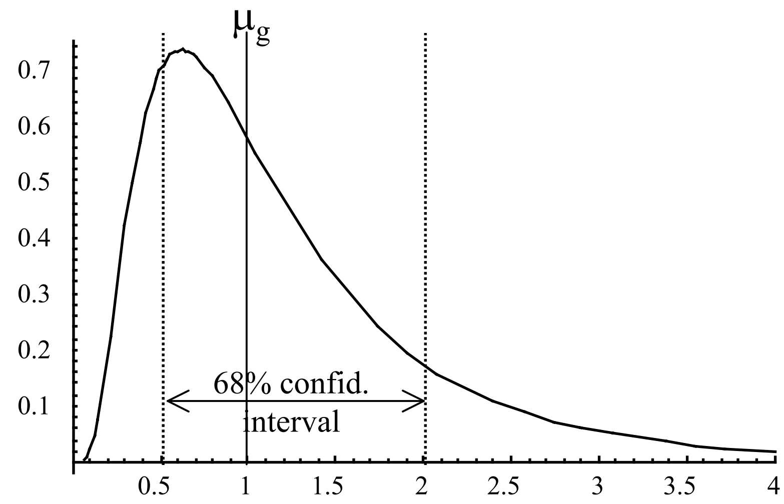
as the probability density of a new distribution between 0 and ∞. This is the **lognormal distribution**. The geometric mean μ_g and geometric standard deviation σ_g are related to μ and σ by

$\mu_g = \exp(\mu)$ or $\ln(\mu_g) = \mu$ and $\sigma_g = \exp(\sigma)$ or $\ln(\sigma_g) = \sigma$

The lognormal distribution, cont'd

The lognormal distribution is **asymmetric**, with a long tail and its mean is larger than its median.
Its **median is equal to μ_g**.

Example, with $\mu_g = 1$ and $\sigma_g = 2$



When plotted vs ln(x) it looks just like an ordinary normal.
For confidence intervals, note that
68% of the distribution is in the interval $[\mu_g/\sigma_g, \mu_g \sigma_g]$
and
95% of the distribution is in the interval $[\mu_g/\sigma_g^2, \mu_g \sigma_g^2]$.

Combination of errors for a general function of terms

For a sum the standard deviation, **and for a product** the standard geometric deviation, can be calculated exactly, **regardless how wide the distributions of the individual terms**. Furthermore, this often determines the entire distribution approximately (central limit theorem). ⇒ **Explicit closed form solution!**

But for general functions $y = f(x_1, x_2, \ldots, x_n)$ a closed form solution can be obtained only in the limit of small uncertainties (narrow distributions)

$$\sigma_y^2 = \left|\frac{\partial f}{\partial x_1}\right|^2 \sigma_{x1}^2 + \left|\frac{\partial f}{\partial x_2}\right|^2 \sigma_{x2}^2 + \ldots \left|\frac{\partial f}{\partial xn}\right|^2 \sigma_{xn}^2$$

This is not appropriate for the very large uncertainties involved in damage costs.

⇒ Use **Monte Carlo** approach, i.e. perform a very large number of numerical simulations, each calculating the result y for a specific choice of $\{x_1, x_2, \ldots x_n\}$, and look at the resulting distribution of the y.

Impact pathway analysis for calculation of damage cost

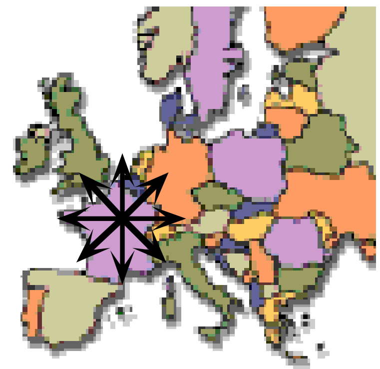
SOURCE
(specification of site and technology)

⇒ **emission**
(e.g., kg/yr of particulates)

↓

DISPERSION
(e.g. atmospheric dispersion model)

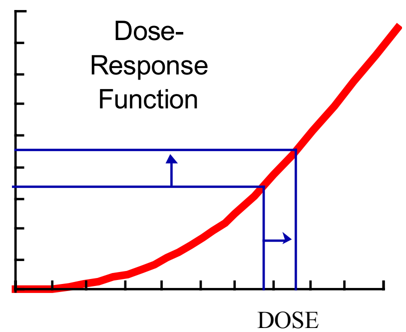
⇒ **increase in concentration at receptor sites**
(e.g., $\mu g/m^3$ of particulates in all affected regions)



↓

DOSE-RESPONSE FUNCTION
(or concentration-response function)

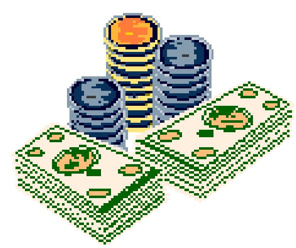
⇒ **impact**
(e.g., cases of asthma due to ambient concentration of particulates)



↓

MONETARY VALUATION

⇒ **cost**
(e.g., cost of asthma)



Formula for calculation of damage cost

The **damage cost C** [€/yr] of a particular impact caused by a continuous **emission** of a pollutant at rate **q** [kg/yr] is the integral over the entire region where receptors (people, agricultural crops, buildings, …) are affected by the pollution source

$$C = \int dx\, dy\; p(\mathbf{x})\; s_{CR}(\mathbf{x})\; \rho(\mathbf{x})\; c(\mathbf{x}, q)$$

where

$\mathbf{x} = (x,y)$ is a point in the impact region,

$p(\mathbf{x})$ = unit cost ("price") of the impact that is evaluated [€/case],

$s_{CR}(\mathbf{x})$ = slope of concentration-response function [cases/(yr·μg/m^3)],

$\rho(\mathbf{x})$ = receptor density [receptors/m^2], and

$c(\mathbf{x}, q)$ = concentration of pollutant in air at $\mathbf{x}$ [μg/m^3)].

In most cases $p(\mathbf{x})$ does not vary with $\mathbf{x}$ and $s_{CR}(\mathbf{x})$ is also approximately constant. With constant p and s_{CR} the equation becomes a **product of three terms**

$$C = p\;\; s_{CR} \int dx\, dy\; \rho(\mathbf{x})\; c(\mathbf{x}, q)$$

each of which can make a large contribution to the total uncertainty. Thus it suffices to carry out a Monte Carlo analysis of the integral to determine its geometric standard deviation; then the overall geometric standard deviation can be found by the rule for products.

The "Uniform World Model" (UWM)

The concentration can be related to the rate at which the pollutant is removed from the atmosphere by wet deposition, dry deposition and/or decay or chemical transformation. This rate can be defined as a flux F_{dep} [$\mu g/(m^2 \cdot s)$]; it is the product of concentration and removal velocity v_{dep}

$$F_{dep}(\mathbf{x},q) = c(\mathbf{x},q)\ v_{dep}(\mathbf{x})$$

Further simplification is possible if $v_{dep}(\mathbf{x})$ and $\rho(\mathbf{x})$ do not vary with x. Replacing $v_{dep}(\mathbf{x})$ and $\rho(\mathbf{x})$ by constants one obtains

$$C = (p\ \ s_{CR}\ \rho/v_{dep}) \int dx\ dy\ F_{dep}(\mathbf{x}, q)$$

By conservation of mass the integral of the removal flux equals the emission rate q. Thus the damage cost is simply a product of these factors

$$C = p\ \ s_{CR}\ \rho\ q/v_{dep}\ .$$

This is the "uniform world model" (UWM).

The “Uniform World Model” (UWM), cont’d

The **UWM** can easily be **generalized to secondary pollutants**. Indicating quantities referring to the primary pollutant by the subscript 1 and the secondary by 2, one can show that the damage cost C_2 of the secondary pollutant due to the emission q_1 of the primary pollutant is

$$C_2 = p_2 \; s_{CR2} \, \rho \, q_1 \, v_{1\text{-}2}/(v_{dep1} \, v_{dep2})$$

where $v_{1\text{-}2}$ is the transformation velocity of the primary to the secondary pollutant, defined as $v_{1\text{-}2} = F_{1\text{-}2}/ c_1$ where $F_{1\text{-}2}$ is the transformation flux [$\mu g/(m^2 \cdot s)$].

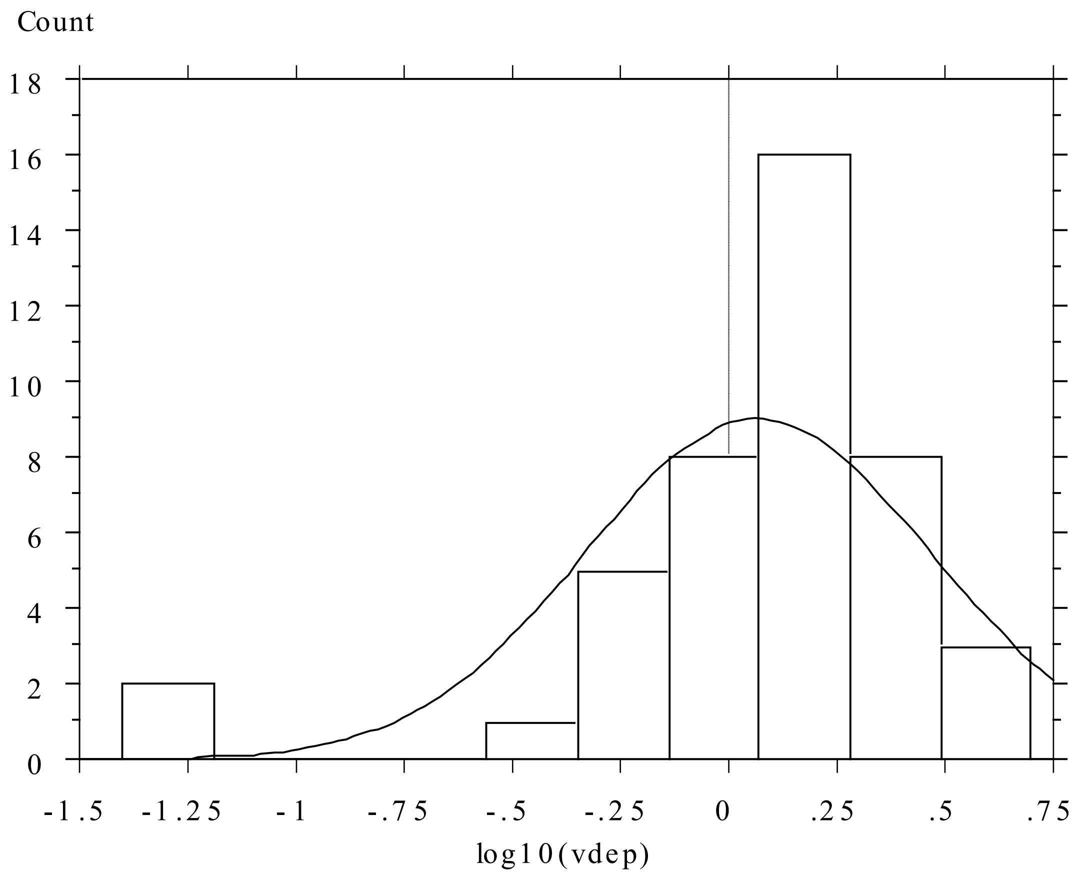
Comparison of UWM with detailed site specific calculations for about a hundred installations in many countries of Europe, as well as China, Thailand and Brazil: **UWM** is so close to the average that it can be **recommended for typical damage costs** for emissions from tall stacks (>~ 50 m); **for specific sites the agreement is usually within a factor of two to three.**

Note: typical values = average over emission sites, equivalent to averaging over receptor distributions $\Rightarrow \rho \sim$ *uniform.*

Uncertainty of the components

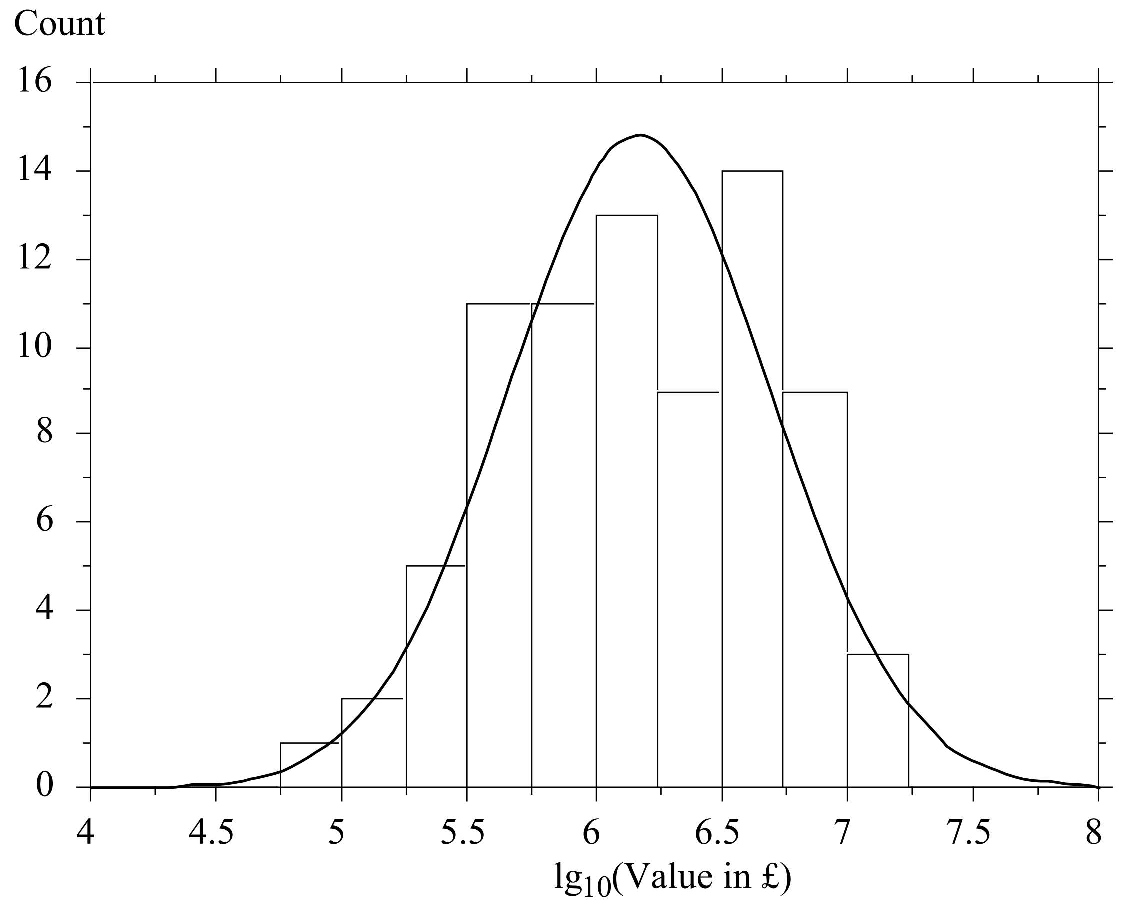
For Dispersion: deposition velocities.

An example: Distribution and lognormal fit of maximum values, in the review of Sehmel [1980], for dry deposition velocity [in cm/s] of SO_2 over different surfaces.



Uncertainty of the components, cont'd

Example of lognormal distribution for monetary valuation: **"value of statistical life",** in $£_{1990}$, in 78 studies reviewed by Ives, Kemp and Thieme [1993], histogram and lognormal fit.



μ_g = 1.5 M£

σ_g = 3.4

Uncertainty of the components, cont'd

Epidemiological studies report their errors as 95% confidence intervals, usually approximately symmetric about the mean; that corresponds to two ordinary standard deviations, 2 σ.

For a simplified uncertainty analysis in terms of lognormal distributions and geometric standard deviations, it is convenient to express the reported symmetric confidence intervals as approximately equivalent geometric standard deviations. Assume that the one-standard deviation (68% probability) interval $[\mu - \sigma, \mu + \sigma]$ corresponds to the interval $[\mu_g/\sigma_g, \mu_g \sigma_g]$ of the lognormal distribution. Thus

$$\sigma_g = \sqrt{\frac{\mu + \sigma}{\mu - \sigma}}$$

Some uncertainty estimates

Example: **uncertainty of mortality cost** due to PM emissions (if emission has been specified in terms of TSP rather than PM_{10}). *Note that the calculation of* s_{CR} *and p is broken down into several factors.*

Component	**lognormal?**	**sigG**	**ln(sigGi)^2**
Emission (TSP/PMx)	probably yes	1.4	0.113
Dispersion	yes	2	0.480
sCR regression	no	1.3	0.069
sCR transfer (composition)	?	1.5	0.164
YOLL, given RelRisk	probably yes	1.5	0.164
VSL (value of statistical life)	yes	2	0.480
Value of YOLL, given VSL	?	1.5	0.164
latency&discount rate	probably yes	1.2	0.033
Total		**3.64**	**3.64**

Recall $[\ln(\sigma_{gy})]^2 = [\ln(\sigma_{g1})]^2 + [\ln(\sigma_{g2})]^2 + \ldots + [\ln(\sigma_{gn})]^2$

Some uncertainty estimates, cont'd

For cancers due to radionuclide emissions

Component	sigG	ln(sigGi)^2
Emission data	1.5	0.164
Dispersion	1.8	0.345
factor for noninhalation pathways	2	0.480
dose-response function (unit risk)	2	0.480
dose and dose-rate effectiveness factor	1.5	0.164
fraction that is fatal	1.3	0.069
YOLL (years of life lost)	1.5	0.164
VSL (value of statistical life)	2	0.480
Value of YOLL, given VSL	1.3	0.069
latency&discount rate	1.3	0.069
Total	**4.84**	**4.84**

Some uncertainty estimates, cont'd

Uncertainty of damage costs for buildings and for agricultural losses.

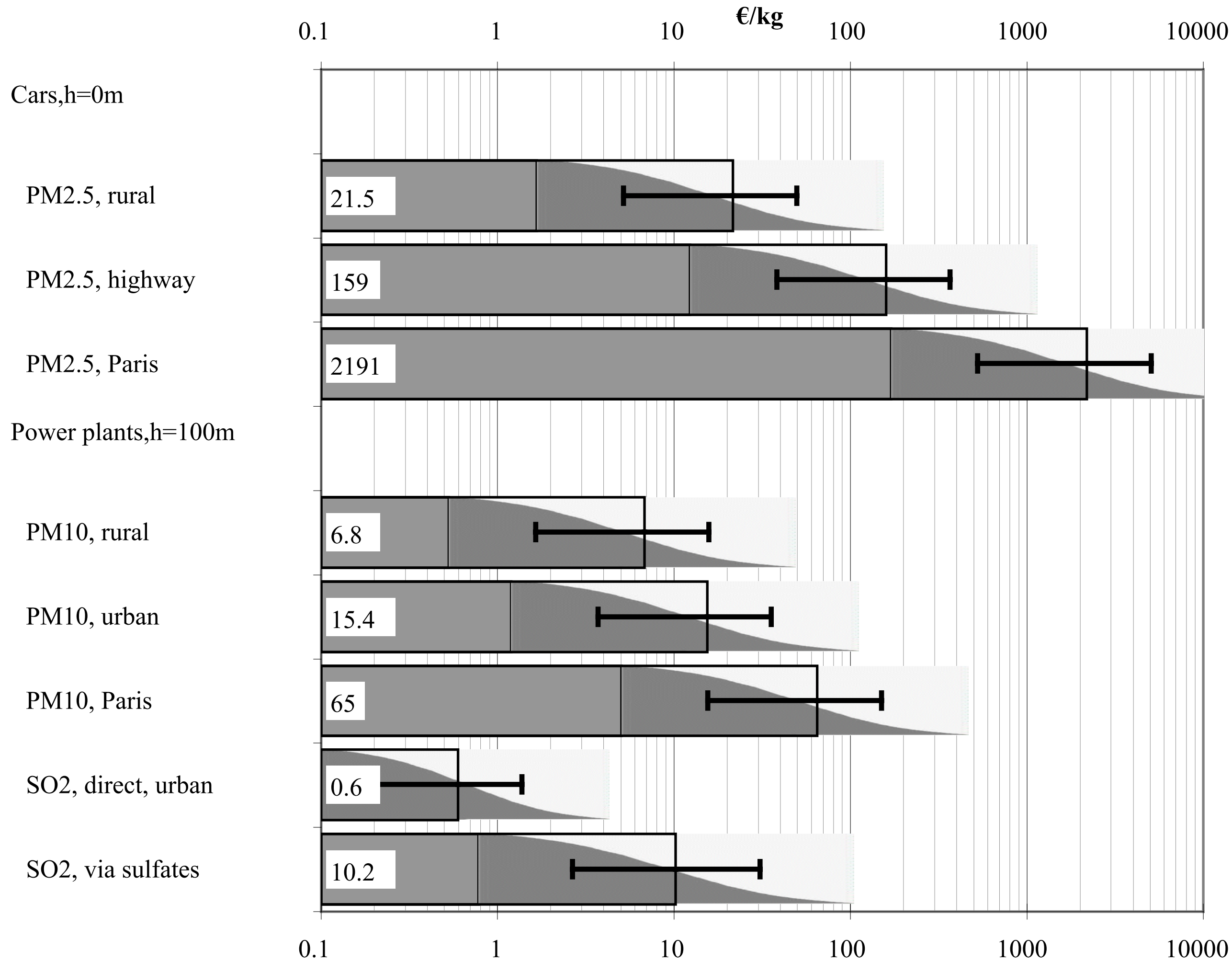
	Lognormality	buildings due to SO_2	crop losses due to NO_2 via O_3
Emission	approximately	1.2	1.2
Dispersion	yes	2	3
f_{CR} regression	no	1.3	1.3
f_{CR} other	?	2	2
inventory	?	2	1.2
cost per m^2 of building surface	?	1.3	
repair frequency	probably yes	2	
value of crops	probably yes		1.2
Total		**4.2**	**3.9**

Monte Carlo $\leftrightarrow$ simplified analysis

Monte Carlo	Simplified analysis
No limit on accuracy	Approximate
Can handle any combination of errors sources and distributions	Only for products (sums); OK only if the distributions with the largest widths are not too far from lognormal (normal)
Requires large number of computer calculations	Simple hand calculation
Opaque	Transparent

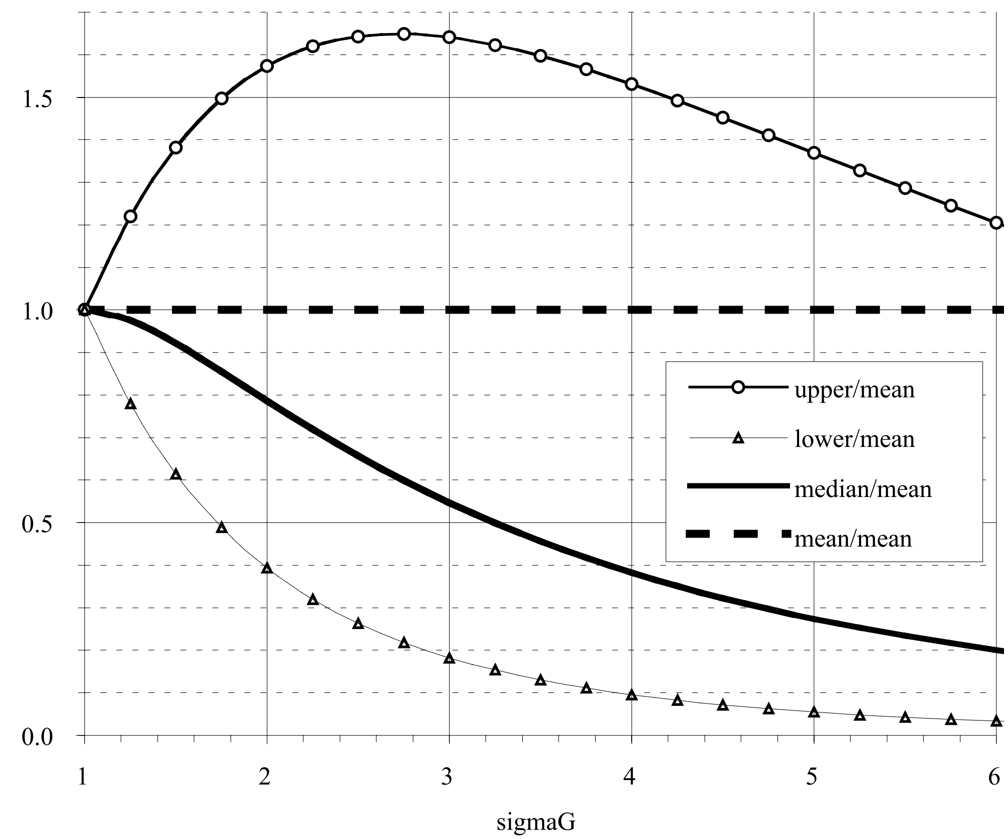
For damage costs the two approaches are complementary

Presentation of uncertainty

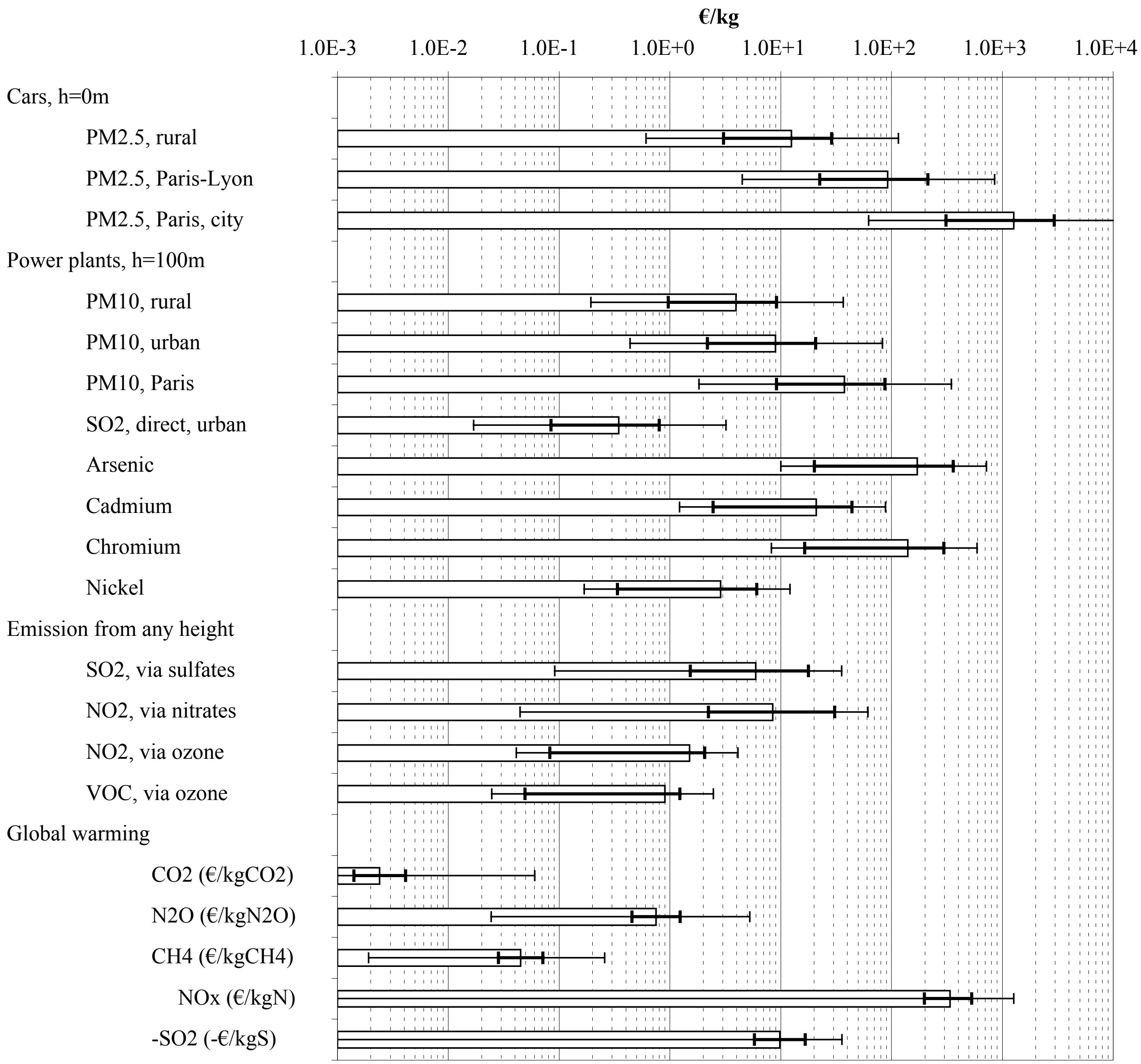


Mean and median

For the lognormal distribution the **confidence intervals in a log plot** are **symmetric about** the geometric mean (=**median** of lognormal distrib.), not the mean. But **damage cost estimates are means**, not medians.



Quantified uncertainties and comprehensive uncertainties (subjective)



Note: these numbers are from ExternE 2000, significantly lower than those of ExternE 1998 shown elsewhere in these notes.

Conclusions

The **uncertainties** of damage costs are **very large**,

- typically geometric standard deviations σ_g **around 3 to 5**
- σ_g around 3 for primary pollutants PM, NO_x,SO_2;
- somewhat larger for secondary pollutants, especially O_3, than for primary pollutants;
- very large for greenhouse gases, σ_g around 5.

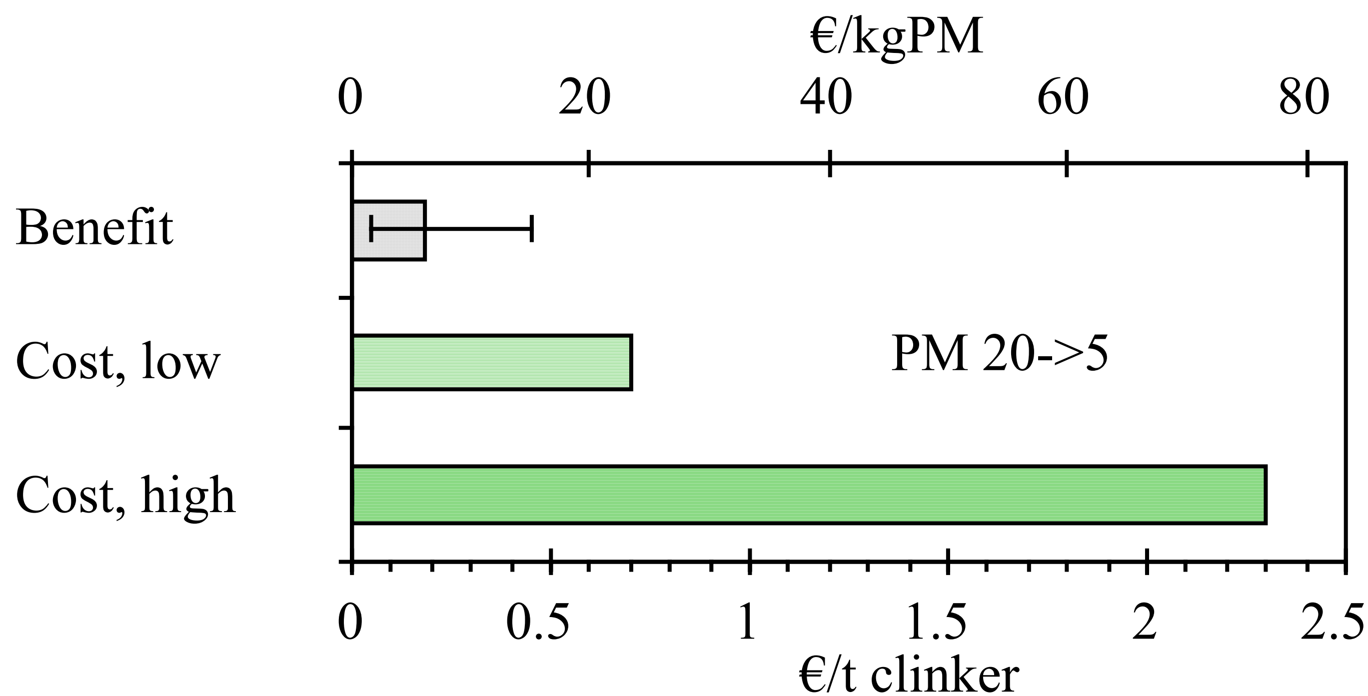
Comprehensive uncertainties can be much larger than the ones that have been quantified, but they involve a great deal of subjective judgment with little or no factual basis.

… but even an estimate with large uncertainty is better than none at all

Conclusions, cont'd

Effect of uncertainty depends on the use of the cost estimate.

Example: was the reduction of PM emission limits from 20 to 5 mg/m^3, proposed for the Directive of 2000 for cement kilns that co-incinerate waste as fuel, justified?



Answer: no, even in view of the uncertainties.