PEER REVIEW OF THE METHODOLOGY OF COST-BENEFIT ANALYSIS OF THE CLEAN AIR FOR EUROPE PROGRAMME

Alan Krupnick, Editor and Contributor
Bart Ostro and Keith Bull, Contributors

As part of the Clean Air for Europe Programme (CAFE), DG Environment selected Alan Krupnick, Bart Ostro and Keith Bull to review the cost-benefit analysis methodology that was proposed to be used in the programme. Specifically, they were to examine to what extent the methodology is based on scientific knowledge and fit for the purposes such analyses will be used for, namely to guide policy development under this programme (Annex 1). This methodology covers the estimation of the effects of air pollution on health, materials, agriculture, ecosystems and cultural sites, and the valuation of such effects in carrying out welfare and macroeconomic analyses. In addition, the reviewers were to assess whether the plans for aggregating results to compare costs and benefits are reasonable, whether suggestions for sensitivity analyses are reasonable, and the extent to which multi-criteria assessment analysis and social effects are feasible and provide value-added. In the course of the reviews, they were also to make suggestions on how the CBA methodology could be improved in the short and medium term.

The reviews have been based primarily on *Methodology for the Cost-Benefit analysis for CAFE Consultation – Issue 3 – July 2004, Draft for Consultation and Peer Review,* by a consortium lead by AEA Technology Environment. In addition, the reviewers were sent stakeholder comments from a meeting on the CAFE CBA methodology on 16 July 2004. Each reviewer was responsible for certain tasks and worked on their own. Alan Krupnick was responsible for coordination of the effort and the integration of the reports into a coherent whole. He also led the effort to build a consensus on the overall report and on components that were not assigned to any one individual for review.

Reviewers, their affiliations and their responsibilities were as follows:

Alan Krupnick, PhD, Resources for the Future: Overall Coordination and Team Leader; Valuation of affected endpoints

Bart Ostro, PhD, California Office of Environmental Health Hazard Assessment: Health Effects estimation

Keith Bull, Ph.D., UNECE, Secretariat for the Convention for Long-Range Transboundary Air Pollution: Estimation of all non-health effects

Plan of the Report

Chapter 2 covers general comments on the overall CAFE CBA methodology. Chapters 3, 4 and 5 are the heart of the report covering comments on methodologies for estimating health effects, non-health effects and monetary values for these. Chapter 6 covers the approach for addressing uncertainties. Chapter 7 comments on the plan for carrying out multi-criteria assessment and Chapter 8 provides comments on miscellaneous elements of the CAFE CBA methodology. For chapters where the group is responsible (Chapters 2, 7 and 8), we use the term "we." For chapters 3, 4 and 5, primary authorship is identified; thus we use the term "I."

Chapter 2
General Comments

The overall methodology follows the *damage function* (what some in the EU have coined as the "impact-pathway") approach. This approach is the approach of choice in addressing air pollution problems and policies using cost-benefit analysis in developed countries around the world. The EXTERNE project, supported by the European Union's DG Research since its inception in the early 1990's and by other groups more recently, applied and pioneered many aspects of the damage function approach (ExternE, 1999; ExternE, 1996). Clearly, the CAFE CBA methodology is based directly on the EXTERNE project and uses many of the same fine and respected researchers. Models using the damage function approach are common in the U.S., the most recent being the USEPA's BENMAP model (Abt Associates, 1993) and the most notable in this context being the effort by RFF and Oak Ridge National Lab to estimate the Social Cost of Electricity (Lee, et al., 1995), a project supported by the US Department of Energy and designed in tandem with the EXTERNE team. Recent examples of cost-benefit analyses in the U.S. of the type appropriate for the EU to consider include: the Regulatory Impact Analyses (RIAs) for the PM and ozone National Ambient Air Quality Standards (NAAQS) (1997) and the Off-Road Diesel Standard (2004)). These were also reviewed by the National Academy of Sciences (2002)

Despite many inherent uncertainties and subjective decisions, we consider the proposed methodology a reasonable basis for generating information to policy makers and the public about the expected health benefits from improved air quality. As proposed, the analysis can help identify the type, general magnitude and relative importance of the health and welfare improvements that might be expected from decreased emissions and concentrations of air pollution. In addition, the analysis will be useful in identifying the critical factors in the assessment of impacts, the sensitivity of the results to certain assumptions, and the key uncertainties. The impact assessment, however, should only be one part of the informational package used in the deliberative and policy assessment. There are other issues – technologic feasibility, resource and political constraints, equity and environmental justice issues – that may also be relevant.
Beyond the general endorsement of this approach by the panel, the reviewers like very much the term "stock at risk," because it is more general than the currently used term in the U.S. "target population." The emphasis on performing uncertainty assessment on both the benefit and cost sides of the effort is also noteworthy because to date uncertainty analyses have been confined to the benefits side. The lack of specificity concerning the uncertainty analysis needs to be remedied, however.

The overall framework is not above criticism. The interface between the RAINS model and the Benefits Analysis, as well as how the general equilibrium modelling (efforts fit into these models is not at all clear from the report.

A final, minor point. The framework as described on pg. iii is confusing, as it shows RAINS feeding into scenario development, when the direction should be the other way, or both ways if the process is to be iterative.

References

Abt Associates Inc. 2003. U.S. Version BenMAP: Environmental Benefits Mapping and Analysis Program User's Manual. (November). Prepared for Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency. Research Triangle Park, NC. http://www.epa.gov/ttn/ecas/benmapdownload.html

ExternE. 1999. Externalities of Energy, Volume 7: Methodology 1998 Update. (Brussels: European Commission), 518 pp.

ExternE. 1996. Externalities of Energy, prepared by ETSU, UK and Metroeconomica, LTD. UK for European Commission, DG XII EUR 16521 EN.

Lee et al. 1995. *Estimating Externalities of Electricity Fuel Cycles* (Oak Ridge National Laboratories/Resources for the Future), 1994-1995, seven volumes, McGraw-Hill/Utility Data Institute, Washington, D.C.

National Academy of Sciences. 2002. *Estimating the Public Health Benefits of Proposed Air Pollution Regulations*.

USEPA. 2004. *Final Regulatory Analysis: Control of Emissions from Nonroad Diesel Engines*. EPA 420-R-04-007. Office of Transportation and Air Quality. Washington DC: US EPA

USEPA. 1997.*Regulatory Impact Analysis for Ozone and Particulate Ambient Air Quality Standards.* Washington, DC: EPA.

Chapter 3
Estimating Health Effects
By Bart Ostro

Below, I comment on the general methodology, whether it is based on current scientific knowledge (based on the Terms of Reference), and if it is "fit for purposes". Next, I comment on the outcomes assessed and the study selection, in the order in which they are presented.

General methodology, whether it is based on scientific knowledge, and if it is "fit for purposes

Over the last several years, several multidisciplinary and international committees have reviewed the methodology similar to that being proposed. For example, the World Health Organization (WHO, 2001), the National Academy of Science (NRC, 2002) and the independent Science Advisory Board for the U.S. Environmental Protection Agency (SAB, 2004) have all reviewed this methodology. The consensus regarding estimation for particulate matter (PM) and ozone has been that the general methodology is based on and supported by current scientific knowledge regarding the health effects of air pollution.

In addition, the reviewers have indicated that the methodology, although uncertain and often sensitive to some crucial assumptions, provides a reasonable basis for the purposes of quantitative estimation of health effects. Therefore, it is "fit for purpose". Most of the estimates are based on epidemiologic studies. Over the last decade these studies have provided direct evidence of adverse health outcomes associated with both short-term (daily or weekly) and long-term (annual or multi-year) exposures. These studies are now also provided mechanistic information about these effects and helping to address the issue of biological mechanism and causality. For example, recent epidemiologic studies have reported associations between air pollution and heart rate variability, arrhythmias, and several markers of inflammation including subclinical atherosclerosis (carotid intima media thickness).

Overall, therefore, the proposed methodology is a reasonable framework for conducting health benefits analysis and appears to be based on the latest available science. However, as indicated below, many issues remain unresolved and await further specification.

Health Outcomes and Study Selection

The overriding issue regarding the development of concentration-response (CR) functions should be that the selection criteria used for studies be clear, balanced, and consistent across endpoints. The role of subjective choice in the selection should be acknowledged and enough information provided in the document so others can appreciate the thought process that went into the selection. Additional sensitivity analyses for the factors having the largest impacts on the overall results are usually warranted and to the extent that uncertainties, in addition to the normal sampling uncertainty, can be characterized and quantified, the final product can be enhanced.

A good model for this can be found in the ozone benefits analysis conducted by Levy et al. (2001). Although one might not agree with all of the decisions by these authors, the criteria and decisions are carefully explained. This allows others to appreciate the arguments but alter the assumptions if they wish.

PM mortality

The authors propose using an estimate of about a 6% increase in mortality hazard rates per 10 $\mu g/m^3$ based on results from a prospective cohort study, Pope et al. (2002). This estimate is obtained by using all years of exposure (1979-83 and 1999-2000) and is applied to all-cause mortality. It is also proposed to apply this function to anthropogenic PM with no threshold. Finally, it is proposed that to avoid double counting no estimates from the time-series studies will be provided.

Firstly, I support the proposed reliance on the prospective cohort studies as the primary basis for estimating mortality effects related to air pollution, in this case measured as PM. These studies evaluate the health effects in a specific population over a period of years. Compared with time series studies, which provide estimates of health effects due to recent exposure, the cohort studies give a more complete assessment of the impact of air pollution since it includes long-term, cumulative effects. Furthermore, these studies provide information necessary to estimate the number of life-years lost in populations, not just the number of premature deaths, thereby allowing for several valuation methods to be used.

I find the use of this estimate appropriate given the results of Pope et al. (2002). In response to the comments of CONCAWE (8/11/2004), it is important to note that this study was used in the WHO Global Burden of Disease (GBD) effort, supported by the National Academy of Science committee examining the estimation of health benefits from pollution control, and by U.S.EPA's Science Advisory Board. Although there are few cohort studies using ambient monitoring in Europe, for several reasons, it is reasonable to extrapolate the results from a sample of roughly 500,000 individuals in 50 U.S. cities to Europe. The mix of PM will be approximately the same, as are housing stock, seasonality, lifestyle, range of weather conditions, background health status, etc. It should be noted, however, that higher estimates are obtained from Dockery et al. (1993) and lower estimates are obtained from Abbey et al. (1999). In addition, traffic-based effects on mortality have been developed by Hoek et al. (2002). Also, there is preliminary evidence that higher effects occur when the population is less mobile (Jarrett, 2004), which presumably would fit the European situation. Therefore, the assessment could include some subjective weighting of these studies to obtain a central estimate and confidence bound even if the weight is "one" for Pope et al. (2002) and zero for others. This could be made more explicit.

In addition, it is of note that recent research (only presented in conference abstracts) support higher estimates. For example, Jarrett et al. using the American Cancer Society (ACS) cohort for the Los Angeles air basin (presented at HEI) and Miller et al. (ISEE, Epi S28) using the Women's Health Initiative Observational Study (WHI-OS) both reported effect estimates closer to that of Dockery et al. -- **some 3 or 4 times larger than the Pope et al. (2002) estimates**. The WHI-OS involved 67,000 postmenopausal women from throughout the U.S. with no history of heart disease at inception. A related question that arises from this is whether this project will utilize evidence from the "grey" literature or only from published studies.

I do not concur with the comment #3a of van Bree and Buringh (17 Aug 2004) where they recommend using a conservative estimate of no effect. Given the number of new cohort studies coming on line, a zero estimate is not sufficiently representative of the current science. To argue that it is one would have to argue that the populations and conditions in Europe are so different from that of the entire U.S., that extrapolation of cohort estimates would be clearly wrong. Given the similar results in the U.S. and Europe for time-series mortality, hospitalization, asthma and respiratory symptoms, this would be a difficult case to make. However, I do agree with their recommendations #3c,d, and e regarding discussion of uncertainties in application of these studies, calculating the impacts from the time-series studies and explicit description of the life table calculations.

Several questions arise concerning the use of thresholds. The current proposal is to generate benefits down to an anthropogenic PM concentration with no threshold. It is correct to indicate that for all-cause and cardiovascular mortality there is no evidence of a threshold at the population level and that the CR function looks fairly linear. However, different analysts have used varying assumptions about the low end of the extrapolation. Some analysts, like those for the Global Burden of Disease effort for WHO used a low end of the PM2.5 distribution from the Pope study of around **7 μg/m^3** as the low end for extrapolation. This conservative approach was taken since the extrapolation was to be for all countries in the world and therefore, the assumptions for extrapolation from the U.S. are more extreme. Others have carried benefits down to an assumed background PM2.5 concentration of approximately **2.5 μg/m^3**.

Contrary to the comments from CONCAWE (8/11/2004), both the recent analysis of the ACS cohort by Pope et al. (2002) and the reanalysis of the earlier study by Kreswki et al. (2000) suggest that a linear, no-threshold model provides a good fit of the data. For example, in the analysis of 16 years of data on approximately 350,000 people, Pope et al. (2002) state: Goodness-of fit tests indicated that the associations [for all-cause, cardiopulmonary and lung cancer mortality] were not significantly different from linear associations ($P > 0.20$)."

Additional evidence for a non-threshold assumption is provided by the time-series studies and the fact that some of the deaths from the cohort study are related to relatively recent and short-term exposures. Many of the time-series studies include very low concentrations of PM; some studies have mean PM10 concentrations below 15 μg/m^3 which are roughly equivalent to about 7.5 μg/m^3 PM2.5. So, two questions arise:

(1) Will the analysts consider an admittedly conservative approach of using 7 µg/m3 and
(2) I if benefits are extrapolated down to pollution background levels, what concentration will be used for background level and will it be the same for every country/region?

My recommendation, given the current science, is to extrapolate down to background concentrations, attempt to estimate different backgrounds for different parts of Europe as appropriate, and possibly conduct a sensitivity analysis of what happens if the benefits end at 7 $\mu g/m^3$ and no lower. In addition, beside calculating life-years gained, I agree that some estimates of the number of cases of premature mortality reduced be provided as well.

Finally, the proposal indicates that no separate quantitative estimates will be provided for mortality based on the short-term exposure, time-series studies. At this point, I would still recommend that the quantification implication of these studies be provided for informational purposes, with the understanding that these effects be not added to those based on the cohort studies. Since there are few cohort studies conducted in Europe that provide CR functions based on ambient monitors, and since there have been many time-series mortality studies conducted in Europe, it seems reasonable that estimates be generated based on the latter studies. Some may find these studies more plausible since no extrapolation from the U.S. is necessary. If this suggestion is taken up, I would also suggest that the study team examine in detail some of the assumptions of the WHO meta-analysis. For example, I think that they only considered the same lag in every study. If different cities have different lag structures due to the nature of PM or the host population in that area, this would lead to an underestimate of the effects.

In addition, nothing is said about cumulative exposures over five or seven days which have been shown in U.S. studies to double or triple the effect estimates relative to those based on a single day's exposure. Basing such estimates on PM10, rather than PM2.5 would ensure that effects of coarse particles were included in the estimate. This was a concern emphasized by Larry Gephart, identified as speaking on behalf of industry representatives at the expert workshop on 16th July 2004.

There is also no clear decision in the proposed methodology about the "cessation lag" – that is, how quickly will the benefits of long-term exposure reduction occur. The U.S. EPA in their assessment of benefits has varied their assumptions over the previous five years. At one point the benefits were assumed to occur within a distributed lag of 5 years. They have also considered, and may adopt, a scenario in which some percent of the benefits (roughly 20to 30%) occur within the first 2 years, 40 to 50% occur within 2 to 5 years and the rest occur in a lagged fashion up to 30 years. However, new research presented as an abstract at the ISEE meetings by Schwartz and Laden (#376) and Roosli et al. (#493) suggested that a lot of the relevant exposure occurs within the year before death. This suggests that the benefits, therefore, would also occur quite quickly. It is not clear what assumptions will be made for the current analysis. Appendix 5 does broadly discuss this issue. I think the assumptions concerning this issue would be greatly enhanced by a careful review of the existing literature on heart disease, lung function and lung cancer, regarding the reversibility of effects after exposure (usually smoking) ends. Since PM2.5 contains some of the same constituents as tobacco smoke, some useful information can be gleaned from this literature to sharpen the assumptions about the lag.

I concur with the study team's discussion regarding SO2 in Appendix 5. Both the study team and the original authors indicate that it is unlikely that the effects are related to SO2 *per se.* SO2 may be a proxy for regional effects or even other pollutants but researchers in the field have indicated that it is extremely unlikely that SO2 itself is the causal agent. Also, as mentioned, SO2 was also found to be associated with non-cardiopulmonary mortality – again, indicating that it may be proxying some non-pollution effect. I would recommend against estimating SO2 effects related to long-term exposure and against using the Krewski models where ecological variables and SO2 were included. Rather, I prefer the more recent and larger follow-up results reported by Pope et al. (2002).

Regarding the "static" approach outlined in Appendix 5, it would perhaps be more appropriate to talk about "premature" cases rather than simply number of deaths. The study team should also note that if long-term exposure is pushing people into the frailty pool, that each year new people may move into the pool to "replace" those who have perished. Thus, one could argue that it may not be the case that the pool is being diminished and as a result, it may be that in future years, the effects will not be lower.

As an aside, it would be useful for the study team to indicate how they will use, if at all, the recently conducted expert solicitation from the U.S. on this subject and the forthcoming solicitation of experts in Europe.

Acute exposure mortality from ozone

The authors propose using a risk estimate for ozone based on the WHO-sponsored meta-analysis of European time-series studies (Anderson et al, 2004). An estimate of 0.3% per 10 $\mu g/m^3$ ozone (8 hour) is suggested, with an implicit threshold of 35 ppb (8-hour average). Impacts will be expressed as number of cases. Later, the number of deaths will be expressed in terms of life years lost assuming an average loss of life of 6 months.

In general, I support the use of these WHO-based estimates. It includes 15 cities with a wide range of climates, co-pollutants, seasonal patterns, health care provision and background health conditions. Moreover, the suggested magnitude of effect is very similar to those obtained in meta-analyses by Levy et al. (2001) and Steib et al. (2002). Specifically, the WHO meta-estimates indicated a relative risk of 1.003 (95% CI = 1.001 – 1.004) for a 10 $\mu g/m^3$ change in 8-hour ozone. For standard atmospheric pressure and temperature, 1 ppb ozone equals 1.96 $\mu g/m^3$. In addition, the average ratio between 1-hour and 8-hour ozone is 1.33 (Schwartz 1997).

Making the conversions, the WHO estimate implies a 0.44% change in daily mortality (95% CI = 0.15 – 0.59%) per 10 ppb change in 1-hour maximum ozone. The meta-analysis by Levy et al. (2001) began with 50 time-series analyses from 39 published articles. A set of very strict inclusion criteria was applied, which eliminated all but four studies. Reasons for exclusion included: studies outside the US, use of linear temperature terms (versus non-linear and better modeled temperature), lack of quantitative estimates, and failure to include particulate matter (PM) in the regression models. Ultimately, their analysis generated an estimate of 0.5% (95% CI = 0.3 – 0.7%) per 10 $\mu g/m^3$ change in 24-hour average ozone. Based on the ratio between 24-hour average and 1-hour daily maximum ozone concentration of 0.4, this converts to a 0.39% change in daily mortality per 10 ppb change in daily 1-hour maximum ozone (95% CI= 0.24-0.55%). If the criteria are loosened to include eleven more studies, the pooled estimate decreases to 0.31% per 10 ppb change in 1-hour ozone. Stieb et al.

(2002) also reported a similar effect estimate (0.51% per 10 ppb change in daily 1-hour maximum ozone). Therefore, based on the currently published data, the WHO analysis provides a reasonable estimate of the effect of ozone.

In contrast, a lower effect estimate is provided by the National Morbidity, Mortality, and Air Pollution Study (NMMAPS). The revised analysis of this large study, conducted in 90 US cities (Dominici et al. 2003), found an effect estimate of 0.17% per 10 ppb change in 1-hour maximum ozone after conversion from the 24-hour average reported in the published study. This estimate is similar to the lower bound of the WHO estimate. The NMMAPS study may underestimate the impact of mortality due to the modeling methodology used to control weather factors. Specifically, this effort included four different controls for temperature, where most other times-series analyses used only two or modeled extreme weather events more carefully. In comparing their results for a given city with studies of individual cities by other researchers, the NMMAPS results are usually lower. It is important to note that the U.S. EPA funded three new meta-analyses on this issue, using slightly different approaches and that results from these studies should be available by the end of 2004.

Based on all of this information, it is incumbent on the study team to provide a stronger rationale for their study selection. Conceivably a Bayesian analysis could be undertaken, which puts greater weight on the European studies for obvious reasons, but factors in the results of many of these other studies. Also, the study team should examine whether there is a stronger effect from cumulative exposure to ozone as is the case, as reported above, for PM.

I also question the assumption of an implicit threshold at 35 ppb 8-hour average. Many of the times-series studies include 8-hour averages (either directly or converted from studies using 1- or 24-hour averages), which feature concentrations well below 35 ppb. It would be preferable to be consistent and use a methodology similar to that used for the PM estimates and, unless a clear threshold is indicated, calculate benefits down to background ozone concentrations in Europe. In California, for example, background concentrations are thought to be around 30 ppb for an 8-hour average. Regardless, the rationale for using 35 ppb, if it is ultimately used, should be fully discussed,

The use of an assumed six months per life lost needs more discussion. One possibility is to examine some of the disease-specific results as well as the pathophysiology of the response to ozone to inform this assumption. If ozone exposure can be plausibly linked to sudden MIs and strokes, then a significant number of life years lost may result. I concur with the suggestion #4a of van Bree and Buringh (17 Aug 2004) indicating that some sensitivity analysis around this issue along with including subjective probability may be warranted.

Another issue to consider is the use of studies of the long-term exposure to ozone. In the analysis of the American Cancer Society cohort, Pope et al. (2002) do not find any association between annual concentrations of ozone and life expectancy. However, when exposure to ozone is constrained to be measured as third quarter average, a weak association emerges (with P ~ 0.07). Given the wealth of evidence indicating a possible inflammatory response to ozone exposure and evidence of long term effects on lung function, the possibility of effects associated with long-term exposure to ozone cannot be ruled out. Perhaps the implications of this study should be examined. In addition, the implications of long-term lung function changes (Kunzli, et al, 1997) could be investigated since these effects are good predictors or death and disease.

Infant mortality from PM

The study team proposes to use the cohort study by Woodruff et al. (1997) which links exposure to PM to post neonatal mortality. This endpoint has been used in the WHO estimates in Global Burden of Disease, although in this case quantitative estimates were based on time-series studies conducted in developing countries. A recent U.S. EPA Science Advisory Board suggested that this effect be included in the national estimates of the benefits of pollution control (SAB, 2004). These studies do not have to be used to calculate premature mortality from the time-series studies, since the latter include all age groups.

If the study team plans on calculating life-years lost, separate estimates for infants and children are necessary since this age group is not considered in the prospective cohort studies. I would recommend an attempt to combine information from the cohort study of Woodruff, the time-series mortality studies, and some more recent studies on neonatal mortality (for example Ritz (2004 –ISEE Since most of the mortality is likely respiratory-related and associated with susceptibility from pneumonia and other respiratory infections, it is quite plausible that effects that have been reported outside of Europe and North America would hold in these areas as well. As an additional exercise, the study team may also consider estimating mortality and morbidity in a two-step process where first, they use the studies that relate air pollution to birth weight and prematurity and second, they relate the latter outcomes to morbidity (asthma onset, COPD, etc.) and mortality. This could serve as an interesting crosscheck on the estimates developed from direct application of the studies.

Morbidity from PM and ozone

It is not clear from the current text (Methodology Paper Issue 3) which endpoints and morbidity studies will be estimated. The Hurley attachment of the presentation on July 16th provides more detail but is not definitive. Appendix 5 indicates that the final determination of studies will come at a later date. Regardless, it is important to recall my suggestion that the selection criteria used should be clear, balanced, and consistent across endpoints. I again question the assumption about only calculating ozone effects down to an 8-hour average of 35 ppb. I would suggest either using the lower end of the existing studies, a threshold level if one has been reported (as in the case of emergency room visits for pediatric asthma), or the background concentrations. If 35 ppb is used, the rationale for its use should be fully explained and some sensitivity analysis with lower threshold examined.

For asthma morbidity endpoints, I would suggest the following where an effect that is independent from PM appears likely: hospital admissions for respiratory conditions, pediatric emergency room visits for asthma, minor restrictions in activity, school loss days, and perhaps asthma exacerbation. The latter of course faces the difficult proposition of combining different endpoints, baseline rates and subgroups across studies but the outcome has certainly received a lot of attention and should be calculated if possible. Care should be taken to preclude double counting among outcomes. For example, asthma ER visits should be subtracted from MRADs since the latter would include the former. However, this double counting is likely to have a very small impact on the total estimates.

I don't feel that there is strong enough evidence to estimate new onset of asthma, although I may not be aware of all of the recent studies in this area.

For PM, endpoints should include both cardiovascular and respiratory hospital admissions, acute bronchitis, ER visits, restrictions in activity (maybe combined with symptom days or days with lower respiratory illness), asthma attacks, and physician visits. In addition, with proper caveats and indications of uncertainties, estimates could also be provided for non-fatal heart attacks, chronic bronchitis, and maybe some other outcomes that are now being reported. Data should be available for Europe to establish reasonable baseline rates – a problem for the Global Burden of Disease efforts when extrapolating results to developing countries with poor surveillance data. If necessary, additional sensitivity analysis could be used to examine the impact of alternative baseline rates for the adverse health outcomes.

The Hurley July 16^{th} presentation appears to suggest that only European studies will be used for many of the endpoints such as hospital admissions. I suggest that additional information and reductions in uncertainty can be obtained by using published studies from the U.S. and Canada. At a minimum, the quantitative results of these studies can be reviewed and provided for context for the European studies. However, the study team should seriously considered using the North American studies either fully weighted or using a Bayesian averaging approach where local studies get greater weight but are informed by studies conducted elsewhere.

Additional Sensitivity Analysis

The study team proposes several sensitivity analyses including:

i. the impact of different toxicities of PM2.5 for sulfates and nitrates
ii. source-specific estimates separating out combustion-based particles from our sources.
iii. additional analyses for other pollutants, such as SO2 and NO2.

At this point, I would agree with the WHO (2004) evaluation that there is not enough current information to quantify the contributions of different PM sources. This is especially true for the results of the long-term exposures. For time-series mortality, one could, as an exercise, do the following two things: (1) carefully examine all of the time-series evidence on sulfates, nitrates and black carbon and attempt to develop summary estimates, with significant caveats regarding uncertainties; BART?? and (2) use the results of Laden et al. (2000) and other recent efforts which attempt to estimate the impact of combustion versus other sources of particles (but not nitrates versus sulfates). This should be viewed as only an exercise and not used for any definitive quantitative estimates.

By the way, I disagree with the EURELECTRIC comment on page 2 where they state that the current epidemiological studies provide an estimate of the effect of primary

PM. In fact, the estimates for the most part are based on PM10 and PM2.5 and therefore are not measuring, for example, the effect of black carbon and other emitted primary pollutants. The early work on this issue appears to indicate that per unit estimates based on black or elemental carbon are larger than that of generic PM2.5. However, this information has not been well developed in studies yet and is therefore not ready for use in the impact assessment. Regardless, the assessment could reflect the possibility that some primary pollutants may have higher toxicity than generic PM2.5. In addition, there is no evidence to support the EURELECTRIC suggestion that the present procedure is an overestimation. However, it is incumbent on the project team to indicate, and quantify when possible, the potential sources of both under and overestimation in the analysis. Finally, I agree with their comments that the empirical functions for SO2 and NO2 need to be detailed, since it is not clear when the study team will use in this regard.

Regarding SO2 and NO2, I would be concerned about double counting of health effects, if both PM and ozone effects have already been estimated.

Other issues

Uncertainty. There was no clear indication of whether there would be a formal quantification of uncertainty. At a minimum, I would assume that uncertainty could be propagated from the emissions-exposure and exposure-health pathways. Subjective probabilities, beyond normal confidence intervals could be applied to the low, medium and high estimates and the ultimate distributions of benefits could be calculated using software packages such as @risk.[1] In addition, value of information assessments could be made to guide future efforts and address potential research needs.

Accountability. The issue of what has been labeled “accountability” could be addressed in this report. By this term, it is meant the likelihood of actual improvements given reductions in air pollution. The study team could and should provide a brief review of the studies in this area which help address the issue of causality and magnitude of effect. My take on these studies is that they support the concept of causality and, in fact, suggest that the benefits per unit reduction may be greater than that estimated from the usual time-series and other studies. Examples of this work include, but are not limited to the studies in the Utah Valley, Dublin, Hong Kong, and (the former East) Germany.

Calculating restricted activity days (RADs). At this request of the study team, I have a few comments on this issue as presented in a powerpoint presentation to Stakeholders on 16 July 2004. First, the series of papers reporting this association are not really from “one study” but use six years of annual data collected by the National Center for Health Statistics. Second, the design has both cross-sectional and time-series elements since subjects were sampled in each quarter in each city. Thus, a fixed effects model was used to control for cross-sectional aspects among cities and then focus on changes within cities. Finally, it would be acceptable to estimate both work loss and RADs from these series of studies with the caution to avoid double

[1] Contact info: Palisade Europe, The Blue House, Unit 1, 30 Calvin StreetLondon E1 6NW UK.

counting since work loss would be also counted as a RAD. The general results of this study have been reported in other studies of work loss and lower respiratory symptoms by several researchers including Krupnick et al. (1990). The latter study was further reanalyzed by Ostro et al. (1993) where the outcomes were more broadly classified as either upper or lower respiratory symptoms.

Long-term exposure and morbidity. This is a large omission since, if one believes that long-term exposure affects mortality and long-function, it is extremely likely to affect the incidence of cardiovascular disease. In fact, one unpublished analysis of the ACS cohort reports such an association. The study team may want to carry out the quantification of this endpoint to appreciate the implications of this study. In the Global Burden of Disease, analysts assumed that, at a minimum, the morbidity effect is equal to the mortality effect.

Summary of Conclusions and Recommendations

1. The selection criteria of studies for the concentration-response functions should be clear, balanced, and consistent across endpoints. The role of subjective choice in the selection should be acknowledged and enough information provided in the document so others can appreciate the thought processes that went into the selection.

2. We support the reliance on the prospective cohort studies as the primary basis for estimating mortality effects. We also find the use of an estimate of approximately 6% per 10 $\mu g/m^3$ of PM2.5 to be reasonable.

3. Uncertainties in these estimates from the cohort studies should be discussed and possibly quantified using subjective probability. Both life years lost and cases of prematurity should be calculated. Mortality effects from the time-series studies should be quantified for information purposes, and calculations from the life tables should be carefully described.

4. Estimates should be extrapolated down to background concentrations with some attempt to determine the actual background concentrations for PM and ozone in different countries or regions. The rationale for the 35 ppm lowest effects level for ozone should be fully documented. In general, the lowest effects level should mimic the concentrations in the original study.

5. The assumptions on cessation lags could be better developed by consulting with the existing literature on heart disease and lung cancer.

6. We support the use of the WHO-based estimates for mortality related to short-term ozone exposure. However, the study team should consider analysis that includes other studies and lag structures. In addition, the assumed life years lost per case needs more discussion and validation. Finally, consideration should be given to using the Pope et al (2002) results for long-term summertime exposure to ozone.

7. Estimates for infant mortality could make use of both the cohort and time-series studies.

8. At this time, I don't feel that the evidence for new onset of asthma is strong enough for quantitative estimates.

9. Regarding SO2 and NO2, we are concerned about the likelihood of double counting the effects from PM and ozone.

10. The study team should consider developing a CR function for morbidity associated with long-term exposure to PM. If not, this endpoint is likely to be the largest omitted component of the health assessment

References

Abbey DE, Nishino N, McDonnell WF, Burchette RJ, Knusten SF, Beeson WL *et al*. 1999. "Long-term inhalable particles and other air pollutants related to mortality in nonsmokers." *Am J Respir Crit Care Med* 159:373-82.

Anderson HR, Atkinson RW, Peacock JL, Marston L, Konstantinou K. 2004. Meta-analysis of time-series studies and panel studies of particulate matter (PM) and ozone (O3). Report of a WHO task group. World Health Organization. (http://www.euro.who.int/document/e82792.pdf)

Dockery DW, Pope CA III, Xu X, Spengler JD, Ware JH, Fay ME *et al*. 1993. "An association between air pollution and mortality in six U.S. cities." *N Engl J Med* 329:1753-9.

Dominici F, McDermott A, Daniels M, Zeger SL, Samet JM. 2003. Mortality among residents of 90 cities. Revised analyses of time-series studies of air pollution and health. Boston, MA: Health Effects Institute, p. 9-24.

Jerrett M, Burnett R, Ma R, Thurston G, Newbold B, Finkelstein N, Shi Y, Krewski D, Thun M. 2004. Air Pollution and mortality in Los Angeles (abstract). Health Effects Institute Annual Conference.

Krewski D, Burnett R, Goldberg MS, Koover K, Siemiatycki J, Jerrett M *et al*. 2000. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. Res Rep Health Eff Inst (A special report of the Institute's Particle Epidemiology Reanalysis Project).

Laden F, Neas LM, Dockery DW, Schwartz J. 2000. "Association of fine particulate matter from different sources with daily mortality in six U.S. cities." *Environ Health Perspect* 108(10):941-7.

Levy JI, Carrothers TJ, Tuomisto JT, Hammitt JK, Evans JS. 2001. "Assessing the public health benefits of reduced ozone concentrations." *Environ Health Perspect* 109:1215-26.

Miller, K, Siscovick D, Sheppard L, Shepherd K, Anderson G, Kaufman J. 2004. Air pollution and cardiovascular disease events in the Women's Health Initiative Observation Study (abstract) The Sixteenth Conference of the International Society for Environmental Epidemiology

National Research Council. 2002. *Estimating the public health benefits of proposed air pollution regulations.* Washington, D.C.: National Academy Press.

Ostro BD, Lipsett MJ, Mann J, Krupnick A, Harrington W. 1993. "Air pollution and respiratory morbidity among adults in Southern California." *Am J Epidemiol* 137:691-700.

Pope CA III, Burnett RT, Thun MJ et al. 2002. "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution." *JAMA* 287:1132-1141.

Ritz B, Wilhelm, M. 2004. Air pollution and infant mortality in the South Coast Air Basin, 1989-2000 (abstract) The Sixteenth Conference of the International Society for Environmental Epidemiology.

Roosli M, Künzli N, Braun-Fahrländer C. 2004. Use Of Air Pollution "Intervention-Type" Studies In Health Risk Assessment (abstract) The Sixteenth Conference of the International Society for Environmental Epidemiology

SAB. 2004. Advisory on Plans for Health Effects Analysis in the Analytical Plan for EPA's Second Prospective Analysis – Benefits and Costs of the Clean Air Act, 1990-2020 : Advisory by the Health Effects Subcommittee of the Advisory Council on Clean Air Compliance Analysis

Schwartz J, Laden F. 2004. Dose, Time and Death: Associations with PM2.5 in a Cohort Study (abstract) The Sixteenth Conference of the International Society for Environmental Epidemiology.

Schwartz J. 1997. "Air pollution and hospital admissions for cardiovascular disease in Tucson." *Epidemiology* 8:371-7.

Stieb DM, Judek S, Burnett RT. 2003. "Meta-analysis of time-series studies of air pollution and mortality: Update in relation to the use of generalized additive models." *J Air Waste Manag Assoc* 53:258-261.

WHO .2001. Health impact assessment of air pollution in the WHO European Region. Technical Report from WHO/ECEH Project. WHO European Centre for Environment and Health, Bonn, Nov 2001.

Woodruff TJ, Grillo J, Schoendorf KC. 1997. "The relationship between selected causes of postneonatal infant mortality and particulate air pollution in the United States." *Environ Health Perspect* 105:608-12.

Chapter 4
Estimating Other Physical Effects

Overview of methods and background to the data used

The methodologies for agriculture/horticulture, ecosystems and materials described in AEAT/ED51014/Methodology Paper Issue 3 are a mix of established and recent approaches, drawing upon scientific results that have become well-established over the last decade.

The methodology team is familiar with the science. In the past they have used similar, though less developed scientific results. As a result of this experience they not only have useful background knowledge but also they are very much aware of the need for close collaborations with the various scientific groups providing the data.

Much of the science used has been developed under the Convention on Long-range Transboundary Air Pollution (CLRTAP), in particular by its international cooperative programmes (ICPs) some of which have been operating since the mid-1980s (NOTE: ICPs are not UNECE bodies; they operate under CLRTAP. While CLRTAP is open to Parties from the UNECE area, the Convention is independent of UNECE (which simply provides the Convention's secretariat). The UN Economic Commission for Europe is a UN international body that has its own programmes and subsidiary bodies. The text on the methodology should be changed to, for example, CLRTAP ICP Forests or just ICP Forests). The science proposed for use in this work is similar to that used already to assist the development of air pollution abatement strategies both for the EU and for the Convention. The scientific bodies increasingly target their work towards practical applications for policy purposes.

Even so, there is a natural tendency for scientists to be cautious in the interpretation of results, especially when models are being used and results need to be extrapolated across many countries. They are keen to ensure that results are not misinterpreted and that policy makers are not misled, hence caveats are often employed to ensure that there is a better understanding of what results might or might not mean.

The authors of this methodology are, from past experience, well aware of this and appear to have taken much trouble to discuss their approaches with the relevant European scientific experts and get agreement on acceptable methods. From reading the proposal and discussing the matter with the consultant involved I am reassured that this collaboration with scientific experts will continue when the methodology is applied to the analysis.

Agriculture/horticulture

I am in agreement with the focus on ozone in the proposed work. While the other effects no doubt occur, and are not forgotten in the overall MCA, the potential impacts of ozone on crops are widespread and may increase in the future.

The authors demonstrate that they have a good grasp of the problems of using an AOT40 approach and recognize the preferred use of flux-based models. However, as they also indicate, flux-based models are still being developed and therefore the extent of their application for estimating crop yield losses may be limited.

The stock at risk methodology (para. 5.3) is a reasonable attempt using the available data. The Stockholm Environment Institute (SEI) at York has developed a European GIS database (map) for Europe over many years, building in the best data sets as they have become available. CLRTAP has relied on this database on many occasions though there are plans to extend it for ecosystem definitions in the Convention's work. The method of reclassifying and using agricultural statistics I believe is the best available approach to defining crops at risk across Europe, though there are inevitably some uncertainties in the methodology.

The use of the flux-based approach (para 5.4) is noted as being dependent upon the availability of flux data for the CAFE scenarios, and it will only be applied to crops, such as wheat, for which the necessary C-R functions have been calculated. This is a reasonable position to adopt. The flux-based functions for wheat and potato have now been agreed in the CLRTAP Mapping Manual so there is potential for these to analysed using the flux-based approach. The use of AOT40 for an MCA for other crops is a safe approach to adopt.

I note that information on the derivation on ozone data are omitted, this being generated by the EMEP and the RAINS models. The uncertainties in such data should not be forgotten since they both operate at the European scale at fairly coarse resolution. However, they are the tools that are used for developing EU and CLRTAP policy and it is appropriate that they should be the sources of data. It is not clear to me if the uncertainties to be considered in the DEFRA project (para 5.7) will include those resulting from the EMEP and RAINS models.

Ecosystems

The ecosystem impacts (para. 7.1) are based upon the well-known development of ecosystem effects in Europe and the development of the critical loads approach, an accepted methodology for both CAFE and CLRTAP. The methodology recognizes that risks differ across Europe and that there are risks from acidification as well as eutrophication. However, it is particularly important for the reader to understand what an exceedance of a critical load means in terms of damage/potential damage. Concern about ozone effects are noted, though the importance of this pollutant could be further stressed with regard to possible effects on forests.

The authors note that eutrophication effects are likely in areas where rainfall is high and areas low in nitrogen. They could also note that such eutrophication effects may be underestimated since, for example, there are effects in predominantly agricultural areas where very high deposition of nitrogen occurs locally, which can have significant impacts on ecosystems such as woodlands that may not normally be considered as nitrogen deficient. That is, nitrogen effects could be more widespread than the European maps of critical loads might suggest.

The methodology proposes to use the CCE data that are linked to the RAINS database. It should be noted that this data set is built from predominantly national data submitted by countries to CCE, so EU Member States should be reassured that these are the best data available for the purpose.

Regarding the methods "agreed with CCE" (para. 7.3) I assume these are those agreed by Parties to the CLRTAP, which include the Member states of EU and the European Community itself. These methods are consistent with what is recorded in the ICP Modelling and Mapping "Mapping Manual" and with those approaches that were used for the development of the Gothenburg Protocol and the parallel EU Directives.

I believe it entirely proper that, as noted in para. 7.3, the CAFE CBA does not attempt to go further in quantifying ecological impacts outside of agriculture than simply using the results from RAINS regarding the area over which critical loads are expected to be exceeded and the associated accumulated exceedence" (NOTE: for consistency with most current literature the authors should use the spelling exceedance not exceedence). Although there is plenty of evidence of other effects on ecosystems, e.g. decreases in biodiversity, these are not built into the internationally accepted critical loads methodology as much as they could be. It would be proper for the authors to draw attention to such possible effects; however, they are correct in not attempting to apply them in the analysis.

I also support the exclusion of ozone effects of trees from the analysis (para. 7.3) but would suggest that further emphasis on its possible importance be made (in keeping with my comments on para. 7.1 above). I would be reluctant to make use of the recent results of IVL until they have been presented to and accepted by the international (CAFE and CLRTAP) scientific community.

Regarding valuation, the authors point to the CLRTAP NEBEI output and propose that their approach be consistent with this. I believe that this is the safest possible option at present as this approach has been accepted by CLRTAP Parties (and hence EU Member States) when it was reported to them. The authors are clearly familiar with other work that has attempted to go further and recognize the limitations of these extended approaches. However, further discussions on ecosystem valuation will be appropriate, as suggested in the report, to see if it is acceptable to extend the currently agreed approaches.

Materials

CLRTAP's ICP Materials has done much work in recent years, some of it funded through EU projects, to develop the damage functions needed for this work. The methodology has taken account of the developments in damage functions and lists the functions it proposes to use (par. 6.2.3).

Stock at risk data (para. 6.2.2) have always been difficult to put together, though the ExternE project made some good progress in this area and this work will build on that work.

The list of uncertainties (para. 6.2.5.) covers the main areas of uncertainty, and some of these may be quite significant. They are acknowledged by the authors, but it is not clear how these might be built into the final methodology.

For acid damage (para. 6.3) it is recognized that estimating stock at risk for cultural heritage poses problems. It is noted that "Conclusions on the precise approach to be used for dealing with cultural heritage in CAFE are still to be finalized". I believe that it is important to try to answer this question since cultural heritage damage is becoming increasingly important to the public; more and more throughout Europe attention is being paid to our cultural heritage and the public have an increasing awareness of damage inflicted and greater expectations for repairs/preservation.

For non-cultural heritage damage (para 6.3) the acceptable damage concept has received much attention for many years. It is useful that the methodology will give this concept due consideration. Acceptable damage has not been seen as a rigid approach to define unacceptable damage, but rather as a measure to aid the understanding that deterioration of materials happens in the absence of pollution.

Proposals for work on ozone and particles (paras 6.4 and 6.5) build upon past methods and appear to be reasonable in their approach.

Regarding further action (para. 6.6), I note the proposed further collaboration with ICP Materials and agreed this should be undertaken early in the analysis. The comparison of this study with the Dutch benefit study would seem a useful check on methodologies, though the differences for the Netherlands may be rather different than those for some other countries where data on stock at risk are much poorer.

Conclusions

Overall the approaches proposed for the cost-benefit analysis for agriculture, materials and ecosystems cautiously adhere to scientific methods already developed under CLRTAP and under CAFE. The authors do not ignore other developments, and the methodology does not exclude them from use at a later stage if it is so decided, but as a first approach the methodology is closely linked to RAINS and the work of the ICPs of CLRTAP.

I am comfortable with this approach, and while it may not go far enough for some, it is likely to be agreeable to most Member States and to most of those who have been involved in such European-scale evaluation exercises in the past.

Chapter 5
Valuation of Physical Effects
By Alan Krupnick

The charge to peer reviewers is to comment on the general methodology, whether it is based on current scientific knowledge (based on the Terms of Reference), and if it is "fit for purposes," as well as to provide recommendations for improving the methodology. In the area of valuation of physical effects, the methodologies differ to such an extent that general comments about methodology and fitness, as well as comments on specific choices made by the CAFE team must primarily be made for each endpoint. The most important endpoint is the valuation of reduced mortality risks to adults, so considerable attention is given to this topic relative to the others, which include valuation of infant mortality, morbidity, building materials, crops, ecosystems, visibility, and climate change. Nevertheless, a few general points can be made about valuation as handled in the CAFE methodology plan.

General methodology, whether it is based on scientific knowledge, and if it is "fit for purposes

Despite many inherent uncertainties and subjective decisions, the idea of attaching monetary values to improvements in the endpoints evaluated in CAFE is laudable and appropriate to performing benefit-cost analysis. By taking such a step, as long as the valuation estimates are based primarily on preferences of the European public, either revealed by their behavior, gauged through various survey techniques, or appropriately transferred from studies in other countries that rely on these approaches, the resulting BCAs can help identify efficient policies for reducing air pollution and serve as a transparent and comprehensive framework for subsequent debates about the merits of these policies.

There are ample examples of the use of governments of such frameworks. The U.S. OMB developed and codified principles for the conduct of BCA's, paying particular attention to the estimation and valuation of health effects (OMB,2000; 1996). The USEPA has its own set of guidelines that are consistent with those from OMB while going into greater detail (USEPA, 2000). These efforts were themselves guided by earlier efforts to develop the damage function approach to benefits estimation, both in the U.S. (Lee, et al., 1995) and in Europe, through the EXTERNE program. For air pollution specifically, intense academic peer review has been given to Congressionally mandated studies of the costs and benefits of the Clean Air Act (the so-called "812 studies") (USEPA, 1999), in which the damage function approach has been strongly endorsed. In addition, Canada now has a history of the use of BCA to help examine its ambient air quality standards (Canada-Wide Standards, 1999), and many developing countries, at the behest and encouragement of the World Bank and the Asian Development Bank, to name two international agencies, use BCA for examining air pollution policies (for instance, Silva and Pagiola, 2003). The consensus regarding the valuation of health and other endpoints is clearly that these efforts are based on and supported by scientific knowledge and provide a useful basis for policy decisions. Therefore, it is "fit for purpose".

At the same time, given the large uncertainties about various steps in this process, some of the specific choices made by the CAFE team can be criticized, as will be presented below. Each endpoint will be discussed in the order that they appear in the methodology document.

Mortality Valuation – Chronic mortality from PM

Planned Approach

The EU team proposes in its main analysis to feature a Life-Years-Gained approach to estimating mortality risk improvements, where each additional life year gained is estimated and multiplied by the measure Value of a Life-Year (VOLY). This approach is in contrast to the traditional approach, where statistical lives lost are estimated and then multiplied by the value of a statistical life (VSL). CAFE also plans to conduct sensitivity analyses using the VSL and estimates of the reduction in lives lost.

The rationale for this decision is primarily that health impacts are to be expressed in years of life lost, which is justified as "our own established practice" and a "wider emerging consensus," and therefore, a VOLY approach is needed to match it. An additional reason is that the VSL approach "is not consistent with WHO guidance." This decision is also said to be practical because "recent empirical studies," namely the NewExt study (Alberini, Hunt and Markandya (AHM), 2004) and well as Clinton, et al (2004), "provides direct evidence" of this value. In addition, the AHM study, which was designed specifically to measure WTP for mortality risk reductions in the context of air pollution, was conducted in three EU countries (France, UK, and Italy).

The approach followed in AHM (2004) to estimate the VOLY is being followed here, i.e., to divide WTP from this study by the measure of life expectancy gain implied by the 5 in 10,000 per year for ten year risk reduction and then compute the mean and median WTP to provide a range of values.

Below, I discuss the literature and the logic underlying the choice of methods, and the choice of specific values, concluding that the proposed approach could be modified in a number of ways to better reflect the uncertainties associated with estimating VOLY as well as VSL.

Critique of General Methods

Define VSL, VOLY. The value of statistical life (VSL) is a convenient concept involving dividing some estimate of the willingness to pay for a mortality risk reduction by that risk reduction. This measure is convenient because it can be multiplied by the statistical deaths averted by a policy to arrive at the benefits of that policy for this health endpoint.

The value of a statistical life year (VOLY) is conceptually similar, involving dividing some estimate of the willingness to pay for an improvement in life expectancy by that life expectancy improvement. This measure is convenient because it can be multiplied by the statistical life-years averted by a policy to arrive at the benefits of that policy for this health endpoint.

The VOSL measure is the traditional one. It can be applied from estimates of lives lost taken directly from the epidemiological literature, which typically offers estimates of deaths averted by a change in pollution. It is based on a huge literature of original studies and meta-analyses (see Krupnick and Alberini (2002) for an overall discussion and the recent Viscusi and Aldy (2003) for a review of hedonic studies, Alberini (2004) is also very useful and most recent.). It has been repeatedly endorsed by peer review panels in the U.S. (USEPA, 1999). At the same time, it is being argued more frequently that much of the supporting literature is for a context not related to environmental policy -- wage premiums paid for risky jobs, involving people of better than those at risk from pollution, in the prime age groups, and taking voluntary risks of mostly immediate death (Krupnick et al, 2002).

The VOLY measure does not have the lineage enjoyed by the VSL , but has risen in prominence because it is undeniable that most of the deaths delayed from environmental policy are to the elderly and to treat deaths in these two groups as equivalent for valuation purposes seems inappropriate because so many fewer life-years are lost when the elderly die. At the same time, the epidemiological literature is not as robust in estimating life-years gained. Additional assumptions are usually needed, such as that life years reduced follows a proportional hazard model (i.e., where life years reductions are proportional to death rates by age, following from an assumed proportional shift in the hazard function in response to an air pollution reduction). In addition, the VOLY literature is very thin, involving only a few studies that directly ask for the WTP for additional life expectancy (e.g., Johannesson and Johansson, 1996; Hammitt and Liu, 2004; Clinton, et al, 2004).

Conceptual underpinnings. Most conceptual studies derive results for the VSL. They all use a life-cycle model of individual choice that has individuals maximizing utility over a lifetime in which they have an earnings and consumption stream (Shepard and Zeckhauser, 1982). These models have direct implications for the choice of a VSL or VOLY measure because they take up the issue of how the VSL should change by age. Such models generally find that the VSL expressed at different ages does not fall from birth, as it would if only expected remaining lifetime drove this estimate, but takes an inverted U-shaped function, generally peaking in the 40's during one's high earning period, and falling thereafter and at earlier ages, with the fall at higher ages less steep than the fall to earlier ages. Kniesner, Viscusi and Ziliak (KVZ) (2004) is only the latest of a long list of studies to find evidence for this shape empirically. This study is particularly interesting because it is one of the first to use hedonic wage methods to add consumption information to the model, finding that rising consumption later in life results in a relatively moderate (and less-than-proportional drop-off of the VSL relative to its 50 year old peak) for people in the 57-65 year age range The ratio of the VSL of this group to the 52-56 year old group is 1.94/2.25 and to the 47-51 year old group is 1.94/2.36. The VSL for the older age group is actually larger than that for the 32-36 year olds. Overall, KVZ find that the VSL of older people exceeds that of the average over all working age adults. Rabl (in Desaigues et al, 2003) looks at the VOLY issue directly, finding that the VOLY increases with age, strongly so for the highest ages.

Empirical Issues. Unlike the VSLs, which are computed from estimates of the WTP for risk reductions using hedonic wage models, stated preference surveys, and consumer behavior surveys, the VOLYs have been computed mainly through computational adjustments of existing VSLs.

The simplest of these approaches is to divide the VSL by life-years remaining, although whose life-years is certainly an issue. That issue was addressed by taking the average age of those for which the VSL estimate originally applied (generally labor market studies, so an age of say 45), recognizing that a 45 year old would have a life expectancy of about 40 years on average and so dividing the VSL for this group by 40 years. This approach is simply wrong because it doesn't recognize that these improvements occur over time, which leads immediately to a discounting procedure for calculating life-years Moore and Viscusi (1988) . From an assumed discount rate, an assumed default life expectancy and an assumption that the value of a life year is constant over the 40 years but for the time value of money, an exponential discounting procedure was developed. For a VSL of about $6 million, the resulting VOLY at 3% for a 40 year loss in life expectancy would be around $270,000.

A variety of other *ad hoc* approaches followed this approach, but all work off of a VSL estimate. The *ad hoc* approach followed to generate the estimates in the CAFE CBA methodology is in the tradition of making computational adjustments. However, this approach uses the WTP for a risk change, but divides it by the life expectancy change calculated from the risk change (from Alberini, Hunt and Markandya, 2004).

At the same time, a few studies asked for WTP estimates for VOLY directly. The first direct efforts to examine this issue were from Johannesson and Johansson (1996) who found exceedingly low VOLYs. Another survey asking about the WTP for life expectancy changes is the DEFRA study (Clinton, et al, 2004), which found a VOLY of L 27,000 scaled from questions asking about one month life extensions in normal health, down to L 6,000 when scaled from questions for a six-month extension. This is not the place to critique this study. However, note that the study specifically evoked air pollution as the cause and the way in which benefits would be delivered. This may have reduced WTP because people do not think it should be their responsibility to pay for air pollution reductions. In addition, the study protocol called for WTP estimates covering four different effects, including acute and chronic mortality events, hospitalization and shortness of breath, and then asked for an allocation of these funds across the four types of benefits. This unorthodox treatment needs careful testing and analysis. And third, the benefit of reduced air pollution is clearly a public good but respondents were asked their WTP for the improvement as it affected themselves and their immediate family. I assume VOLY is based on an average payment across all family members. But if people bid as a public good, this would tend to raise their bid above what it would be if the good were a private good.

I note that the Krupnick et al (2002) survey (which underpins the EXTERNE valuation efforts) originally had a work plan which called for asking life expectancy (LE) questions in a modified version of the survey. However, they found that respondents in focus groups had very little understanding of this complex concept, generally thinking that an additional life year was tacked on at the end of one's life rather than being a shift in the survival function affecting the hazard rate over all ages. They further found that WTP for very small changes in LE, of the order of several weeks or a month, was simply not seen as valuable by most participants.

Nevertheless, the French team (headed by Brigitte Desaigues) included questions eliciting WTP for life expectancy improvements on the Krupnick et al survey administered in France. Although the sample size was very small and there are other issues, the results include a VOLY ranging from €20,000 to €220,000, depending on the basis for the scaling to a life-year (e.g., one month or six months of life expectancy improvements) (Desaigues et al, 2003).

Much of the newer literature on the topic of VSLs and VOLYs has focused rather on the effect of age on WTP for risk reductions. Krupnick et al (2002) and Alberini et al (2004), to which the New Ext effort owes its lineage, was designed to test whether the VSL differs across different age groups and those in different health states. In fact, the results suggest that the VSL does not differ significantly across age groups. If the VSL is identical for different age groups faced with identical mortality risk reductions then it is inappropriate to use a VOLY approach because the standard VOLY approach assumes life-years are additive in preferences (i.e., that the VOLY is constant for all life-years, regardless of the age of those benefiting), while the valuation studies show that they are not; indeed, it shows that life years become more precious as there are fewer left to lose.[2]

VSL and VOLY and Health. Another important factor in deciding on the appropriate concept for valuing mortality improvements is the affect of poor health on WTP. The existing literature, albeit only a few studies, including the NEW EXT studies, shows that, if anything, people in poor health are willing to pay more for a given reduction in their mortality risk that healthy people would pay. This means that the VSL and, by inference, the VOLY, should, on efficiency grounds, be differentiated by health status.

Use of different VSLs and VOLYs for different groups more generally. There are a potentially wide variety of covariates, in addition to health status and age, which could influence one's WTP for a given mortality risk reduction (or life expectancy improvement). These include most notably, income, as well as gender, family size, and education to name a few. Indeed, each of these variables has been found to be significant in predicting WTP by at least one study. In addition, the survey developed by Alberini, Cropper, Krupnick and Simon (ACKS) (Alberini, et al, 2004a; 2004b), which has now been applied in seven countries (the three in the EU plus U.S., Canada, Japan and South Korea) find that such values differ after making various adjustments for differences in identifiable variables.

At the same time, whether one actually makes use of these findings in a cost-benefit analysis is an open question. Currently, there appears to be (in my judgment) consensus among policymakers that income differences in WTP should not be grounds for using different VSLs for different income groups. In this case, the VSL reflects average incomes.

---

[2] Another relevant recent study is Cameron and DeShazo (2004). This study is unique in combining mortality and morbidity by asking for choices over a wide variety of health effects described in terms of a life cycle profile, such as 3 months of disease X, followed by hospitalization and then death, all at given ages. It uses a conjoint survey format and is administered over the web to 1619 individuals given 7,520 choice sets containing15,040 illness profiles. The paper is exceedingly complicated and results are hard to interpret, but there is some evidence that the VSL falls with age.

Whether different VSLs should be used for different age groups was a huge policy debate in the U.S., resulting in a pronouncement from the Bush Administration (*Washington Post,* 2003) and a provision passed by the House of Representatives against any use of the "senior discount," effectively prohibiting use of a VSL lower for the elderly than for younger people.[3] While direct translation of this prohibition to the VOLY would imply simply that VOLYs could not vary by age group, the spirit of the debate is clearly that VOLYs should not be used in policymaking because this would discriminate against the elderly, who have fewer life-years to live.

In the U.S., there has never been a debate about whether different VSLs should be used in different political jurisdictions. The presumption is that the same VSL is to be used in all places when federal regulation is at issue. This position would presumably apply to VOLYs. For the EU the situation is more complicated because the political jurisdictions are sovereign to a significant extent. VSLs do appear to be different across the three EU countries that implemented the ACKS survey. Whether such distinctions should be made in policy decisions is a political decision ultimately. But on efficiency grounds, they should be.

Using different or same values over time. In the U.S. there is an emerging consensus among EPA peer reviewers that the VSL should be revised upwards to reflect growing per capita incomes, as studies show the income elasticity of WTP is positive. (Any source(s) for this?) These issues would also apply to VOLYs. This should be of particular interest in the context of air pollution given there is a time lag between the introduction of legislation, their effectiveness on emissions reduction and finally on air quality. For instance the value of life in 2020 is likely to be higher than in 2000 and thus, it would be good to have some estimates of the change.

Latency and VSLs and VOLYs. So far, the discussion has been entirely in the context of a reduction in air pollution causing an immediate reduction in death risks. There are many reasons to suspect that exposure to air pollution has cumulative impacts and that therefore, reductions in pollution would take some time to reduce deaths risks, certainly in terms of chronic PM exposure. In the U.S., this issue is termed the "cessation lag." The longer into the future the benefits are felt, the lower their value today, because of time discounting. The literature on this point is of two types. One makes mechanical adjustments to the VSLs or VOLYs, using an assumed discount rate, to reflect the cessation lag, which reduced these values (USEPA, 2004). Another, including the ACKS series of studies, estimates the WTP for mortality risk reductions experienced in the future, specifically beginning at age 70. This literature reveals that the latent VSLs (applying to those between 40 and 60) are about 30% lower than VSLs for equal contemporaneous risk reductions. Presumably, such considerations would also apply to VOLYs.

Other Issues left to Examine. Even the new class of VSL estimates does not incorporate some important factors affecting valuation. One is the fact that most pollution reductions are public goods. The literature is primarily about an individual's valuation of mortality risk reductions experienced by an individual, not an

[3] http://thomas.loc.gov/cgi-bin/bdquery/z?d108:HZ00338:

individual's valuation of mortality reductions experienced by the community. Where feelings of altruism are important and of a certain type (i.e., paternalistic), it would be appropriate to add up WTP across individuals. Presumably, although the literature is very thin on this point, WTP for a public good would exceed WTP for the private benefits, cet. par (Strand, 2001). Thus, the resulting VSL would also be larger (as would the VOLY). And a recent paper (Strand, 2004) shows conceptually that even without altruism VSLs for the public good could be larger than those for the private good.

Another important factor is the involuntary nature of exposure to pollution. The basic studies underlying the VSLs are hedonic wage studies where workers surely have some control over the risk they take on the job, both through their behavior on the job and through the type of job they accept. While behavioral change can reduce exposure to air pollutants (e.g., by staying indoors) it is likely that air pollution exposures are perceived as being less voluntarily accepted. According to the qualitative risk literature (Slovic, 1987, 1992) involuntary risks rank higher to be avoided than voluntarily assumed risks, so it may be the case that VSLs with this context taken into account would be larger. Again, the literature on this point is exceedingly thin and inconclusive.

Specific Choices

So far this narrative argues that the VOLY is not constant over ages when an additional life-year is experienced. This still leaves the question of what the appropriate VOLY's would be by age or, alternatively, what the appropriate VSLs would be, by age.

As noted above, the New Ext team developed estimates of VOLY that are not age-specific. The approach taken, as stated in AHM (2004) (but not in the CAFE Methodology document) is to convert the risk reductions being valued into life-expectancy changes (using work by Rabl, 2001) and apply the same WTP estimates to these changes. Thus, in addition to computing a VSL as WTP/risk change, they computed a VOLY as WTP/life expectancy change, normalizing this change to a year or month.[4] The VOLY implied a mean VOLY of €125,250 and a median of €55,800.

The choice of these VOLYs is interesting because AHM also found that the VOLY varied by age, being highest among those 60 and above (27% above the 40-49 group). They also recognized that the life tables used by Rabl may not reflect an individual's own assessment of their life expectancy. Fortunately our survey asked about this, so substituting respondents own assessments yields slightly higher estimates of VOLY: mean €142,000 and median €58,200. For comparison, the VSLs for this 5 in 1,000 risk change over 10 years were mean €2,258,000 and median of €1,052,000. The group over 60 has a VSL 20% lower than the 40-60 group but not statistically significant. This implies that a VOLY is much higher for the older group.

There are several additional issues with the way this study was used to generate the VOLY estimates. The first is the choice of median versus mean, the second is the use

[4] Note that the conversion of our 5 in 1,000 risk change over 10 years, which was given to each individual in the survey, converts to a different life expectancy change for each age/gender cohort. These changes ranged from 0.64 to 2 months, averaging 37 days.

of the results for 5 in 1,000 rather than some other change, the third is the use of WTP for people in a "normal health state" and the fourth is the decision to discount VOLY for future changes at a 4% rate.

There is no consensus on the appropriate statistic for describing the average sample response. Conceptually, the mean is the appropriate measure for use in CBA because they fully summarize the heterogeneity of values in the sample. The median is sometimes used instead (or other variants such as a trimmed mean, which drops outlying responses), because it is felt that the median is a more "robust" statistic (being not influenced by outliers, which can be prevalent in stated preference studies) and because it is a conservative measure, being lower than the mean because, in part, the WTP responses are bounded at zero but not bounded from above. Much of this practice was developed for Natural Resource Damage Assessments. The value of natural assets (and changes to them) carried defaults of zero. Thus, a median estimate was thought to be a vast improvement, even though inappropriate and inconsistent for CBA. In the health valuation area, however, the default value, at least in the U.S., has been $6 million VSL (a mean estimate) (USEPA, 1999). Hence using a median estimate appears more inappropriate because it doesn't match the prevailing default measure.

Concerning the issue of using the 5 in 1,000 risk change, unfortunately, the European applications of the Krupnick et al survey (2002) did not ask the 1 in 1,000 WTP question first, as was done in the U.S. and Canada. Based on the latter results, however, I feel confident in predicting that the implied VSLs for this smaller risk change would be at least two times and maybe three times larger. This factor would raise the VOLYs by comparable amounts.

Concerning the plan to use VOLYs for people in normal health states rather than diseased health, I am in basic disagreement with this approach. Most people affected by air pollution probably are not in normal health, although the epidemiological literature should be consulted more directly on this point to justify this position. If there is a political reluctance to use a different VOLY for groups of different health status, then the use of VOLY (or VSL) for the diseased health state seems reasonable, although it would lead to inefficiency where any normals lose life expectancy by being exposed to pollution.

In CAFE, latency is to be handled through a discount rate adjustment of 4%. An alternative would have been to use the AHM results for the WTP to avoid latent health effects. In addition, the AHM study (also in draft form as Alberini et al, 2004) finds implied discount rates of 5% in France and 10 % in the UK., which leads to questions about using a 4% rate and even the use of one rate across all countries.

The involuntary risk and public goods issues were not raised.

For the sensitivity analysis, a VSL will be used of E1 million, which is broadly consistent with the AHM study results for the mean WTP for a 5 in 10,000 per year risk reduction. It is to be noted that the VSL most in use by the US government is about $6 million, considerably larger. While the $6 million figure is unduly influenced by the hedonic wage studies and would be lower if based on other types of studies, it should not be lightly dismissed. At the same time, the use of results from the Alberini et al (2004) study for a smaller risk reduction would probably lead to a higher VSL.

In terms of the effects of health on the VSL, Alberini et al (2004) find a WTP for people who have been hospitalized for cardiovascular or respiratory illness over the last five years about twice that of all others, other things equal. Thus, a VSL of €2 million could be supported by the results in the European context.

Conclusions and Recommendations

From the above discussion, I conclude the following:

1. Use of a single VSL from hedonic wage studies is problematic, even though traditional
2. There is pertinent evidence for VSLs in Europe that vary by age and could be used in the report or transferred from U.S. and Canada studies (where the U.S. estimate for the "over 70" effect is a 20% discount but the coefficient is insignificant and where the Canada estimate is 30% and significant).
3. Direct, credible estimates of the VOLY are lacking. The estimates to be used in CAFE are derived computationally.
4. Being mindful of the very careful effort by the CAFE team to sift through the complicated issues surrounding these choices, the implication of 3. and the above discussion is that there is little justification for use of a single VOLY (irrespective of the age and health status of those experiencing this gain) to represent all life-years gained. This conclusion is in line with the comments from NERI.
5. Use of VOLYs by age and health status might be a reasonable option.
6. There is not much justification for ignoring results in the literature for a 1 in 10,000 annual change in risk, as opposed to the plan, which focuses entirely on the 5 in 10,000 annual change. The latter results are far more trustworthy given the protocols followed by the NewExt group; however, results in the U.S. and Canada, suggest that the 1 in 10,000 results are also credible.
7. Consideration should be given to using latent VSLs from the New Ext study.
8. Focus should be given to using mean WTP results. Perhaps the range of unit values for mortality should encompass the results for 5 in 10,000 risk changes and 1 in 10,000 risk changes.
9. Mortality values for diseased health states should be used instead of those for normal health states, if only one health state must be picked.
10. Following U.S. practice, unit values could be increased for increases in per capita incomes.
11. Varying unit values across countries is important for the integrity of cost-benefit analysis, recognizing that political concerns might obviate against this approach.
12. Based on the New Ext study, implied discount rates of 5% in France and 10 % in the UK were discovered. Perhaps sensitivity analyses with these rates and the common rate of 4% should be conducted.
13. However, it should be noted that it is likely that the willingness to pay for life (years) saved increases over time with incomes and thus could offset the above trends.

I would make an additional suggestion. Cost-effectiveness analyses using life-years gained as the effectiveness measures should be given more prominence in the plan. Net cost-effectiveness analysis, where the monetized benefits are subtracted from costs and then divided by life-years gained could be a useful measure for discriminating across alternative regulatory options. To do this would give prominence to the life-years approach without using the questionable VOLY estimates. Alongside such analyses should be CBA analyses using both VSLs and VOLYs, as outlined above.

Mortality Valuation – Acute Mortality from Ozone

For ozone, the identical methodology is to be used as that for PM above. So, all my comments apply. In terms of recommendations, the Pope study and subsequent reanalyses do not find evidence for a chronic mortality effect for ozone. Thus, assuming effects are acute only, there would be less justification for counting latency. Without cumulative effects, the case is probably better that the effects are concentrated among the older, diseased population. This suggests that the VOLYs and VSLs used should also reflect these values.

Infant mortality Valuation

The proposed method is to assume, as a central case, that the VSL for reducing child mortality risk is the same as that for reducing adult risk, but that this would rise to 1.5 of the adult VSL in a sensitivity case. The logic of the choice of the 1.0 factor for the base case could be clearer. I think it reflects the uncertainty about the credibility of the studies in the literature – an admittedly small literature -- which suggests that a central estimate might be the adult VSL times the factor of 1.5 or even 2. Such estimates come from studies of parental valuation of their children's health and death risks (Hoffmann et al, 2003). I think that the choice of 1.0 may be conflating what the studies show with uncertainty about what they show. I would recommend that a factor greater than 1.0 be chosen for the base case, with a sensitivity case of 1.0 and another sensitivity case as high as 2.0. I would also note that I support the EC project (VERHI) to develop such estimates for Europe, as the literature is thin and benefits transfers are always questionable.

Morbidity Valuation

The valuation of morbidity endpoints is less important than that for mortality, thus they merit less scrutiny. The exception is the value of a statistical case of chronic morbidity. This estimate is E200,000, based on U.S. studies. This estimate is a reasonable one for the current $:E exchange rate of 1.25:1, as the U.S. value is $260,000 per case avoided. A few issues should be noted, however. First, there is an unusually thin literature underlying this value, basically two studies that are related through use of the same methodology, and are used complementarily. The VMH study (Viscusi, Magat and Huber, 1991)is the foundation study, based on about 300 people interviewed in a shopping mall in North Carolina. As the case of chronic bronchitis being valued was extreme relative to a typical case, the VMH value is adjusted using a severity factor found in Krupnick and Cropper's (1992) study using the same methodology as VMH, but applied to about 300 relatives of adults with chronic lung disease, the relatives all living in the Washington D.C. area. Second, the benefit transfer should properly be based on purchasing power parity rather than on the exchange rate.

The other morbidity valuation estimates are unremarkable. I would note that the Navrud acute morbidity WTP estimates vary by country (Navrud, 2001; Ready et al, 1999), and therefore, it is reasonable, for an efficiency perspective, that this country variation be preserved in the CAFE methods. I would also note that there is new work to place the benefits of hospitalization – usually calculated using the cost of illness (COI) approach – on a willingness to pay footing (Chestnut et al, 2003). This approach involves estimating a very comprehensive COI measure, then subtracting any COI-type benefits that might be captured with a WTP estimate (such as leisure time lost) and then adding in WTP estimates, in this case obtained from a special survey, for a total value of reductions in hospitalization events.

Valuation of Building Materials

The EU has been a pioneer in collecting data and developing methods to estimate materials damages and value them. The latter is based on repair and replacement costs. Generating defensible estimates of such costs is very difficult because of the behavioral issues one must take into account. For instance, if air pollution hastens the damage to a building material, it might be thought that this would lead to more rapid replacement or repair of that material. However, it may be that the normal repair and replacement cycle is too rapid to be affected by reduced air pollution damage, in which case the incremental savings in repair and replacement costs would be zero. This concern is even greater for the PM-soiling damage impact pathway where cleaning costs rather than WTP estimates are being used due to an absence of the latter information.

One overarching data question concerns the information in Table 10. What are critical thickness loss numbers based on? Since these numbers embody the behavioral elements of the maintenance/replacement choice, more information is needed on their lineage to sensibly critique the approach.

Cultural Heritage

The Methodology states that valuation of cultural heritage will not be attempted because there are only 25 studies on valuation of cultural heritage, 4 for air quality. If the studies were of reasonable quality, this seems like an ample record to me. So more clarification of this literature would be helpful.

Crop Valuation

Calculating the welfare benefits of reduced crop damage is easy to do in the way suggested by the CAFE Methodology Document and, for very small changes improvements in crop yields, would not be inaccurate, with one exception – the effects of subsidies on welfare calculations. Kopp and Krupnick (KK) (1987) show that the dead weight losses associated with many types of agricultural policies change when the crop supply function changes, as it would with an air pollution policy. As Lichtenberg and Zilberman (1986) demonstrate, policies that increase yields subject to price supports exacerbate deadweight losses, reducing the overall welfare gains from the policies. KK show that such changes in deadweight losses can offset more than one-third of the produce plus consumer surplus gains from an increase in yields. Whether the CAFE team wants to take on contentious issues about agricultural policies across the EU is another matter. But, ignoring them can significantly change the benefit estimates for this sector.

Ecosystem Valuation

I applaud the New EXT researchers who resisted the idea of using control costs as a substitute for preference-based WTP estimates to value ecosystem services and assets. In the absence of the latter information and multicriteria approach based on physical effects seems as good as any.

I would note, however, that there is a lively and growing ecosystem valuation literature in both the U.S. and Europe and wonder if this literature was reviewed during the methodology debate. See, for example, Bateman, et al (2003), Jakobsson and Dragun (1996), Banzhaf et al (2004), Kopp and Smith (1993) and Moran and Pearce (1997). An excellent resource is USEPA (2003).

Visibility Valuation

I concur with the conclusion that it would be unwise to transfer U.S. values to Europe because Europeans seem, in general, to be insensitive to reduced visibility. The idea to review such effects and possibly include them as a sensitivity analysis seems reasonable and may eventually increase awareness of this effect in Europe.

Valuation of Avoided Climate Change Damages

I concur that the literature is too thin amidst a huge set of possible effects to justify a monetary valuation approach.

Valuation Through General Equilibrium Methods

The plan is to somehow integrate GEM-E3, a general equilibrium model for European states, with the BCA to capture welfare losses throughout the economy. This is a very ambitious task, with little in the literature to guide it. The CAFE report is silent on plans for this component of the study.

Bringing the Analysis Together

This section is uncontroversial except for two items. The first is the statement that "The impacts and monetary results may also be expressed for different pollutants." My concern is with the word "may." If this effort is to mean anything for policy, it must estimate benefits for pollutants separately, to the extent possible given the jointness in damage that may exist. An aggregate valuation approach has been used all too often in the U.S. EPA, with the result that little guidance can be offered on whether one pollutant or another needs more regulatory attention.

The second concern is the statement that "Account needs to be taken of uncertainty in the comparison of the costs and benefits." This statement is very important and clear, but it is not carried through in the report to discuss exactly how uncertainties are to be represented and, beyond this, what types of representations are needed to help decision makers make more informed decisions.

Conclusions

The overall approach followed for valuation of health and other endpoints is a sound one and clearly much effort and careful thought went into the decisions made by the CAFE team. I endorse many of these decisions, as noted above, while questioning some others, notably, the choice of a single VOLY for the base case and the use of an equivalent VSL to that of adults for valuing children's premature death. I offer some advice here and provide some perspective and ideas for improving the presentations in other areas as well as pointing out some areas of where the team should probe more deeply before the methodology is finalized.

References

Alberini, A. 2004. Robustness of VSL Values from Contingent Valuation Surveys. submitted to EARE Annual Meeting, June 25-28, Budapest.

Alberini A., M. Croppper, A. Krupnick, and N.B. Simon. 2004a. Does the Value of a Statistical Life Vary with Age and Health Status? Evidence from the U.S. and Canada. *Journal of Environmental Economics and Management* 48: 769-792.

Alberini A., M. Cropper, A. Krupnick, and N.B. Simon. 2004b. Willingness to Pay for Mortality Risk Reductions: Does Latency Matter? Working Paper, University of Maryland (revise and resubmit to *Journal of Risk and Uncertainty)*

Alberini, A., A. Hunt, A. Krupnick, M. Cropper, N. Simon, A. Markandya, R. Scarpa and B. Desaigues. 2004. Age and Health and the Value of Statistical Life: Evidence from a Contingent Valuation Survey in the UK, France, and Italy. Working Paper.

Alberini, A., A. Hunt and A. Markandya. 2004. Willingness to Pay to Reduce Mortality Risks: Evidence from a Three-Country Contingent Valuation Study. Working Paper.

Alberini A. and A. Krupnick. 2002. "Valuing Health Effects of Pollution," *International Yearbook of Environmental and Resource Economics* 2002/2003 edited by Folmer H. and T. Tietenberg. Amsterdam: Kluwer Publishers.

Banzhaf, S., D. Burtraw, D. Evans and A. Krupnick. 2004. *Valuation of Natural Resource Improvements in the Adirondacks.* Washington D.C.: Resources for the Future.

Bateman, I., P. Cooper, S Georgiou, et al. 2003. Scope Sensitivity Tests for Preference Robustness: An Empirical Examination of Economics Expectations Regarding the Economic Valuation of Policies for Reducing Acidity in Remote Mountain Lakes, CSERGE Working Paper EDM 04-03.

Cameron, T.A. and J.R. DeShazo. 2004. An Empirical Model of Demand for Future Health States when Valuing Risk Mitigating Programs.

Canada-Wide Standards Development Committee for PM and Ozone. 1999. Particulate Matter and Ozone, Canada-Wide Standards for Consultation. Discussion Paper presented at National Multi-Stakeholder Consultation Workshop. May 26-28 Calgary, Alberta.

Chestnut, L., et al. 2003. *The Economic Value of Preventing Respiratory and Cardiovascular Hospitalizations*, prepared for California Air Resources Board, Sacramento, CA (May).

Clinton, S. et al. 2004. *Valuation of Health Benefits Associated with Reductions in Air Pollution: Final Report* for the Department of Environment, Food and Rural Affairs, UK (March).

Desaigues, B., et al. 2003. Monetary Valuation of Air Pollution Mortality: Current Practice, Research Needs and Lessons from Contingent Valuation. (unpublished paper).

Hammitt, J.K. and J.T. Liu. 2004. Effects of Disease Type and Latency on the Value of Mortality Risk. *Journal of Risk and Uncertainty* 28: 73-95.

Hoffmann, S, V. Adamowicz and A. Krupnick. 2003. Economic Uncertainties in Valuing Reductions in Children's Environmental Health Risks. Paper presented at the Annual meetings of the American Agricultural Economics Association. July 27-30 Montreal, Canada.

Jakobsson, K.M. and A.K. Dragun. 1996. *Contingent Valuation and Endangered Species: Methodological Issues and Applications.* Cheltenham, UK: Edward Elgar.

Johannesson, M and P-O Johansson. 1996. To Be, or Not to Be, That is the Question: An Empirical Study of the WTP for an Increased Life Expectancy at an Advanced Age. *Journal of Risk and Uncertainty* 13: 163-174.

Kneisner, T.J, W. Kip Viscusi and J.P. Ziliak. 2004. Life-cycle Consumption and the Age-Adjusted value of Life. Discussion Paper No. 459 Boston, MA: Harvard's John M Olin Center for Law, Economics and Business. (February).

Kopp R.J. and A.J. Krupnick. 1987. Agricultural Policy and the Benefits of Ozone Control. *Journal of the American Agricultural Economics Association*. 69(5): 956-962.

Kopp, R.J., and V.K. Smith. 1993. *Valuing Natural Assets: The Economics of Natural Resource Damage Assessment.* Washington, DC: Resources for the Future.

Krupnick, A.J., and M. Cropper. 1992. The Effect of Information on Health Risk Valuation. *Journal of Risk and Uncertainty*. 5: 29-48.

Krupnick, A., et al. 2002. Age, Health, and the Willingness to Pay for Mortality Risk Reductions: A Contingent Valuation Survey of Ontario Residents, *Journal of Risk and Uncertainty*. 24(2): 161-186.

Lee R.et al. 1995. *Estimating Externalities of Electricity Fuel Cycles* (Oak Ridge National Laboratories/Resources for the Future), 1994-1995, seven volumes, Washington, D.C.: McGraw-Hill/Utility Data Institute,

Lichtenberg, E and D. Zilberman. 1986. "The Welfare Economics of Price Supports in Agriculture," *American Economic Review* **76**: 1135-41.

Moore, M.J. and W.K. Viscusi. 1988. The Quantity-Adjusted Value of Life. *Economic Inquiry* 26: 369-388.

Moran, D., and D. Pearce. 1997. The Economics of Biodiversity. *The International Yearbook of Environmental and Resource Economics 1997/1998—A Survey of Current Issues,* edited by H. Folmer and T. Tietenberg. Cheltenham, UK: Edward Elgar Publishing Company, 82-113.

Navrud, S. 2001. Valuing Health Impacts from Air Pollution in Europe. *Environmental and Resource Economics* 20: 305-329.

OMB (Office of Management and Budget). 1996. *Economic Analysis of Federal Regulations Under Executive Order 12866*. Washington D.C.: OMB. Available at http://www.whitehouse.gov/omb/inforeg/riaguide.html. Last accessed September 2004

OMB. 2000. *OMB Guidelines to Standardize Measures of Costs and Benefits and the Format of Accounting Statements* (M-00-008). Washington DC: OMB. Available at http://www.whitehouse.gov/omb/memoranda/m00-08.pdf. Last accessed September 2004.

Rabl, Ari. 2001. Relationship between life expectancy and probability of dying, unpublished paper, Paris, France: Ecole des Mines, (April).

Ready, R.C. et al. 1999. Benefit Transfer in Europe: Are Values Consistent Across Countries? Paper presented at the EVE Workshop on Benefit Transfer, October Lillehammer.

Shepard, D.S. and R.J. Zeckhauser. 1982. Lifecycle Consumption and Willingness to Pay for Increased Survival, in *The Value of Life and Safety,* edited by M.W. Jones-Lee. New York: North Holland Publishing.

Silva, P. and S. Pagiola. 2003. *A Review of the Valuation of Environmental Costs and Benefits in World Bank Projects.* Environmental Economics Series, Paper No. 94 (December).

Slovic, P. 1987. Perception of Risk. *Science* 236: 280-285.

Slovic, P. 1992. Perception of Risk: Reflections on the Psychometric Paradigm, in S. Krimsky and D. Golding, eds., *Social Theories of Risk* Westport, Conn.: Praeger.

Strand, J. 2001. Public and Private Goods Values of Statistical Lives: Results from a Combined Choice Experiment and Contingent Valuation Survey. University of Oslo, Working Paper.

Strand, J. 2004. Individual and household value of mortality reductions with intra-household bargaining. Oslo, Norway: Department of Economics, University of Oslo, (January).

USEPA. 1999. *The Benefits and Costs of the Clean Air Act 1990 to 2010*, U.S. EPA, Report No. EPA-410-R-99-001. Springfield, VA: NTIS.

USEPA. 2000. *Guidelines for Preparing Economic Analyses*. EPA 240-R-00-003 Washington DC: USEPA

USEPA. 2003. *Ecological Benefits Assessment Strategic Plan, Appendix B: Literature and Information Resources for Ecological Benefits Assessment*: Draft (September).

USEPA. 2004. *Final Regulatory Analysis: Control of Emissions from Nonroad Diesel Engines*. EPA 420-R-04-007. Office of Transportation and Air Quality. Washington DC: US EPA

Viscusi, W.K. and J. Aldy. 2003. The Value of Statistical Life: A Critical Review of Market Estimates throughout the World. *Journal of Risk and Uncertainty* 27(1): 5-76.

Viscusi, W. K., W. A. Magat, and J. Huber. 1991. Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Trade-Offs for Chronic Bronchitis. *Journal of Environmental Economics and Management* 21(1): 32-51.

Washington Post. 2003. "Under fire, EPA drops the 'senior death discount'." C. Skrzycki, reporter, May 13, E01.

Chapter 6
Addressing Uncertainty

General approach in CAFE

The CAFE methodology for addressing uncertainty is still evolving, with the methodology document doing little more than listing ideas. Issues of addressing uncertainty are also related to issues for the reporting of results as well as issues for the use of multi-criteria assessment techniques to capture the non-quantifiable effects. Therefore, the Panel cannot comment on the plan for presenting uncertainties.

However, the Panel can offer their suggestions for the treatment of uncertainty.

(1) We believe that a distinction should be made between how uncertainty should be addressed on the technical side and how uncertainties should be reported. This distinction is important because reporting uncertainty is complicated and should be done judiciously. But the necessary information to make such judgments must be available.
(2) On the technical side, the model for estimating benefits and costs (as well as any effectiveness measures used for cost-effectiveness analysis) should be capable of fully addressing statistical uncertainty, in the sense of capturing standard errors around all key parameters and promulgating these distributions through the analysis to yield probability distributions of benefits, costs, effectiveness measures and net benefits
(3) The technical treatment of uncertainty should also address what is termed "model uncertainty," e.g., uncertainties about the appropriate C-R functions, the appropriate valuation study or approach. The CAFE Methodology appears to favor several different types of sensitivity analyses to do this, which we support. However, while there has been an effort to identify some key model uncertainties, there has not been an effort to decide how these different choices about sensitivities will be organized. The spirit of this point was stated in some stakeholder comments calling for specifications of scenarios.
(4) Another technical point is to highlight and consider the differences between variability (or process uncertainty) and uncertainty. Capturing variability in air quality modeling, for instance, requires modeling different types of episodes, while capturing uncertainty requires developing and implementing distributions for rate constants.
(5) It may be useful to subdivide the factors that drive differences between models, such an endogenous processes (discrepancies among input parameters across models), exogenous processes (different model structures), differences in value judgments (discount rates)
(6) It would be useful to elucidate the different types of analyses one can do, once uncertainties have been estimated, i.e., importance analysis, expected value of information analysis, etc.
(7) It may be useful to consider different types of probabilistic models (e.g., Monte Carlo simulation)
(8) In the U.S. we have adopted the approach of including tables on non-quantifiables that indicates how important they might be and the directional effect (which could be unknown) of their inclusion of the results (e.g., benefits would increase). The MCA approach would appear to address this issue,

perhaps with a larger edifice than is needed.On the reporting side, the EU should obtain some feedback from its key policy makers, as well as key stakeholders, about how they would like to see uncertainties portrayed and analyzed. Information overload is a potential problem. ADDin practice how to do this.

(9) A related concern is the cost and scope of the effort to address uncertainties. Major rulemaking processes demand more attention in this regard than smaller ones. What is the minimum scale of such efforts? How should efforts be ramped up, i.e., out of all the possible ways of estimating and expressing uncertainties, which are more valuable than others?

(10) We agree with the viewpoint expressed in the third paragraph of section 3.4.6 that the slope of the health function is only one of the important areas of uncertainty. On the benefit side, the valuation and air quality modeling are also important areas of concern. On the cost side, it would be instructive to examine *ex ante* versus *ex post* cost estimates for various pollution abatement strategies.

Comments on Key Areas of Uncertainty

The CAFE team has identified several key areas of uncertainty. These include:

Mortality impacts of PM
Valuation of the above
Chronic bronchitis and PM
Speciated PM effects
Valuation of ecosystem benefits
Cost uncertainty

The Panel is in broad agreement that these are important areas of uncertainty, although we feel that the highest value of information could be obtained by understanding speciation and cost uncertainty, because so little is known here. The other areas identified by the CAFE team have uncertainties that have been much discussed and where the element of "surprise" would be relatively small. Such elements could be handled through sensitivity analysis.

In addition, the panel suggests that the study team consider explicitly the following important areas of uncertainty:

1. Infant mortality
2. Ozone mortality associated with both short- and long-term exposure
3. Morbidity associated with long-term exposure

Chapter 7
Multi-Criteria Assessment (MCA)

We read with interest the approach proposed for presenting qualitative results using the "modified MCA" method (para. 3.3.2). The aim of the methodology – to provide a greater understanding of all impacts – is a goal that many of us have sought in the past with varying degrees of success. A systematic and thorough, but concise, attempt would be a valuable contribution to future decision-making.

The example data sheets provide a good indication of how results might be presented, and these have the potential to be useful tools. However, one omission from the example data sheets appears to be a discussion of uncertainties. While these are partly covered throughout the various sub-heads, bringing together the current understanding of uncertainties under a separate subhead could, we believe, be useful.

We are pleased to see the list of consultations planned and feel that in the time available these will provide adequate review of the results from the analysis.

The method of using *, ** or *** as rating for the MCA is, on first reaction, a useful one. It provides broad definitions for describing impacts in very general terms. It is possible to imagine some variation to the words proposed as descriptors but those chosen are a useful starting point. However, we do wonder how significance is defined. Would it be large enough to alter a choice? Change the balance of benefits and costs? Also the choice of the 3-class system is, we believe, the minimum possible number of categories and therefore provides the simplest possible interpretation. It would be possible to envisage a large number of classes but we are not sure that these would add to an overall understanding of the issues.

The Panel also believes that the MCA should mix quantified and unquantified costs and benefits. Doing so would give visibility to the nonquantified effects, and in our view, would not be overly confusing.

An alternative rating system – and one that combines quantitative and qualitative endpoints -- is something like that used in *Consumer's Reports* and appears in Burtraw et al. (1998). Ratings are a five-point scale from "high" to "low," capturing answers to specific questions about the nature of uncertainties. These questions would head the columns of a table while the rows would be endpoints (e.g., mortality). The questions in the cited article include:

(1) Links between science and economics: Are benefit endpoints well established? Does science provide the information needed for economic analysis?
(2) Economic Methods: Are economic methods adequately developed?
(3) Data availability: Are sufficient data available to do a benefit assessment?
(4) Expected benefits: Are benefits likely to be large?
(5) Value of additional information: With the goal of improving benefit estimates, what is the relative short-term return on investment in information acquisition, R&D, etc.?

It is worth noting that this article also has what has become the standard treatment of uncertainty in USEPA regulatory analyses – a table listing "Major Uncertainties, Omissions and the Direction of Bias." This table has a column to name the uncertainties and omissions, another column indicating direction of bias in its exclusion and another describing the nature of the issue.

Chapter 8
Miscellaneous Comments

Social Effects

This section lacks focus and is not well developed. It is not clear what is being proposed as a work product. There are basically three issues, at least, that could be addressed. These are:

1. Are groups with lower socio-economics scores (SES) exposed to higher levels of pollution? Based on the data presented, in three of the four cities this appears to be the case. It would be useful, however, if the study team examined and discussed how general these findings are to other cities in Europe given the proclivity of many higher SES groups to live closer to the city center in many European cities. The association of SES and exposure to traffic from to major roadways as well as stationary sources using GIS and other techniques could be useful areas of inquiry.

2. For the same levels of exposure, do lower SES groups (that presumable have related problems with access to health care and other difficulties related to poverty) exhibit greater responses to air pollution? From the U.S., there is preliminary, but inconclusive data to suggest this may be the case. The Krewski al. reanalysis of the ACS and Harvard cohorts did show that those with lower education exhibited a stronger response to long-term exposure. However, lower education is associated with greater exposure to air pollution so this may be reflecting greater exposure rather than greater intrinsic sensitivity. In addition, there is some evidence that reduced residential mobility in this subgroup (and therefore lower exposure measurement error) may be responsible for the higher effect estimate.

3. What is the expected incidence of the predicted change in pollution exposure and associated health effects?