

RISK-BENEFIT ANALYSIS METHODS FOR VEHICLE SAFETY DEVICES

Kimberly M. Thompson
John D. Graham
Harvard Center for Risk Analysis
John W. Zellner
Dynamic Research, Inc.
United States of America
Paper Number 01-S4-O-340

ABSTRACT

Vehicle safety devices, similar to new pharmaceuticals and medical devices, may be associated with injury risks as well as injury benefits. Available analytical methods from the public health, medical and vehicle safety fields are described. A literature review is provided that includes an overview of relevant principles of risk analysis, risk-benefit terminology, fields of application, types of risk-benefit analysis, methods of quantification, assumptions, data needs, treatment of uncertainties, and risk-benefit criteria. Several applicable quantification methods are further described, including Quality Adjusted Life-Years, Disability Adjusted Life-Years, and Normalized Injury-Fatality Costs. Data input sources are described, including accident sampling and analysis, and paired comparison test and simulation methods. Example applications are presented for car seatbelts, head restraints, driver and passenger airbags; motorcycle leg protectors and airbags; and all terrain vehicle rollover protection structures. In the context of historical trends in the public health, medical and transport safety fields, typical risk-benefit criteria are presented and described. Discussion and recommendations regarding potential applications, further development and standardization issues are provided.

INTRODUCTION

Like new pharmaceuticals and medical devices, vehicle safety devices may be associated with risks as well as benefits. An illustration of the risks of a safety device can be found in the growing number of nonfatal injuries and fatalities that caused by the installation of frontal-crash airbag systems in new passenger cars and light trucks. The sobering experience with airbag systems has generated renewed interest in the application of analytical

methods, such as risk-benefit analysis, to decisions about the installation of new safety devices.

In this paper, we describe several analytical methods that are currently employed in medicine, public health and vehicle safety research and that have applicability to risk-benefit assessment in the motor vehicle safety field. We also review related literature regarding risk-benefit principles, terminology, typical applications, types of risk analysis and quantification methods, and discuss methods of treating uncertainties and risk-benefit criteria. Finally, we provide example applications in the vehicle safety field and recommendations regarding the use of risk-benefit analysis in establishing standards or in research.

OVERVIEW OF RISK-BENEFIT ANALYSIS

The vast literature of risk analysis continues to expand rapidly and grow in its diversity with regard to both fields of application and types of analyses. The former includes fields such as:

- Medicine (pharmaceutical and medical device approvals)
- Environmental protection
- Consumer safety
- Transport safety
- Occupational safety
- Military systems
- Legal, litigation
- Casualty-loss (insurance)

with broadening involvement of government agencies, organizations, corporations, universities and research institutes.

While risk analysis can involve either qualitative or quantitative components, or both, very little standardization exists in the field as a whole, and

remarkably, even within subdisciplines. The International Organization for Standardization (ISO) is drafting a risk assessment terminology, and this is a task that is made more difficult by the wide diversity of applications, methodologies, and existing terminologies.

Professional societies, like the Society for Risk Analysis (SRA) provide forums at both national and international levels for discussion and publication of risk-benefit methodologies, policies, and issues, although no single society can currently claim coverage of all elements of the wide and diverse field.

During the last three decades, the motivation for using risk assessment in the US resulted primarily from legal developments (ie, both regulatory and tort) in the previously noted fields. For example, both legislative measures (e.g., 1990 Clean Air Act) and judicial rulings (e.g., 1980 Supreme Court occupational chemical exposure ruling) have increasingly required Federal agencies to conduct risk assessment in support of agency decisions. Even when this is done, an agency decision can be overruled by Federal courts if the risk assessment "is judged to be of insufficient technical quality to meet prevailing standards of judicial review" (Center for Risk Analysis, 1992). In addition, legislative efforts (Committee on Science and Technology, 1979, 1980) and more recently, ad hoc inter-agency efforts (i.e., Federal Coordinating Council on Science, Engineering and Technology), have been directed toward harmonization of US government risk assessment methods.

Other US agencies (eg, NHTSA) are chartered by Congress to base decisions considering the costs and benefits of potential actions, which may include risk analysis as a component but is not synonymous with risk analysis to some analysts. In general, cost-benefit analysis differs from risk analysis in terms of focus and methods. For example, in a cost-benefit analysis the risk component might or might not be included in an analysis of the "net benefit" of a device. Further, while cost-benefit analyses may focus on analyzing both the positive and negative effects of an intervention that is intended to improve the health or safety of the public, they may not consider the treatment of risk and uncertainty about risk.

Overall, risk-benefit analyses can be categorized according to the situation and may include cases where:

- Neither risks nor benefits can be readily quantified
- Risks can be quantified, but benefits cannot readily be quantified (or vice versa)
- Risks and benefits are quantified in very different terms or units (e.g., risk of radiation exposure versus lower cost of energy)
- Risks and benefits are quantified in similar terms (e.g., change in costs of illness or injuries, and the related unintended side effects)

The latter category is the focus for analysis of pharmaceutical and medical devices as well as vehicle safety devices.

Early work in the pharmaceutical and medical device area examined the tradeoffs between risks and benefits where both were expressed in similar units (e.g., Walker and Asscher, 1985). Yet, despite the enormous numbers of drugs and devices being approved in each country each year, most risk-benefit evaluation is still done on a subjective, ad hoc basis, by panels of experts (e.g., Doug-Tyson, 1994; Lasagna, 1994).

Table 1 summarizes a set of principles for risk analysis (Center for Risk Analysis, 1992). Many of these principles indeed are applicable to the vehicle safety device area, and should be considered by both industry and government policy and decision makers.

Fischhoff et al. (1981) discuss several typologies (including market-based approaches, decision analysis, and historical precedent) and provide criteria for quantitative risk-benefits decisions. Fischhoff et al. (1981) discuss the limitations of each type in detail.

At a minimum, any analytical approach to weighing the risks and benefits of a new safety device should satisfy the following minimal criteria:

- Incorporate information about both the fatal and nonfatal injuries that are caused and prevented by a safety device
- Incorporate information about nonfatal injuries of varying degrees of severity into the risk-benefit determination

- Have an explicit procedure for weighing risks and benefits that occur in different subpopulations of users, where the subpopulations may be defined according to objective characteristics such as age, gender, physical stature and weight of the user
- Be preference-based in the sense that the health-state preferences of at-risk users play a significant role in the ultimate risk-benefit determination.

Preference-based approaches are attractive in those societies where the principles of consumer sovereignty and citizen participation are given some degree of deference in decision making.

In the sections that follow, we describe three analytical methods that have some applicability to safety device design. The first two are currently widely employed in the medical and public health literature. The third is an example of a new technique that has been used in vehicle safety research. Each of the methods presumes that information is available on the effectiveness and risks of the technology under study. Such information may be in the form of clinical, epidemiological, experimental, or simulation data, for example.

THE QUALITY-ADJUSTED LIFE YEAR

When a pharmaceutical, medical device or surgical technique is evaluated, the quality-adjusted life year (QALY) is often used as the metric for comparison (Zeckhauser and Shepard, 1976). QALYs combine information on duration of life and health-related quality of life into a numerical index. New (or existing) medical technologies are analyzed to determine whether their application in specified patient populations will increase or decrease QALYs (Weinstein and Stason, 1977), where the net change in QALYs is defined as the number of QALYs saved or preserved by the intervention ("benefits") minus the number of QALYs lost due to application of the intervention ("risks"). In principle, this approach should be directly applicable to analysis of safety devices.

The logic behind QALY measurement is rooted in the decision analysis. Consider a scale for the rating of life years that ranges from 0 to 1.0, where 1.0 is perfect health and 0 is the worst health state (usually death). Any health state between these extreme values can then be rated somewhere between 0 and 1.0. If a year of life in a wheelchair is rated 0.6

Table 1.
Key Principles of Risk Analysis (Center for Risk Analysis, 1992)

- | |
|--|
| <ol style="list-style-type: none"> 1. Estimates of attributable health [or safety] risk should make use of the best available science. 2. Since reputable scientists often do not agree about how to assess risk, scientific disputes should be acknowledged in risk assessments. 3. When hard data are lacking, risk assessments should be explicit about any assumptions and indicate how sensitive results are to plausible changes in assumptions. 4. Good risk assessments usually develop a central estimate of risk as well as upper and lower bounds on risk that acknowledge the extent of scientific uncertainty. 5. Public policy decisions about acceptable risk require public participation and the application of democratic principles. 6. There is no quantitative level of risk that is universally acceptable or unacceptable; it depends on the circumstances, the people affected, and the decision context. 7. Good decisions about health risk require consideration of other cherished values such as quality of life, equity, ecological health, personal choice, and economic welfare. 8. Programs to reduce risk should be designed to avoid unintended side effects that may increase risk. 9. When risk reduction is desired, economic incentives and information should be considered in addition to conventional command-and-control regulation. 10. The context in which a risk occurs (e.g., voluntary versus involuntary risk) may influence public reaction to risk as much as the magnitude of the risk in question. |
|--|

on the QALY scale, that means that a person would be indifferent between the known outcome of living in a wheelchair for a year and a 60%-40% lottery where the two possible outcomes are living a year in perfect health and experiencing the worst possible health state (usually death). Alternatively, 10 years in a wheelchair is rated the same as 6 years in perfect health (Torrance, 1986).

Consider a safety device that will, once applied in the entire fleet of vehicles, prevent 1,000 deaths and 50,000 long-term injuries each year in the United States. The number of QALYs saved by this device is computed in a three-step process.

The analyst first computes the number of years of life preserved by the device. If the average death causes a loss of 40 life years and the average long-term injury reduces life expectancy by 0.5 years, then the total gain in duration of life is 65,000 life years (1,000 x 40 plus 50,000 x 0.5).

The second step is to apply appropriate quality adjustments to the years of life that are saved by the device. Suppose that the safety device does not restore people to perfect health status but instead adds years of life at a 0.9 quality level. Under these assumptions, the lifesaving effects of the device create an additional 36,000 QALYs (1,000 x 40 x 0.9) rather than the 40,000 computed above.

The third step accounts for the effect of the device on the quality of life among people who otherwise would experience long-term nonfatal injuries. For those long-term injuries that are mitigated by the device, suppose that the average quality of life improves from 0.6 to 0.9 on the scale. That means that the 50,000 people who experience less severe nonfatal injuries will each gain 0.3 QALYs per year for the rest of their life, for a total of 600,000 additional QALYs gained (assuming the benefit persists for 40 years per person). In this hypothetical example, the number of QALYs gained from the increase in duration of life (36,000) is smaller than the number of QALYs gained (600,000) from the increase in health-related quality of life among people experiencing long-term nonfatal injuries.

If we add the additional complexity that the safety device causes some fatalities and long-term injuries, the extension of the method is straightforward. The number of QALYs lost by these "side effects" is computed and then subtracted from

the gross number of QALYs gained by the device. An evaluation of driver and passenger airbags that employed this approach provides an example (Graham et al, 1997).

Using QALYs as the metric for evaluation requires the analyst to make some implicit assumptions about the nature of preferences for duration of life and quality of life, including the aggregation of those preferences (across people and time). This requires making some very important assumptions (Pliskin et al, 1980; Johannesson, 1996).

First, the QALY method assumes that each QALY is of equal value, regardless of which member(s) of society experience(s) the benefit. Under this egalitarian assumption, details like where these people live, their income or asset positions, and their genders, ages, and ethnic backgrounds are irrelevant. This assumption is in striking contrast to some economic methods that might give more weight to the lives of specific populations that are judged to be economically more productive or alternatively more in need of protection.

Another key and closely related assumption is that QALYs can be aggregated, regardless of how they are distributed among people. For example, saving 10 QALYs for one person is considered equivalent to saving 1 QALY each for 10 people. The willingness to aggregate QALYs across people reflects the influence of utilitarianism on the QALY method.

Second, the QALY method assumes that people are willing to trade off years of life in a given impaired health state for fewer years in an ideal health state at a constant rate, irrespective of the number of years of life spent in the state. Thus, if 10 years of life with chronic back pain is judged to be equivalent in value to 5 years of life in perfect health, then it is assumed that 20 years with chronic back pain is equivalent in value to 10 years in perfect health. Analysts call this the "constant proportionality tradeoff assumption."

Third, the QALY method assumes that people are indifferent between survival curves that have the same life expectancy. Analysts call this the "risk neutrality assumption" concerning life years and quality of life. For example, people are assumed to be indifferent between an opportunity to gain 5 additional life years for certain and a 50-50 lottery on a gain of 10 to 0 life years. Although it may seem

unreasonable to assume that individuals are risk-neutral toward QALYs, Arrow and Lind (1970) provided a rigorous argument that society should be risk neutral when allocating lifesaving resources, because that posture toward uncertainty saves the most QALYs in the long run.

Some analysts have argued that some or all of these assumptions are unreasonable (at least for particular applications). For example, the Healthy-Year Equivalent (HYE) has been proposed as a viable alternative to the QALY because it is more general and thus does not rely on the assumptions concerning constant proportional tradeoff and risk neutrality (Gafni et al, 1993). Yet HYE has not had much impact on practice in medicine and public health because the data requirements for implementation are too severe to be practical in most applications.

This is not to imply that using QALYs is easy. The data requirements for the QALY method are substantial. The analyst must have access to data on the impact of the intervention of interest on life expectancy and health-related quality of life. In the field of trauma, life expectancy for immediate deaths can be approximated by use of life tables, a standard tool in public health. Life expectancy losses associated with nonfatal impairments have been the topic of some specialized studies concerning head and spinal cord injury, but the database contains significant gaps.

Until recently the quality-of-life decrements associated with nonfatal trauma were unknown and thus analysts were forced to rely on QALY weights derived from general health utility scales. However, MacKenzie and colleagues at Johns Hopkins University recently created the Functional Capacity Index (FCI), a preference-based scoring system for nonfatal injuries that permits a mapping of injuries on the Abbreviated Injury Scale into QALY values. As currently designed, the FCI is aimed at assessing the impact of injuries that are known to have functional limitations one year post-injury. Applications to adult populations are complete and pediatric applications are in progress. Clinical validation studies of the FCI are also nearing completion and are likely to cause revision of some of the initial published values. Although the FCI is still in the developmental stages, it is already at least as sophisticated as the health-utility scales that are now widely used by clinicians and public health professionals interested in disease prevention.

In summary, it is now feasible to apply the QALY method to any safety device whose effectiveness and side effects can be described in terms of frequency of nonfatal injuries by severity-level and fatalities. At least in the US, the influence of QALYs in medicine and public health is likely to grow in the future because an Expert Panel commissioned by the US Public Health Service of the Department of Health and Human Services recently recommended use of QALYs as the primary measure of net effectiveness in studies intended to measure the cost-effectiveness of medical and public health interventions (Gold et al., 1996).

THE DISABILITY-ADJUSTED LIFE YEAR

Although the QALY metric is commonly employed by health professionals in many developed countries (particularly Northern Europe, Canada, and the United States), a close cousin to the QALY called the disability-adjusted life year (DALY) is promoted by the World Bank and the World Health Organization and is widely used in developing countries around the world (Murray, 1994). Like the QALY, the DALY combines information on duration of life and quality of life into a numerical index that can be aggregated across people. However, we note some important differences between DALYs and QALYs.

First, the DALY approach evaluates health states in terms of a single dimension, degree of disability, whereas the QALY approach incorporates information on a wide range of dimensions of quality of life that clinicians and patients believe are important to health status. It is likely however that many of these dimensions of health ultimately have a discernible impact on functional status and thus may be captured by a disability-oriented approach such as DALYs. The disability scale has six levels with expert judgment employed to assign weights to the different levels of the scale. For example, a disability that causes limited ability to perform at least one activity in life is assigned a value of 90.4 on a 0 to 100-point scale, where 100 is good and 0 is bad. A case of disability that requires a patient to have assistance in all aspects of daily living is scored 8 on the 100-point scale.

Second, the DALY approach weights some life years differently than others depending upon the age of the affected person. The weighting function is an inverted "U," with the highest weights assigned to years of life in the middle of the lifespan and the

lowest weights assigned to years of life at the beginning and end of the lifespan. The rationale for this particular weighting function can be found in the theory of human capital and life-cycle approaches to the monetary valuation of human life (Becker, 1993; Shepard and Zeckhauser, 1984). The current age weights used with the DALY method cause the value of a healthy life year to peak at age 25, where a year of life is assumed to be 50% more valuable than a health year of life at either age 10 or age 55. The function assumes that a healthy life year at age 5 or 85 has one-third the value of a healthy life year at age 25. It is important to realize that the DALY approach applies the age-weighting function in addition to any health-related differences in quality of life that may be a function of age. Thus, any disabilities associated with chronic illnesses at older ages are scored first, before the additional impact of the age-weighting function is applied.

Like the QALY method, the DALY method has been the target of significant criticism in the medical and public health literatures (Nord, 1992). Yet there has been no practical alternative proposed to the DALY and thus the DALY method is dominating burden-of-illness and cost-effectiveness applications in developing countries throughout the world.

NORMALIZED INJURY-FATALITY COSTS

Normalized Injury-Fatality Costs (referred to as NIC) is a research-based tool with some initial applications in the vehicle safety field. The NIC is based on data originally published by Miller, et al., (1990), which currently is the only known injury database that defines both medical and ancillary costs, by both the Abbreviated Injury Scale (AIS) (eg, AAAM, 1990) and by body region.

The NIC model was formulated by Newman et al. (1992) in order to provide a tool for examining vehicle safety devices that may reduce injuries to one body region in some situations, but increase injuries to the same or other body regions in other situations.

The basic assumption of NIC is that all health-state values and preferences are reflected in the actual lifetime costs of medical treatment, "ancillary" costs (which include costs of permanent partial incapacity), and fatality costs to society.

The normalization of NIC is by the average total economic costs of one fatality, including the loss of wages and household productivity. Consequently,

similar to QALY and DALY, an accident with no injuries is said to have an NIC equal to zero and a fatal accident is said to have an NIC equal to 1.0. In the current NIC formulation, a medical and ancillary cost based on AIS is defined for head, chest, abdomen, and lower extremities. Currently, two critical injuries (spinal and upper extremity) that should be added are excluded, because of yet to be solved problems in monitoring for these injuries in crash tests.

NIC intentionally neglects so-called "pain and suffering" costs, as these were believed to be subject to enormous variations (eg, from zero in many regions to millions or even billions of dollars awarded by some juries), depending on local, state and national legal and socio-economic factors, and because these were not considered to be tangible economic costs.

Potential tradeoffs in NIC between, for example, frequent minor injuries and rare fatal injuries are handled separately under risk-benefit criteria, described subsequently. A standardized formulation of NIC based on Newman et al. (1992) is formalized in International Standard ISO 13232-5 (1996), intended for research evaluation of rider crash protection devices fitted to motorcycles.

The costs of injuries derived by Miller et al., (1990) were based on data from the U.S. National Highway Traffic Safety Administration's National Accident Sampling System (NASS), National Crash Severity System (NCSS), and Fatality Analysis Reporting System (FARS), and from the Detailed Claims Information of the National Council on Compensation Insurance.

NIC is a function of the AIS injury severities to each body region, and is a combination of several components that we describe briefly below:

- Cost of survival, which is a function of:
 - Medical cost
 - Ancillary cost including
 - Household productivity
 - Lost wages
 - Work place costs
 - Legal costs
 - Permanent partial incapacity
 - Probability of survival
- Cost of fatality

Medical Costs

Medical costs include medical care and initial and subsequent hospitalization associated with the injuries. Injury severity probabilities (or observed normalized frequencies) are multiplied by the associated medical cost for each AIS level and summed to produce the medical probability costs of the entire injury distribution for each body region. The medical costs for such injuries are then directly extracted from the cost data, without any need for further manipulation. The highest medical cost of all injured body regions is selected because it is assumed that medical costs are controlled by the most costly injury.

Medical probability costs are calculated independently of ancillary probability costs to allow for more specific cost definition. It may be of interest, for example, to know what proportion of the total cost arises from the medical component.

Ancillary Costs

Ancillary costs include household productivity, lost wages, workplace costs, and legal costs but exclude "pain and suffering," which is not a tangible cost and is subject to large regional and case-to-case variations.

The calculations of injury probability ancillary costs for the head, thorax, and abdomen are performed using the same methods as described for the medical costs. Ancillary costs for the lower extremity injuries are derived from the Permanent Partial Incapacity (PPI), described next. Since PPI is a function of disability, ancillary costs are more significantly affected by PPI, than are medical costs.

Permanent Partial Incapacity- Injuries of the same AIS may involve significantly different disability outcomes, and costs that cannot be recognized if lower extremity injuries are differentiated by AIS alone. The permanent partial incapacity index is used for lower extremity injuries to address the problem of diminished sensitivity to the AIS. The PPI approach, developed by Farisse et al. (1983), assumes that injuries involving the joint are more serious than injuries of the shaft and that multiple leg injuries are more serious than single injuries. Leg injuries are assigned a PPI value representative of the expected percentage of disability to result from the injuries. This simple solution increases sensitivity to injury type as well as

to the number of injuries. Since the PPI represents a percentage of disability it can be directly applied to disability costs.

A correlation between AIS and PPI must be established if costs are to be evaluated as a function of PPI, because cost data are based on AIS. The closest cost equivalent to a complete loss of function in the lower extremities is assumed to be equivalent to the ancillary cost of an AIS 4 spinal injury causing motor/sensory loss in both legs. A partial disability, reflected by a PPI value of 20%, is estimated to cost 20% of the AIS 4 spinal injury. The highest ancillary cost of all body regions is selected because as with medical costs, the ancillary costs often are driven by the most costly injury.

Fatality Cost

The cost of fatality, obtained from Miller, et al., (1990) is defined independently of body region. The current methodology treats fatality costs separately to provide greater insight into the injury scenario.

Probability of Fatality- All AIS 6 (i.e., "maximum" injuries), regardless of body region are assumed to be fatal. Although the proportion of survivors may be growing with improved medical care, the cost data, which was collected in 1982-85 by Miller, defines costs for AIS 1 through AIS 5 and for fatalities, AIS 6 level injuries, and/or from combinations of less severe injuries (AIS < 6) in addition to the probability of death from injuries of AIS < 6. The probabilities of all these possibilities are accounted for in the formulation.

Due to Non-AIS 6 Injuries- Several methods for calculating the probability of death have been proposed, the most well-known being the Probability of Death Score, PODS (Somers, 1981). The odds of death or PODS, is the ratio of the probability of death to the probability of survival. Unfortunately, using this metric implies that the probability of survival for the given population must be known. In the case of crash test and computer simulations, these data may not be known. PODS and other models base the fatality estimates on only the two most serious injuries. This provides the greatest improvement in goodness of fit, but in no way precludes the use of multiple injuries.

The method used in NIC for estimating the probability of death from AIS < 6 injuries is defined by Ulman and Stalnaker (1986). The rates proposed

by Ulman and Stalnaker are based on data obtained from the National Crash Severity System (NCSS) database, and are derived from separate regression equations for each AIS level. Fatality estimates are based on the three most severely injured body regions. Ulman and Stalnaker define a probability of death for each combination of three AIS injuries from 1-0-0 to 5-5-5. This approach integrates well with the crash test and simulation data. The injury severity probabilities are used directly to calculate the probability of occurrence of each triple AIS combination. No other description of the population is required.

The use of three body regions allows for the differentiation of single and multiple injury outcomes. Since cost data do not distinguish between single and multiple injuries and since the proportion of costs attributable to multiple injuries is not known, the fatality rate is the best measure.

The data needs for NIC, as for QALY and DALY, are quite extensive, but similar to what is typically collected in AIS-based vehicle accident data, or as in ISO 13232, paired comparison crash tests or simulations.

RISK-BENEFIT CRITERIA

Regardless of whether injury risks and benefits of a safety device are calculated using QALY, DALY, NIC or some other formulation, ultimately some quantitative criteria are needed to evaluate the outcome.

Such risk-benefit criteria need to be founded on principles such as those listed in Table 1, which allow for democratic and situational variations.

An important and fundamental aspect of risk-benefit criteria is the mathematical form of the comparison between risks and benefits. The two basic forms of comparison are the *difference* between risk and benefit (sometimes referred to as “net benefit”); and the *ratio* (or proportion of risk to benefit. These two forms are not equivalent, but rather are fundamentally and philosophically different, and to a large extent reflect major differences in social priorities.

That a device should have a “net benefit” (i.e., benefits minus risks is a positive number) reflects a social priority on “the greater good” for society as a whole. As a perhaps extreme example, this might correspond to a device that in saving 60 lives and

causing the loss of 40 lives, is considered to be acceptable (the net benefit of 20 lives being a positive quantity).

That a device should have a “very small” risk-to-benefit ratio (i.e., the risks divided by the benefits) reflects a social priority on “the rights of the individual.” An extreme example of this would be a device that saves a million lives and causes the loss of one life being considered unacceptable (the ratio in this case being 1/1,000,000 or 0.0001 percent).

Each of these forms of criteria has been used in the past to evaluate devices and other health or safety interventions. Regardless of whether the two examples cited would be judged to be ethically appropriate by a given society at a given time, they do represent the two great socio-political forces of the past 300 years. Therefore, rather than giving preference to either one, we suggest that both forms of criteria (i.e., difference and ratio) should be considered when evaluating any risk-benefit outcome.

In addition to the form of the criteria, in some applications it may be useful to express risk-benefit results on a “per accident” basis, as well as on a “per beneficial case” and “per harmful case” basis. The latter distinction is important in order to address to situations in which a large number of small injury benefits may outweigh a small number of serious (i.e., life threatening) injury risks. This method of quantifying results again reflects a social concern for “the rights of the individual.”

As it seems likely that the historical tension between these two different social priorities will continue into the future, we suggest that all of these criteria and their respective priorities be considered in the evaluation of a safety device. As stated in a recent proposed amendment to ISO 13232 (Kebschull, et al, 2000):

- the average injury benefit per accident should be greater than the average injury risk per accident (i.e., benefits minus risks are positive), *and*
- the risk/benefit ratio should be less than [a very small number]¹

¹ Data from Malliaris et al (1982) for car safety belts using the HARM index suggest that 7 percent risk/benefit ratio is acceptable to the public for that device. Data from Iijima et al (1998) for pre-1998

- the average injury benefit per beneficial case should be greater than the average risk per harmful case

Each of these suggested, general criteria can have their own mathematical formulation (see example formulation provided by Kebschull et al, 2000).

Risk-Benefit Criteria with Varying Population Segment

In clinical applications, it is sometimes sufficient to demonstrate that a proposed treatment (even with known side effects) will increase net QALYS or net DALYS in a defined patient population.

For a mandatory or uniform safety device standard applicable to the entire population who ride in passenger cars, it may not be sufficient that net QALYS or DALYS for the population as a whole will increase (or that the risk/benefit ratio is acceptably small), since this may vary substantially for different segments of the population. In other words, a complication arises if a safety device increases net QALYS or DALYS for some users, but reduces net QALYS or DALYS for other users.

Recent experience with passenger (as opposed to driver) airbags in the US has demonstrated that a mandatory performance or design standard may not be acceptable to the public even though it increases net QALYS or DALYS for the population as a whole. Graham et al. (1997) estimated that passenger airbags in the U.S. (i.e., those designs sold from 1990 to 1997, prior to airbag depowering) have saved roughly 5 life years for every year of life lost due to airbag-induced injuries. Yet the lost life-years have been concentrated in an identifiable subpopulation, children under the age of 10. When the effects of airbags were evaluated for this subpopulation, net mortality increased and net life expectancy reduced following the installation of fully powered airbags (Graham, et al, 1997).

US public and Congressional concern about the airbag was not mollified by information on the passenger airbag's overall ratio of benefit to risk for all passengers. People demanded that the ratio of

car airbags suggest that 3 percent is acceptable when considering all occupants, but 12 percent is unacceptable when considering right front passengers only, which include substantial numbers of children and small adults.

benefit to risk for children be improved, even though such steps might compromise some of the benefits of the device for mid-sized adult male users. Interim steps taken included permission for manufacturers to depower airbags by 20 to 30 percent and permission for parents with large families with carpooling needs to purchase a manual airbag cutoff switch that can be used to turn off the passenger airbag whenever a child must be seated in the front seat (Graham et al, 1998). More recently, the US government has implemented phased-in advanced airbag regulations that require for out-of-position small sized occupants either non-injurious airbag deployment, or airbag deactivation by means of occupant sensors.

As additional real-world crash experience accumulates, it is becoming increasingly clear that the driver airbag does not offer the same ratio of benefit to risk for all drivers. Certain subpopulations of drivers, including the elderly, women, and adults of short stature may experience less favorable ratios of benefit to risk than are experienced by large adult males. Concerns about these subpopulations has stimulated regulators and safety engineers to investigate a variety of advanced airbag concepts that may improve the airbag's ratio of benefit to risk for these drivers.

COPING WITH UNCERTAINTY ABOUT INPUTS TO RISK-BENEFIT ANALYSIS

When there is scientific uncertainty about the inputs to a risk-benefit calculation, the tools of formal uncertainty analysis and value-of-information analysis can be employed to help regulators and engineers cope with the resulting ambiguity about the proper course of action (Morgan and Henrion, 1990).

At a minimum, a deterministic risk-benefit analysis of a proposed safety device should include analysis of uncertain inputs to determine how stable the results of the analysis are when subjected to reasonable changes in the numerical values of inputs. Graham et al. (1997) found, for example, that a few percentage points change in the airbag's effectiveness rate for adult passengers can have a profound impact on the passenger airbag's ratio of benefit to risk (i.e., as measured in QALYS).

When there are multiple uncertain inputs to a risk-benefit calculation, it may be appropriate to perform two-way and three way sensitivity studies. In the case of the passenger airbag, Graham et al. (1997) published a two-way sensitivity analysis

involving the airbag's effectiveness rate for adults and its effectiveness rate for children under the age of 10. The interaction of these two uncertain inputs can produce major changes in the airbag's overall ratio of benefit to risk.

When the stakes in a decision are large and additional scientific research or data collection can be expected to reduce uncertainty, it may be useful to undertake a formal probabilistic uncertainty analysis and/or a formal value-of-information analysis. The probabilistic uncertainty analysis can be used to isolate the quantitative impact of each uncertain input on the overall degree of uncertainty about the ratio of benefits to risks. This type of analysis can highlight whether, for example, the imprecision about the airbag's effects on adults is more or less important than the imprecision about the airbag's effects on children.

If it is feasible to reduce imprecision through additional experiments or data collection, the cost and delay associated with such studies need to be compared to the possible benefits in reduced uncertainties and more well-informed design decisions. Although this type of analysis can be performed intuitively or judgmentally, there are cases where a more formal analysis may be appropriate to estimate the ultimate monetary or health impact of alternative approaches to data collection and decision making (Weinstein, 1983; Finkel and Evans, 1987; Thompson and Graham, 1996). Available computer software packages make value-of-information much more tractable than it was 10 years ago. An increasing number of VOI applications are being published in the public health and clinical literature, and the available tools are suitable for application to analyses of proposed safety devices.

EXAMPLE APPLICATIONS

This section lists several examples of the application of risk-benefit analysis methods to vehicle safety devices.

Car Passenger Airbags

As previously mentioned, Graham et al (1997) investigated the effects of pre-1998 US passenger airbags on children under the age of 10 using QALY, and found that net mortality was increased and net life expectancy was reduced by the installation of fully powered airbags.

Iijima et al (1998) calculated US car airbag fatality risk-benefit ratios for all occupants and found them to be 3 %; and for passengers only and found them to be 12%, associated with the prevalence of children and small adult airbag-induced injuries.

Car Safety Belts

Rogers and Zellner (1998) cite the HARM injury cost data from Malliaris et al (1982), which suggest a risk/benefit ratio of 6% for car safety belts (when both risks and benefits are extrapolated to 100% belt usage rate).

Car Head Restraints

Rogers and Zellner (1998) cite the HARM injury cost data from Malliaris et al (1982), which indicate that the risk from head restraint contact is 0.58% of all occupant injury costs. Huelke and O'Day (1975) indicate that the benefit from head restraints is 10 to 15% of injury costs, but the basis for this is unclear and is certainly not HARM. However, if it had been a HARM-like cost index, this would suggest a risk-benefit ratio of about 6 % for head restraints.

Motorcycle Leg Protector Feasibility

Rogers and Zellner (1998) describe results of extensive crash testing and computer simulations of a motorcycle rider leg protector prototype system, which using NIC indicated a risk-benefit ratio of 116% (i.e., the risks were greater than the benefits). This occurred because, although the device reduced lower leg fractures in many cases, it induced more serious head injuries and upper leg fractures in other cases.

Motorcycle Airbag Feasibility

Iijima et al (1998) describe extensive crash tests and computer simulations of a prototype motorcycle airbag, which using NIC indicated a risk-benefit ratio of 25%, due in part to increased head and neck injuries in some ground impacts, when an airbag was fitted. Further investigation of and attempt to ameliorate this effect were recommended.

All Terrain Vehicle Roll Over Protection System (ROPS) Feasibility

Van Auken et al (1997) describe rollover crash tests and computer simulations of two prototype ATV ROPS devices, which indicated a risk-benefit ratio of

more than 100% for NIC and for most other body region forces and predicted injury severities, due to impacts between the ROPS and the unrestrained rider during rollover events. It was recommended not to introduce these particular devices.

SUMMARY AND RECOMMENDATIONS

Taken together, these example applications of risk-benefit analysis methods to vehicle safety devices appear to have been useful for indicating systems which have favorable risk-benefit ratios, those which have unfavorable risk-benefit ratios, and those for which additional research and development would be needed. We see that to date, risk-benefit analysis methods have been applied to a variety of safety devices, and these have provided useful insights and direction.

Risk-benefit analysis principles include the need for health preference-based indices and the need to consider various subpopulations when evaluating vehicle safety devices which may be used by a wide variety of persons. Risk-benefit criteria are discussed including the importance of considering the net injury benefit (i.e., the differences between injury benefits and risks), the risk-benefit ratio (i.e., the magnitude of the injury risks compared to the injury benefits) and the magnitudes of the average risk and the average benefit (which allows due consideration of the magnitude of the potential adverse and beneficial effects independently from their frequency). Methods are also available and should be applied for assessing the level of uncertainty in risk-benefit analyses (Thompson and Graham, 1996).

Overall, while risk-benefit analyses and general criteria can provide important and invaluable information to regulators, safety engineers, and researchers, due attention must be paid to risk-benefit principles and the importance of subpopulations and uncertainty estimates when a safety device is considered for application to the entire user population of a given vehicle type. Case-by-case judgments will inevitably be necessary and special consideration will be given to vulnerable subpopulations if valid concerns are raised that these subpopulations will be made less safe by a proposed safety device.

REFERENCES

- Anon., The Abbreviated Injury Scale, 1990, Association for the Advancement of Automotive Medicine, Des Plaines, Illinois, 1990.
- Anon., Harvard Center for Risk Analysis: Annual Report 1992, Harvard School of Public Health, Cambridge, Massachusetts, 1992.
- Anon., Motorcycle—Test and analysis procedures for research evaluation of rider crash protective devices fitted to motorcycles, ISO 13232, International Organization for Standardization, Geneva, 1996.
- Anon., Risk Benefit Analysis in the Legislative Process, First Session, Joint Science and Technology Subcommittee Hearings, 96th Congress, U.S. Government Printing Office, Washington, D.C., July 1979.
- Anon., Risk Benefit Analysis in the Legislative Process, Second Session, Joint Science and Technology Subcommittee Hearings, 96th Congress, U.S. Government Printing Office, Washington, D.C., March 1980.
- Arrow, K. J., Lind, R. C., "Uncertainty and the Evaluation of Public Investment Decisions," *American Economic Review*, Vol 60, 1970.
- Becker, G. S., Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, University of Chicago Press, 1993.
- Doug-Tyson, M., Telephone conversation with J. Zellner of Dynamic Research, Inc., Centers for Drug Evaluation, U.S. Food and Drug Administration, Washington, D. C., 27 July 1994.
- Farisse, J., Bennoit, B., Seriat-Gautier, B. Daou, N. and Lang, O., "A Detailed Injury Severity Scale for Lower Limbs Based on Analysis of Injuries Sustained by Pedestrians," *Proceedings of the 27th Stapp Car Crash Conference*, San Diego, 1983.
- Finkel, A. M. and Evans, J. S., "Evaluating the Benefits of Uncertainty Reduction in Environmental risk Management," *JAPCA*, Vol 37, 1987.
- Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S. L. and Keeney, R. L., Acceptable Risk, Cambridge University Press, 1981.

- Gafni, A., Birch, S. and Mehrez, A., "Economics, Health and Health Economics, HYE's versus QALYS," *Journal of Health Economics*, Vol 12, 1993.
- Gold, M.R., Siegel, J.E., Russel, L.B., and Weinstein, M.C., Cost-Effectiveness in Health and Medicine, Oxford University Press, New York, 1996.
- Graham, J. D., Thompson, K. M., Goldie, S. J., Segui-Gomez, M. and Weinstein, M. C., "The cost-effectiveness of airbags by seating position," Journal of the American Medical Association, 1997.
- Iijima, S., Hosono, S., Ota, A. and Yamamoto, T., "Exploratory Study of an Airbag Concept for a Large Touring Motorcycle," 16th International Technical Conference on the Enhanced Safety of Vehicles, Windsor, Ontario, May 1998.
- Johannesson, M., Theory and Methods of Economic Evaluation of Health Care, Kluwer Academic Publishers, 1996.
- Kebschull, S. A., Van Auken, R. M. and Zellner, J. W., "Proposed Amendment to ISO 13232 Part 5, Injury Indices, Risk/Benefit Calculations," ISO/TC22/SC22/WG22 N 255, July 2000.
- Lasagna, L., Telephone conversation with J. Zellner of Dynamic Research, Inc., Center for Study of Drug Development, Tufts University, Boston, 21 July 1994.
- Miller, T. R., Pindus, N. M., Leon, T. S. and Douglas, J. B., "Motor Vehicle Injury Costs by Body Region and Injury Severity," 34th Annual Proceedings, Association for the Advancement of Automotive Medicine, Scottsdale, Arizona, 1990.
- Morgan, M. G. and Henrion, M., A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, Cambridge University Press, 1990.
- Newman, J. A., Tylko, S. and Miller, T., "Toward a Comprehensive Biomechanical Injury Cost Model," 36th Annual Proceedings, Association for the Advancement of Automotive Medicine, Portland, Oregon, October 1992.
- Nord, E., "An Alternative to QALYS: The Saved Young Life Equivalent," *British Medical Journal*, Vol 305, 1992.
- Pliskin, J. S., Shepard, D. S., and Weinstein, M. C., "Utility Functions for Life Years and Health Status," *Operations Research*, Vol 28, 1980.
- Rogers, N. M. and Zellner, J. W., "An Overall Evaluation of UKDS Motorcyclist Leg Protectors Based on ISO 13232," 16th International Technical Conference on the Enhanced Safety of Vehicles, Windsor, Ontario, May 1998.
- Shepard, D. S. and Zeckhauser, R. J., "Survival versus Consumption," *Management Science*, Vol 30(4), 1984.
- Somers, R. L., "The Probability of Death Score: an Improvement of the Injury Severity Score," Proceedings of the 1981 Association for the Advancement of Automotive Medicine Conference, San Francisco, 1981.
- Thompson, K. M., and Graham, J. D., "Going beyond the single number: Using probabilistic risk assessment to improve risk management," *Human and Ecological Risk Assessment*, 1996.
- Torrance, G. W., "Measurement of Health State Utilities for Economic Appraisal," *Journal of Health Economics*, Vol 5, 1986.
- Ulman, M. S. and Stalnaker, R. L., "Evaluation of the AIS as a Measure of Probability of Death," Proceedings of the 1986 IRCOBI Conference on the Biomechanics of Impact, Zurich, 1986.
- VanAuken, R. M. and Zellner, J. W., ATV ROPS Tests and Simulations, DRI-TR-98-2, Dynamic Research, Inc., Torrance, California, October 1998.
- Walker, S. R. and Asscher, A. W., Medicines and Risk/Benefit Decisions, MIT Press Limited, Lancaster, BG, 1987.
- Weinstein, M. C. and Stason, W. B., "Foundations of Cost-Effectiveness Analysis for Health and Medical Practices," *New England Journal of Medicine*, Vol 296, 1977.
- Weinstein, M. C., "Cost-Effective Priorities for Cancer Prevention," *Science*, Vol 221, 1983.
- Zeckhauser, R. and Shepard, D., "Where Now for Saving Lives?" *Law and Contemporary Problems*, Vol 40, 1976.