This is the pre-peer reviewed version of the following article: MacKenzie, C.A. (2015). Summarizing risk using risk measures and risk indices. *Risk Analysis*, 34(12), 2143-2162, which is published in final form at http://dx.doi.org/10.1111/risa.12220. (The final version of this article is substantially different than this pre-peer reviewed version.)

Summarizing Risk Using Risk Measures and Risk Indices

Cameron A. MacKenzie

Defense Resources Management Institute

Naval Postgraduate School

Monterey, CA 93943

camacken@nps.edu

ACKNOWLEDGMENTS

The author would like to thank two anonymous reviewers, Elisabeth Paté-Cornell of Stanford University (who first suggested the topic), Kash Barker of the University of Oklahoma, David Caswell of the Air Force Academy, and Brad Powley of Strategic Decisions Group for their helpful comments on this paper. The majority of this research was conducted while the author was a student at Stanford.

ABSTRACT

Our society is fascinated with risk in many different areas and disciplines. One of the main ways to describe and communicate the level of risk is through risk indices, which summarize risk using numbers or categories such as words, letters, or colors. These indices are used to communicate risks to the public, understand how risk is changing over time, compare among different risks, and support decision making. This paper emphasizes the importance of describing risk with a probability distribution, developing a numerical risk measure that summarizes the probability distribution, and finally translating the risk measure to an index. Measuring the risk is the most difficult part and requires the analyst to summarize a probability distribution into one or possibly a few numbers. The risk measure can then be transformed to a numerical or categorical index. We apply the method outlined in this paper to construct a risk index that compares the risk of fatalities in aviation and highway transportation.

Keywords: risk measure, risk index, transportation, risk communication

1. INTRODUCTION

Our society is preoccupied with the subject of risk. From financial risk to the risk of a terrorist attack to the risk of disease, individuals want to understand how a disastrous event may affect them and how they can avoid it. With so many different types of risks present in modern society, risk indices have increasingly become an important way to communicate the seriousness of a risk. A search for the term “risk index” in Google Scholar returns 980 results for the year 2001 compared to 4,460 results for the year 2011 even though a search for the term “risk” returns more than three times as many results for 2001 as for 2011. It would appear that risk indices are outpacing the risks that they measure. The increasing popularity of risk indices demands a careful study of risk indices and the best ways to construct a risk index.

Three well-known examples of risk indices, which summarize the risk of an event via a single number or category, demonstrate their wide variety. The American Society of Civil Engineers(1) uses a letter grade to rate the quality of U.S. infrastructure; value-at-risk is a single number that describes how much a portfolio could lose at a specified probability level; and the now-defunct Homeland Security Advisory System used a color and word categorization to describe the national security threat level. First, each index summarizes risk differently: one uses letters, another uses real numbers, and the third uses colors with word descriptors. Second, each risk index covers a different discipline: engineering, financial, and security. Third, the purposes of the indices vary: the letter grades are used to motivate investment and attention on U.S. infrastructure, an investment firm usually calculates value-at-risk for its own internal benefit, and the Homeland Security Advisory System was designed to communicate the risk of terrorism to an entire nation.

As these three examples show, risk indices can have varied purposes. A risk index usually attempts to fulfill at least one of the following objectives.

- Describe risk accurately
- Communicate the level of risk
- Compare among different risks
- Identify the most serious risk
- Translate how risk changes over time
- Measure the effectiveness of risk-reduction strategies
- Recommend actions for a given level of risk

Despite the different goals, audiences, and means of constructing risk indices, this paper offers general recommendations to aid in the construction of a risk index for any discipline. We argue that the specific goals and intended audience of a risk index should determine the final form of the index. Because risk indices summarize risks into a single metric, careful decision making where risk is present should eschew the index in favor of a more complete description of risk. As Cox[(2)] demonstrates, selecting among different risk reduction strategies should rely on both the cost and effectiveness of each strategy.

In order to arrive at guidelines for constructing a risk index, Section 2 first defines a risk index and distinguishes it from a risk measure. An analyst who wants to create a risk index should first calculate a numerical risk measure and second translate that measure to an index. Section 3 categorizes the different choices for a numerical risk measure and explores the possibility of creating a measure composed of multiple numbers. We discuss some of the advantages and disadvantages of each type of risk measure. Section 4 develops options for how an analyst can use the numerical measure to create a risk index. Section 5 applies the variety of risk measures and indices to an illustrative example focusing on the risk of transportation. The conclusion describes how the above objectives can help determine the most appropriate type of risk index.

2. STEPS TO CREATING A RISK INDEX

The term risk index is frequently used but rarely defined. In economics, where the subject of indices has been explored both theoretically and practically, indices are "statistical summaries used to compare different situations, times, conditions, or objects."[(3)] Similarly, a risk index summarizes the risk of an event or situation by using numbers or categorical values for the purpose of identifying and comparing risks. Summarizing information appeals to all of us who want to feel informed but do not have time to fully digest the entirety of the information.

Before exploring how risk can be summarized, one must understand the most complete way to describe risk. In the tradition of probabilistic risk analysis, risk is best defined as a combination of the probability of an unwanted event occurring and the severity or consequences if that event occurs.[4,5] Fully describing risk involves calculating a probability distribution over the range of possible consequences for an event. A risk index is therefore a summary of the probability distribution over the range of possible consequences.

The terms risk measure and risk index are often used synonymously, and both attempt to summarize risk through a single value. This paper makes an important distinction between the two. A risk measure is a numerical summary of risk, and that number is a real number. As will be discussed later, we also allow the measure to be a vector of real numbers. A risk index can also be a real number, but it is often composed of ordinal numbers, letters, words, or colors, thus widening the means of communicating risk to the user. We view risk measures to be a subset of risk indices.

To construct a risk index, an analyst should first develop a probability distribution over the range of consequences that fully describe the risk of an event, investment, or project. Next, the analyst should develop a numerical measure that accurately summarizes the probability distribution. Finally, the risk measure can be translated to a different scale (like a color or ordinal number) to communicate the risk to a larger audience. Many experts[6,7,8,9] have described how to conduct probabilistic risk analysis, and this paper focuses on the last two steps. The next section discusses and comments on some of the commonly used approaches for calculating a risk measure.

3. CHOICES FOR RISK MEASURES

We classify the many different risk measures in use today into four general categories: moments of a probability distribution, quantiles of a distribution, disutility functions, and combined factors. Each of these risk measures has positive and negative aspects. We also discuss how a risk measure can also be composed of multiple numbers such as the first two moments of a distribution or three different quantiles.

This paper uses a random variable X to represent the consequences of an event whose risk we are interested in measuring. $X < 0$ implies negative or adverse consequences (e.g., money lost on an investment, deaths resulting from an event); $X > 0$ implies positive consequences; and $x_1 < x_2$ indicates that x_1 has worse consequences than x_2.

3.1. Moments of a Distribution

In statistics, moments are the most accurate ways to summarize a probability distribution, and they can serve as good measures of risk. The mean value is perhaps the most frequently used risk measure, and the risk of an event may be stated as the mean or expected value.[2,10,11] Standard deviation can also serve as a measure of risk, especially in finance, and the greater variability, the more risk to which an individual is exposed. Relying on just one of these measures can ignore important information. The mean fails to convey any information about the spread of the distribution, and only relying on standard deviation as a risk measure is problematic because it gives no indication of the central tendency. Using standard deviation as a measure of risk could lead to situations where a decision maker prefers a certain outcome of no gain or loss to a highly variable situation where only positive outcomes are possible.[12]

A risk measure may combine the first two moments into a single number in order to incorporate both the location and spread of the probability distribution. Based on Vrijling et al.,[13] we present a measure $M(X)$ in Eq. (1) that is a weighted sum of the mean $\mu(X)$ and standard deviation $\sigma(X)$ where k is a risk aversion parameter. Because the risk measure is a function of the negative of the mean, the measure gets larger as consequences worsen.

$$M(X) = -\,\mu(X) + k\sigma(X) \tag{1}$$

Risk neutrality implies $k = 0$, and risk aversion implies that $k > 0$. The k parameter can be tailored to reflect an individual's willingness to trade off between the mean and the standard deviation.

3.2. Quantiles of a Distribution

Risk analysis often focuses on extreme events, which are low-probability, high-consequence events. The quantile of a distribution can measure these high-consequence events, where the q-quantile is the value of the consequences x at probability q such that $q = P(X \leq x)$. Perhaps the best example of using a quantile as a risk measure is value-at-risk (VAR). With origins dating to the early twentieth century but popularized in the 1990s, VAR is defined as the worst loss that might be expected from holding a security or portfolio given a specified level of probability.[14] VAR does not, however, describe the maximum possible loss or take into account losses that exceed VAR. Because VAR is not a convex function, it does not necessarily encourage diversification of risky investments, which is a desirable characteristic of financial risk measures.[15,16]

Conditional value-at-risk (CVAR) calculates the expected value given that losses exceed VAR, and it better captures the extreme tails and encourages portfolio diversification.[17,18]

Although CVAR was introduced as remains most popular as a financial risk measure, it has also been proposed to measure the risk of transporting hazardous materials,[19] industrial production under uncertain demand,[20] and wait times at hospitals.[21] This measure has been suggested as a useful tool for analyzing engineering risk as part of the partitioned multiobjective risk method which seeks to find Pareto optimal sets of alternatives based on different measures of risk.[9,22]

Although measures based on a quantile of a distribution can describe the risk of extreme events to a greater degree than mean or variance, they do not account for risks that are less extreme than the chosen quantile. Two distributions could be identical for all values less than the value corresponding to q but differ for values greater than VAR, and a risk measure based on quantiles would be the same for these two distributions. Using a quantile of a distribution as a risk measure seems most suitable if it is used in concert with a measure representing the average risk or if the analyst is primarily concerned with extreme events and can ignore less extreme risk scenarios.

3.3. Disutility Functions

The third type of risk measure is based on disutility functions. Because utility functions provide a solid foundation for making decisions under uncertainty, measuring risk with these functions is a natural extension.[23,24] Expected disutility is given in Eq. (2) where $V(x)$ is the disutility function and $f_X(x)$ is a probability density function over the potential outcomes.

$$M(X) = \int_{-\infty}^{\infty} V(x) f_X(x) dx \tag{2}$$

Disutility functions are decreasing functions where one of the parameters usually represents an individual's tolerance for (or willingness to accept) risk.[25] If the risk measure is designed to only capture potential losses (downside risk), the upper limit of integration should be zero instead of positive infinity. Risk measures based on a disutility or utility function have been advocated in finance because these functions encourage portfolio diversification,(18) provide a consistent measure of risk over time,(26,27) and can be modified to incorporate uncertainty in the probability distribution for an investment.(28)

Although many different disutility functions may be appropriate, this paper will focus on exponential disutility and the α-t model. The exponential disutility function is given in Eq. (3) where γ reflects an individual's tolerance for risk.

$$\begin{aligned}&\text{If risk averse, } V(x) = e^{-\gamma x} - 1,\ \gamma > 0\\&\text{If risk neutral, } V(x) = -x\\&\text{If risk seeking, } V(x) = 1 - e^{-\gamma x},\ \gamma < 0\end{aligned} \tag{3}$$

All of these functions are positive when $x < 0$, and the disutility function equals 0 if $x = 0$.

Another possible disutility function defines risk as a failure to meet a target outcome or objective. Fishburn(29) generalized this viewpoint in the α-t model, which is defined in Eq. (4) where t is a fixed upper bound so that only consequences below this upper bound are included in the risk measure, and α characterizes the risk attitude.

$$M(X) = \int_{-\infty}^{t} (t - x)^{\alpha} f_X(x) dx \tag{4}$$

The parameter $\alpha < 1$ implies risk seeking, $\alpha = 1$ implies risk neutrality, and $\alpha > 1$ implies risk aversion. A typical value for t is 0, which indicates that only outcomes where $x < 0$ will be included in the risk measure.

Both the exponential disutility and the α-t model can be transformed so that the units of the risk measure are in the same units as the consequences x. This can be accomplished by taking the inverse function of the disutility function: for exponential disutility, $V^{-1}\big(M(X)\big)$ where $V^{-1}(\cdot)$ is the inverse function of Eq. (3) and $t - M(X)^{1/\alpha}$ for the α-t model.

Estimating the risk attitude parameter poses the most difficult obstacle for using a risk measure based on disutility. If multiple individuals are going to use the risk measure, as is usually the case, determining a single risk attitude parameter that can be used for the entire group may be impossible. Even if everybody's risk attitude could be assessed, individuals have different tolerances for risk,[30,31] and one measure will likely not accurately reflect each individual's attitude toward risk.

3.4. Functions of Indicators or Factors

The fourth category of risk measures are those based on factors or indicators which are aggregated together. Several factors that are related to the risk of the event are aggregated together, often through a weighted linear combination of the factors.[32] Some risk measures or indices based on factors describe the risk of earthquakes,[33] hurricanes,[34] climate change,[35] and terrorist activity.[36]

One reason indices or measures based on aggregating different factors are popular is because these factors or indicators can usually be measured or easily assessed from experts. When data are available, regression can be deployed to calculate the weighting parameters. Aggregating these factors through a functional form makes the measure appear objective, and the challenging task of estimating a probability distribution is usually not necessary.

As deterministic functions, these measures frequently fail to account for the uncertainty that is inherent in risk. Even when they incorporate uncertainty, they treat uncertainty or the frequency of the event as another indicator or factor rather than using a true probability distribution. For example, failure mode and effect analysis (FMEA), a risk measure used frequently in industry to identify points of failure in a process or system, measures the frequency of a problem, the likelihood the problem will not be detected, and the severity on ten-point scales. The risk priority number is calculated as the product of the three factors. As Gilchrist[37] and Ben-Daya and Raouf[38] discuss, this method can generate some serious inconsistencies. FMEA assumes each factor is linear with respect to the risk priority number, which may not be the case. According to FMEA, if severity and frequency have the same score on the ten-point scale, reducing the severity by one point is equivalent to reducing the frequency by one point. However, if frequency were measured with probability and severity were measured with a real quantity like cost, reducing either frequency or severity might be a much better risk-reduction strategy.

3.5. Risk Measures of Multiple Values

Using a risk measure composed of two or three values conveys more information than a measure based on a single value. A risk measure based on the moments of the distribution can include both the mean and standard deviation of the probability distribution. Investment science has traditionally used mean-variance analysis to evaluate and compare different investments. This measure conveys both the location and the spread of the distribution, and the closer the distribution is to a normal distribution, the more the mean and standard deviation fully describe the distribution.

In order to emphasize extreme events without ignoring less extreme scenarios, a three-number measure may be composed of different quantiles, like the 0.05, 0.50, and 0.95 quantiles. Such a description could communicate a fairly accurate picture of the probability distribution. People interested could compare the risks of two different events on a quantile-by-quantile basis.[39]

Because the major difficulty with a disutility-based measure is determining a risk attitude parameter appropriate for several individuals, a risk measure could be composed of multiple values, each of which corresponds to a different risk attitude. For example, one value could be expected disutility for a moderately risk seeking person, another value for a risk neutral person, and the final value for a moderately risk averse person. A person using this risk measure could use the measure that best corresponds to his or her risk preference.

One difficulty with using a risk measure composed of multiple values is that the measure does not instruct a user how to rank or compare risks. For example, if a risk measure is composed of the mean and standard deviation, anyone using the measure must still decide whether the mean or standard deviation is more important. If the measure should determine that one event carries more risk by weighting the importance of the standard deviation vis-à-vis the mean, the risk measure in Eq. (1) provides an appropriate option. Alternatively, an analyst could present the two- or three-value risk measure and realize that different people will make different conclusions about the riskiness of the event based on their own beliefs.

Despite the ability to communicate more information with multiple values, most risk measures appear to be single value. It is unclear as to why a single-value measure might be preferred: because analysts prefer to summarize their research into a single value or because users of these measures want to interpret a single value. We encourage analysts to think more seriously about whether a risk measure composed of multiple values better meets the goals behind the research.

4. FROM A RISK MEASURE TO A RISK INDEX

After choosing a risk measure, the analyst may want to transform the measure to a risk index. As this paper has defined a risk measure and risk index, a risk measure can be considered a risk index without any further calculations or operations. However, transforming the scale of the risk measure may help facilitate a better understanding of the index. For example, a value such as 20.2 might be difficult to interpret, but if that measure was transformed to a value such as 80 on a scale from 0 to 100, it is easier to understand that the number represents a serious risk. If a measure is composed of multiple values, each value could be mapped to the new scale. Whether a single or multiple-value measure is used, the new scale can be composed of continuous real numbers (a numerical index) or discrete categories or numbers (a categorical index).

4.1. Numerical Index

A continuous numerical scale is appropriate if the risk index is going to be used to compare different risks that are quite similar or to see how risk is changing over time because such a scale reveals subtle changes and differences. If decision makers are going to use the measurement or index as a basis for decisions, a numerical scale is preferable to a categorical scale because the numerical scale can provide greater detail and more levels of differentiation than a general category.

A 100-point or 10-point scale is probably the most natural type of scale for a numerical index, as humankind has relied on a decimal numerical system for millennia. A linear transformation of the risk measure to a new scale requires identifying the maximum and minimum possible values of the risk measure—M_{max} and M_{min}—which correspond to 0 and 100 on the 100-point scale. Eq. (5) demonstrates how the risk measure can be mapped to a 100-point risk index $I(X)$.

$$I(X) = \frac{M(X) - M_{min}}{M_{max} - M_{min}} * 100 \tag{5}$$

Determining a value for M_{max} presents a greater challenge than selecting M_{min}, which will usually equal 0. One could choose a value for M_{max} that represents the minimum value of risk that is considered unacceptable. Some governmental agencies have developed thresholds for individual fatality risks[40] which can help guide the selection of M_{max}.

Such a numerical risk index may raise questions about whether the intervals between numbers on the index have any meaning. For example, if one event has a value of 20 on the risk index and another event has a value of 40, is the latter twice as risky as the former? Does a difference of five between two values mean the same thing whether or not the values are in the twenties compared with whether the scores are in the eighties?

These questions can be answered by investigating if a risk index has the following mathematical properties: positive homogeneous, additive, and translation invariant. An index with positive homogeneity means that if all the consequences worsen by a factor of $\lambda > 0$, the index number is multiplied by λ, as Eq. (6) shows.

$$I(\lambda X) = \lambda I(X) \tag{6}$$

Positive homogeneity implies that an event with an index value of 40 is twice as risky as an event with an index value of 20. If $M_{min} = 0$ and the risk measure is positive homogeneous, the numerical index obeys this property.

An additive index implies that if two events are combined, the risk of these two events equals the sum of the risks of each event, as shown in Eq. (7) where X and Y are random variables representing two events.

$$I(X + Y) = I(X) + I(Y) \tag{7}$$

For the risk index to be additive, M_{min} must equal 0. The appealing feature of this property is that an organization can separately measure multiple risks and add the risks together to obtain a numerical measure of the organization's risk.

Finally, a translation invariant index means that improving the consequences by some real number β, the riskiness of the event decreases by that same amount, as given in Eq. (8).

$$I(X+\beta) = I(X) - \beta \tag{8}$$

A translation invariant risk index means that the numerical difference between two index values can be interpreted using the units used to measure the consequences. If the risk measure is translation invariant, i.e., if $M(X+\beta) = M(X) - \beta$, mapping the measure to a 100-point scale as shown in Eq. (5) implies the relationship in Eq. (9).

$$I(X) - I(X+\beta) = \frac{100\beta}{M_{max} - M_{min}} \tag{9}$$

Thus, the index is translation invariant if $M_{max} - M_{min} = 100$ and the risk measure is translation invariant. The first condition that $M_{max} - M_{min} = 100$ seems unlikely. If the risk measure is translation invariant but the first condition is not met, the difference between two values on the index can still be used for comparisons among index values. For example, a difference of five between two scores means the same thing whether the scores are in the twenties or the eighties.

The mean value is the only risk measure considered that satisfies the positive homogeneity, additive, and translation invariant properties. VAR, CVAR, and the measure combining mean and standard deviation as in Eq. (1) are positive homogeneous and translation invariant[18] but not additive. Standard deviation and the *α*-*t* model are positive homogeneous but neither translation invariant nor additive. Exponential disutility is additive if $P(X \le 0) = P(Y \le 0) = 1$ and X and Y are independent, and it is translation invariant but not positive homogeneous. (For any of these properties to hold for the two disutility models, the inverse function of the disutility model must be used.)

The type of risk measure chosen may be influenced by whether the analyst or decision maker believes it is more important that the risk index should have any of these properties. If separate risks are going to be combined into an overall risk measure, the additive property is necessary. If risks are going to be compared against each other, a positive homogeneous measure might be better because saying something like one risk is twice as risky as another event has a clear mathematical meaning. If all three properties are important, using the mean as the measure might be the best choice.

4.2. Categorical Index

If the analyst is uncomfortable about communicating specific numbers or is worried that these numbers might give an inaccurate impression, he or she might want to transform the measure to a categorical rather than a numerical risk index. A categorical risk index is most appropriate when the goal is to give people a general sense of the level of risk. Especially if a large number of people are going to be using a risk index, using categories can facilitate broader communication because categories can be understood more easily than a numerical scale.

A categorical risk index provides an ordering relationship among the categories, which can be composed of numbers, words, letters, or colors. Using words as categories like “severe” or “moderate” has an explicit communication purpose, and using letters implies a grading system in the United States. Another popular type of risk index is a color-coded chart. If a risk is classified in the “red” category, it sends a clear signal that the risk is very serious and may automatically trigger mitigating actions.

Regardless of the categorical scale chosen, the index should still rely on a careful analysis of the risks, and the level of a category should be based on a numerical risk measure. After determining the type of category, the analyst would choose the number of categories. A general rule of thumb for determining how many categories is to minimize the number of categories but have enough categories to so that distinct risks are categorized differently.

The next step is to decide how to translate the numerical risk measure to a categorical scale. The simplest method is to divide the range of measures into equally divided intervals. Another method is to choose the categories based on the actual calculated risk measures. For example, 10 different events may have the following measures: 2.0, 2.5, 6.4, 7.0, 7.3, 9.2, 10.0, 11.0, 13.0, and 13.5. The analyst may want to put the first two events in the same risk category, the next three events in another category, the next three events in one or possibly two categories, and the final two events in another category.

Another method of creating categories is based on where a distinction between categories will most likely be noticed. Gustav Fechner, a German psychologist, postulated the intensity of a sensation increases as the logarithm of the stimulus increases, and several risk scales—including the Richter scale for earthquakes and the Torino scale for near-Earth objects like asteroids—use a logarithmic scale.[(41)] Dividing the categories according to the base-10 logarithm of the risk measure (i.e., 10^{-2}, 10^{-1}, 0, 10) may follow most closely with how people perceive changing intensities.[(42)]

One benefit of using a categorical scale is that recommended actions can be attached to each category. If the risk falls into a certain category, the public can be advised to take certain actions. For example, the National Weather Service issues tornado watches and warnings. Each advisory has a specific meaning and carries recommendations for people who receive a watch or a warning.[(43)] Because these advisories have been used for a long time and are issued frequently in areas where tornadoes appear, people understand what these advisories mean. A person subconsciously attaches his or her own risk attitude to these advisories when deciding whether to follow the recommended actions.

5. ILLUSTRATIVE EXAMPLE: TRANSPORTATION RISK

An illustrative example comparing the risk of fatal accidents between aviation and highway demonstrates the construction and interpretation of risk measures and indices. Approximately 35,000 people die in transportation-related incidents in the United States each year, and the vast majority of fatalities result from accidents on the road.[44] Whether driving or flying is safer depends on the assumptions used to calculate the probability of a fatal accident.[45,46] Communicating the risk of fatalities of different transportation alternatives may help people make more informed choices, and policy makers are interested in reducing transportation-related fatalities.[47,48,49,50] Risk measures and indices for aviation and highway risk can inform both the general public and policy makers. These indices could satisfy several purposes: to inform the public about the relative risk of each mode of transportation, to understand if one mode is becoming safer or more dangerous over time, and to help determine if additional safety measures are needed.

Consequences of a highway or aviation incident can include fatalities, injured persons, and the monetary cost of accidents. Because our interest is to demonstrate how consequences can be incorporated into a risk measure and index as opposed to generating a complete picture of transportation risk, the consequences in this example are composed solely of fatalities. Aviation risk is divided into two categories based on U.S. Federal Aviation Regulations: air carriers or commercial aviation, and air taxi and commuters or on-demand and chartered flights. The data reveal a large difference in the risk of fatalities between these two categories of aviation. Most of the flying public use air carriers, and the risk of a fatal accident from an air carrier may be independent of the fatal risk from air taxis and commuters. The number of fatalities from each fatal accident in aviation and highway travel in the United States is recorded from 2005 through 2009 in Tables I – III. The data are derived from the Bureau of Transportation Statistics,[51,52,53,54] the National Transportation Safety Board,[55] and the National Highway Traffic Safety Administration.[56]

Table I. Fatalities from air carrier incidents in the U.S., 2005-2009.

Fatalities per incident	Count of incidents	Count per million miles
1	4	0.0000983
20	1	0.0000250
49	1	0.0000250
50	1	0.0000250

Table II. Fatalities from air taxi and commuter incidents in the U.S., 2005-2009.

Fatalities per incident	Count of incidents	Count per million miles
1	18	0.0724
2	14	0.0563
3	7	0.0282
4	7	0.0282
5	4	0.0161
6	2	0.00843
7	2	0.00843
8	2	0.00843
9	1	0.00402

Table III. Fatalities from highway incidents in the U.S., 2005-2009.

Fatalities per incident	Count of incidents	Count per million miles
1	165,614	0.0111
2	12,177	0.000814
3	1,858	0.000124
4	509	0.0000340
5	144	0.00000962
6	38	0.00000254
7	14	0.000000935
8	5	0.000000334
9	4	0.000000267
10	2	0.000000134
11	1	0.0000000668
12	1	0.0000000668
17	1	0.0000000668
23	1	0.0000000668

We assume that the data perfectly represent the probabilities for fatalities per incident and that the number of incidents per one million miles equals the probability of an incident per one million miles. In reality, there is certainly a non-zero probability that an air carrier incident could result in two, three, or even several hundred fatalities.

The different formulas discussed in Section 3 are used to measure the risk of fatalities for the three different transportation modes. Table IV depicts the mean, standard deviation, VAR for two different quantiles, and CVAR. Figure 1 displays risk measures using Eq. (1), which combines the mean and standard deviation into a single risk measure, for different values of k. Figures 2 and 3 plot risk measures for exponential disutility and the α-t model (where $t = 0$) for different risk attitudes. The risk measure for the exponential disutility function and the α-t model are transformed by taking the inverse of $V(x)$ so that the measure is in units of fatalities. If $\alpha < 1$, raising $M(X)$ to the power $1/\alpha$ results in an extremely small number. Consequently, the risk measure for the α-t model is not in units of fatalities for a risk seeking attitude.

Table IV. Summary risk measures for transportation

	Mean	Standard deviation	VAR $q = 0.05$	$q = 0.01$	CVAR $q = 0.05$	$q = 0.01$
Air carrier	0.0037	0.36	0	0	0.061	0.31
Air taxi and commuter	0.67	1.57	4	8	6.17	8.40
Highway	0.013	0.13	0	1	0.27	1.12

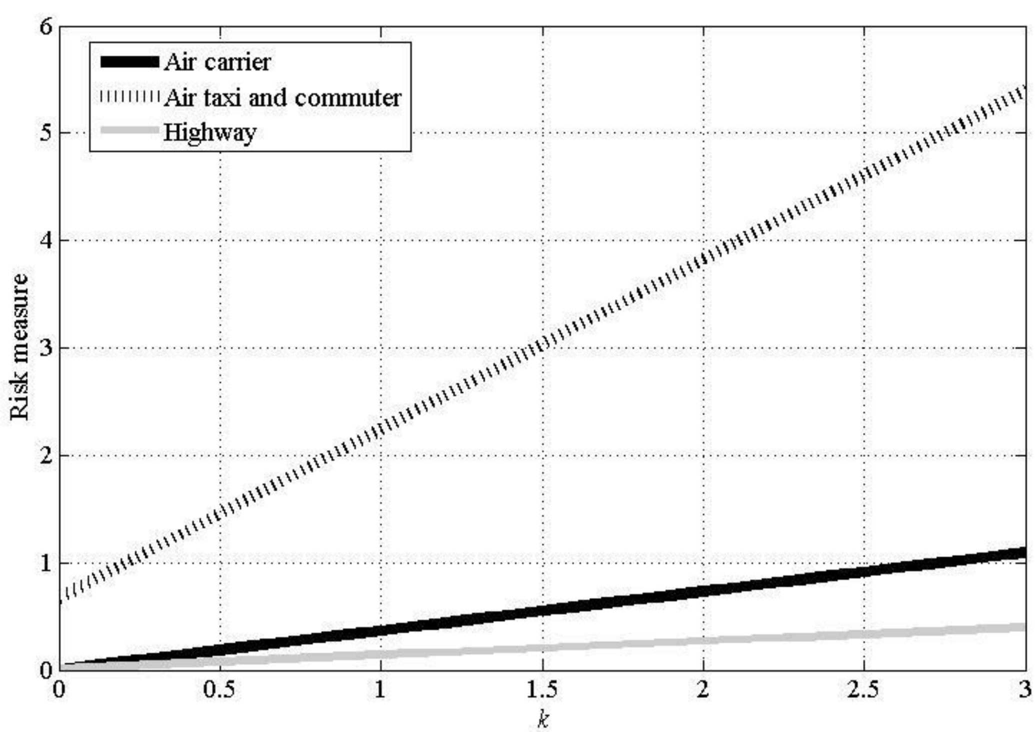

Fig. 1. Risk measure as a weighted sum between the mean and standard deviation as given in Eq. (1).

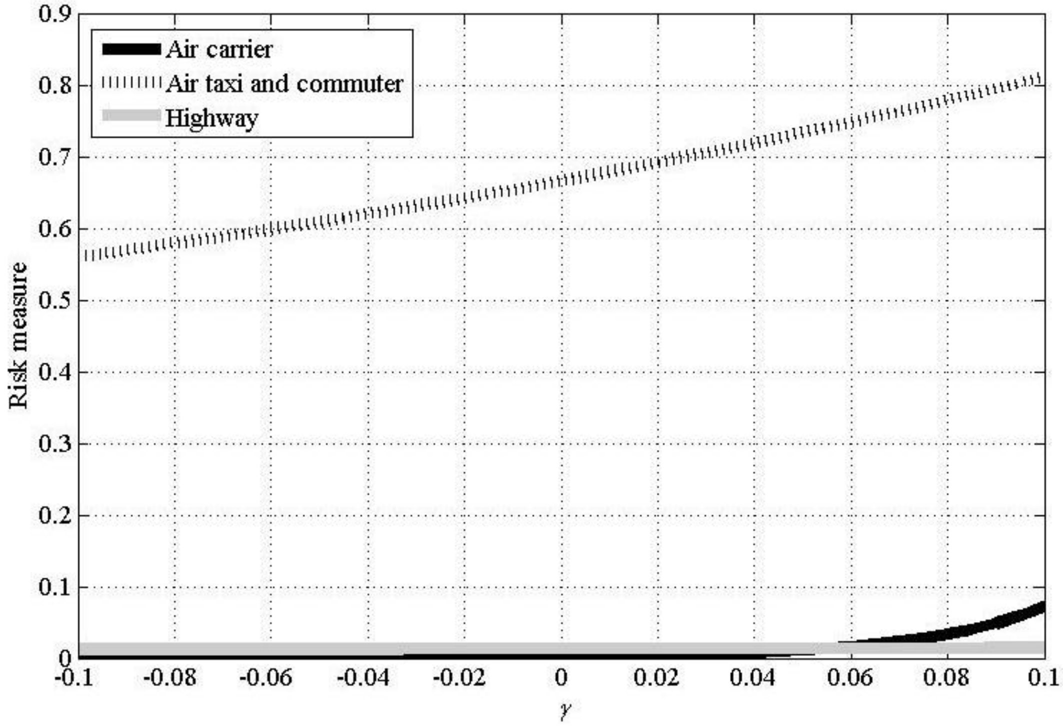

Fig. 2. Risk measure using an exponential disutility function.

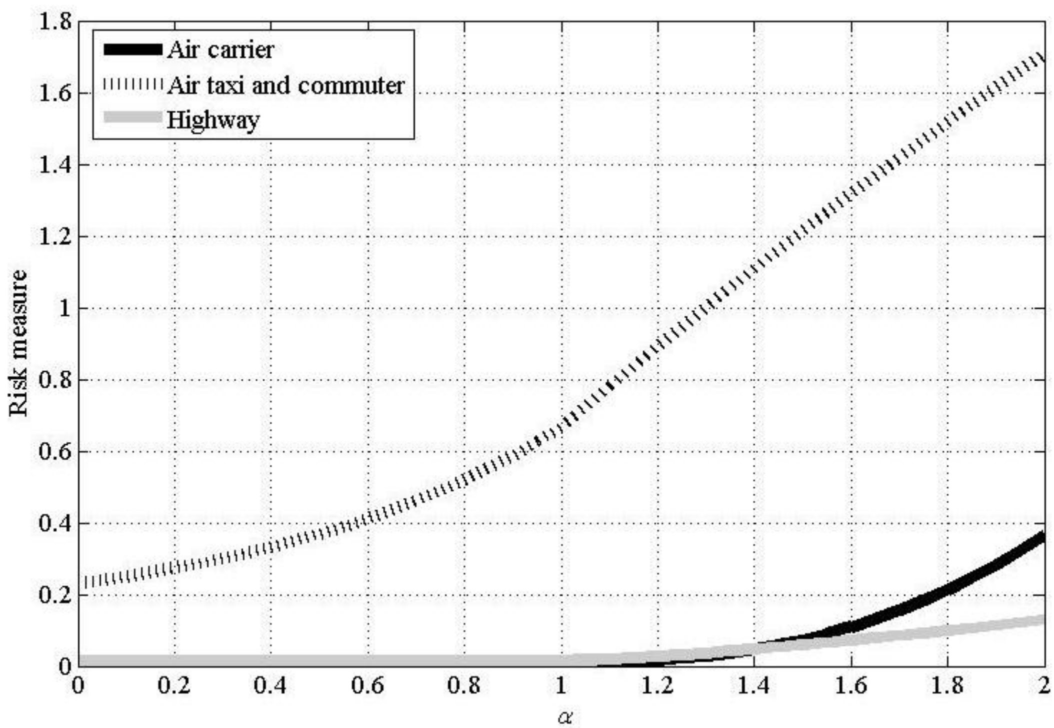

Fig. 3. Risk measure using the α-t model where $\boldsymbol{t = 0}$.

Air taxi and commuter carries the largest risk of fatalities according to all the risk measures under consideration. If the mean (or a risk neutral attitude) is used as a risk measure, highway transportation carries more risk than air carrier transportation. The mean number of deaths from highway transportation is 3.5 times greater than the mean number of deaths for air carrier transportation. VAR and CVAR, both of which attempt to measure extreme values in the tails of the probability distributions, also produce risk measures in which highway carries more risk than air carriers.

The risk of fatalities from highway transportation also exceeds that from air carrier travel for moderate risk averse attitudes (Figures 2 and 3), but as risk aversion increases, air carriers become riskier than highway. Because the probability of 49 or 50 fatalities is non-zero for air carriers, this mode of transportation has a larger standard deviation than highway. The standard deviations for both modes of transportation are much larger than their respective means, and putting a small amount of weight ($k \geq 0.1$) on the standard deviation in measure in Eq. (1) changes the ranking of these two risks. Similarly, if $\gamma \geq 0.055$ in the exponential disutility or if $\alpha \geq 1.45$ in the α-t model, air carrier risk exceeds highway risk. The larger values for γ and α emphasize crashes with scores of fatalities, which have occurred with air carriers.

Whether a user interested in transportation fatalities would exhibit risk aversion depends on the goal of the risk measure and the user's preferences. If the goal is to help inform policy makers about reducing transportation fatalities, any risk attitude may be justified. Debate exists within the decision and risk analysis literature over the appropriate risk attitude for fatalities in a public policy context.[57,58,59] Many risk assessments of potential fatalities assume moderate risk aversion.[60,61,62] People tend to prefer a situation with a large probability of a small number of fatalities to one with a smaller probability of a larger number of fatalities even if the expected number of fatalities is identical in both situations.(63,64) However, Slovic et al.(58) question the risk aversion assumption due to an experiment in which people chose to reduce the number of single-fatality incidents rather than reduce the probability of a catastrophic incident with a large number of fatalities. Another study(65) revealed that people may exhibit both risk seeking and risk averse attitudes, depending on how the situation is framed. Because incidents with a large number of fatalities receive more media and public attention, a policy maker may want to focus on minimizing the number of accidents with many fatalities.

The risk measure may also be used to inform an individual whether he or she is more likely to die traveling in an airplane or in a motor vehicle. For this purpose, the probability of a fatal incident is the best measure, which corresponds to $\gamma = -\infty$ and $\alpha = 0$. As seen in Figure 3, the probability of a fatal incident per one million miles is 0.23 for air taxi and commuter, 0.012 for highway, and 0.000175 for air carriers. *Ceteris paribus*, an individual should travel with air carriers to minimize his or her chances of being killed while traveling.

Translating the above risk measures to a different numerical scale may help with communication and give people a more intuitive sense of risk. A 100-point numerical scale is created where a value of 0 on the scale is equivalent to no fatalities ($M_{min} = 0$) and a value of 100 is equivalent to exactly one fatality per one million miles ($M_{max} = 1$). Table V depicts risk indices using this scale for the measures based on the mean values, probabilities of a fatal incident, exponential disutility where $\gamma = 0.05$, and α-t model where $\alpha = 1.25$. (Decimals are rounded to the nearest integer for these four indices.) The logarithmic index in the last column of Table V uses a different transformation and will be analyzed shortly

Table V. Transportation risk index

	100-point scale Mean	Probability of fatal incident	Exponential disutility $\gamma = 0.05$	α-t model $\alpha = 1.25$	Logarithmic 10-point scale
Air carrier	0	0	1	2	7.5
Air taxi and commuter	67	23	73	95	9.8
Highway	1	1	1	3	8.1

The severity of the risk as communicated by the index differs according to the risk measure. The index value of 95 based on the risk averse α-t model makes air taxi and commuter risk appear extremely severe, but the index value of 23 based on the probability of a fatal incident makes the risk appear much more acceptable. When compared with air taxi and commuter, the indices for air carrier and highway are much less severe, and the air carrier index is zero or close to zero for indices based on the mean, probability of a fatal incident, exponential disutility, and the α-t model.

The logarithmic index is the most unique index depicted in Table V. This index is calculated by adding 10 to the base-10 logarithm of the expected number of fatalities. Thus, one fatality per one million miles corresponds to an index value of 10, and 10^{-10} expected fatalities corresponds to 0. (If 10^{-10} expected fatalities is too small to serve as an effective lower bound for the index, the base of the logarithm can be changed to reflect a different lower bound.) Taking the logarithm of the expected values as opposed to linear scaling—as in the other indices—compresses the differences between the expected number of fatalities. The risks of fatalities from each of the three modes of transportation appear more similar with a logarithmic index, but this index may more accurately reflect how people naturally perceive the differences among these risks.(66)

The analyst or the decision maker needs to determine which index best serves the objective. If the goal is to inform the public about the risk of each mode of transportation, the risk index based on the mean or probability of a fatal incident or the logarithmic index may be most appropriate. An index based on either of the disutility functions or a measure involving CVAR may best communicate how risk is changing over time to a policy maker that wants to emphasize incidents with a large number of fatalities. One of the risk measures from Table IV (probably the mean or a measure composed of two or three values) can help a decision maker identify if more safety measures are needed for one of the transportation modes; however, that decision should also incorporate the cost and effectiveness of the proposed safety measure.[2]

6. CONCLUSION

This paper has explored the different uses and types of risk indices and measures. We argue that a risk index should be based on a numerical analysis of the risk using probabilities and consequences. A risk measure summarizes the probability distribution and can be categorized into one of four types: moments, quantiles, disutility functions, and weighted indicators or factors. Risk measures can also be composed of two or three values, and decision makers who want to rely on risk measures should consider measure composed of multiple values.

Once the risk measure has been calculated, the analyst can map that measure to a numerical or categorical index, or the measure itself can serve as a risk index. Recall the list of objectives for risk indices outlined in the introduction. The most important goals for a situation should determine the type and structure of the risk index. If the most important goal is to describe the risk of an event in the most accurate way possible, the analyst should use the probability distribution or rely on a measure with multiple numbers. If a single number is desired, using a disutility function that incorporates the decision maker’s preferences to the greatest possible extent is best. A numerical measure is also most appropriate when the goal is to assess how a mitigating action impacts the risk. The measure can provide a metric to judge the effectiveness of different mitigation strategies if it is known how these mitigation strategies impact the probability distribution. In this case, the risk is first measured under the assumption of no mitigation strategy and then assessed assuming the mitigation strategy was enacted. The difference between the two measures describes the benefit of the mitigation strategy.

A numerical risk index can best achieve the objectives of comparing between different risks, determining the most serious risk, and understanding how a risk changes over time. A numerical risk index can enable a user to clearly see which risk is most serious and how close the risk is to the maximum level of risk, although determining the maximum level of risk can pose serious challenges.

A categorical scale can be most useful when the primary goal is to communicate risk to a large group of people and recommend actions if the risk falls into a certain category. Simple and clear messages can be most effective, especially in moments of crises, and a categorical scale that uses colors or words may be the clearest communication tool. If more than one goal is important as is usually the case, the analyst or policy maker should choose the index that best meets those goals and the situation.

The audience must also be considered when choosing the type of risk measure or index. If the audience is limited to a group of people who understand probability distributions, the analyst may forgo a risk measure and present the group with a probability distribution over the consequences. If the audience is composed of people who regularly work with numbers but who do not have the expertise or the time to work with probability distributions, a risk measure composed of multiple numbers may be the most appropriate. If the audience is an entire nation, a categorical index that is simple but informative may be ideal.

In all of these cases, an explanation should accompany the risk index. The analyst needs to explain what the number or category means, what it means if the risk increases or decreases, what it means if two risks have different risk numbers or are categorized differently, and the limitations of the risk index or measure. It takes time for people to trust and learn how to react to a new risk index. If the categories are specific and people understand why a risk is categorized at a given level, the index can provide useful information about the risk of an event to the users.

REFERENCES

1. American Society of Civil Engineers. America's infrastructure report card. Reston, VA, March 25, 2009. Available at: http://www.infrastructurereportcard.org/, Accessed on September 25, 2012.
2. Cox LA Jr. Evaluating and improving risk formulas for allocating limited budgets to expensive risk-reduction opportunities. Risk Analysis, 2012; 32(7):1244-52.
3. Hand DJ. Measurement Theory and Practice: The World through Quantification. New York: Oxford University Press, 2004.
4. Kaplan S, Garrick BJ. On the quantitative definition of risk. Risk Analysis, 1981; 1(1):11-27.
5. Hubbard DW. The Failure of Risk Management: Why It’s Broken and How to Fix It. Hoboken, NJ: John Wiley & Sons, 2009.
6. Paté-Cornell ME. Uncertainties in risk analysis: Six levels of treatment. Reliability Engineering and System Safety, 1996; 54(2-3):95-111.
7. Bedford T, Cooke RM. Probabilistic Risk Analysis: Foundations and Methods. Cambridge:

Cambridge University Press, 2001.

8. Cox LA Jr. Risk Analysis of Complex and Uncertain Systems, International Series in Operations Research and Management Science. New York: Springer, 2009.

9. Haimes YY. Risk Modeling, Assessment, and Management. 3rd ed. Hoboken NJ: John Wiley & Sons, 2009.

10. Paté-Cornell ME. Fire risks in oil refineries: Economic analysis of camera monitoring. Risk Analysis, 1985; 5(4):277-88.

11. Willis HH. Guiding resource allocations based on terrorism risk. Risk Analysis, 2007; 27(3):597-606.

12. Cox LA Jr. Why risk is not variance: An expository note. Risk Analysis, 2008; 28(4):925-928.

13. Vrijling JK, van Hengel W, Houben, RJ. A framework for risk evaluation. Journal of Hazardous Materials, 1995; 43(3):245-261.

14. Crouhy M, Galai D, Mark R. Risk management: comprehensive chapters on market, credit, and operational risk. New York: McGraw-Hill, 2001.

15. Artzner P, Delbaen F, Eber J-M, Heath D. Coherent measures of risk. Mathematical Finance, 1999; 9(3):203-28.

16. Föllmer, H, Schied A. Convex measures of risk and trading constraints. Finance and Stochastics, 2002; 6:429-447.

17. Acerbi C, Tasche D. Expected shortfall: A natural coherent alternative to value at risk. Economic Notes, 2002; 31(2):379-388.

18. Giesecke K, Schmidt T, Weber S. Measuring the risk of large losses. Journal of Investment Management, 2008; 6(4):1-15.

19. Toumazis I, Kwon C, Batta R. Value-at-risk and conditional value-at-risk minimization for hazardous materials routing. In: Batta R, Kwon C, editors. Handbook of OR/MS models in hazardous materials transportation. Springer; forthcoming.

20. Verderame PM, Floudas C. Operational planning of large-scale industrial batch plants under demand due date and amount uncertainty: II. Conditional value-at-risk framework. Industrial & Engineering Chemistry Research, 2010; 49(1):260-275.

21. Dehlendorff C, Kulahci M, Merser S, Andersen KK. Conditional value at risk as a measure for waiting time in simulations of hospital units. Quality Technology & Quantitative Management, 2010; 7(3):321-6.

22. Asbeck EL, Haimes YY. The partitioned multiobjective risk method (PMRM). Large Scale Systems 1984; 6:13-38.

23. Bell DE. Risk, return, and utility. Management Science, 1995; 41(1):23-30.

24. Jia J, Dyer JS. A standard measure of risk and risk-value models. Management Science, 1996; 42(12):1691-1705.

25. Liu Y-W, Kapur KC. New models and measures for reliability of multi-state systems. In Misra KB, editor. Handbook of performability engineering. London: Springer-Verlag; 2008. p. 431-445.

26. Weber S. Distribution-invariant risk measures, information, and dynamic consistency. Mathematical Finance, 2006; 16(2):419-441.

27. Klöppel S, Schweizer M. Dynamic indifference valuation via convex risk measures. Mathematical Finance, 2007; 17(4):599-627.

28. Föllmer, H, Schied A. Robust preferences and convex measures of risk. In: Sandmann K, Schönbucher PJ, editors. Advances in finance and stochastics: essays in honour of Dieter Sondermann. New York: Springer-Verlag; 2002. p. 39-56.

29. Fishburn, PC. Mean-risk analysis with risk associated with below-target returns. The American Economic Review, 1977; 67(2):116-126.

30. Donkers B, Melenberg B, van Soest A. Estimating risk attitudes using lotteries: A large sample approach. Journal of Risk and Uncertainty, 2003; 22(2):165-195.

31. Hallahan TA, Faff RW, McKenzie MD. An empirical investigation of personal financial risk tolerance. Financial Services Review, 2004; 13:57-78.

32. Fischhoff B, Watson SR, Hope C. Defining risk. Policy Sciences, 1984; 17:123-139.

33. Davidson RA. An urban earthquake disaster risk index [dissertation] Stanford, CA: Stanford University; 1997. 280 p.

34. Davidson RA, Lambert KB. Comparing the hurricane disaster risk of U.S. coastal cities. Natural Hazards Review, 2001; 2(3):132-142.

35. Anemüller S, Monreal S, Bals C. *Global Climate Risk Index 2006: Weather-Related Loss Events and Their Impacts on Countries in 2004 and in a Long-Term Comparison*. Bonn: Germanwatch e.V., 2006. Available at: www.germanwatch.org/klak/cri06.pdf, Accessed on September 25, 2012.

36. Piegorsch WW, Cutter SL, Hardisty F. Benchmark analysis for quantifying urban vulnerability to terrorist incidents. Risk Analysis, 2007; 27(6):1411-1425.

37. Gilchrist W. Modelling failure modes and effects analysis. The International Journal of Quality & Reliability Management, 1993; 10(5):16-23.

38. Ben-Daya M, Raouf A. A revised failure mode and effects analysis model. The

International Journal of Quality & Reliability Management, 1996; 13(1):43-47.

39. Long J, Fischhoff B. Setting risk priorities: A formal model. Risk Analysis, 2000; 20(3):339-51.

40. Paté-Cornell ME. Quantitative safety goals for risk management of industrial facilities. Structural Safety, 1994; 13(3):145-157.

41. Plum MM. Rational and development of a scale to communicate environmental and other community risks. In: Linkov I, Ferguson E, Magar VS, editors. Real-time and deliberative decision making, Dordrecht, The Netherlands: Springer; 2009. p. 103-126.

42. Paulos JA. Innumeracy: Mathematical Illiteracy and Its Consequences. New York: Hill and Wang, 1988.

43. Novy, CH, Edwards R, Imy D, Goss S. SPC and Its Products, March 2010. Available at: http://www.spc.noaa.gov/misc/about.html#Severe Weather Watches, Accessed on September 25, 2012.

44. Bureau of Transportation Statistics (US), Research and Innovative Technology Administration. Table 2-1: Transportation fatalities by mode. In: National transportation statistics, Chapter 2 - transportation safety, section A – multimodal, April 2012. Available at: http://www.bts.gov/publications/national_transportation_statistics/html/table_02_01.html, Accessed on September 25, 2012.

45. Evans L, Frick MC, Schwing RC. Is it safer to fly or drive? Risk Analysis, 1990; 10(2):239-246.

46. Sivak M, Weintraub DJ, Flannagan M. Nonstop flying is safer than driving. Risk Analysis, 1991; 11(1):145-148.

47. McCarthy PS. Public policy and highway safety: A city-wide perspective. Regional Science and Urban Economics, 1999; 29:231-44.

48. Horn MET, Fulton N, Westcott M. Measures of societal risk and their potential use in civil aviation. Risk Analysis, 2008; 28(6):1711-26.

49. Li G, Baker SP. Crash risk in general aviation. JAMA, 2007; 297(14):1596-1598.

50. Bhalla K, Ezzati M, Mahal A et al. A risk-based method for modeling traffic fatalities. Risk Analysis, 2007; 27 (1):125-36.

51. Bureau of Transportation Statistics (US), Research and Innovative Technology Administration. Table 2-9: U.S. air carrier safety data. In: National transportation statistics, Chapter 2 - transportation safety, section B – air, July 2011. Available at: http://www.bts.gov/publications/national_transportation_statistics/html/table_02_09.html, Accessed on June 27, 2012.

52. Bureau of Transportation Statistics (US), Research and Innovative Technology Administration. Table 2-10: U.S. commuter air carrier safety data." In: National transportation statistics, Chapter 2 - transportation safety, section B – air, July 2011. Available at: http://www.bts.gov/publications/national_transportation_statistics/html/table_02_10.html, Accessed on June 27, 2012.

53. Bureau of Transportation Statistics (US), Research and Innovative Technology Administration. Table 2-13: U.S. on-demand air taxi safety data." In: National transportation statistics, Chapter 2 - transportation safety, section B – air, October 2011. Available at: http://www.bts.gov/publications/national_transportation_statistics/html/table_02_13.html,

Accessed on June 27, 2012.

54. Bureau of Transportation Statistics (US), Research and Innovative Technology Administration. Table 2-18: Motor vehicle fatalities, vehicle-miles, and associated rates by highway functional system." In: National transportation statistics, Chapter 2 - transportation safety, section C – highway, July 2012. Available at: http://www.bts.gov/publications/national_transportation_statistics/html/table_02_18.html, Accessed on September 25, 2012.

55. National Transportation Safety Board (US). Aviation: Accident Database & Synopses, 2012. Available at: http://www.ntsb.gov/aviationquery/index.aspx, Accessed on June 28, 2012.

56. National Highway Traffic Safety Administration (US). Fatality Analysis Reporting System (FARS) Encyclopedia. NCSA Data Resource Website, 2012. Available at: http://www-fars.nhtsa.dot.gov/Main/index.aspx, Accessed on June 25, 2012.

57. Fischhoff B, Lichtenstein S, Slovic P, Derby SL, Keeney RL. Acceptable Risk. Cambridge: Cambridge University Press; 1981.

58. Slovic P, Lichtenstein S, Fischhoff B. Modeling the societal impact of fatal accidents. Management Science, 1984; 30(4):464-474.

59. Keller LR, Wang Y, Feng T. Measures of risk equity. In: JJ Cochran, editor. Wiley encyclopedia of operations research and management science. Hoboken, NJ: John Wiley & Sons; 2011.

60. Griesmeyer JM, Okrent D. Risk management and decision rules for light water reactors. Risk Analysis, 1981; 1(2):121-136.

61. Vatn J. A discussion of the acceptable risk problem. Reliability Engineering and System Safety, 1998; 61:11-19.

62. Vrijling JK, van Hengel W, Houben RJ. Acceptable risk as a basis for design. Reliability Engineering and System Safety, 1998; 59(1):141-150.

63. Ferreira J Jr, Slesin L. Observations on the social impact of large accidents. Massachusetts Institute of Technology; October 1976. 46 p. Technical Report No. 122.

64. Keeney RL. Equity and public risk. Operations Research, 1980; 28(3):527-534.

65. Fischhoff B, Slovic P, Lichtenstein S. Knowing what you want: Measuring labile values. In: Wallsten TS, editor. Cognitive processes in choice and decision behavior. Hillsdale, NJ: Lawrence Erlbaum Associates; 1980. p. 117-141.

66. Calman KC, Royston GHD. Risk language and dialects. British Medical Journal, 1997; 315:939-942.