Value of GCR Information: Cost Effectiveness-Based Approach for Global Catastrophic Risk (GCR) Reduction

Anthony M. Barrett*,† (tony@gcrinstitute.org)
†Global Catastrophic Risk Institute
* http://tony-barrett.com * http://gcrinstitute.org

Abstract

In this paper, we develop and illustrate a framework for determining the potential value of global catastrophic risk (GCR) research in reducing uncertainties in the assessment of GCR risk levels and the effectiveness of risk-reduction options. The framework uses the decision-analysis concept of the expected value of perfect information (EVPI) in terms of the cost-effectiveness of GCR reduction. We illustrate these concepts using available information on impact risks from two types of near earth objects (asteroids or extinct comets) as well as nuclear war, and consideration of two risk reduction measures. We also discuss key challenges in extending the calculations to all GCRs and risk-reduction options, as part of an agenda for comprehensive, integrated GCR research. While real-world research would not result in perfect information, even imperfect information could have significant value in informing GCR reduction decisions. Unlike most value of information approaches, our equation for calculating value of information is based on risk reduction cost effectiveness, to avoid implicitly equating lives and dollars e.g. using a value of statistical life (VSL), which may be inappropriate given the scale of GCRs. Our equation for value of information may be useful in other domains where VSLs would not be appropriate.

1. Introduction

Global catastrophic risks (GCRs) are risks of events that could significantly harm or even destroy human civilization at the global scale (Hempsell 2004; Baum 2010). GCRs presently posing hazards to humanity include nuclear war (Sagan 1983; Turco, Toon et al. 1983; Robock, Oman et al. 2007; Cirincione 2008; Hellman 2008; Barrett, Baum et al. 2013) and pandemic diseases (Nouri and Chyba 2008). In the near to longer term future, GCRs could include climate change (Weitzman 2009; Travis 2010; Baum, Maher Jr. et al. 2013) and misuse or accidents involving technological developments in areas such as artificial intelligence (Yudkowsky 2008; Chalmers 2010; Sotala 2010) and nanotechnology (Phoenix and Treder 2008). Proposed interventions to reduce GCR include nuclear disarmament (Robock, Oman et al. 2007), development and distribution of vaccines and antiviral medications (Osterholm 2005), reducing greenhouse gas emissions through public policies (Aldy, Barrett et al. 2003) and various individual behaviors (Dietz, Gardner et al. 2009), and abstaining from developing certain technologies (Joy 2000).

A growing body of work makes the case that reducing GCR, or certain types of GCR, is of very high value and thus should be one of the highest objectives for society (Ng 1991; Bostrom 2002; Posner 2004; Matheny 2007; Tonn 2009; Ćirković, Sandberg et al. 2010; Beckstead 2013). Published estimates of the value of preventing global catastrophe range vary wildly, from $10 billion (Bostrom and Cirkovic 2008) to infinity (Weitzman 2009; Baum 2010), depending partly on the definition used for "global catastrophe". Even the low end of this suggests a large allocation of resources towards GCR reduction.

However, setting GCR reduction as a high priority is not a sufficient guide for action: there are many open questions regarding how best to allocate resources for GCR reduction. One basic question is how much to allocate toward direct risk-reducing interventions, and how much to allocate to research to inform these interventions. The decision analysis concept of expected value of information (Clemen and Reilly 2001; Keisler 2004; Bhattacharjya, Eidsvik et al. 2013) can inform decisions about how much to spend on information (i.e., reduce uncertainties) prior to making other resource allocation decisions. Usually in value-of-information calculations, decision options are evaluated using utility functions, money, or functionally similar metrics that have implicit commensurability between option tradeoffs, e.g. lives saved vs. dollars spent. However, equating lives and dollars, e.g. using a typical value of statistical life saved (VSL), may be inappropriate given the potentially vast scale of GCRs. (Moreover, quantifying total event consequences of global catastrophe in conventional benefit-cost terms would be complicated by uncertainties about direct event impacts, indirect impact factors such as public behavioral responses, and the levels of such impacts that could be borne before reaching civilizational-collapse tipping points.) We take a different, cost effectiveness based approach in this paper instead. A cost effectiveness based equation for value of information also may be useful in other domains where typical VSLs would not be appropriate.

In this paper, we argue that value of information based on cost-effectiveness is a useful tool for analysis of GCR to inform risk-reduction decisions, and we show that it can be defined in a practical manner. We argue that such an approach would be most valuable if applied in a comprehensive, integrated fashion to all major types of GCR, rather than one at a time. We describe a number of challenges that would arise in such efforts, and argue that these challenges can be addressed. We also provide an illustrative, though highly idealized, example that shows how a practical value of information calculation can work; it also provides support for our argument that such calculations can have considerable value, and it provides further support for our argument that value of information can provide additional insight when more than one GCR is under consideration.

In Section 2 of this paper, we give a brief overview of the basics of the approach, and how to apply it to GCRs and risk-reduction interventions in a comprehensive, integrated fashion. In Section 3, we discuss key challenges in real-world implementation of this paper's framework, and argue that these challenges can be addressed. In Section 4, we illustrate the basic framework using a simple notional model of GCR from two types of near earth objects (NEOs, i.e. asteroids

and extinct comets) as well as nuclear war, and consideration of two related risk reduction measures; the illustrative example shows that such calculations can have considerable value, especially when considering multiple GCR. We conclude in Section 5. (In Appendix I, we provide a detailed derivation of our formula for the expected value of information in terms of the cost-effectiveness of GCR reduction.)

2. Overview of Framework for Value of GCR Information

In this section, we briefly discuss ways to approach three linked sets of quantitative issues: First, representing the probabilities of multiple GCRs; second, assessing the overall cost effectiveness of GCR reduction measures and calculating the value of information for GCR reduction; third, contrasting perfect and imperfect information. More details of our approaches and assumptions are given in the following sections.

2.1 GCR Probabilities

Figure 1 is a fault tree or logic tree illustrating that there are multiple types of global catastrophic risks, occurrence of which is assumed to be causally independent of the others, at least at the level of detail used in the fault tree (e.g. nuclear war does not cause asteroid impact). The event "Global Catastrophe" is the top event, with round-corner nodes for a series of GCR types branching out below, all connected by an OR gate. The fault tree graphically indicates that a global catastrophe will occur if any of the following types of event occur with global catastrophe-level consequences: a large NEO impact (either an asteroid or a comet impact), or large nuclear war, or a combination of smaller events (small NEO impact plus small nuclear war), etc. In addition, Figure 1 includes square-corner decision nodes risk management for two types of GCR reduction options: (i.e. NEO redirection, and food stockpiling) that could reduce the probabilities of global catastrophe level outcomes. Grey arrows from the square-corner decision nodes to the round-corner fault tree nodes indicate that the risk management decisions can influence the risks of global catastrophe level events. Figure 1 also illustrates that some risk reduction measures, e.g. food stockpiling, can have benefits in reducing multiple types of GCR. Although the fault tree portion of Figure 1 is quite simple, it is intended to underline the main motivation for considering GCRs as a whole, and not just individual types such as asteroids, comets, or nuclear war: in order to assess and reduce the total probability of global catastrophic risk, ideally we would assess all types of GCRs and GCR reduction measures in a comprehensive way. The framework also can account for interactions between GCR events, such as when occurrence of one type of event reduces society's resilience to or even causes another type of event (Baum, Maher Jr. et al. 2013). Such interactions between GCRs could be represented using larger, more detailed fault trees (e.g., by adding branches for scenarios in which both NEO impact and nuclear war events occur around the same time, either just by chance of timing or because a NEO impact somehow causes nuclear war), though it could be difficult to explicitly account for many GCR-interaction scenarios, and important uncertainties could remain about unmodeled GCR-interaction dynamics.

3
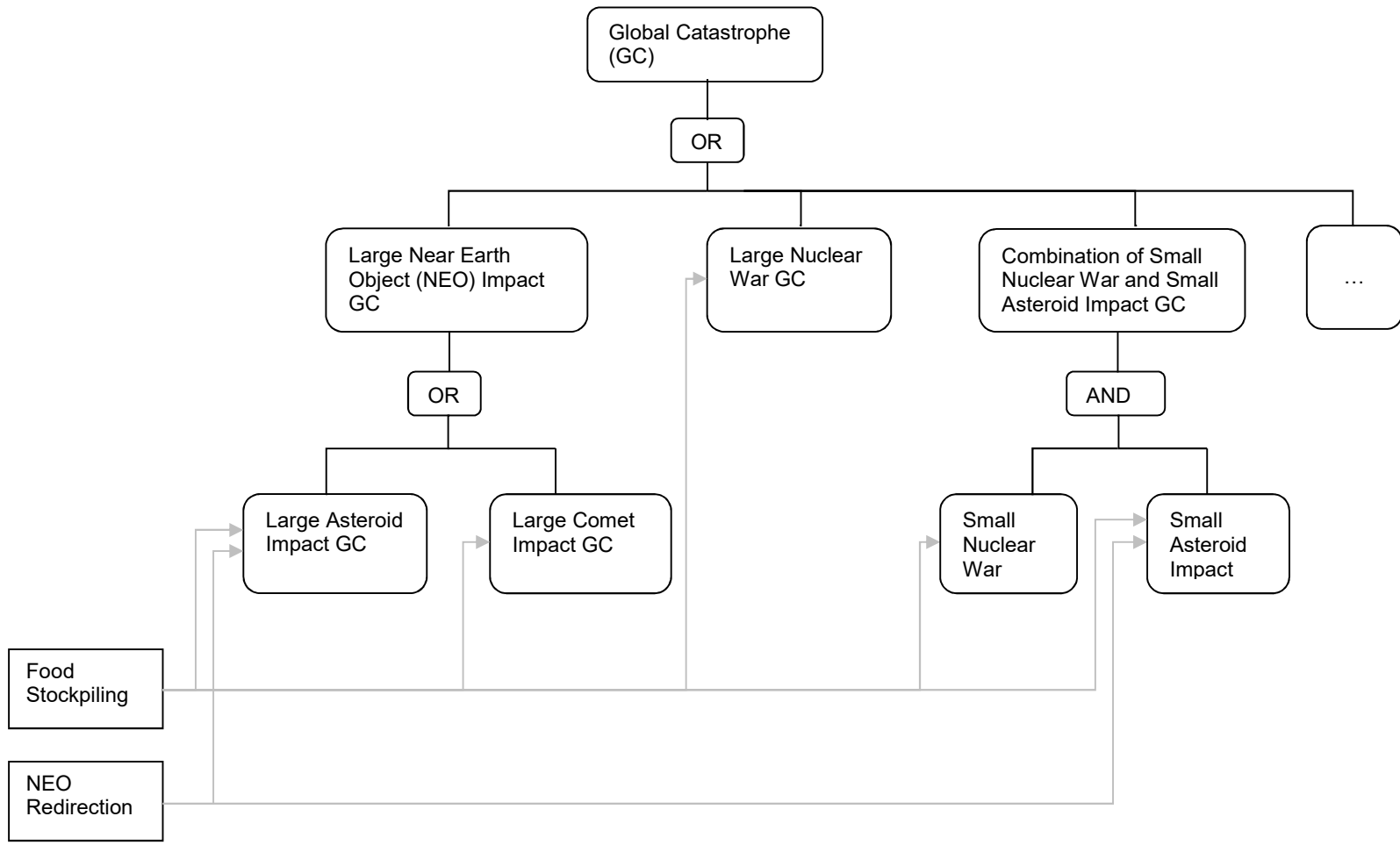
Figure 1: High-Level Global Catastrophe Fault Tree and Risk Management Decision Influence Diagram

Figure 2 is a generic consequence exceedance probability plot for some type of event (e.g. NEO impacts) with curves showing relationships between event consequence and probability of events with consequence exceeding that level, for both initial and reduced event risks. The figure illustrates that reduction in probability of global catastrophe can be achieved either by reduction of probability of events, or by reduction of consequences. Starting from the upper right of the figure, the point where the initial event probability-consequence curve intersects with the global catastrophe consequence threshold indicates the initial probability of global catastrophe. The figure also includes two reduced-risk curves, one for reduced event probability and another for reduced event consequence. The curves for reduced probability and reduced consequence have been placed where they result in the same reduction in probability of global catastrophe, partly to keep the figure simple and partly to emphasize the idea that GCR reduction can be achieved by reducing either event probability or consequence. For example, NEO impact risk reduction measures could reduce the probabilities of global catastrophe level outcomes either by shifting the curve downwards with reduced NEO impact probabilities (e.g., via NEO redirection) or by shifting the curve leftward with reduced NEO impact consequences (e.g., increasing societal resilience to NEO impact via food stockpiling). Thus, probabilities of global catastrophe for a particular GCR event type could be calculated as a function of global catastrophe consequence threshold, using consequence exceedance probability models for that event type. Of course, development of appropriate consequence exceedance probability models would often require substantial research, especially when focusing on rare or unprecedented events, for which a lack of data often leads to substantial uncertainties and biases (Taylor 2008).
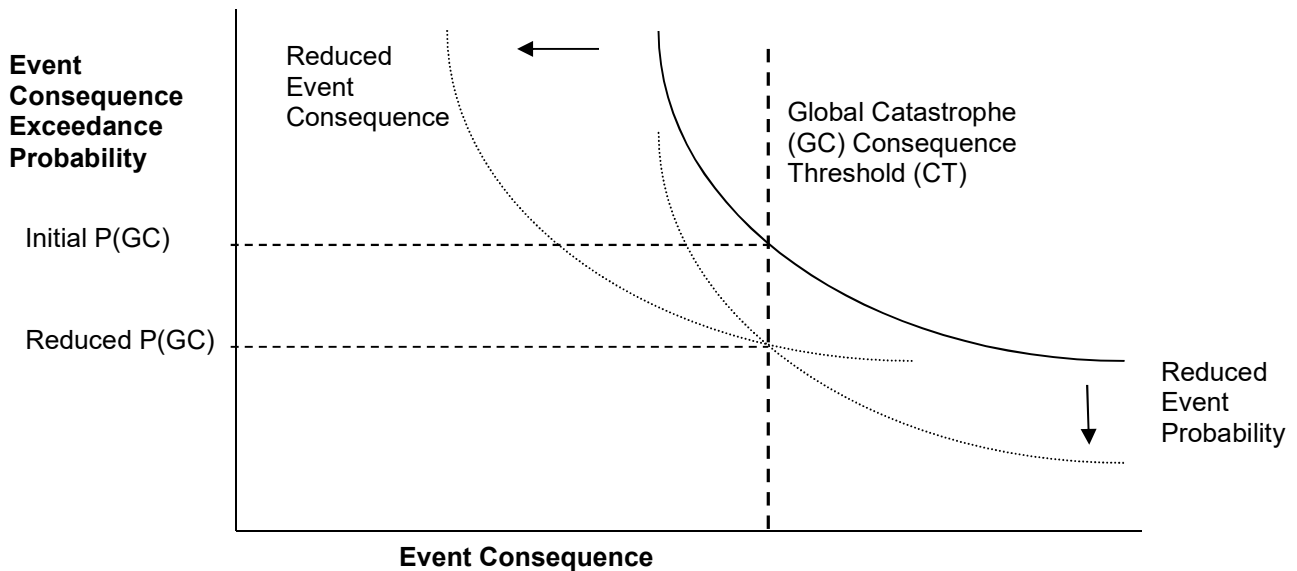
Figure 2: Global Catastrophe Probability as Function of Event Consequence and Exceedance Probability

In many cases, there would be large uncertainties for both the direct consequences of an event (e.g. in terms of atmospheric soot loading from various nuclear war or NEO impact scenarios), as well as what threshold level of consequences would result in global catastrophe (e.g. in terms of the effects of atmospheric soot loading on agricultural productivity and other indirect effects on human society, which could be highly nonlinear if stresses could reach civilizational resilience-exceedance tipping points). Such uncertainties could be modeled using probability distributions for the global catastrophe consequence threshold and exceedance probability function. One way to represent uncertainties is to display 5th and 95th percentile value lines in addition to the mean value lines (Garrick 2008), as shown in Figure 3.
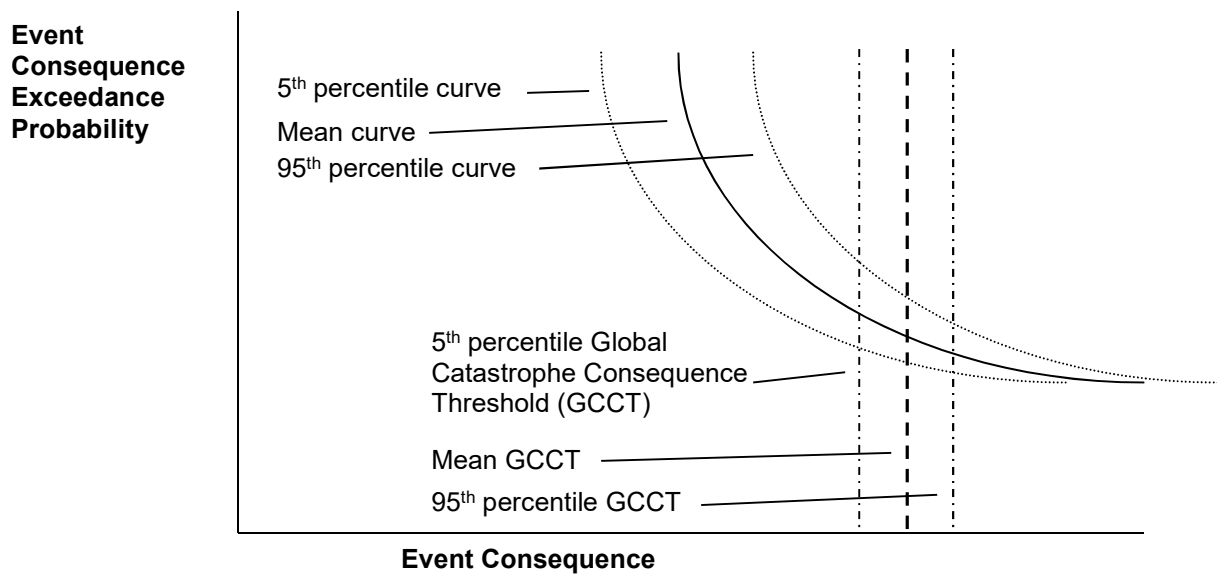
**Figure 3: Uncertainties in Global Catastrophe Probability Modeling**

Given the previously mentioned assumption of causal independence, Equation 1 gives the total probability of a global catastrophe level event within some time period, p_{total}, as a function of the independent probabilities p_j of catastrophe events of each GCR type, j, for a total of y GCR types. Equation 1 is mathematically consistent with the previous statement that a global catastrophe will occur if any of the following types of global catastrophe occurs: a large asteroid impact, or a comet impact, or nuclear war, etc.

$$p_{total} = 1 - \prod_{j}^{y}(1 - p_j) \qquad (Eq. 1)$$

2.2 Cost Effectiveness and Value of Information for GCR Reduction

Figure 4 is a high-level decision tree, consistent with typical trees used in calculating value of information. (The tree does not include specific quantitative values for probabilities, costs and benefits, but it does indicate the general sequence of decisions and events.) The

6

leftmost square decision node represents a decision to be made on whether to invest in research to inform decisions on risk reduction measures; other decision nodes represent addition to a basic decision on whether to measures to reduce risks. In Figure 4, the research decision is simple: conduct research to better understand whether risks are currently high or low, or do not conduct such research. The risk reduction decision options are also simple: invest to reduce risks, or do not invest to reduce risks. The decision on whether to conduct research is made before the decision on whether to invest in reducing risks. If the decision maker chooses not to conduct research, then they make the risk reduction decision with some amount of uncertainty about whether risks begin as high or low. (That uncertainty is represented by circular chance nodes, and the outcomes of chance nodes are represented by diamonds.) If the decision maker does choose to conduct research, then they have more information and less uncertainty about whether risks begin as high or low, and the decision maker can use that information when making their decision on whether to invest in reducing risks.
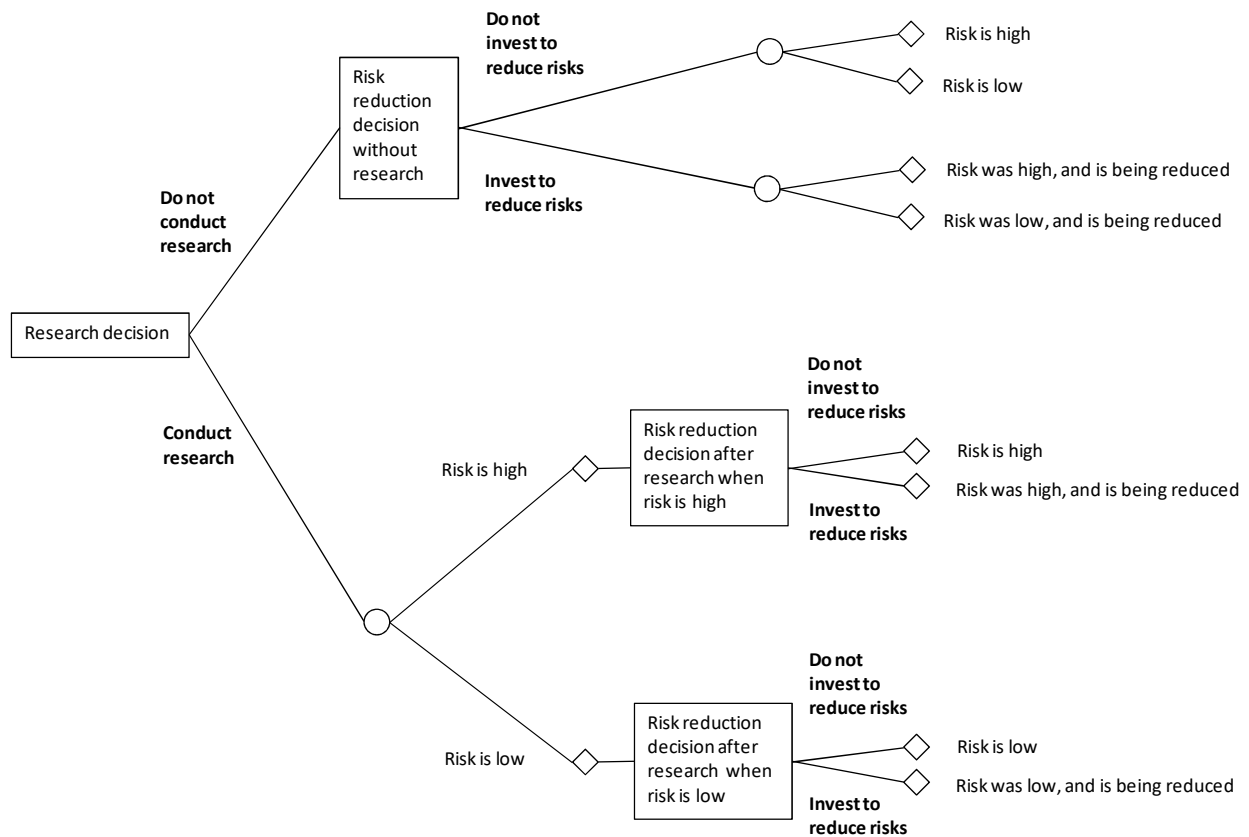
Figure 4: High-Level Decision Tree for Research and Risk Reduction Decisions

A full valuation of GCR reduction interventions, including research to gain information, requires some evaluative metric. Typically, decision options are evaluated using utility functions or functionally similar metrics.[i] Such metrics have implicit commensurability between option tradeoffs, e.g. lives saved vs. dollars spent. Use of such approaches allows for a relatively

simple equation for expected value of options with various attributes (Clemen and Reilly 2001), including tradeoffs between GCR reduction and other objectives.

In this paper, we avoid full valuations and instead conduct partial valuations in terms of cost effectiveness, measured in GCR reduction per unit cost. We focus on cost effectiveness for two reasons. First, a full valuation for GCR is complicated by the widely varying estimates for the value of preventing global catastrophe; which can range from $10 billion to infinity, as mentioned in Section 1. Second, many GCR reduction decisions involve allocating resources, such as money.

However, equating lives and dollars, e.g. using a value of statistical life saved (VSL), may be inappropriate given the scale of GCRs. Therefore, our equation for calculating value of information is based on risk reduction cost effectiveness, which incorporates estimates of the performance and costs of risk reduction options without use of VSLs. Our cost effectiveness based equation for value of information may be useful in other domains where VSLs would not be appropriate.

We assume that there are one or more decisions to be made about the allocation of resources to some combination of options for risk reduction and options for research, and that the decision rule is to choose whatever combination of options has best overall expected GCR reduction cost effectiveness among options considered in the analysis. (Such considerations could occur in a series of risk-reduction decisions, in which case the goal could be to identify the most cost-effective interventions first, and then the second-most, and so on until a risk reduction budget or target has been reached.) Then in such decisions, the decision maker should buy as much risk reduction (and risk research enabling better risk reduction decisions) as they can at whatever total cost, as long as that results in the greatest cost effectiveness. Such decisions can arise when considering public policies, as well as the actions of individuals and other non-governmental organizations. (We assume that budgets are not an issue in the context of the risk reduction and research options under consideration, and we do not explicitly account for potential budget constraints in the following. However, consideration of budget constraints could be addressed as an extension of the approach used in the following.)

For the purposes of this analysis we ignore actual costs of research and focus on the amount of resources the decision maker ought to be willing to pay for the value added by the research in the context of the decision the research could inform. In other words, we focus on finding the maximal potential benefits of research. We assume that research ought to be invested in up to the point where a funder would obtain no further benefit from investing in additional research (because up to that point, they would get a better overall cost effectiveness by investing in additional research). At that point, the expected cost effectiveness of the best risk-reduction option before research is equal to the expected cost effectiveness of the best risk-reduction option after research, including the cost of research.

Equation 2 gives the value of research as the cost-effectiveness based expected value of perfect information, *CEEVPI* (see Appendix I for derivation):

$$CEEVPI = E\left[\frac{(c_s^b)(p_0^a - p_s^a)}{(p_0^b - p_s^b)} - c_s^a\right] \qquad \text{(Eq. 2)}$$

The equation assumes the following: There exists a set of n available risk reduction options numbered 0, 1, ... i... n. Option number 0 is the status quo case, where no new (or non-"business-as-usual") risk reduction option is implemented. c_i is the cost of implementing risk reduction option i. (It costs nothing to do nothing, so $c_0 = 0$.) p_i is the annualized total probability of global catastrophe if implementing option i. (We make the simplifying assumption that p_i values are static, or unchanging over the relevant time period. Consideration of dynamic, or time-varying p_i values, could be addressed as an extension of the approach used in the following.)

Each c_i is treated as a random variable with some probability distribution reflecting uncertainty about the true cost of implementing intervention i. Each p_i is also treated as a random variable, with a probability distribution reflecting plausible estimates of the annual probability of global catastrophe given intervention i.[ii]

Computationally, the uncertainty is represented using Monte Carlo simulation, where in Monte Carlo simulation iteration m, there are sampled values c_{im} and p_{im}. The risk reduction option s is the option with the "best" or highest risk reduction cost effectiveness in Monte Carlo iteration m.

In addition to decisions on which risk reduction option to choose, there are also decisions on whether to first spend some resources on research to reduce uncertainties (and to more accurately identify which risk reduction option would be most cost effective) before making decisions on risk reduction options. We denote cases whether research is conducted to reduce uncertainty on a particular factor using superscript b for "before" research or without information from research, and superscript a for "after" research or with information from research.

Generally, research will have the greatest expected value if it has substantial possibility of informing a decision, i.e. a choice between risk reduction options. However, the *CEEVPI* formula also implies that if it is expected that the best option after research is the same as the best option before research (i.e., if $s^a = s^b$), then the research still can have positive expected value if it is inexpensive enough and also provides sufficient reduction of uncertainties in p and c factors.

We provide an example, calculating *CEEVPI* for illustrative catastrophic NEO impact risks and risk reduction options, in Section 4. The example suggests that the value of GCR information could be quite substantial.

2.3 Perfect and Imperfect Information

In the context of a decision analytic model, the value of information is based on the extent to which information reduces the uncertainty about the value of a particular parameter in the model. Perfect information eliminates that uncertainty. The expected value of perfect information (EVPI) is the difference between the expected value of a decision with perfect

9

information (where the new information influences the decision we make) and without additional information (where we make the decision with our initial level of uncertainty) (Clemen and Reilly 2001).

We do not expect real-world GCR research to yield perfect information in the sense of eliminating all uncertainties. In general, EVPI calculations are used to set an upper limit to how much should be spent on reducing uncertainty. On their own, EVPI calculations cannot predict how valuable specific research will be in reducing uncertainty. However, even imperfect information can have great value in reducing decision model parameter uncertainties by some amount. Straightforward extensions of the approach to EVPI calculations used in this paper (based on cost effectiveness calculations) could provide methods to assess the Expected Value of Imperfect Information (EVII) (Clemen and Reilly 2001) and Expected Value of Including Uncertainty (EVIU) (Morgan and Henrion 1990).

3. Key Challenges of Integrated Assessment of GCR

In this section, we discuss important challenges for implementation of our framework for calculating value of information, and for comprehensive, integrated assessment of GCR to inform risk-management decisions. We have already mentioned some of these challenges, which we discuss further here; we also discuss others that we have not previously mentioned.

One challenge is that in the real world, there would often be complex interactions between GCRs, not all of which could be modeled. As previously mentioned, one important simplification of our approach is the assumption of independence of GCRs except where indicated in the model. In principle, many types of interactions could be accounted for by building them into fault trees or other model components, but that could require substantial efforts. As with modeling of any complex system, there would be large uncertainties about how much of the real-world dynamics would remain unmodeled. A similar set of challenges (and irreducible uncertainties) would be encountered in attempting to define global catastrophe consequence thresholds.

Another challenge would be in setting appropriate thresholds for catastrophe. An important simplification of our approach is that we use a binary threshold for catastrophe (i.e. an event is only regarded as a global catastrophe if the event's consequences exceed the global catastrophe consequence threshold, however that is defined). In reality, events of a range of magnitudes could be regarded as global catastrophes, either because different stakeholders have different definitions of what constitutes a global catastrophe, or because of uncertainties about what levels of direct effects from catastrophe events would reach civilizational tipping points. (Those uncertainties would stem partly from the difficulty of predicting indirect effects of catastrophe events, which involve complex factors such as the behavioral responses of large human populations. However, the analytic challenges and uncertainties would be even greater if the aim were to quantify total event consequences in conventional benefit-cost terms, which is another reason to use a simpler cost-effectiveness approach.) Differences between global catastrophe thresholds can have important implications for decision making.[iii] Decisions should

favor preventing higher-magnitude global catastrophes, or to decrease the severity of any given global catastrophe. Furthermore, ideally decisions would be robust (not highly sensitive) to placement of global catastrophe threshold. As always, sensitivity analysis can usefully examine the decision implications of varying global catastrophe thresholds, and uncertainty analysis can suggest ranges to use in sensitivity analysis.

There also would be challenges in defining what decision procedures to actually use, and how to incorporate considerations such as budget constraints and timing decisions. It seems unwise to take a perfectionist approach to assessing risks and risk-reduction optimality, because the complexity and scale of all potential risks and intervention options (including all interactions and combinations) could make that approach intractable. A more practical approach could be to make a series of risk-reduction decisions, either at regular or irregular intervals, that would first implement the most cost-effective interventions (or combination of interventions), then the second-most, and so on until a risk reduction budget or target risk level is reached (essentially a greedy algorithm solution to a knapsack problem, in operations research terms). We believe the latter approach would be roughly consistent with our basic framework, though our current framework does not attempt to explicitly account for budget constraints, nor decision sequentiality. It also should be noted that our basic approach implicitly assumes the goal is zero probability of global catastrophe, but other targets could be used; for example, Tonn (2009) suggests a 10^{-20} annual probability of global catastrophe as an "acceptable risk" target. Finally, accounting for timing of events and interventions could present substantial complications. For some issues, it could be important to account for decisions of exactly when to research, when to implement measures, and in what sequence; the urgency of implementing various measures also could be important. Although time dependencies are not explicitly reflected in the level of detail given in this paper, implicitly they could be incorporated into the model parameter values for effects of the risk reduction measures. (For example, if considering implementation of an intervention today, vs. some years from now, and if the GCR minimization objective is to minimize the probability of global catastrophe over the next century, then for many GCRs types such as NEO impact, presumably analysis would show greater GCR reduction benefits from implementing interventions sooner rather than later.) At least in principle, time dependencies could be accounted for in modules whose outputs are fed to the model structure shown in this paper.

Another challenge is the fact that in the real world, there is not a single very well-funded actor whose prime objective is to reducing GCR cost effectively. Instead, there are many potentially important decision makers, each with limited budgets and responsibility for GCR factors, and with various objectives that compete with GCR reduction. Potentially important decision making entities include government agencies such as the US National Aeronautics and Space Administration (NASA), which have programs to address specific categories of GCR such as NEO impact risks; nongovernmental organizations such as the Open Philanthropy Project, which have programs to address either specific categories of GCR or all GCR broadly; corporations such as Walmart, whose product management decisions can have implications for

societal resilience, emerging technologies and other GCR factors; and individuals such as researchers, whose work can improve understanding of GCR factors and that have decisions to make about where to focus their own research efforts. Nevertheless, if credible integrated assessment has identified some GCR reduction options as clearly being more cost effective than others, that could influence decisions by various means, especially where actors already have some incentives to reduce societal risks. For government agencies, integrated assessment could inform budget reallocations, e.g. taking funding from low-value areas to fund higher-value risk-reduction programs, and incentives could be provided via government rules that encourage cost-effective risk-reduction benefits to society. At the other end of the size spectrum, for individual researchers, integrated assessment could suggest which kinds of research could best lead to risk-reduction societal impacts, which are encouraged by both formal funding reviews and informal norms. Nongovernmental organizations and corporations also often combine efforts on voluntary stewardship initiatives and other programs to reduce societal risks, and thereby gain reputational rewards.

Some of the most important challenges concern the scope of analysis, such as what GCRs and risk reduction measures to consider initially (given that starting-point estimates or at least bounding ranges would be needed for all associated modeling parameter values).[iv] One approach that should be relatively tractable would be breadth-first: begin by taking a broad-but-shallow approach to modeling GCRs and risk-reduction options relatively comprehensively, but with little detail, and with quantitative parameter estimates aimed only at bounding ranges of uncertainties; then a series of subsequent, repeated model-improvement steps could iteratively add depth (i.e. to add detail and better quantitative estimates using the best available empirical data, expert judgment, etc.); decisions on where to focus model-improvement efforts via research could be guided by value-of-information calculations.

4. Illustrative Example: Notional Model of NEO Impact Risk and Mitigation

In this section, we illustrate our concepts using information in the literature on impact risks posed by two types of near earth objects (NEOs) as well as nuclear war. We also provide illustrative modeling of two types of impact risk reduction measures (i.e. NEO redirection, and food stockpiling) that could reduce the probabilities of global catastrophe level outcomes. These are very simple, notional models of risks and decisions, intended only to illustrate our value-of-information concepts. The example does not attempt to reflect all the latest references, such as the information on asteroid and comet impact risks yielded by the WISE and NEOWISE survey programs (such research has often resulted in downward revisions in catastrophe probability estimates for both high- and low-albedo NEOs). The example also does not attempt to estimate the risks, nor risk-reduction benefits, related to GCRs besides NEOs and nuclear war, although considering those would affect overall GCR reduction cost effectiveness estimates. For example, food stockpiling could have benefits in reducing the effective consequences of pandemics, which is a category of GCR that is not considered in this illustrative example.

4.1 Illustrative Model GCRs, Risk Reduction Measures, and Assumptions

The first type of NEO we model is "bright" or easily visible asteroids/comets, which can be observed and tracked long before impact using current astronomical capabilities. The Spaceguard Survey is believed to have detected most such NEOs with greater than 1 km diameter (NRC 2010). The second type of NEO we consider is "dark", or low reflectivity Damocloids, which current identification and tracking systems may not see until the objects are already headed directly toward impact. Partly because of the difficulty of observing Damocloids using optical telescopes, there are large uncertainties about the frequencies of Damocloid impact (Napier 2008; NRC 2010). Before the Spaceguard Survey, such objects were thought to be a small risk relative to other NEOs, but Damocloids and long-orbit objects have more recently been viewed as potentially posing the majority of remaining impact risk (NRC 2010). The NEOWISE survey program has been using infrared to better identify Damocloids. Presumably, additional investments in research using infrared, radar or other technologies could provide better observations of Damocloids; perhaps such Damocloid observation systems would be deployed as some combination of Earth-based systems, satellites, and probes.

We also model two types of impact event risk reduction measures. First are near earth orbit (NEO) object redirection measures that offer good and relatively inexpensive reduction of risks of asteroids that are identified and thought to impact years or decades away (Matheny 2007). The NEO redirection measures would reduce the probability of impact of a large asteroid. However, we assume that they would not reduce the probability of impact of Damocloids (at least, not without additional investments to identify Damocloids, which is beyond the scope of this illustrative example). The second type of risk reduction measure is food stockpiling to provide significant food reserves for a large number of people in case of a period of reduced food production (Rampino 2008). The impact effects of large asteroids and comet could be broadly similar to nuclear winter and super-volcanism in their negative impacts on global food production. Food stockpiles may help humanity to survive either event. Rampino (2008) mentions one potential super-volcanism survival strategy would be to stockpile enough food (e.g. grain) to last several years until agricultural productivity goes back up. Rampino notes that current inventories are only equivalent to about two months' consumption. While difficult to accomplish in many parts of the world, it still might be relatively feasible without advanced technology, should be relatively uncontroversial (especially if production is handled in a way that does not drive up global food prices very much), and could have some value across a number of GCR hazards including war, quarantine after pandemic, etc. In addition, unlike some other GCR mitigation measures, stockpiled food should retain near its purchase value in normal usage even in time periods where no GCR scenarios arise (i.e. if no emergencies arise before the stored food expires, the food can be eaten when rotated out of the stockpile and replaced with new reserves).

We implement calculations for the illustrative example in a computational model using the software package *Analytica* by Lumina Decision Systems. The computational model

incorporates all the defined equations and parameters. To estimate probability distributions of outputs, the model performs Latin hypercube sampling, with a model sample size of 10,000 iterations. The model varies continuous-valued inputs according to the previously given probability distributions, and the model produces probabilistic values of its outputs.

For more on relevant distributions, e.g., Uniform() and Triangular() distributions, see Morgan and Henrion (1990) or the Analytica User Guide (Chrisman, Henrion et al. 2007).

4.1.1 Assumptions for Baseline P(Global Catastrophe)

Our estimate for the probabilities of bright object and dark object impact risks is based partly on first estimating the total bright object asteroid impact risk, and then estimating how large the dark object comet impact risk is in comparison specifically to bright objects. We estimate the total probability of impacts of asteroids of size at least 1 km as corresponding to a frequency of one in 3×10^5 years. That is based on Figure 2.4 in NRC (2010) or the equivalent at NRC (2010), which indicate a 1 km object impacts approximately once every 4×10^5 years.

We assume that only 15% of the total population of bright NEOs remain undiscovered (NRC 2010). For bright NEOs that have already been discovered, we also assume negligible impact risk: "none of those detected objects has a significant chance of impacting Earth in the next century" (NRC 2010). The simple way we reflect that in the model is to say the impact risk from visible/bright NEOs is $0.15 \times (1/(3 \times 10^5))$.

We assume that impact frequency of Damocloids have a probability distribution of Uniform$(0,4) \times (1/(3 \times 10^5))$, based on statement of Napier (2008) that the hazard from Damocloids of 1 km diameter "is unknown; it could be negligible, or could more than double the risk assessments based on the objects we see". Some corroboration is provided by the statement of Napier (2008) that at the time of his writing, for 1 km objects, there is an "expected impact frequency of about one such body every 500,000 years". Once every 500,000 years is about the same as the once every 3×10^5 years we assume for over-1km visible objects, but for better consistency and comparability with bright NEOs, we use 3×10^5 instead of once every 500,000. Napier (2008) also observes: "Estimates based on the mean impact cratering rate indicate that, on the long-term, a 1 km impactor might be expected every half a million years or so. Again, modelling uncertainties to do with both excavation mechanics and the erratic replenishment of the near-Earth object (NEO) population yield an overall uncertainty factor of a few. A rate of one such impact every 100,000 years cannot be excluded by the cratering evidence."

All of the above numbers also have additional uncertainty factors (coefficients) of Triangular(0.5, 1, 2), which is loosely based on the statement in NRC (2010) that the uncertainties in intervals between impacts are "on the order of a factor of two". We assume that these uncertainties in the visible/bright NEO frequencies are uncorrelated with the uncertainties in the Damocloid frequencies.

Some corroboration of the relative risks of bright vs. dark objects, and associated uncertainties, is provided by NRC (2010): "With the completion of the Spaceguard Survey (that is, the detection of 90 percent of NEOs greater than 1 kilometer in diameter), long-period comets will no longer be a negligible fraction of the remaining statistical risk, and with the completion of the George E. Brown, Jr. Near-Earth Object Survey (for the detection of 90 percent of NEOs greater than 140 meters in diameter), long-period comets may dominate the remaining unknown impact threat."

Finally, for an extremely simple estimate of the annual probability of nuclear war, based loosely on estimates given in the literature (Hellman 2008; Barrett, Baum et al. 2013; Lundgren 2013) we simply use Triangular(0, 0.0001, 0.001). It seems likely that the annual probabilities of global catastrophe events are orders of magnitude higher for large-scale nuclear war than for large NEO/comet impacts.

In our calculations, we use a simplifying approximation of annual probability as being equivalent to annual frequency (e.g. a frequency of 1 event in 500,000 years implies annual probability of 1/500000).

Table 1 contains summaries of the assumed annual probabilities of global catastrophe level events from each considered GCR type. The expressions reflect the substantial uncertainties.

Table 1: Expressions for Assumed Baseline Annual Probabilities of Global Catastrophe

GCR Types	Baseline P(Global Catastrophe)
Visible Near Earth Objects	Triangular(0.5, 1, 2) * 0.15 * (1/(3*10^5))
Long-Period Comets (Damocloids)	Triangular(0.5, 1, 2) * Uniform(0,4) * (1/(3*10^5))
Nuclear War	Triangular(0, 0.0001, 0.001)

4.1.2 Assumptions for Reduction in P(Global Catastrophe) if Implementing Each GCR Reduction Measure

Although the lower bound of effectiveness of NEO detection and redirection might seem to be quite low, it is based partly on the idea that NEOs that haven't already been discovered might be significantly more difficult to detect than the ones that have already been detected. This may be suggested by Napier (2008, p. 226) regarding the success of NEO detection efforts to date: "There is a caveat: extremely dark objects would go undiscovered and not be entered in the inventory of global hazards."

For food stockpiling, the assumed probability distribution for the reduction in probability of global catastrophe level NEO/comet impacts is Uniform(0.1,0.9). It assumes that the

stockpile would be comprised of extremely inexpensive sources of calories and nutrients (see below on cost assumptions), for which there would be large uncertainties about risk reduction performance.

Table 2 contains summaries of the assumed effects and costs of GCR reduction measures. (A status-quo option, which adds no cost and does not reduce GCR, is omitted from the table but is an additional option in the model.) The effects of the measures are given in terms of their assumed reduction in probability of global catastrophe from each GCR type. The costs of the measures are given in terms of their present value of their costs in 2012 dollars.

Table 2: Assumed Effects and Costs of Risk Reduction Measures

GCR Reduction Measures	Reduction in P(Global Catastrophe) from Each GCR Type			Costs ($ Billion)
	Visible Near Earth Objects	**Long-Period Comets (Damocloids)**	**Nuclear War**	
NEO tracking and redirection measures	Uniform(0.1,0.9)	0	0	7.5
Food stockpiling for all of humanity	Uniform(0.1,0.9)	Uniform(0.1,0.9)	Uniform(0.1,0.9)	1800

### 4.1.3 Assumptions for Costs of GCR Reduction Measures

The cost estimate for food stockpiling assumes a world population of 7 billion, 1 year stockpile, and per person-year stockpile cost based on the food expenditures of the world's poorest people, which is approximately $0.70 per day (GiveWell 2013).

The cost for tracking and redirection capability assumes 30 years of costs, with $250 million annual costs (NRC 2010).

4.2 Example Results

In this section, we give results from the computational model for the illustrative example, using the previously stated assumptions. Figure 5 gives the probability density function (PDF) of the base-case annual probability of global catastrophe from both visible and dark NEO impacts. (On the horizontal axis, u is "mu" or micro, i.e. 10^{-6}.) Contemplating probabilities of probabilities can be confusing, but it is easy to see in PDF figures where there are broad spreads of probability (corresponding to great uncertainties) or narrow spreads (for less uncertainties). The figure shows that there are substantial uncertainties about dark-object Damocloid risks and even greater uncertainties about nuclear war risks; both of the latter could be much greater than visible-NEO impact risks.
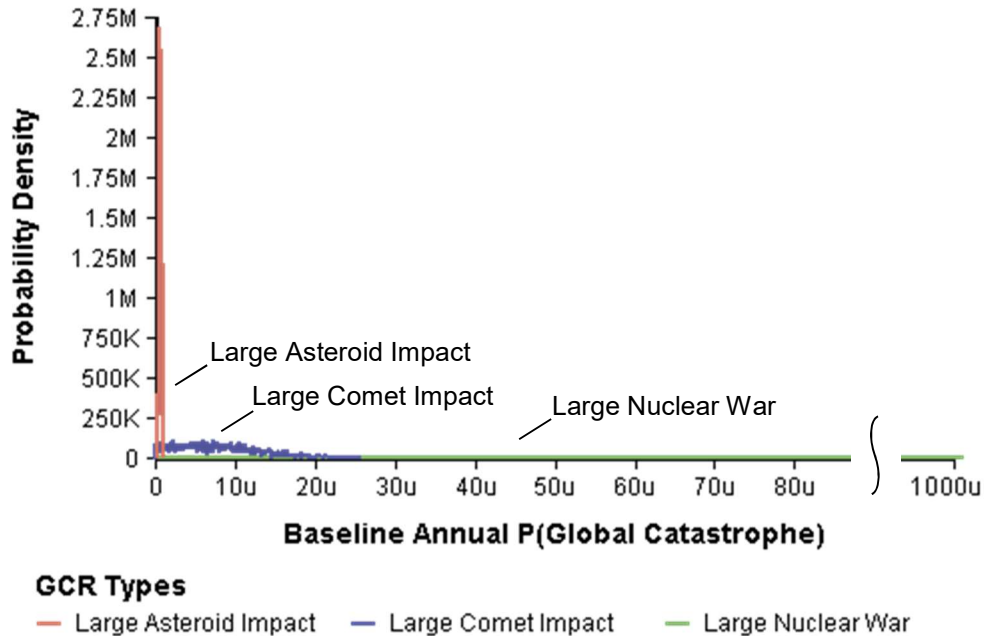
**Figure 5: PDF of Baseline Annual Probabilities of Global Catastrophe**

Table 3 gives the mean cost effectiveness of GCR reduction measures (without research to reduce uncertainties) in terms of how much the measure reduces the average total probability of global catastrophe per dollar spent on the measure. (Recall that in these terms, a high number for cost effectiveness is desirable, because it indicates a large reduction in global catastrophe probability for the dollars spent. The calculations incorporate the global catastrophe probability distributions shown in Figure 5, which showed that nuclear war risks could be much greater than NEO impact risks.)

Table 3: Mean Cost Effectiveness of GCR Reduction Measures (Reduction in Total Global Catastrophe Probability per $)

Include Nuclear War In Scope?	NEO Tracking and Redirection Measures	Food Stockpiling for All of Humanity	Both food stockpiling and NEO tracking/redirection
Yes	4×10^{-17}	1×10^{-16}	1×10^{-16}
No	4×10^{-17}	2×10^{-18}	2×10^{-18}

Table 3's mean cost effectiveness comparison would seem to suggest spending on food stockpiling if nuclear war risk is included in the scope of analysis, but instead would suggest spending on NEO tracking if nuclear war risk is *not* included in the scope of analysis. Moreover, as previously mentioned, there are substantial uncertainties about the risks and cost effectiveness, and food stockpiles might actually be more cost effective than NEO tracking even if nuclear war risk is not included in the scope of analysis. Figure 5a and 5b gives the PDF of the cost effectiveness for each GCR reduction measure if nuclear war risk is / is not included in analysis, respectively. (The status-quo option has cost effectiveness of zero because it does not change

17

GCR probability.) The figure indicates the overlapping ranges of the probability distributions of cost effectiveness of food stockpiling and NEO redirection measures. According to the assumptions used in the Monte Carlo model, if nuclear war risk is included in the scope of analysis, there is a 0.8 probability that food stockpiling will be the most cost effective measure, and there is a 0.2 probability NEO tracking and redirection will be most cost effective; conversely, if nuclear war risk is *not* included in the scope of analysis, there is a 0.999 probability that NEO tracking and redirection will be the most cost effective measure, and there is a 0.001 probability that global food stockpiling will be most cost effective.
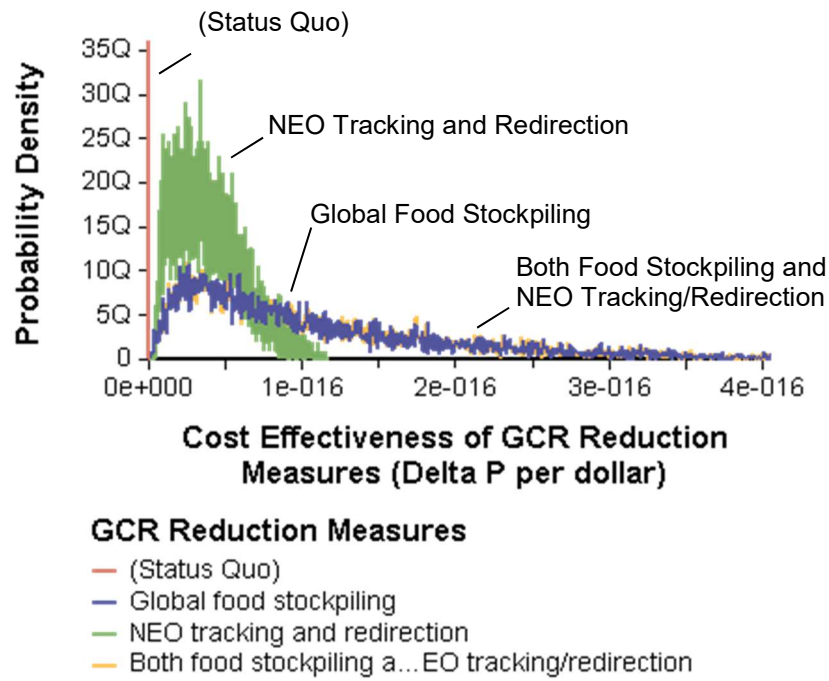
Figure 5a: PDF of Cost Effectiveness of GCR Reduction Measures if Nuclear War Risk is Included in Scope
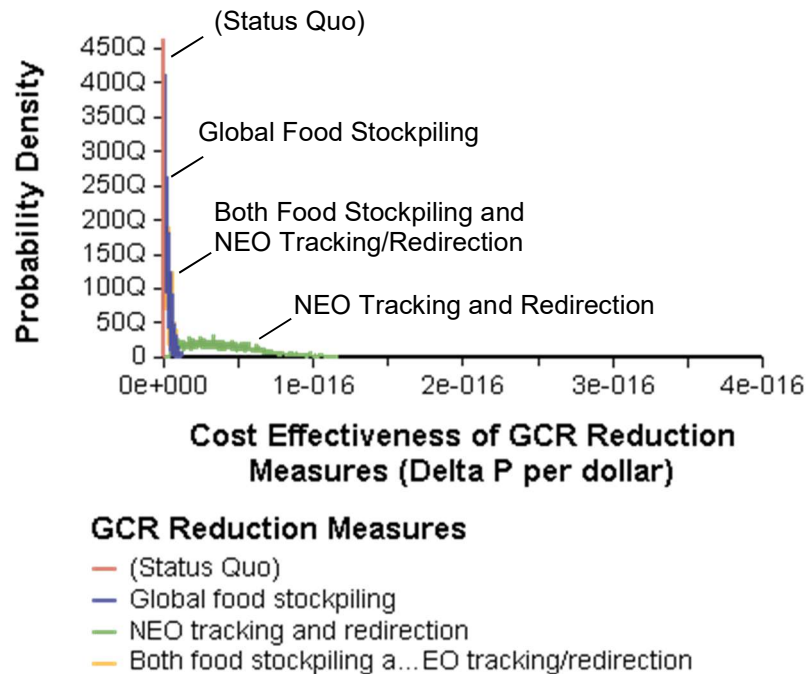
Figure 5b: PDF of Cost Effectiveness of GCR Reduction Measures if Nuclear War Risk is *Not* Included in Scope

Further research could reduce uncertainties to better determine which risk reduction measure would really be more cost effective. According to the Monte Carlo model's assumptions and use of Equation 5, the cost effectiveness-based expected value of perfect information (*CEEVPI*) in the illustrative examples in this paper is $2 billion if nuclear war is included in the scope of analysis, and $400 million if nuclear war is *not* included in the scope of analysis. In this illustrative example, research on the risks and risk reduction effectiveness would have a substantial expected value, largely because of the huge uncertainties about the baseline risks and about the effectiveness of risk reduction measures. The example also supports the argument that we can learn something valuable by doing the analysis for more than one type of GCR at a time.

5. Conclusion

In this paper, we argue that value of information based on cost-effectiveness is a useful tool for analysis of GCR to inform risk-reduction decisions, and show how to apply it to GCRs and risk-reduction interventions in a comprehensive, integrated fashion. We discuss key challenges in real-world implementation of this paper's framework, and argue that these challenges can be addressed. We then illustrate these concepts with simple example models of impact risks from both visible and "dark" near earth objects as well as nuclear war effects, and consideration of related risk reduction measures; the illustrative example shows that such calculations can have considerable value, and also supports looking at more than one GCR at a time.

19

Unlike most value of information approaches, our approach for calculating value of information is based on risk reduction cost effectiveness, to avoid implicitly equating lives and dollars e.g. using a value of statistical life (VSL), which may be inappropriate given the scale of GCRs. Our equation for value of information may be useful in other domains where VSLs would not be appropriate.

Our suggested approach could be used generally to work toward a comprehensive rigorous assessment of GCRs and risk-reduction options. A useful step could be to expand and update this paper's illustrative model (e.g. to reflect more recent NEO research[v] and other NEO risk management options[vi]). However, it would be more valuable to work towards a broader agenda for integrated assessment to inform GCR reduction decisions. Ideally, the scope of such assessment would address all important GCRs over key time periods (e.g. the next century) and also key risk-reduction options of relevant stakeholders (including, but not limited to, public policy options of governments). This paper's framework could help guide steps in such assessment by prioritizing pieces of research in terms of value of information for reducing the total probability of GCRs. While real-world GCR research would not result in perfect information, even imperfect information could have significant value in informing GCR reduction resource allocation decisions.

Our approach could have great value in comprehensively, rigorously assessing GCR and risk-reduction options. Prior GCR research is of only limited value to informing GCR reduction decisions. Much of the work to date has focused on specific GCRs, leaving great uncertainty about which GCRs are most important to focus on. Notable exceptions include research finding that GCRs from cosmic events are small relative to GCRs from human actions (2005); an informal survey of GCR researchers providing estimates of the probabilities of human extinction from a small number of GCR types (Sandberg and Bostrom 2008); analyses of interacting sequences of GCRs (Tonn and MacGregor 2009; Baum, Maher Jr. et al. 2013); and several largely-qualitative surveys (Bostrom 2002; Rees 2003; Posner 2004; Smil 2008; Cotton-Barratt, Farquhar et al. 2016). These studies are insightful but do not provide rigorous quantitative recommendations for risk-reduction resource allocations. We are aware of only one study, that of Leggett (2006) which attempts to quantitatively evaluate GCR reduction measures across a broad space of GCR, but that study had shortcomings such as not considering all GCR categories nor all potentially valuable GCR reduction measures. The modest literature available does not come close to resolving the large uncertainties surrounding both the GCRs themselves and the effectiveness of possible risk reducing interventions. Our work suggests that comprehensive, integrated assessment of GCRs could be quite valuable for informing GCR reduction decisions, and tools can be developed for making comprehensive, integrated assessments for informing GCR risk reduction decisions.

Appendix I. Derivation of Cost Effectiveness-based Formula for Expected Value of Information

In this Appendix, we provide the detail on our derivation of the *CEEVPI* formula.

Our following calculations are aided by two simplifying assumptions, as previously discussed: a binary threshold for global catastrophes, and independence of different GCRs (except to the extent that GCR event interactions and dependencies are accounted for in the fault trees or other model components). Then let X_j equal the value loss due to global catastrophe j. Due to the binary threshold assumption, X_j is the same for all j. Let $R(t)$ be the risk of an event occurring during time period t, with R = probability × magnitude. Then, given the independence of different GCRs, the total risk for y GCRs is:

$$R_{tot}(t) = p_{tot}(t)X = \left(1 - \prod_j^y (1 - p_j(t))\right) X \qquad (Eq.\,3)$$

We further assume that all GCRs have sufficiently low probabilities-per-time-period $p_j(t)$ that the total probability of global catastrophe in that time period can be approximated as the sum of the independent probabilities, such that:

$$R_{tot}(t) \approx \left(\sum_j^y p_j(t)\right) X \qquad (Eq.\,4)$$

In this paper, we evaluate possible GCR-reducing interventions in terms of their cost-effectiveness, i.e. their reduction in GCR per unit cost. We favor cost-effectiveness for two reasons. First, cost-benefit analysis is hampered by the challenge of quantifying the value loss due to global catastrophe, X. The benefit of interventions is the reduction in risk, which also depends on X. While X is generally believed to be very large, quantitative estimates span a huge range, as previously mentioned. In contrast, cost-effectiveness analysis does not depend on X. Let c_i and CE_i be the cost and cost-effectiveness of intervention i. Then:

$$CE_i = \frac{(R_{0,tot}(t) - R_{i,tot}(t))}{c_i} = \frac{(p_{0,tot}(t) - p_{i,tot}(t))}{c_i} X \qquad (Eq.\,5)$$

Since X is equal for all global catastrophes, comparisons of the cost-effectiveness of different interventions are the same regardless of the value of X.

We assume that there is a decision to be made about allocation of resources to some combination of direct risk reduction and research, and that the main decision rule is to choose whatever combination of options has best overall expected GCR reduction cost effectiveness among options considered in the analysis. Then the decision maker should buy as much risk reduction (and risk research enabling better risk reduction decisions) as they can at whatever total cost, as long as that results in the greatest cost effectiveness. (We assume that budgets are not an issue in the context of the risk reduction and research options under consideration, and we do not explicitly account for potential budget constraints in the following. This implicitly

assumes that sufficient total resources are either being provided by a single entity or are coordinated in some fashion.)

If the information's expected effect and cost are such that even with the research cost included, it would achieve a better cost effectiveness than whatever would have been the optimal investment before the research based on expected values, then the research information is worth its cost. This is true for investments in information up to the point where the expected cost effectiveness with research is the same as without research. We use that relation to make an equation to solve for the expected value of the information in terms of everything else.

For the purposes of this analysis we ignore actual costs of research and focus on the amount of resources the decision maker ought to be willing to pay for the value added by the research in the context of the decision the research could inform. In other words, we focus on finding the benefits of research, which would result if research yields information that reduces or eliminates uncertainties in a decision model (i.e. turns model variable probability distributions into either tighter, more-accurate distributions, or into maximally accurate point values).

We define several terms:

There exists a set of n available risk reduction options numbered 0, 1, … i… n. Option number 0 is the status quo case, where no new (or non-"business-as-usual") risk reduction option is implemented.

c_i is the cost of implementing risk reduction option i. (It adds no cost to do nothing new, so $c_0 = 0$.)

p_i is the annualized total probability of global catastrophe if implementing option i.

Each c_i and p_i is assumed to be a random variable with some probability distribution reflecting uncertainty about the variable's true value or value in a particular instance. p and c terms are assumed to be uncorrelated, i.e., covariance(p,c) values are assumed to be zero. Computationally, probability-distribution uncertainty is represented using Monte Carlo simulation, where in Monte Carlo simulation iteration m, there are sampled values c_{im} and p_{im}. The expected value of any variable x is E[x], which is found computationally by finding the mean value of variable x across the set of Monte Carlo iterations.

c_r is the cost of conducting research that reduces uncertainties by some amount.

In any particular Monte Carlo iteration m, the cost effectiveness of risk reduction measure i is the ratio of risk reduction to cost, or CE_{im}, where:

$$CE_{im} = \frac{(p_{0m} - p_{im})}{c_{im}} \qquad \text{(Eq. 6)}$$

Then the risk reduction option s with the "best" or highest risk reduction cost effectiveness in Monte Carlo iteration m is the option where:

$$s = \arg\max_{i} \left[\frac{(p_{0m} - p_{im})}{c_{im}} \right] \qquad \text{(Eq. 7)}$$

In other words,

$$\frac{(p_{0m} - p_{sm})}{c_{sm}} = \max_{i} \left[\frac{(p_{0m} - p_{im})}{c_{im}} \right] \qquad \text{(Eq. 8)}$$

We denote cases whether research is conducted to reduce uncertainty on a particular factor using superscript b for "before" research or without information from research, and superscript a for "after" research or with information from research. (Thus, before research is conducted on p_i, it is p_i^b, and after research is conducted, it is it is p_i^a.) Again, we ignore actual costs of research and focus on the amount of resources the decision maker ought to be willing to pay for the total value added by the research. We use the term w to denote the amount of resources the decision maker ought to be willing to pay for the total added value of conducting research. (It adds no value to do no research.) For the purposes of this derivation, we do not provide more detailed break-downs of the amount of resources the decision maker ought to be willing to pay for the value added by performing specific pieces of research that comprise total value added by research v, which actually could consist of separate pieces of research on different uncertain factors. (The amount of resources the decision maker ought to be willing to pay for the value added by each piece of research could be assessed using an extension of the derivation provided here.)

Note that the best option after research, s^a (which has cost effectiveness in Monte Carlo iteration m of $\frac{(p_{0m}^a - p_{sm}^a)}{c_{sm}^a}$) is not necessarily the same as the best option before research, s^b (which has cost effectiveness in Monte Carlo iteration m of $\frac{(p_{0m}^b - p_{sm}^b)}{c_{sm}^b}$). Research that reduces but does not eliminate uncertainty about a factor yields imperfect information. In a case where research produces perfect information about a factor, all uncertainty is eliminated about the factor after research. In terms of Monte Carlo iterations, after perfect information, one of the Monte Carlo iterations will have randomly-sampled factor values whose are closest to the actual real-world factor values.

As long as doing more research adds more value, and if we ignore the actual costs of performing the research, we assume that resources for research ought to be invested in up to the point where research would be so expensive that a funder would obtain no further benefit from investing in additional research (because up to that point, they would get a better overall cost effectiveness by investing in additional research). At that point, the expected cost effectiveness of the best risk-reduction option before research is equal to the expected cost effectiveness of the best risk-reduction option after research, including the amount of resources the decision maker ought to be willing to pay for the total value added by research:

$$\max_i E\left[\frac{(p_0^b - p_i^b)}{c_i^b}\right] = \max_i E\left[\frac{(p_0^a - p_i^a)}{(c_i^a + v)}\right] \tag{Eq. 9}$$

Using the best-option notation,

$$E\left[\frac{(p_0^b - p_s^b)}{c_s^b}\right] = E\left[\frac{(p_0^a - p_s^a)}{(c_s^a + w)}\right] \tag{Eq. 10}$$

The $E[\cdot]$ terms can be distributed and re-gathered because all the relevant calculations (i.e., both the expected-value calculations and the cost-effectiveness calculations) involve linear operations and because p and c variables are assumed to be uncorrelated. (To be more specific, manipulation of the numerator and denominator are allowed because they are linear operations, and multiplicative operations are allowed because the covariance is assumed to be 0.) Distributing and rearranging the terms to solve for the expected value of the amount of resources the decision maker ought to be willing to pay for the total value added by research, E[w],

$$\frac{(E[p_0^b] - E[p_s^b])}{E[c_s^b]} = \frac{(E[p_0^a] - E[p_s^a])}{(E[c_s^a] + E[w])} \tag{Eq. 11}$$

$$E[w] = E\left[\frac{(c_s^b)(p_0^a - p_s^a)}{(p_0^b - p_s^b)} - c_s^a\right] \tag{Eq. 12}$$

Thus, the above expression for E[w] gives the value of research as the cost effectiveness-based expected value of information, *CEEVI*:

$$CEEVI = E\left[\frac{(c_s^b)(p_0^a - p_s^a)}{(p_0^b - p_s^b)} - c_s^a\right] \tag{Eq. 13}$$

This formula for *CEEVI* actually applies to both perfect information and imperfect information cases. However, our focus in this derivation is on the limiting case where the research yields perfect information, which provides the upper limit to the value of research, i.e. the cost effectiveness-based expected value of perfect information *CEEVPI*. It turns out that when used in *Analytica* software by Lumina Decision Systems, the above *CEEVI* formula can be used in a straightforward fashion to set up the Monte Carlo simulation computations for *CEEVPI* (by directly using each factor's Monte Carlo sampling values in each Monte Carlo iteration) and that is what we use in the illustrative *Analytica* model accompanying this paper. (Computation of the cost-effectiveness based expected value of imperfect information *CEEVII* would require an extra step to simulate after-research imperfect-information probability distributions for each factor, instead of after-research perfect-information point values.)

References

Aldy, J. E., S. Barrett, R. N. Stavins. 2003. Thirteen plus one: A comparison of global climate policy architectures. *Climate Policy* **3**(4) 373-397.

Barrett, A. M., S. D. Baum. 2017. A model of pathways to artificial superintelligence catastrophe for risk and decision analysis. *Journal of Experimental & Theoretical Artificial Intelligence* **29**(2) 397-414.

Barrett, A. M., S. D. Baum, K. R. Hostetler. 2013. Analyzing and reducing the risks of inadvertent nuclear war between the United States and Russia. *Science & Global Security* **21**(2) 106-133.

Baum, S. D. 2010. Is humanity doomed? Insights from astrobiology. *Sustainability* **2**(2) 591-603.

Baum, S. D. 2012. Value typology in cost-benefit analysis. *Environmental Values* **21**(4) 499-524.

Baum, S. D., T. M. Maher Jr., J. Haqq-Misra. 2013. Double catastrophe: Intermittent stratospheric geoengineering induced by societal collapse. *Environment, Systems and Decisions* **33**(1) 168-180.

Beckstead, N. 2013. On the overwhelming importance of shaping the far future. Department of Philosophy. New Brunswick, NJ, Rutgers University.

Bhattacharjya, D., J. Eidsvik, T. Mukerji. 2013. The value of information in portfolio problems with dependent projects. *Decision Analysis* **10**(4) 341-351.

Bostrom, N. 2002. Existential risks: Analyzing human extinction scenarios and related hazards. *Journal of Evolution and Technology* **9**(1).

Bostrom, N., M. M. Cirkovic, eds. 2008. *Global catastrophic risks*, Oxford University Press.

Chalmers, D. 2010. The singularity: A philosophical analysis. *Journal of Consciousness Studies* **17** 7-65.

Chrisman, L., M. Henrion, R. Morgan, B. Arnold, F. Brunton, A. Esztergar, J. Harlan, L. Korsan, R. Mulford, R. Sonnenblick, B. Sterlin, E. Wainwright. 2007. Analytica user guide. Los Gatos, California, Lumina Decision Systems.

Cirincione, J. 2008. The continuing threat of nuclear war. N. Bostrom and M. M. Cirkovic, eds. *Global catastrophic risks*. Oxford University Press, Oxford, 381-401.

Ćirković, M. M., A. Sandberg, N. Bostrom. 2010. Anthropic shadow: Observation selection effects and human extinction risks. *Risk Analysis* **30**(10) 1495-1506.

Clemen, R. T., T. Reilly. 2001. *Making hard decisions*. Duxbury, Pacific Grove, CA.

Cotton-Barratt, O., S. Farquhar, J. Halstead, S. Schubert, A. Snyder-Beattie. 2016. Global catastrophic risks 2016. *Global Challenges Foundation and Global Priorities Project.*

http://globalprioritiesproject.org/wp-content/uploads/2016/04/Global-Catastrophic-Risk-Annual-Report-2016-FINAL.pdf.

Denkenberger, D., J. M. Pearce. 2015. *Feeding everyone no matter what: Managing food security after global catastrophe*. Academic Press, Waltham, Massachusetts.

Dietz, T., G. T. Gardner, J. Gilligan, P. C. Stern, M. P. Vandenbergh. 2009. Household actions can provide a behavioral wedge to rapidly reduce us carbon emissions. *Proceedings of the National Academy of Sciences of the United States of America* **106**(44) 18452–18456.

Garrick, B. J. 2008. *Quantifying and controlling catastrophic risks*. Academic Press, Burlington, Massachusetts.

GiveWell. 2013. Standard of living in the developing world. http://www.givewell.org/international/technical/additional/Standard-of-Living.

Hellman, M. 2008. Risk analysis of nuclear deterrence. *The Bent of Tau Beta Pi*(Spring) 14-22.

Hempsell, C. M. 2004. The investigation of natural global catastrophes. *Journal of the British Interplanetary Society* **57**(2-13).

Joy, B. 2000. Why the future doesn't need us. http://www.wired.com/wired/archive/8.04/joy.html.

Keisler, J. 2004. Value of information in portfolio decision analysis. *Decision Analysis* **1**(3) 177-189.

Lee, R. C., T. D. Jones, C. R. Chapman. 2014. A decision analysis approach for risk management of near-earth objects. *Acta Astronautica* **103** 362-369.

Leggett, M. 2006. An indicative costed plan for the mitigation of global risks. *Futures* **38** 778-809.

Lundgren, C. 2013. What are the odds? Assessing the probability of a nuclear war. *Nonproliferation Review* **20**(2) 361-374.

Matheny, J. G. 2007. Reducing the risk of human extinction. *Risk Analysis* **27**(5) 1335-1344.

Morgan, M. G., M. Henrion. 1990. *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press, Cambridge.

Napier, W. 2008. Hazards from comets and asteroids. N. Bostrom and M. M. Cirkovic, eds. *Global catastrophic risks*. Oxford University Press, Oxford, 222-237.

Ng, Y.-K. 1991. Should we be very cautious or extremely cautious on measures that may involve our destruction? *Social Choice and Welfare* **8** 79-88.

Nouri, A., C. F. Chyba. 2008. Biotechnology and biosecurity. N. Bostrom and M. M. Cirkovic, eds. *Global catastrophic risks*. Oxford University Press, Oxford, 450-480.

NRC. 2010. Defending planet earth: Near-earth object surveys and hazard mitigation strategies: Final report. Committee to Review Near-Earth Object Surveys and Hazard Mitigation Strategies; US National Research Council, Washington, DC.

Osterholm, M. T. 2005. Preparing for the next pandemic. *Foreign Affairs* **84**(4) 24-37.

Phoenix, C., M. Treder. 2008. Nanotechnology as global catastrophic risk. N. Bostrom and M. M. Cirkovic, eds. *Global catastrophic risks*. Oxford University Press, Oxford, 481-503.

Posner, R. A. 2004. *Catastrophe: Risk and response*. Oxford University Press, New York, NY.

Rampino, M. R. 2008. Super-volcanism and other geophysical processes of catastrophic impact. N. Bostrom and M. M. Cirkovic, eds. *Global catastrophic risks*. Oxford University Press, Oxford, 205-221.

Rees, M. 2003. *Our final century: Will the human race survive the twenty-first century?* William Heinemann, Oxford.

Reinhardt, J. C., X. Chen, W. Liu, P. Manchev, M. E. Paté-Cornell. 2015. Asteroid risk assessment: A probabilistic approach. *Risk Analysis*.

Robock, A., L. Oman, G. L. Stenchikov. 2007. Nuclear winter revisited with a modern climate model and current nuclear arsenals: Still catastrophic consequences. *Journal of Geophysical Research* **112**(D13107).

Sagan, C. 1983. Nuclear war and climatic catastrophe: Some policy implications. *Foreign Affairs* **62**(2) 257-292.

Sandberg, A., N. Bostrom. 2008. Global catastrophic risks survey. *Future of Humanity Institute, Oxford University.* http://www.fhi.ox.ac.uk/__data/assets/pdf_file/0020/3854/global-catastrophic-risks-report.pdf.

Smil, V. 2008. *Global catastrophes and trends: The next fifty years*. MIT Press, Cambridge, Massachusetts.

Sotala, K. 2010. From mostly harmless to civilization-threatening: Pathways to dangerous artificial general intelligences. *Eighth European Conference on Philosophy and Computing (ECAP10)*,

Taylor, P. 2008. Catastrophes and insurance. N. Bostrom and M. M. Cirkovic, eds. *Global catastrophic risks*. Oxford University Press, Oxford, 164-183.

Tegmark, M., N. Bostrom. 2005. How unlikely is a doomsday catastrophe? *Nature* **438**(7069).

Tonn, B. E. 2009. Obligations to future generations and acceptable risks of human extinction. *Futures* **41**(7) 427-435.

Tonn, B. E., D. MacGregor. 2009. A singular chain of events. *Futures* **41** 706-714.

Travis, W. R. 2010. Going to extremes: Propositions on the social response to severe climate change. *Climatic Change* **98**(1-2) 1-19.

Turco, R. P., O. B. Toon, T. P. Ackerman, J. B. Pollack, C. Sagan. 1983. Nuclear winter: Global consequences of multiple nuclear explosions. *Science* **222**(4630) 1283-1292.

Weitzman, M. L. 2009. On modeling and interpreting the economics of catastrophic climate change. *Review of Economics and Statistics* **91**(1) 1-19.

Yudkowsky, E. 2008. Artificial intelligence as a positive and negative factor in global risk. N. Bostrom and M. M. Cirkovic, eds. *Global catastrophic risks*. Oxford University Press, Oxford, 308-345.

[i] For an example of a canonical utility function based decision analysis framework for one GCR category, asteroid and comet impact risk, see Lee et al. (2014).

[ii] For an illustrative example of a probability distribution reflecting uncertainty about an annualized global catastrophe probability, see Figure 5 in Section 4.2. For more on such probability distributions, see Chs. 4 and 5 of Morgan and Henrion (1990).

[iii] For some GCR types, it may not be most useful to think in terms of consequence-exceedance thresholds, but in terms of probabilities of various possibilities, such as in future "artificial superintelligent catastrophe" scenarios. However, modeling approaches such as fault trees could be useful for some such scenarios (Barrett and Baum 2017).

[iv] There are also related challenges in selection of metrics, such as for event consequences: whether to focus on estimated fatalities over some specific time scale, or to also consider economic impact, etc. Even choosing cost metrics for use in cost-effectiveness analysis presents challenges; in this paper we assume cost is defined in monetary (dollar) terms but those have limitations (Baum 2012), and scarcities exist for other resources such as labor capacity.

[v] See, e.g., Reinhardt et al. (2015).

[vi] For example, there are a number of options for alternative food sources during a crop-failure crisis (Denkenberger and Pearce 2015). Those potentially could be more cost effective than food stockpiling, but we believe their effectiveness also would have greater uncertainty because of complexity, etc.