The Mathematics Enthusiast

Volume 12
Number 1 *Numbers 1, 2, & 3*

Article 28

6-2015

Adults' Perceptions of Risk in the Big Data Era

Theodosia Prodromou

Follow this and additional works at: https://scholarworks.umt.edu/tme

Part of the Mathematics Commons
Let us know how access to this document benefits you.

Recommended Citation
Prodromou, Theodosia (2015) "Adults' Perceptions of Risk in the Big Data Era," *The Mathematics Enthusiast*: Vol. 12 : No. 1 , Article 28.
Available at: https://scholarworks.umt.edu/tme/vol12/iss1/28

This Article is brought to you for free and open access by ScholarWorks at University of Montana. It has been accepted for inclusion in The Mathematics Enthusiast by an authorized editor of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.

Adults' Perceptions of Risk in the Big Data Era

Theodosia Prodromou
University of New England, Australia

Abstract: The present digital era has seen rapid growth in the availability of big data; we were curious about whether such availability of data changes perceptions and assessments of risk. In this paper, we investigate adults' (35-63 years old) perceptions of risk in the big-data era and how it figures in their everyday life. We developed decision-making scenarios for socio-economic, environmental and health topics that involve modelling with personal value systems alongside Gapminder word map data. Going beyond the idea of risk in statistical theory, we attempt to gain an understanding of the processes by which adults assess risks.

Keywords: risk, assessments of risk, big data, informal risk-based reasoning, gapminder.

Introduction

The concept of risk is all-pervasive presence in our daily lives since it permeates decision-making processes in individuals' daily lives, whether personal, professional, or social. Indeed, risk, and the way it is managed, is a critical aspect of decision making at all levels. For instance, we evaluate profit opportunities in business with respect to the engendered countervailing risks, and we make decisions for ourselves or for our family members by analysing the risks involved in areas such as healthcare, sports and exercise, etc.

It is highly significant to make a "risk analysis" when planning any new project or business project and evaluate solutions of types of problems on a risk-cost or cost-benefit basis, such problems might be global, political, financial, and individual. Risk appears to be an extraordinarily diverse area, being studied in a wide range of disciplines. More specifically, the notion of risk is commonly used nowadays in science, medicine, and technology, having its roots in the area of decision-making (Edwards & Tversky, 1967). "Making important decisions in the face of uncertainty is unsettling and difficult and so is a vital area" (Spiegelhalter, 2012). Making decisions about problems from all the aforementioned range of disciplines, demands basing judgements on data to balance effect and likelihood, along with other values-based considerations, of likely hazards.

The word "risk" is hard to pin down (BSON, n.d.); there is still no broad consensus on the meaning of this term. In references to risk, the necessary meaning of the idea of risk is subject of a variety of epistemological perspectives (Adams, 1995; Stirling, 1999; Weale, 2002). Indeed, various national and international standards and guidelines mention risk, and there are many different definitions of the underlying concept. "Even among risk practitioners in the various professional bodies there is an ongoing debate about the subject matter at the heart of their discipline. And of course there is huge variation in the general literature, reflecting the lack of official agreement on the basic definition of risk" (Hilson & Murray-Webster, 2004, p. 2). Nevertheless, all definitions agree that risk is related to uncertainty, and it has consequences (Hilson et al., 2004). Hilson et al., (2004) defines risk in terms of "an uncertainty that could have a positive or negative effect on one or more objectives".

In this paper, I use elements of Hilson's et al. (2004) definitions of risk. I am particularly interested in defining risk as a consequence of uncertainty (Hilson et al., 2004) and relate risk to ideas of probability and likelihood (Pratt, Ainley, Kent, Yogui, & Kapadia, 2011).

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 12, no. 1,2&3**, pp. 364-377
2015© The Author(s) & Dept. of Mathematical Sciences-The University of Montana

In the following section we will discuss the concept of risk in terms of uncertainty and differentiate between risk and uncertainty, building the definition of risk as a consequence of uncertainty during decision-making.

Risk as a Consequence of Uncertainty

We all use personal intuitions to evaluate and interpret "risk" when faced with a new and uncertain event. While there is concern because risk does not appear to underlie people's intuitive and every day rationalizations (Lupton, 1999), these concerns strengthen when specific events such as property damage or loss after a fire event are brought to the public attention.

The notion of "risk" is inextricably linked to the notion of uncertainty. Risk depends on uncertainty in a potential fact, event, outcome, or scenario, etc. For example, uncertainty causes mortgage issuers to demand property purchase insurance. The person or corporation occupying the mortgage-funded property must purchase insurance on real estate if we intend to lend them money. It would be a risk to lend them money because we did not know if apprehension or dread was about to occur. The future is always unknown, and we are faced with numerous uncertain outcomes (from the mundane—e.g., "we don't know when it will start raining"—to the extreme—e.g., "we don't know when or where the hurricane/typhoon will make landfall").

People's perceptions of risk arise from perceptions of the quantification of uncertainty. Nevertheless, it is important to differentiate between mathematical uncertainty (i.e., a quantification of probability) and epistemological uncertainty (lack of knowledge about outcomes).

Literature concerned with the relation of risk and uncertainty to ideas of probability and likelihood has attempted various methods to quantify risk and uncertainty. Indeed, both risk and uncertainty are defined by probabilities or based on discrete and continuous probability distributions. Risk isn't the same as the underlying prerequisite of uncertainty.

My particular concern is about the relation of risk to ideas of probability and likelihood, and statistical interpretation of everyday situations that involve risk. Risk may be quantified using diverse ways, or we may refer to risk without any quantification. For example, the risk of slipping and falling when it is raining is higher than that of slipping and falling when it is a sunny day.

In both official government literature and in media reports, risks reported are often expressed as percentages, or as 1 in *N* cases. So, for example, the risk of cancer disease resulting from drinking, or the risk of cybercrime, or the risk of influenza to air travel is mentioned in terms of mathematical probabilities.

However occasionally increased risks reported by media can be presented more dramatically through the quotation of relative risks. Indeed, the media may state when reporting risk, of, say, exposure to cigarette smoke significantly increases the (relative) risk of developing of lung cancer by 30% in smokers. What exactly does that mean? Say the absolute risk of developing a disease is 10 in 100 in non-smokers. The 30% relates to the 10—so the absolute increase in the risk of 30% of 10 is 3 and the absolute risk of smokers developing lung cancer is 13 in 100.

The media are crucial players in the communication about risk and media often misrepresent risk statistics and report risk involved in some medical cases in a misleading manner (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007). While some studies show that "the key components of risk studies, such as probabilities, seem to have merely a minor impact in the media reporting" (Renn, 1991); other studies that look at media reports show quite the opposite. For instance, Freudenburg and colleagues after analysing 128 hazard events, they concluded that, "the amount of coverage is predicted only by the objective characteristics of the hazard events" (Freudenburg, Coijemb, Conzale, & Helgeland, 1996, p. 38).

Ultimately, the factors affecting media coverage of risk involve representations of diverse hazards. The identification of risk for decision making is valid when it is first approached in terms of the impact of hazards on risk. In decision theory the notion of risk can be expressed in the form of the "subjective expected utility" (SEU). The idea of subjective expected utility (SEU) suggests that for every hazardous event there is both a likelihood of the event happening, and a numerical utility that indicates the impact that he hazard would have on the target individual or organisation or a group of people. Decision-making theory defines risk as the product of the likelihood and utility. Hazardous events usually involve a number of different hazards and their summation gives rise to total hazard. The risk is, therefore, calculated as the sum of all the individual risks across all the events.

When the risk is expressed in the form of the SEU, individuals who are required to make a risk decisions in their everyday lives, first have to acknowledge a set of different possible courses of action, analyse those courses of actions into hazard events, and give values to the likelihood of every hazard and utility. They then calculate the risk of each course of action, add these risks and choose the course of action that gives the minimum total risk.

Making decisions in everyday life when expressing risk in the form of the SEU by analysing many micro-problems, gives rise to many obstacles related predominantly to the human nature limitations for employing subjective expected utility and making actual decisions in any literal way (Simon, 1990, p. 13; Simon, 1997). Another essential problem that contributes to the complexity of decision making is uncertainty about the reliability of what we know. We are not sure that we are aware of all the different parts of knowledge required about the set of different events that might give rise to the hazard.

The majority of individuals, when making decisions about their everyday lives, do not minimise total risk as in standard theory. Individuals often make risk-based decisions from a rational perspective in which the data are often personal. Lay-people's decisions about risk are predominantly dependent on personal interpretations of situations and decisions are taken from an intuitive and informal viewpoint. Nonetheless, as Slovic has argued: "lay-people sometimes lack certain information about hazards; however their basic conceptualization of risk is much richer than that of the experts and reflects legitimate concerns that are typically omitted from expert risk assessments" (Slovic, 2000, p. 191). The aim of this project is to understand more clearly the reasoning involved in making such risk-based decisions and how adults (35-63 years old) reason about risk in the big-data era.

The process of informal risk-based reasoning and risk interpretation is likely to call upon a coordination of a variety of different kinds of data involving explicit and implicit, quantitative and qualitative, and frequently value-laden.

Personal heuristics account for inconsistencies between an individual's judgment of a risk event (Brandstätter, Gingerenzer, & Hertwig, 2006) and the objective risk calculated by experts (Crossland, Bennett, Ellis, Gittus, & Godfrey, 1992). For example, availability—people's tendency to draw on knowledge that pops into mind quickly (Tversky & Kahneman, 1973; Kahneman, Slovic, & Tversky, 1982; Folkes, 1988)—and affect—emotional responses to highly unlikely events such as terrorist attacks on the airflight compared to most common events such as car accidents—both of which are shaped by personal experiences, contribute to cognitive heuristics for estimation of the likelihood of events are incorporated into individuals' judgment of risk (Greening, Gollinger, & Pitz, 1996).

In an effort to understand the complexity of informal risk-based reasoning, we acknowledge that real problems are extremely complex in their context-dependence, and they involve different levels of decision–making: individual (e.g., deciding about personal life insurance), social (e.g., deciding about the economic crisis in Europe), and mixed situations where there is an interplay between individual and social levels (e.g., in national child vaccination programs where parental choice is in interplay with official policy).

My interest lies in understanding how people call upon knowledge about the context throughout the process of risk-based reasoning and risk interpretation, and whether and how people coordinate likelihood and impact in their risk-based reasoning. I try to understand the processes used as individuals attempt to assess risk in various situations when presented with big data.

Purpose/Goals of the Study

In the study presented here, the project I set out underpins the investigation of adults' perceptions of risk in the big data era. In particular, going beyond the idea of risk in statistical theory, we attempt to gain an understanding of adults' perceptions and assessments of risk when they deal in our contemporary society with large scale data. The overall research study has involved working with a small group of adults 35-64 years old, recruited in pairs or individually from social media (e.g., Facebook, Twitter). This research study will be concerned with investigating adults' (35-64 years old) perceptions of risk in the big-data era and illustrating how risk is involved in decision-making in their everyday lives. An initial assumption is that decision-making involves the coordination of different kinds of information, based on quantitative models and personal value systems and judgements. To research decision-making when risk is involved, I develop decision-making scenarios (Appendix 1) for socio-economic, environmental, and health topics that involve modelling with personal value systems alongside using data from media or other (big) data sources. These scenarios will be used with adults' (35-64 years old) during an interview process.

In this research study, participants discussed various scenarios with the researcher, so that we can gather data on how people understand everyday situations that involve risk.

This research will also lead educators to a better understanding of adults' perceptions of risk when adults have to:

1. Make decisions about hazardous events; and,
2. measure the likelihood of events.

My research question, therefore, is: What are the personal values and models that affect individuals' thinking about risk and their process of reasoning.

Methodology

The total number of participants that was recruited was 40. The participants were initially contacted via social media and invited to participate in this research study. Selection was based on the potential participants who will first respond to the invitation.

The inclusion criteria required the participants to be aged 35-64, can use fractions and percentages to talk about risk and be aware about the dominant impact of risk in their everyday life.

In this paper we will discuss the results from the face – to– face interviews with four participants, each participant aged (35-44 years old). These four participants were mature students of a rural university who are studying for undergraduate degrees via online education. All participants were willing to collaborate with the researcher. The researcher sat with each participant or group of participants and asked participants questions about different scenarios for socio-economic, environment and health topics. Participants made formal or informal inferences while using information from Media or big data databases (sometimes Gapminder word map data). The discussions of the participants with the researcher were likely to be informal and as frank as possible. The researcher encouraged participants to explore and interrogate their own knowledge of the concept of risk in the big-data era and explain how risk figures in their everyday life. While these participants were keen to engage with the research study, the challenges that arose in interaction with the Gapminder world map raised questions about the software. The researcher intervened to demonstrate relevant aspects of the software, to address any technical issues, and to ask questions for clarification. The extensive discussions with the participants about diverse risk situations were recorded.

Camtasia video screen capture software recorded the reasoning process through the participants' extensive dialogues and manipulation of the Gapminder map. Data collected included audio recordings of participants' voices and video recordings of the screen output on the computer activity. The participants were asked to use the mouse systematically to point to objects on the screen when they reasoned about the various quantities presented in the Gapminder World Map graph. Having students to point on the screen helped to supplement the recording of their voices and explained their actions and interactions with the quantities illustrated by the Gapminder visualization tool that may otherwise have been subject to many interpretations. The recorded data were supplemented with notes and memos made by the researcher. The sessions lasted approximately 90 minutes.

The recordings were transcribed and analysed qualitatively. This process started with plain accounts for each participant and comparisons of the reasoning of each participant.

For the rest of this article we present the findings through the cases of two pairs: Tim (a student in environmental sustainability) and Michael (in Psychology); and George (in Sociology) and Steph (in primary Education). After presenting the chronological account of their engagement with the three scenarios, we discuss some key foci themes that emerged through the comparative approaches (Corbin & Strauss, 2007).

An Account of two pairs' activity

Tim (T) and Michael (M) began by trying to make sense of the information provided for the risk and heart attack scenario, interrogating the information and trying to assess the impact of taking medication to reduce the risk of a heart attack or stroke.

> *R*: Okay, and the second question 'Lupitor is a medication that could reduce of, by 36%. Will I buy Lupitor? Will I buy this medication, do you think it will help me or not?
>
> *M*: I think it will help yes.
>
> *R*: And how much will reduce the risk, of a heart attack?
>
> *T*: Umm, if we are just thinking just about the heart attack, I think it will reduce it by 36%
>
> *R*: Yeah, but have in mind that I had only 12% chance, risk of a heart attack so how much my chances will be?
>
> *T*: Well it will go …it will go from 12% to, 8 to 8and a half %
>
> *R*: How do you make this calculation?
>
> *T*: Well I just take 1, 36% is just over one third which is 33%
>
> *M*: So 1/3 of 12%, 1/3 of twelve is 4, so I take 4 off the twelve. Actually it will be more just under twelve because I'm increa…, I'm reducing it by more than 1/3. So it will be 7.8, but the 12% will go down to about 7.8 % or something around there.
>
> *R*: Okay, understand. So is it a good or does it reduce the risk by a lot or?
>
> *T*: I don't think 12% is a huge risk in any case… And I'll, maybe I'll think about not taking because I do not like taking drugs that much, so I would rather, try to do without the Lupitor.

Assessing the likelihood seemed to involve attempting to appreciate the relative size of probabilities. At this point, Tim and Michael felt that although the Lupitor could reduce this risk by 36%, the 12% risk of heart attack or a stroke would be reduced about 7.8%, and such a risk is not especially dangerous, so Tim came to a decision that he will not buy Lupitor.

In attempting to make sense of the information available as provided by the data taken from UK national mortality statistics, Tim began with interpreting the data that show the annual risk of death by age and sex in the UK.

> *T*: Okay, well, the first line “all ages”, I think is everyone in the UK. And it says that in any particular year, one in every 136 men, and one in 193 women will die. And then it just breaks it down into different ages groups so if you’re under 1 years old, you have, it’s actually higher than I thought. 1 in 177 and 1 in 227, men and women respectively will die. The lowest age group that will, where people will die if 5-14. So 1 in 8333 men and 1 in 10417 women, well, at that age I should, 5-14 it’s boys and girls will die. When you get to the 85 and over, it’s because it goes slowly umm, the chances increase as you get older. So once you get to 85 and over, 1 in 6 men and 1 in 7 women will die. Umm but there seems to be a really big difference between 15 to 24.
>
> *R*: Why do you think?
>
> *T*: I guess that because of car accidents
>
> *M*: Drugs… and alcohol and cars and all those sort of things.

Participants attempted to make sense of the annual risk of death by age and sex in the UK for different ages, interrogating the information and connecting it with their everyday life experiences. While participants interpreted the annual risk of death by age and sex in the UK, appeared intuitively to try to explain risk attributed to different hazardous events.

Participants compared annual risk data of death by age and sex in the UK:

> *T*: In the age group of thirty-five to forty-four, one in every 1106 women will die. The men have higher chance of dying, 1 in 663 men. So, by the time we get to say, the number of women, it’d be almost 2 men would have died in the 35-to-44 age group.

Tim after comparing the risk data of death for females and males of the 35-to-44 age group in the UK, he found a relation between the two annual risk data. When the two participants, were asked to discuss the annual risk of death by age and sex in other countries, they were encouraged to use the Gapminder world map data.

> *T*: Ah, some countries had a life expectancy in some years which went down quite a bit. Other countries it went up, but not by much. The down side was quite a bit bigger but umm, I’m not sure I do understand. Do you mean where does the country, or does the world get a constant population?
>
> *R*: No I mean what is happening with regards to the risk of death in other counties.

The above excerpt portrays how the data presentation influenced participants’ decision-making and caused tension between interpreting risks and focusing on the representation of variation derived from uncertainty as portrayed in the graphical representation of the Gapminder world map data. The animated Gapminder graph that shows variation and the change of the behaviour over time perplexed the participants who did not attempt to quantify of risk or uncertainty.

Then the researcher re-introduced the consideration for numerical quantification of risk:

> *M*: I’ve selected population growth, each year for the vertical axis and population total on the horizontal axis. It doesn’t give an age and sex breakdown, we need some sort of measure of how many people are dying to get the risk of death. That’s not the total, what happened? So,

crude death rate on the vertical axis and on the horizontal is still the population total (Figure 1).

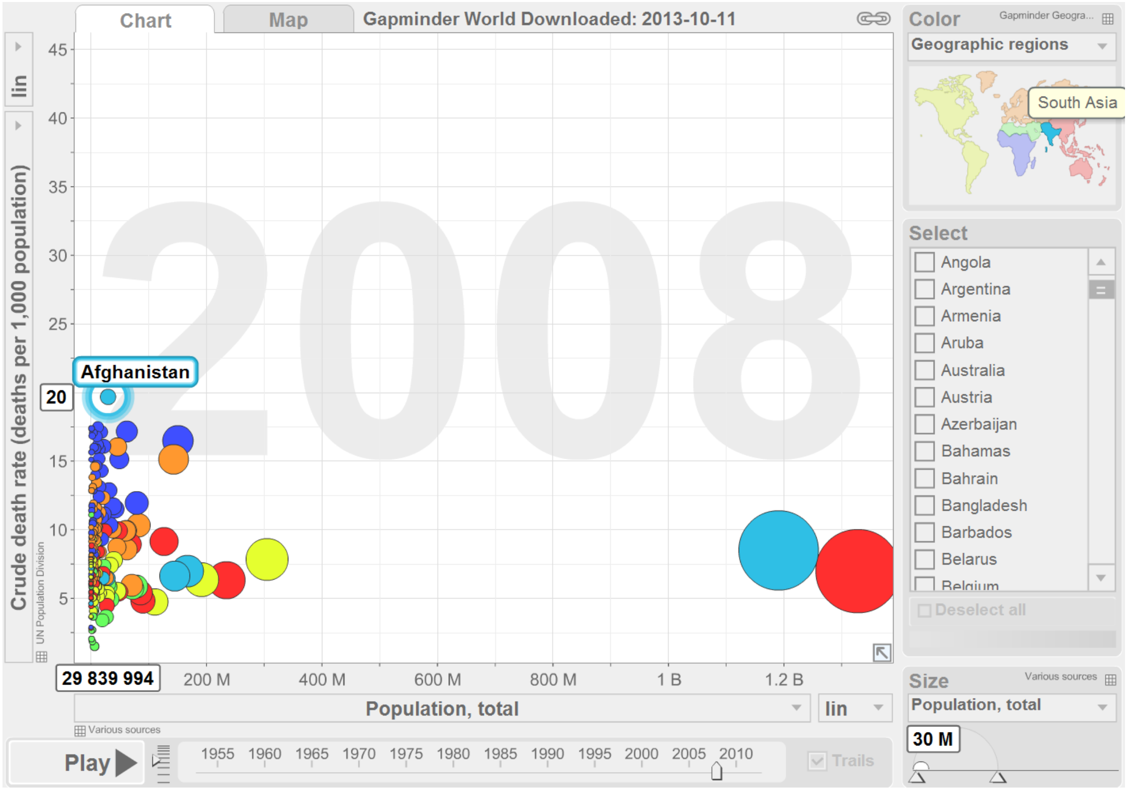

Figure 1. Crude death rate vs. Population total

R: What do you observe there?

T: Well India and China have the big populations in the world so they are on the far right of the axis, of the graph. Umm, and they have …… auh number of deaths per 1000 population

R: Can you see any percentage there?

M: 8.3%

T: So the total, so it's, for every 100 people three are 8.3 deaths

M: So 1 in how many, well 100 divided 8.3, ... , Or one say, divide by 8 is twelve and a half and one in twelve people

T: It sounds kind a high.

R: What about China?

T: 7.2, so one in fourteen. It doesn't sound right.

M: These two countries are the countries that they have the highest populations but not the highest death rate.

T: No, the highest death rate is Afghanistan.

R: Okay, which is how much

T: 19%

R: okay, what that means?

T: Oh no, I've made a mistake, I've made a mistake. It's nineteen deaths per 1000 population. So, one death every, it's close to twenty, twenty-two people or something like that.

R: How did you calculate it?

T: It's nineteen deaths every thousand people. So I divided the left hand side by nineteen and one thousand divided by nineteen. Oh no, I made a mistake because it's about fifty something. I need to get a calculator… So I'm dividing a thousand by nineteen, which is the deaths per one thousand population. And that gets me one death every fifty-two people.

The participants preferred to think in terms of actual numbers of people rather than probabilities or percentages. Often they had to take into consideration other social constructions (e.g., war) with which they were more familiar and interpret the data into their own schemata as shown below.

T: Well, if you're in Afghanistan, you have a much higher risk of dying. Umm, if you're in what was it, India you have fairly similar but you're a little bit, you're after in the United Kingdom than what you are in India.

R: Do you want to talk about Australia?

T: Seven.

R: Is it good?

T: Yes. Oh, one thousand, so in order to find out, three are seven deaths per one thousand people in Australia.

M: To find out how many people it takes for one person to die, I'd divide one thousand by seven…that is how many people are dying. Per one thousand, so it's one hundred and forty-two, umm, but that's both genders so, so it's probably safer if it's about fifty -fifty men and women, it's probably safer in Australia than what it is in the UK.

R: Why?

T: Because one in every one hundred and forty-two people are dying here in Australia. In the UK, it's something between fifty-fifty.

R: Okay, do you want to comment about something else, another country.

M: In the death rate, United Arab Emirates.

R: What is the risk of death by age and sex in the United Arab Emirates?

M: There, you've got umm, one and a half people …dying for every thousand.

R: Is it correct?

T: Yes, it's very, very few people are dying in the United Arab Emirate compared to the rest of the world.

The participants had to think of other countries, observed on Gapminder graph and the risk of death in those countries and compare these risks. The country that has the lowest death risk from the data representation on of data Gapminder attracted the attention of the participants.

When George (G) and Steph (S) engaged with the third scenario, which is about the risk of cancer due to alcohol consumption, they first looked at the two pictures that show the percent of alcohol-related cancer deaths per number of daily drinks consumed:

G: Percentage of high risk of cancer deaths based on daily number of drinks consumed.

S: So the high risk because we have 12% for women and 15% for men.

G: For men it's, umm it's high risk if you have 7 or more drinks per day. That doesn't seem to have a category on this on the bar chart.

S: Except to say more than 3 drinks per day. Uh so, well look… I think, I think umm, I think it's hard to say. Umm, well the questions ask "How do you understand the percentage of high risk of cancer death based on daily number of daily drinks consumed?" So, umm, this doesn't really look at risk in terms of high medium or low. Or high risk, risky or low risk the way the cancer council statistics provided so I think this doesn't even count.

G: The pie chart does not count ….well, it doesn't provide the information I need to answer the question that's provided. The other picture provides information (provided by the Cancer Council) that if you are a man and you have 7 or more drinks a day, then of 15% of men that do that have, will die in each year and for women if they have 5 or more drinks each day, 12% of them. Oh it really, it says are in risky, high risk category. I'm sort of making a bit of a leap that that means, you know, that they'll get cancer, they'll get cancer and not die. Umm 12% of women.

S: Well even that one doesn't provide that much information because what does risky, high risk mean? Does it mean that you'll get cancer? It doesn't say 20%, 12% of women will get cancer… it doesn't define what risk or high risk actually means.

In attempting to make sense of the information available, questions were raised about the actual meaning of the risk classes. It particularly concerned Steph that the high risk and low risk classes had no numbers associated with those classifications.

G: Actually I've just seen the rest of the graph, so you have 43% in this group in the la, for women, 43% … are in the low risk. So, it is 43% in the low risk, 12% in the high risk and 45% in the risky. But it still doesn't talk about what low risk, risky and high risk actually mean … I mean how many in those categories will get cancer. It all it says is that 43% are, have the habit of drinking up to 2 drink a day. And 12% have the habit of drinking of drinking 5 or more drinks a day. It doesn't say those that drink 5 or more drinks per day have a 50% chance of getting cancer.

S: All it says is that 12% of women drink 5 or more drinks per day and I'm starting to think about the pie chart as well. Umm, there different categories. One says up to 2 drinks a day, one says up to 1.5 drinks a day so, umm gain it's difficult to match the two, the pie chart to the glass graph…The more you drink, the higher the risk, it is that you will get cancer, but I don't know the numbers behind all of that.

Although the dialogue shows that the two participants are engaged with this scenario, it carries the implication that they cannot synthesise the information provided in the first graph that is the number of daily drinks consumed and the number of alcohol-related cancer deaths.

When they attempted to look at the graph estimated Alcohol consumption per average adult (15+) versus life expectancy on Gapminder map (Figure 2):

R: So the vertical axis will be alcohol consumption per average adult and the horizontal will be life expectancy.

G: Yeah because, umm well the, ah so you have Japan has the highest life expectancy.

S: What's this country, this has Moldova has the highest amount of alcohol consumption and the life expectancy.

G: Okay, South Korea. So …. Japan has the highest amount of alcohol consumption.

S: I think Japan has the lowest risk of cancer deaths.

R: Why do you think that?

G: Well because they live longer.

S: But what about the consumption of the alcohol.

G: Well it's in the middle. Umm…

R: What is happening in Afghanistan?

S: The life expectancy is low, probably because of conflict and war.

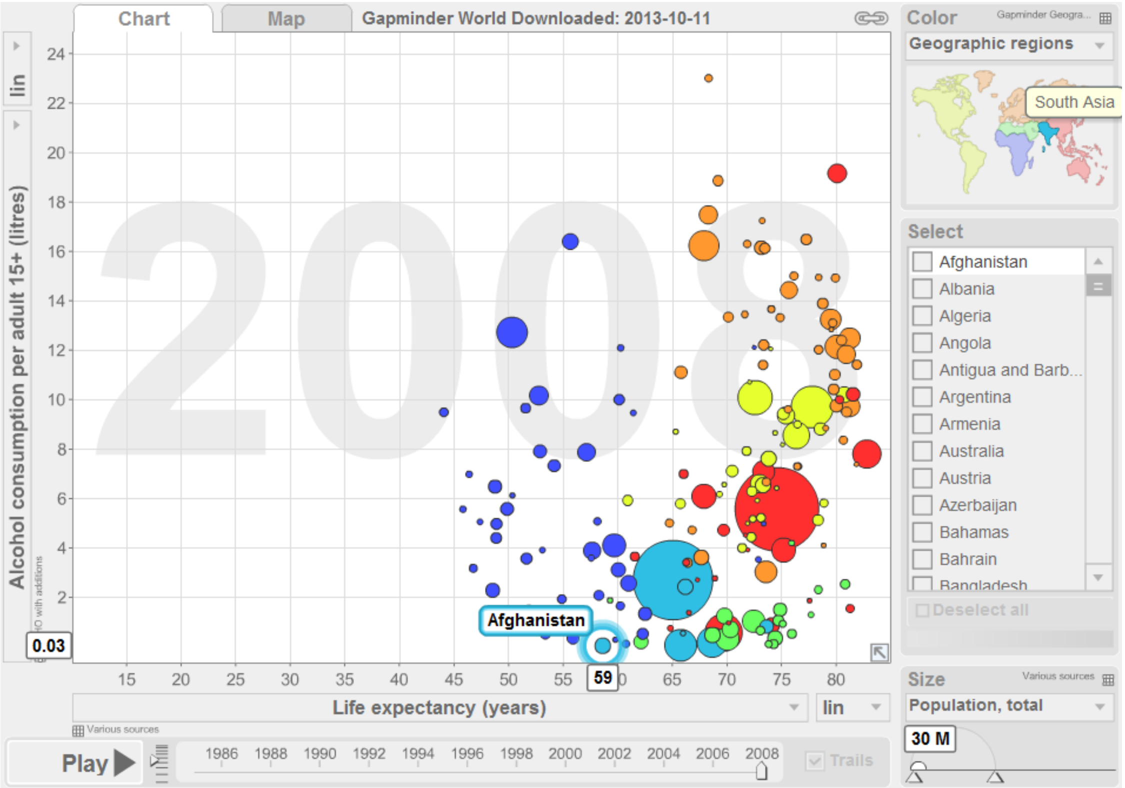

Figure 2. Alcohol consumption per average adult (15+) versus life expectancy

G: The alcohol consumption is very low, 0.03 litres for each adult over 15. Umm so, I'd say, yeah Afghanistan has the lowest risk of cancer deaths based on daily drinks consumed.

S: Ummm, and then I would say that Moldova … has the highest risk of cancer deaths because that's not a country in conflict so there's a lot of alcohol consumption. It's the highest alcohol consumption and it doesn't have a high, it's a medium but it doesn't have high life expectancy. So I think the consumption of alcohol that is occurring in Moldova is probably bringing the life expectancy down because of cancer.

George and Steph were more inclined to focus on the amount of alcohol consumption and the cancer deaths in countries where people consume the highest amount of alcohol as shown on the Gap minder map. When the relation of cancer deaths and the alcohol consumption are not linked, the participants attempted to coordinate alcohol consumption, life expectancy and qualitative data stemmed from the contextual politico-economic situation of the country. For example, they mentioned that the life expectancy in Afghanistan is low, probably because of the war. However, the participants commented that the alcohol consumption is very low, (0.03 litres for each adult over 15) and Afghanistan has the lowest risk of cancer deaths based on daily drinks consumed.

When Tim and Michael looked at the alcohol consumption and life expectancy in other countries, they commented:

T: Well it's interesting now that you say that, that the orange countries ... Europe and Central Asia …tend to have higher alcohol consumption and life expectancies than the African countries.

M: Yes, I see a pattern here

T: I'm going to umm, select GDP per capita. Alcohol consumption is per adult, GDP per capita, well most adults are generally the ones that contribute to the GDP. So … you could sort of say that umm, if income gets higher and higher, in-general you can say that alcohol consumption gets higher….And with that cancer due to caner due to alcohol consumption would be generally be higher but, obviously there's a couple of exceptions. Moldova again… they don't have a high income, but they have a high consumption of alcohol.

M: … but if a country like South Korea, they generally have a higher income and a higher consumption of alcohol.

T: That would be because of umm, Muslim bans on, on alcohol consumption so they have higher incomes but they don't consume much alcohol. So the likelihood is they probably don't have many cancer deaths because of… because they don't have a lot of high alcohol consumption.

M: I mean India and China are always interesting because they've got the biggest circles… India has a lower income and lower alcohol consumption, but China is, is higher income and alcohol consumption. But both are lower than Australia's….

When Tim and Michael introduce another variable that is the GDP per capita, they focus on alcohol consumption and GDP per capita and dismiss to refer to the risk data of cancer deaths because of alcohol consumption. Other features of the data visualisations of Gapminder, such as the size of the dots for China and India, indicating their large populations, attracted the attention of the participants.

Discussion

Despite the complexity of the risk-based reasoning, the participants were engaged with the three scenarios that required them to manage alongside the balancing of trading-off probabilistic evidence and a variety of different kinds data involving qualitative data inextricably derived with the context throughout the process of risk-based reasoning.

The responses of the participants largely became data that revealed how people are assessing and comparing risks. Almost all of the questions are what we might call "assessment questions" – they're questions asking for some sort of assessment or evaluation of a situation (e.g., "What is the income of the people of the countries that have higher and lower risk of cancer deaths based on daily number of daily drinks consumed?"). When the participants engaged with standard questions about risk involved in large-scale data sets, participants' reasoning was essentially based on mathematical calculations—for example the way that the participants worked through the calculation in the Lupitor example. They easily made inferences while reflecting on coordinating likelihood and impact in the context of their evidence.

When the participants were dealing with risk–based decisions about large-scale data, that required making sense of large collections of data, they "often achieve insight into big data by implementing a few principles that would make their data exploration easier:

•Explore the trends over time of fewer variables on the instantiated data structures;

•When faced with a trade-off in which there was clear best choice, they would create a new visualisation to further explore the question" (Prodromou, 2014, p. 71).

Using the interactive visualisations, they attempted to find meaning in the data, but in their visual justification efforts, they seemed to get distracted by other features of the data such as size of the population of a certain country or a pattern that the data seemed to follow, factors that add extra layers of complexity to the stuntedness and construction of risk.

In fact, features of big data appear to be highly salient factors that impinge on decision-making and frame the level of risk in situations that require consideration of large scale data. While the evidence from participants' dialogues suggests that participants made decisions based on salient features of the visual representations of large global datasets, quantification of risk was held back by issues around uncertainty and risk. Especially, participants acknowledged that they could not put numbers on risk when they dealt with uncertainty caused by large-scale data. In fact, participants made no effort to estimate risk using numerical values, or referring to any odds. They merely referred to risk using qualitative statements articulated in subjective terms based on their knowledge about the socioeconomic situation of the country of interest. Much of their discussions about risk were directed towards qualitative descriptors to provide rough quantification of risk.

References

Adams, J. (1995). *Risk*. London, UK: UCL.

BSON. (n.d.). The Notion and Definition of Risk. Retrieved from http://www.web-books.com/eLibrary/NC/B0/B60/006MB60.html

Brandstätter, E., Gingerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, 113(2), 409-432.

Corbin, J., & Strauss, A. (2007). *Basics of qualitative research: Techniques and procedures for developing grounded theory (3rd ed.)*. Thousand Oaks, CA: Sage.

Crossland, B., Bennett, P. A., Ellis, A. F., Gittus, J., Godfrey, P. S., Hambly, C., Kletz, T. A., & Lees, F. P. (1992). Estimating engineering risk. In F. Graham-Smith (Ed.), *Risk, analysis, perception, management*. London: The Royal Society.

Edwards, W., & Tversky, A. (Eds.) (1967). *Decision Making: Selected Readings*. Penguin Books.

Freudenburg, W., Coijemb, C. L., Conzale, J., & Helgeland, C. (1996). Media coverage of hazard events: analyzing assumptions. *Risk Analysis*, 16 (1), 31-32.

Folkes, V. S. (1988). The availability heuristic and perceived risk. *Journal of Consumer research*, 15(1), 13-23.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8(2), 53-96.

Greening, L., Dollinger, S. J., & Pitz, G. (1996). Adolescents' perceived risk and personal experience with natural disasters: An evaluation of cognitive heuristics. *Acta Psychologica*, 91(1), 27-38.

Hilson**.** D. A**..** & Murray-Webster R. (2004). *Understanding and managing risk attitude: Proceedings of 7th Annual Risk Conference, held in London, UK*, 26 November 2004.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge University Press.

Lupton, D. (1999). *Risk*. London: Routledge.

Pratt, D., Ainley, J., Kent, P., Yogui, C., & Kapadia, R. (2011). Role of Context in Risk-Based Reasoning. *Mathematical Thinking and Learning*. 13, 322-345.

Prodromou, T. (2014). Drawing inference from data visualisations. *International Journal of Secondary Education*, 2 (4), 66-72.

Renn, O. (1991). Risk communication and the social amplification of risk. In: R. Kasperson & P. Stallen (Eds) *Communicating Risks to the Public* (Dordrecht, The Netherlands, Kluwer Academic Publishers).

Simon, H. A. (1990). *Reason in Human Affairs*. Stanford, CA: Stanford University Press.

Simon, H. A. (1997). *Administrative Behavior: A study of decision-making processes in administrative organization* (Fourth ed.). New York: Free Press.

Slovic, P. (2000). *The perceptions of risk.* London, UK: Earthscan.

Spiegelhalter, D, (2012). Risk and uncertainty. Retrieved from http://www.cam.ac.uk/research/discussion/risk-and-uncertainty

Stirling, A. (1999). *On science and precaution in the management of technological risk (A synthesis report of case studies)*. Seville, Spain: Institute of Prospective Technological Studies (European commission Joint Research Centre). Retrieved from ftp://ftp.jrc.es/pub/EURdoc/eur19056en.pdf

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), 207-232.

Weale, A. (Ed.). (2002). *Risk, democratic citizenship and public policy*. Oxford, UK: Oxford University Press/ British Academy.

Appendix

Questions

The interview questions will not be of a personal nature but general questions relating to the risk of death.

First Scenario: The risk of heart attack

Visit to my GP

•Recently I went to see my GP

•He told me I had a 12% chance of a heart attack or stroke in the next 10 years. What does this percentage mean?

•Lupitor could reduce this risk by 36%. Will I buy Lupitor?

Second Scenario: The Risk of death

Someone once said that the only certainties in this world were death and taxes. We look at tax rates with some interest, but give death rates much less attention, except when they are forced on us by some catastrophe, societal or personal. For that reason we thought it useful to have a reminder of the major effects on death rates - namely our sex, and our age.

Data were taken from UK national mortality statistics, which provides death rates per million population by age and sex. These have been recalculated to show the results as an annual risk - a chance of 1 in X of dying in the next year (on average), by your age and by your sex.

	Annual death risk 1 in X	
	Men	Women
All ages	136	193
Under 1	177	227
1- 4	4386	5376
5 - 14	8333	10417
15-24	1908	4132
25-34	1215	2488
35-44	663	1106
45-54	279	421
55-64	112	178
65-74	42	65
75-84	15	21
85 and over	6	7

Table 1: Annual risk of death by age and sex in the UK.

•Could you please explain to me the above numbers?

•What is the (percentage) risk of dying next year by your age?

•Which age and sex has the higher risk of death in the UK? Explain the percentages.

•What is happening in other countries? Please refer to Gapminder map.

Third Scenario: The risk of Cancer due to Alcohol Connection

I read online the following articles:

•The Cancer-Alcohol Connection, from the webpage:

http://www.everydayhealth.com/cancer/the-cancer-alcohol-connection-7199.aspx

and use the picture entitled:

"Alcohol-Related Cancer facts

• Cigs war won: Now cancer campaigners set their sights on beer, from the webpage: http://www.news.com.au/national/cigs-war-won-now-cancer-campaigners-set-their-sights-on-beer/story-e6frfkw9-1226088686962

and use the picture entitled:

"Australian Bureau of Statistics data showing the extent of problem drinking / Graphic: Vincent Vergara Source: news.com.au"

to answer the following questions:

•How do you understand the percentage of high risk of cancer death based on daily number of daily drinks consumed?

•Which country has the higher and lower risk of cancer deaths based on daily number of daily drinks consumed? Please refer to the graph estimated Alcohol consumption per average adult (15+) versus life expectancy on Gapminder map.

•What is the income of the people of the countries that have higher and lower risk of cancer deaths based on daily number of daily drinks consumed? Please refer to Gapminder map.