DOI:
10.1002/acp.3381

Citation for published version (APA):
Heard, C. L., Rakow, T., & Spiegelhalter, D. (2018). Comparing comprehension and perception for alternative speed-of-ageing and standard hazard ratio formats. *Applied Cognitive Psychology*, *32*(1), 81-93. Article ACP3381. https://doi.org/10.1002/acp.3381

Comparing comprehension and perception for alternative speed-of-ageing and standard hazard ratio formats.

Applied Cognitive Psychology

Word count: 8574

ABSTRACT

A hazard ratio presents one benefit of exercise as reducing annual mortality risk by 19%. Alternatively, speed-of-ageing metaphors present this as adding 2½ years to one's life expectancy – equating to 1 extra hour each day – or taking 2½ years off one's "effective age". Few studies compare these (increasingly popular) metaphors. Study 1 compared perception and comprehension between speed-of-ageing metaphors and hazard ratios. Study 2 compared the hazard ratio with three versions of effective age (change-in-age, personal, age-matched age). Results revealed a disadvantage to the change-in-age format (behaviour X makes someone Y years older), with unhealthy behaviours perceived as less risky, healthy behaviours as less beneficial, information judged less likely to affect behaviour and harder to understand. The personal format (behaviour makes your effective age X) shows no such disadvantage and is objectively better understood than hazard ratios. These results support the use of personalised effective ages in health and risk communication.

Key words: Risk Communication, Risk Perception, Health Communication; Information Processing, Effective Age, Decision Making

INTRODUCTION

"Just over a can of drink a day could increase the risk of heart failure by 23%, a study found" (Daily Telegraph, 2015). As this newspaper headline illustrates, key results from medical and epidemiological studies are often communicated to the public as hazard ratios (the relative annual chance of an adverse event) converted into a percentage increase (negative behaviours) or decrease (beneficial behaviours). Thus, a hazard ratio of 1.23 becomes a 23% increased risk. However, people often have difficulty understanding these percentage formats, potentially confusing a relative change as an absolute change (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz & Woloshin, 2008) thereby often overestimating risk (Spiegelhalter, 2012a). Such examples of misunderstanding are seen in the media, with newspapers, for example, taking a 10% increase in the annual chance of death to mean that "10% of all deaths could be avoided" if a behaviour is reduced (Spiegelhalter, 2012b; Willey, 2012).

Given this difficulty, and research suggesting that people are less accurate in their risk estimates when information is presented as a percentage (compared to other formats such as natural frequencies; Knapp, Gardner, Carrigan, Raynor & Woolf, 2010), one alternative is to re-express hazard ratios as change in life expectancy or change in one's speed-of-ageing. This converts hazard ratios into time lost or gained on your life by different lifestyle (habitual) behaviours (Spiegelhalter, 2012a, 2012b)[1]. This aims to improve understanding by using a more concrete and familiar representation than probability. To illustrate, 20 minutes of moderate exercise is associated with a 19% reduction in annual mortality risk (hazard ratio 0.81; Woodcock, Franco, Orsini & Roberts, 2010). Assuming this effect throughout adulthood, this represents: adding 2½ years over one's whole life; adding ~1 hour to one's

[1] Simplified, these equivalences are computed by dividing log (Hazard Ratio) by log (1.09) to estimate the expected years added or subtracted; then, each year corresponds to 30 minutes a day (see Spiegelhalter, 2012b).

life each day; or taking 2½ years off one's "effective age" (having the annual risk of death of someone 2½ years younger). These formats above can be split into three subtypes, each focussing on a different time-based metric, as described next.

'Over-life' (Life-expectancy): This presents the benefit (risk) as years gained (or lost) from one's life (e.g. 2½ years over one's whole life), so either improving or reducing one's life expectancy. One potential limitation of this format, as the time discounting literature has identified (Story, Vlaev, Seymour, Darzi & Dolan, 2014) is that by pushing discussion of the impact of behaviours to the distant future, it may be dismissed (less-valued) in favour of more immediate choices (e.g. the enjoyment of a sugary snack). Explanations for this include the abstract nature of far-future events potentially giving them less impact (Liberman & Trope, 2003), reluctance to delay gratification (Loewenstein, Read, & Baumeister, 2003), and risk discounting whereby the future is ignored because one cannot know whether one will survive to receive promised benefits (Patak & Reynolds, 2007).

'Day' (micro-life): A present-focussed speed-of-ageing format, which could potentially overcome this problem, is the *day format*, which transforms the change in life expectancy (in years) into hours gained (lost) each day (e.g. ~1 hour added to one's life per day). Subtly different, this can be presented as a number of 'microlives': each microlife representing one of the million half-hours in the average adult life-span (Spiegelhalter, 2011).

'Effective Age': Chronological age represents our traditional conception of age (i.e. years since birth), and as such, is fixed at any time. In contrast, one's effective age, defined as the age of a 'healthy' person who has the same risk profile (category), increases or decreases depending on the type of life lived[2]. Thus, a 30-year-old heavy smoker may have the same

[2] Thus, generally, a reference trajectory of a typical 'healthy' person's risk level is calculated. Then, by mapping a person's actual risk level (which could be heightened through poor behaviour) onto this, the age of a "healthy" person with that risk level can be found (Spiegelhalter, 2016). Appendix A illustrates this.

annual mortality risk as a ‘healthy’ 38-year-old. (Spiegelhalter, 2015). A similar analogy is used in property appraisal, where buildings have both an actual age (years since construction) and effective age (how old the house appears). Similar to how health-based effective ages take account of behaviour and functioning of an individual’s body, this takes account of condition and functionality of a property (Forsythe Appraisals, 2015).

Two key forms are distinguished in presenting one’s effective age: change and personal age. Similar to the over-life format, this change format also presents the information in terms of years gained or lost: making your effective age younger or older than your chronological age (by ‘X’ years). In education, such change-formats are used to describe reading ability, such as: ‘children had word reading which was three years and six months ahead of their chronological age’ (p.15; Gibbs, 2015). The second form, personal age, uses an individual’s age (e.g. 30 years) and change-in-years information (e.g. behaviour adds 8 years) to calculate a personalised effective age (e.g. 38 years). In healthcare, this has been used to communicate the effective age of particular organs (e.g. heart age, Lopez-Gonzalez et al., 2014; lung age, Parkes, Greenhalgh, Griffin & Dent, 2008). Compared to traditional percentage-based expressions (or raw test statistics), this has been found to lead to improved adoption of healthier behaviours (e.g. quitting smoking), helping reduce an individual’s risk.

While there is research on specific organ-age metrics (Lopez-Gonzalez et al., 2014) there is limited empirical evidence establishing how people view, understand, and react to information presented in these (now widely used) speed-of ageing formats compared to more traditional formats. As interest in using these more recently developed metaphors in “formal” and “official” risk or health communications (e.g. patient information leaflets) develops, it is important that empirical research compares how people perceive and understand information in these formats and therefore begins to assess their effectiveness for communicating risks or benefits. We therefore conducted two studies to address this gap by comparing different ways of communicating a hazard ratio. Study 1 investigated the four main formats (i.e. the three speed-of-ageing metaphors and the "traditional" hazard ratio percentage); then Study 2 compared the traditional format with variants of the effective-age format.

STUDY 1

To address our aims, Study 1 used several self-report measures to assess perception and comprehension for information about the risks (or benefits) of different lifestyle behaviours (e.g. exercise), as well as reactions to that information. The stimulus information was presented in one of four formats: hazard ratio percentage, day, over-life and effective age (change). The selection and phrasing of our scenarios reflected the typical way in which this information is presented when it is used in an applied setting[3]. Four types of self-report measures were used: perception of risk (or benefit, dependent on scenario), understanding, surprise, and behaviour-change belief. These four measures were chosen for their potential relevance for risk communication and decision aid designers, as described next.

Risk perception and understanding were measured because improving people's knowledge and understanding is a core aim of most health or risk communications, decision aids, or patient information sheets (e.g., for informed consent). One issue with risk perceptions to bear in mind is that whether one takes a higher risk (or higher benefit) perception to be more beneficial may be context-specific: for example, dependent upon whether one is seeking to correct a bias to over-estimate, or to under-estimate risk.[4]

While self-ratings of understanding may not match with objective measures, these subjective measures are important because they likely reflect people's willingness to engage

[3] One could express the effect as a relative change in age (e.g. "this increases your effective age by 8%"). However, people rarely use statements like: "you are 10% older than me". Moreover, because effective age changes are invariant on an absolute scale (number of years) but not on a relative scale (unlike many treatment effects) this could only be used if an individual's age is known.
[4] The ethics of this are considered in the Discussion.

with information – which is usually a pre-requisite for improving (objective) understanding. This measure, therefore, allowed identification of formats where people have particular difficulty in understanding the information presented. Moreover, understanding and risk/benefit perception both reflect beliefs and attitudes, which are implicated as important drivers of behaviour in several cognitive theories of decision making and behaviour change (e.g. theory of reasoned action, theory of planned behaviour) – though in-and-of-themselves they do not always predict actions and behaviours as strongly as one might expect (Madden, Ellen & Ajzen, 1992). Additionally, attitudes also have an affective component, and the importance of such emotional reactions in processing and responding to risk communications is well documented (e.g., Rakow, Heard & Newell, 2015). Assessing participants' surprise allowed for further assessment of this important dimension. More specifically, given that violations of expectancy play a key role in learning (Rescorla & Wagner, 1972), surprise should reflect people's propensity to revise their beliefs, and so alter their goals and actions. Because changing behaviour is the ultimate goal of most risk or health communications, our final measure assessed how likely the communication was to change behaviour. This statement was written from a third-person perspective to reduce bias based on people's individual level of a behaviour. For example, for a person already eating 5 portions of fruit and vegetables, their consumption is already "high", so the information in the box might do little to change that individual's behaviour (while it might change the behaviour of those with lower levels of consumption).

Predictions

Three main predictions are made for comparisons between the hazard ratio and speed-of-ageing metaphors. First, given that the higher amount levels deal with either a higher risk or higher benefit (e.g. larger percentages or amount of time), for both perception and behaviour change belief, it is predicted that the higher amount levels will show higher risk (benefit) perceptions and behaviour change belief. For understanding, no difference is expected between the amount levels because although this changes the magnitude of the value, the concepts (and therefore how easy they are to understand) do not change.

Second, given that people often overestimate probabilities when these are presented in (percentage) hazard ratio format (see Introduction), it is predicted that this format should lead to the highest risk/benefit perceptions of all the formats. This is because if the information is misinterpreted as an absolute risk level (Gigerenzer et al., 2008; Spiegelhalter, 2012a, 2012b), the percentage increase (for negative behaviours) or percentage reduction (for positive behaviours) will be larger.

Third, since the three speed-of-ageing metaphors were designed to use a more familiar dimension (i.e., time rather than probability), it is predicted that the three speed-of-ageing metaphors should be perceived by readers as easier to understand than the hazard ratio format.

Because the speed-of-ageing formats are new, no *a-priori* predictions are made for specific differences between the formats at a general level; and given the similar concepts behind these formats, it is expected that ratings across these measures should be similar. One place where differences are predicted between these speed-of-ageing formats are in relation to the individual differences measures where a number of specific predictions are made and are discussed below.

First, numeracy was included because previous research finds that less numerate individuals show greater susceptibility to information presentation effects (Reyna, Helson, Han & Dieckmann, 2009: Peters, 2008). Therefore, we predict that numeracy and format may interact with a greater increase in self-rated understanding for the speed-of-ageing formats in those with low rather than highly numerate individuals (who may have a strong understanding of the traditional numerical format). Second, because the speed-of-ageing metaphors are time-based, consideration of future consequences (measuring the extent to which people are more present- or future-focussed) was included to understand whether this interacts with format. As described earlier, the day format is present focussed while the over-life format is future focussed. Therefore, because the day metaphor appears to bring the time of the effect closer (i.e. “every day”), those low in consideration of future consequences (who are more present focussed) may have high risk/benefit perceptions in this day condition, where time-focus and presentation format are congruent. On the other hand, those with high consideration of future consequences, may either show little difference between the format because they consider the future consequences irrespective of condition (and show higher ratings of risk across all formats); or may show higher risk perceptions in the over-life condition, where the risk is presented in a more future-focussed format which is congruent with their perspective.

Methods

Participants

160 (42 male, 118 female) students from the University of Essex participated in the study for course credits or payment. Age ranged from 18-45 years (mean: 21.92, standard deviation: 4.32).

Procedure and Task Materials

All instructions and tasks were presented via the online *Qualtrics* survey platform. After a short introduction, participants answered demographic questions (gender and age) and then completed a series of tasks, including (in this order): eight format task pages, an 11-item consideration of future consequences (CFC) scale and an adaptive numeracy test.[5] Details for the main tasks discussed in this paper are included below. Further details of these tasks and other tasks which are not discussed in this paper in depth can be found in the Supplementary Material at: http://www.statslab.cam.ac.uk/Dept/People/Spiegelhalter/davids.html

Format Task. Each format task presented information about the change in annual mortality risk associated with an activity (e.g., eating processed meat) at a specified amount of 'exposure' (e.g. 50g per day) in a given format (e.g. effective age). There followed 6 perception and comprehension questions based on this information. From a possible 32 different task pages, made up of 4 behaviours (eating fruit and vegetables, eating processed meat, moderate exercise, and watching TV) each at 2 amount levels (low and high), in 4 formats (hazard ratio percentage, day, over-life, and effective age), participants were shown 8 of these task pages: 2 for each behaviour (i.e. both amount levels), with each behaviour in a different format. Therefore, across the 8 format tasks, each participant had 2 tasks that focussed on each of the 4 behaviours and 2 tasks presented in each of the 4 formats. A worked example for 50g of processed meat of each format used in the experiment (as well as those used in Study 2) can be found in Appendix C. See the online Supplementary Materials for the exact hazard ratio percentage and number of hours (years) statistics that were presented for all four scenarios (at both amount levels) and for each format.

A balanced design for the combination of behaviour type and format was achieved using a Latin square to create four different sets of eight tasks (see Appendix B). Participants were randomly assigned to one of these four sets, and the order of the eight tasks within their designated set was randomised (separately for each participant). Each of the 6 risk/benefit perception and comprehension questions was answered on a 20-point rating scale (labelled at

[5] Two further tasks were also included; Appendix B and the Supplementary Materials provide further details of these.

each end). For the first three questions, the wording of the question differed depending on the valence (positive or negative) and amount level of the behaviour.

For the first question, participants rated how beneficial (for positive behaviours) or how risky (for negative behaviours) a particular behaviour was (at a particular amount level). For example, “In general, how risky is eating 50g of processed meat a day?” (1=Not Risky, 20=Very Risky). Second, participants rated the importance of increasing (for positive behaviours) or decreasing (for negative behaviours) the amount of a particular behaviour. For example: “In general, how important is reducing the amount of processed meat eaten a day?” (1=Not Important, 20=Very Important). Third, participants rated how good (for positive behaviours) or bad (for negative behaviours) a particular behaviour (at a particular amount level) is for someone’s health. For example: “In general, how bad for one’s health is eating 50g of processed meat a day?” (1=Neither Good or Bad, 20=Very Bad).

The last three questions were identical for all tasks and were: “How easy was it to understand the information in the box above?” (1=Very Difficult, 20=Very Easy), “Did you find the information in the box above surprising?” (1=Not Surprising, 20=Very Surprising), and “Do you think the information in the box above would encourage someone to change their behaviour?” (1=Definitely No, 20=Definitely Yes).

For analysis, the format task consisted of four behavioural scenarios with a 4 (format: hazard ratio percentage, day, over-life, and effective age change) by 2 (amount level: low vs. high) design. Four dependent variables were derived from the risk perception and comprehension questions. The responses for the first three questions were combined to form an overall risk/benefit perception rating. Cronbach’s alpha for each 3-item set (for different behaviours and amount levels) ranged from ‘good’ to ‘excellent’ (α-range: .78 to .91) supporting the use of a combined rating. The final three questions were analysed separately.

Individual difference measures (CFC, Numeracy). An 11-item (5-point Likert-scale) Consideration of Future Consequences scale (CFC; Strathman, Gleicher, Boninger and Edwards, 1994) was used. This scale included questions such as "Often I engage in a particular behaviour in order to achieve outcomes that may not result for many years" (1=extremely uncharacteristic of you; 5=extremely characteristic of you). Internal consistency was excellent (α=.84, 7-items reverse-coded), thus items were combined to create a single CFC score.

Based on specifications in Cokely, Galesic, Schulz, Ghazal and Garcia-Retamero (2012), the adaptive Berlin numeracy test was implemented on the online platform. This test places participants into four quartiles for the general population (based on large validation studies). Participants in this study were predominantly in the lower numeracy groups (Q_1 41%; Q_2 31%; Q_3 18%; Q_4 11%).

Research Ethics. Ethical approval for this study was granted by the research ethics committee at the University of Essex and informed written consent gained from all participants in the study. Financial support for both was provided entirely by a collaborative PhD studentship received by the first author from the University of Cambridge (Winton Fund) and the Economic and Social Research Council (ESRC).

Results and Discussion

For this main perception and comprehension task, analysis of the results was originally conducted at the individual scenario level. Analysis at this level revealed a similar pattern of differences between formats across all scenarios for each question. As such, and also for brevity, only the collapsed analysis is presented within this paper[6]. With this in mind, high evaluation responses on these measures are termed "risk/benefit perceptions" as higher scores

[6] Details of these separate analyses for Studies 1 and 2 are in the online Supplementary Materials.

on the perception measure represent stronger evaluations in either case. Thus for negative behaviours, higher responses on the scale represent stronger negative responses (i.e. a higher perception of the risk); while for positive behaviours higher responses on the scale represent stronger positive responses (i.e. a higher perception of the benefit). For all other variables, the wording was the same for all scenarios because these variables were invariant to stimuli valence changes.

To analyse the effect of amount level and format on each measure of perception and comprehension, a series of 2x4 within-subjects ANOVAs collapsed across behaviours were conducted (Figure 1)[7], with further post-hoc (LSD) analyses conducted where appropriate.

*****Figure 1*****

For both risk/benefit perception and behaviour change belief, when the risk (benefit) is larger (i.e. higher amount level), participants rated the risk (benefit) as more risky (more beneficial) and more likely to lead to behaviour change (both $p<.001$). However, for both understanding and surprise, amount level was generally inconsequential, with little difference in understanding and surprise between the amount levels. This is as expected for understanding, since changing amount level changes only the magnitude of a number – leaving the concepts involved (or their presentation) unaltered.

As shown in Figure 1, significant effects of format are found across all measures. Providing partial support for our prediction, the hazard ratio percentage (the relative risk format) shows higher risk/benefit perceptions than the effective age change format ($p<.001$). However, rather than similar risk/benefit perceptions for all three speed-of-ageing metaphors, risk/benefit perceptions were also significantly lower in the effective age change format than

[7] Across both studies, partial eta-squared (η^2) is used as the effect size measure: this estimates the proportion of variance accounted for by an effect. On figures, errors bars represent 95% confidence intervals.

both the over-life ($p<.001$) and day format ($p=.027$), and were lower in the day than over-life format ($p=.004$).

Contrary to our prediction that people would find the hazard ratio percentage most difficult to understand, the effective age change format was rated as most difficult to understand (compared to all other formats: $p<.001$). Of the other formats, ratings of understanding were more similar, although significant differences were found between each comparison pair, with the highest ratings for the over-life format, followed by the day and then hazard ratio percentage format (all $p<.043$).

Similarly, for behaviour change belief, this was also lowest in the effective age change format, with significantly lower behaviour change belief in this format than both the hazard ratio percentage ($p=.004$) and over-life format ($p<.001$), but not the day format ($p=.101$). A similar pattern although reversed in direction is also seen for surprise, with significantly higher ratings of surprise for effective age change than any other format (all $p<.036$).

Individual difference variables. Contrary to previous research, no effect of numeracy (or interaction with format) was found for either risk/benefit perceptions or understanding. However, consideration of future consequences (CFC) scores were a significant predictor of risk/benefit perceptions ($B=1.04$, $z=3.77$, $p<.001$)[8] with higher risk/benefit perceptions for those with higher CFC scores (i.e. more future focussed). However, when the interaction between format and CFC scores was assessed between the clearly present- (day) and clearly

[8] Analysis via Generalised Estimating Equations model (with exchangeable correlation matrix and robust standard errors) was used. Continuous variables were mean-centred. For the categorical variables, where only two levels were used in the analysis (e.g. for the interaction analysis), this was coded as -1 or 1. For the four level format variable (used in the main analysis), this was dummy-coded (i.e. split into three dummy-coded variables).

future- (over-life) focussed formats, no significant interaction between format and CFC scores were found (*B*=0.24, *z*=0.81, *p*=.415).

The relationship between these measures and behaviour change belief. Risk/benefit perceptions, self-rated understanding and surprise were regressed on behaviour change belief, separately for each scenario (8 analyses in total). Both risk/benefit perceptions (all $p<.001$) and surprise (all $p<.05$ except one: $p=.058$) were significant positive predictors of behaviour change belief. Self-rated understanding was not a significant predictor in any analysis (all $p>.088$).

Conclusion

Taken together, compared to any other format, these results revealed a disadvantage to the effective age change format, which showed less extreme risk (benefit) perceptions compared to the other speed-of-ageing and hazard ratio formats, and, importantly, is also rated as harder to understand and less likely to lead to behaviour change.

STUDY 2

These results for the effective age (change) format contrast with previous research arguing that effective age is superior to traditional formats in improving adoption of healthier attitudes and behaviours (Lopez-Gonzalez et al., 2014). This research implies that the effective age format would be rated highest for understanding and behaviour change belief – not lowest, as we found. There are, however, two forms of effective age: change (i.e. years older/younger) and personal age (i.e. your body as “X” years old). Those previous studies (e.g. heart age, Lopez-Gonzalez et al., 2014; lung age, Parkes et al., 2008) which have found a benefit to using effective age have focussed on the personal age format (personalised heart or lung ages). Study 1, however, focussed on the ‘generic’ change format, and its findings were opposite to those previous studies. Two main differences between the formats suggest potential explanations for this difference.

First, the personal effective age benefits from being simpler, requiring less effort on the part of the recipient because the (age) calculation is already done. Perhaps more important than any minimal effort saving is the fact that it ensures that the calculation actually *is* done, and, therefore, cannot be overlooked. We might suppose that, when someone knows their “new” age, it makes their risk exposure more concrete.

Second, this personal age has the advantage of being, as the name suggests, personal to the individual, while the change format may be presented in the third person (i.e. referring to “a person’s effective age”) and applicable to any individual. This personalisation may therefore lead information to have a larger effect on people’s behaviour, because, by tailoring it to the individual, it is difficult for them to ignore the fact the risk relates to them. In more general (third person) formats, they may discount the relevance of the risk to themselves believing “they can take the risk”, but “others should be safe” (Stone, Choi, Bruine de Bruin & Mandel, 2013).

Design Choice and Predictions

To investigate these differences, Study 2 employed an identical task format to Study 1, but this time comparing the traditional hazard ratio percentage format with three variations of effective age: change, personal age and age-matched. This final format, age-matched effective age, was identical to a personal age format, in that it calculated a person’s effective age; however, rather than being phrased in the second person (e.g. with your age, your effective age), it was phrased in the third person (e.g. for someone of your age, their effective age). This was used to ascertain whether any differences between formats are due to the format design itself or to the language used (third person vs. personal). Overall, it is predicted that these new effective age formats (personal and age-matched) will show higher ratings across both understanding and perceived behaviour change belief than the change format. If it is the simpler nature of the 'new' formats which provide the benefit, then both new effective age formats will show an equal benefit. If it is the personalisation of the format, then the largest benefit will be seen for the personal effective age format. As before, both numeracy and CFC were measured, with only the addition of a further numeracy measure.

Second, to understand not only how well people *feel* they understand these risk communications (their subjective understanding), but also how well they actually understand these underlying concepts, participants were asked what they understood from statements about effective age and hazard ratio percentages (open-ended responses). Despite participants giving high subjective understanding ratings for the hazard ratio percentage, it is possible that this confidence may be misplaced. If some participants misinterpret the risks as absolute percentages, this may show in their explanations of what a hazard ratio percentage means.

Methods

Participants

171 (33 male, 138 female) staff and students from King's College London participated for payment. Their ages ranged from 16 to 55 years (mean: 24.86, standard deviation: 6.63). Numeracy performance on the Berlin Numeracy Test was higher than in Study 1 (Q_1 23%; Q_2 27%; Q_3 14%; Q_4 41%).

Procedure and Task Materials

After a short introduction, participants answered demographic questions (gender, age, whether English was their first language); then completed a series of tasks (in this order) including: eight format question tasks, a (free-text) concept understanding task, an (updated) 14-item CFC scale (Joireman, Shaffer, Balliet & Strathman,, 2012) and both a subjective numeracy scale (Fagerlin et al., 2007) and an adaptive (objective) numeracy test (Cokely et al., 2012).

Format Question Task. This task, its implementation and design, was identical to Study 1; differing only in the response scale used (shortened to a 7-point scale) and the presentation formats examined: the day and over-life formats were replaced with the alternate effective age formats (personal and age-matched), with the hazard ratio percentage and effective age change format remaining unaltered. The dependent variables for the format task were identical to Study 1, with responses for the first three questions again combined into a single measure (α-range: .81 to .91).

Concept Understanding Task. Participants typed (free-text) 'write your thoughts' answers to three questions (order randomised):

What does it mean for a 28-year old to have an 'effective age' of 31?

What does it mean for someone to have an increase in annual chance of death of 10%?

What does it mean for someone to have an increase in annual chance of death of 110%?[9]

For each participant, responses for these questions were coded (by two independent coders) for understanding: one score for the calculated effective age concept, and another for hazard ratios concept (both questions taken together). The coders used a marking scheme that identified different levels of understanding ('0' to '4': Not Relevant, Incorrect, Partial, Some, and Full Understanding). For the hazard ratio, a major misunderstanding (coded '1') might be

[9] A final multiple-choice question asking which phrase best describes the concept of "effective age" from the paragraph above (with piped in answers from the effective age question), was included; however further discussion goes beyond the scope of this paper.

explaining the 110% percentage shown as representing death/impossibility; while for the effective age, this may be explaining body age as related to looks not health. For the hazard ratio, full understanding (coded ‘4’) meant showing understanding of: (1) risk as relative (to a comparison group); (2) percentage increase as increase from a “baseline” not total risk percentage; (3) the risk as an *annual* risk/chance. For the effective age concept, full understanding meant showing understanding that speed-of-ageing: (1) related to *mortality* risk (not merely health status); and (2) communicated risk level (i.e. as having same risk level as someone older). Intermediate scores (‘2’ or ‘3’) were given when some of these elements were present, but not all were present/clear. A score of ‘0’ was only given if the answer revealed a misunderstanding of the question rather than of the concept.

In further qualitative analysis, key themes in the responses were identified and coded for each concept, based on *a priori* and emerging themes. The on-line Supplementary Materials give further details of this qualitative analysis of themes and the quantitative coding of understanding.

Individual difference measures (CFC and Numeracy). A revised 14-item CFC (7-point Likert) scale (Joireman et al., 2012) replaced the 11-item version used in Study 1; and an 8-item subjective numeracy scale (Fagerlin et al., 2007) which included questions such as “How good are you at working with fractions?” was added. Both scales had good internal consistency (α=.78 for CFC, with 7-items reverse-coded; α=.83 for subjective numeracy) so responses were combined to create a single score for each scale. The adaptive Berlin Numeracy Test was implemented as per Study 1.

Research ethics. Ethical approval for this study was granted by Psychiatry, Nursing and Midwifery Research Ethics Subcommittee at King’s College London and informed written consent gained from all participants in the study.

Results and Discussion

Responses to the Format Task

To analyse the effect of amount level and format on each measure of perception and comprehension: risk/benefit perceptions, understanding, surprise and behaviour change belief, a series of 2x4 within-subjects ANOVA were conducted collapsed across behaviour (Figure 2), with further post-hoc (LSD) analysis conducted where appropriate.

*****Figure 2*****

As in Study 1, large effects of amount level are found for both risk/benefit perceptions and behaviour change belief, with the higher amount level associated with higher perceptions of the risk (benefit) and a higher belief that this information would change behaviour. For understanding and surprise, although significant, these effects are much weaker than those of the other two measures.

The effect of format. For perceptions of risk/benefit, the hazard ratio format was perceived as significantly more risky (more beneficial) than the effective age change format (p=.008), and any of the new effective age formats (vs. personal p=.017; vs. matched p=.003). To unpack the interaction between amount and format, this effect of format was analysed separately for both amount levels. This revealed a significant effect of format for the high amount level (F(2.40, 408.06)=6.41, p<.001, η^2=.036, small effect) but not the low amount level (F(2.48, 421.43)=1.68, p=.170, η^2=.010, small effect). The pattern of significant differences was the same as for the main effect.

Supporting our prediction that these modified effective age formats (personal and age-matched) would be better understood, significantly higher ratings of understanding were given for both the effective age personal (p=.001) and age-matched (p<.001) format, when compared to the effective age change format. There was no clear difference between the age-matched and personal format (*p*=.702), suggesting that specifying age to provide a simpler and more concrete basis for comparison is more important rather than whether language is “personal”. For the hazard ratio format, ratings of understanding were not significantly above those of the effective age change format (*p*=.123).   For the personal and age-matched formats, only a small advantage in self-rated understanding was found compared to hazard ratio format, which was not significant (both p>.160). Unpacking the interaction between format and amount level revealed only a significant effect of format for the high amount level (F(2.59, 440.30)=6.72, p<.001, η^2=.038; low amount F(2.65, 451.02)=1.85, p=.144, η^2=.011, both small effects). An almost identical pattern of significant differences were found as for the main effect (however, hazard ratio was also significantly higher than the effective age change format, *p*=.010).

This pattern of results, however, did not extend to perceived behaviour change belief, where these differences echoed those of the risk/benefit perception responses. Thus, behaviour change belief was significantly higher for information presented as hazard ratios than for any of the effective age formats (all *p*<.001).

For surprise, significantly higher ratings of surprise were given when information was presented in the hazard ratio percentage format than for both the effective age matched and change format (both *p*=.012). For the new effective age personal format, ratings fell between the hazard ratio and other effective age format (all *p*>.139).

Using the same analyses as Study 1, no significant effects of numeracy were found. However, on risk/benefit perceptions, CFC scores showed a main effect (*B*=0.25, *Z*=2.90, *p*=.004) with higher risk/benefit perceptions for those with higher consideration of future consequences.

The relationship between these measures and behaviour change belief. Risk/benefit perceptions, self-rated understanding and surprise were regressed on behaviour change belief, separately for each scenario (8 analyses in total). Both risk/benefit perceptions (all p<.001) and surprise (all p<.05 except one: p=.077) were significant predictors; but self-rated understanding was not a (consistent) significant predictor of behaviour change belief (all p>.275, except one: p<.038).

Concept Understanding Task

With one exception, all participants gave responses to the effective age question and to both questions about hazard ratios. Responses which showed a misunderstanding of the question rather than the concept (e.g., explaining it in terms of the amount of behaviour – ppt.146: "there're not eating fruit and vegetables") were excluded from the analysis of that concept. This led to responses from 30 individuals for each concept (effective age and hazard ratio) being excluded. Scores for the participant responses given by the two coders were combined into a mean rating score, because analysis revealed good agreement between coders for both the hazard ratio (K=.57, p<.001) and effective age (K=.69, p<.001) concept.[10]

Understanding. As Figures 3a and 3b highlight, there is a striking difference in the levels of understanding between these two formats. While, fewer than 10% of participants misunderstood the effective age concept completely, 34% of participants answered incorrectly for the hazard-ratio concept. A Wilcoxon signed ranks test revealed a significantly higher level of understanding (Z=-4.38, p<.001) across participants for effective age (mean: 2.80) than for hazard ratios (mean: 2.29).

[10] The first author devised the mark-scheme in consultation with the other two authors – then scored responses as first marker. These authors were not blind to the study predictions, but were blind to participant's subjective understanding ratings. The second marker was blind to the predictions of the study, and rated participants' responses independently according to the mark-scheme.

*****Figure 3*****

A number of themes emerged from the qualitative analysis of incorrect responses for the hazard ratio format. In addition to answers that were incorrect because of vagueness (ppt.10: “10% is quite a large number”), two key markers of misunderstanding emerge from responses to the 110% question: death as certain (13 participants) and 110% as impossible (2 participants). In addition, 9 participants mentioned being confused about what this format ‘really’ means.

The relationship between subjective and objective understanding. As both subjective (from format task) and objective (from concept understanding) measures of understanding were recorded, a final auxiliary analysis was conducted to investigate whether participants are aware of their actual level of understanding. Little correlation was found between these measures of understanding for either the hazard ratio concept ($r(170)=.037$, $p=.631$), nor for the effective age format (when all three formats examined, all $r(170)<.062$, all $p>.422$). These results suggests that participants have little awareness of their actual understanding level for the hazard ratio or effective age concepts.

Conclusion

The format task revealed higher risk/benefit perceptions and behaviour change belief for the hazard ratio format compared to all effective age formats. For self-rated understanding, as predicted, both calculated effective age formats (personal and age-matched) were rated easier to understand than the effective age change format. Despite generally high subjective understanding, analysis of objective understanding revealed that over a third of participants clearly misunderstood the hazard ratio concept (compared to only 10% for the effective age concept).

Consistency of Findings across Both Studies

Studies 1 and 2 had two conditions in common for the format task: hazard ratio and effective age change. When these conditions were compared, for risk/benefit perceptions and behaviour change belief in both studies, hazard ratio format is rated significantly above the effective age change format. The effect of these formats on surprise ratings was inconsistent: with opposite directions of effect in the two studies. For self-rated understanding, the same direction of effect was observed in both studies (reduced understanding in the effective age change format) but this effect was much smaller in Study 2 and only significant at the higher amount level. These differences between formats may have been attenuated because seeing the effective age format in another form may have boosted perceived understanding of the change format. In light of this inconsistency, we pooled responses for self-rated understanding from both experiments for these two conditions. To account for the change in scale (Study 1: 1-20; Study 2: 1-7), scores on the 20-point scale were linearly scaled onto a 1-7 scale. A significant effect of format on self-rated understanding was found with higher ratings for the hazard ratio format (Mean: 6.22; Standard deviation: 1.17) than the effective age change format (Mean: 5.76, Standard Deviation: 1.43); $F(1,330)=34.05$, $p<.001$, $\eta^3=.094$ (moderate effect).

GENERAL DISCUSSION

This study is one of the first to directly compare perception and comprehension of public-health information (particularly about lifestyle risks) between the traditional (hazard ratio percentage format) and alternative speed-of ageing formats. Overall, these results reveal that rather than a clear divide in understanding and perception of the risks between the speed-of ageing and traditional hazard ratio percentage format as had been predicted, the reality is more complicated, particularly when considering the effective age metaphor. In fact, across the measures of perceptions and comprehension, the hazard ratio percentage is similar to both the day and over-life format, with often only a slight increase in ratings for the over-life compared to the day format (and for risk (benefit) perceptions, the hazard ratio percentage lying between the two).

For the final speed-of ageing metaphor, effective age, however, it appears that the specific format used (e.g. change or personalised age) makes an important difference to how well the format is understood and information perceived. While presenting information as a change score appears to be difficult to understand, the personal age format is rated by readers as much easier to understand, on a par with, or even slightly better than, the hazard ratio format. This supports the results found in previous research focussed on specific organ-focussed effective ages (Lopez-Gonzalez et al., 2015; Parkes et al., 2008). Such differences may occur because although one can use simple addition or subtraction to calculate one's personal effective age from change information, this requires effort on the part of the listener/reader and is open to misinterpretation; for instance, misinterpreting the gain in time (i.e. losing 2 years off one's effective age) as a loss (e.g. misinterpreting as 2 years off one's life expectancy). In contrast, the personal or age-matched formats remove both effort and potential misinterpretation by completing the calculation, and potentially represents a more readily understood way to present lifestyle-related information to the public or to patients.

Further, analysis of concept understanding revealed that, despite generally high subjective understanding ratings, over a third of participants clearly misunderstood the hazard ratio concept (compared to only 10% for effective age). This further suggests that the personal effective age format may be a better choice than the hazard ratio and suggests that confidence in understanding the hazard ratio concept may be misplaced. As predicted, participants often misunderstood the percentages as an absolute percentage increases rather than relative (to a comparison group), consequently believing that the percentage increase of 110% was impossible or meant death.

Our results confirm that mistaking relative risk percentages for absolute risk percentages – as highlighted by the newspaper headline "10% of all deaths could be avoided" (Spiegelhalter, 2012b; Willey, 2012) – represents a seductive misunderstanding of hazard ratios. For improving lifestyle behaviours, the consequences of this misunderstanding may be considered small and even, positive, given that overestimation "nudges" people towards healthier options (though whether it is ethical to "nudge" in this way, is controversial; Rakow, Heard & Newell, 2015). In other cases, the consequence of such misunderstandings can have serious negative consequences. For example, as Gigerenzer (2009) describes, the publication of evidence in 1995 suggesting that taking the oral contraceptive pill increased the risk of thrombosis twofold (thus, by 100%) caused panic and anxiety, leading many women to stop using the pill, despite the absolute risk increase being small: only 1 in 7000.

Our results add to such findings by Gigerenzer and others because, not only do a proportion of people misunderstand the format (as has been identified previously), but our findings show that these individuals appear unaware of this misunderstanding. Indeed, consistent with such unawareness – or even perhaps misplaced high confidence – there was no clear relationship between self-rated understandings and objective understanding scores. Because confident individuals may be less likely to seek to better understand the format or check their current understanding is correct, such misunderstandings are likely to be perpetuated and false confidence strengthened. Considering the prevalence of the hazard ratio percentage in public news media, this makes the implications of this finding particularly worrisome.

The speed-of-ageing metaphors, particularly the personalised effective age metaphor, do not appear to suffer with such problems. As our results show, both perceived and tested understanding is higher for this calculated effective age, with fewer than 10% misunderstanding this concept. Rather, the majority of participants show a good-to-excellent level of understanding. Such small misunderstandings as do occur represent cases where "incorrect" inferences could be considered reasonable. For example, talking about functioning in explaining speed-of-ageing because factors which increase risk of death may co-occur with factors that underlie ill health. Overall, such results suggest that the personalised effective age may represent a better alternative to the traditional hazard ratio percentage which is prone to serious misunderstandings.

The lifestyle risks and benefits used as our behaviours of interest represent just one context where such findings on the speed-of-ageing metaphors and misunderstanding of the hazard ratio concept may be applied. Not only are such results applicable for other lifestyle behaviours (e.g. smoking, alcohol consumption), such findings are potentially generalizable to discussions of risks/benefits of long-term medications (e.g. for chronic conditions), such as statins for high cholesterol or anticoagulants for deep vein thrombosis. This second area (long-term medications) is a potentially important application because millions of people follow such treatment regimes (e.g. an estimated 38.6 million Americans are on statins; Adedinsewo et al., 2016) and therefore, if even a small proportion stop taking treatment because they misunderstand its benefits or risks, the impact could be substantial. Some consequences of this problem were described by Matthews et al. (2016): intense media coverage of the controversy over the risk benefit balance for statins led those patients already taking statins to be more likely to stop taking them after media coverage. It has been estimated that this could result in more than 2000 extra cardiovascular events across the UK over a ten-year period.

The use of these formats in a range of domains warrants further attention, particularly to determine whether there are particular domains where one metaphor is best or where, as highlighted above, differences between metaphors are amplified. For example, future research could assess whether differences between change and personal formats of the effective age will always appear, or may be situation specific. In domains where change-in-years is already common or intuitive such a disadvantage to the change format may be reduced. For example, within the educational domain (e.g. discussing children's reading abilities), it is common to think of improvements as related to age, perhaps because stages of schooling are often closely tied to chronological age.

Conclusion

The main goal of this paper was to introduce speed-of-ageing formats for communicating mortality risks, and assess them by looking at a range of important perception and comprehension measures. Our results have revealed that, generally, perceptions across the three speed-of-ageing metaphor types (effective age, day format and over-life format) are similar. However, the specific form of the effective age format does seem to make a difference, with understanding rated lower for the change-in-age format, while the personalised and age-matched formats – which each specify an actual age – are rated much easier to understand. Finally, from Study 2, our evidence suggests a particular benefit to this personalised age format over the traditional hazard ratio format, where people display poorer understanding. It is our hope that these results will prompt further research investigating these formats across other health information domains (e.g. medication risks and benefits) and other communication domains (e.g. education).

References

Adedinsewo, D., Taka, N., Agasthi, P., Sachdeva, R., Rust, G., & Onwauanyi, A. (2016). Prevalence and factors associated with statin use among a nationally representative sample of US Adults: national health and nutrition examination survey, 2011-2012. *Clinical Cardiology*, *39(9)*, 491-496.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: the Berlin numeracy test. *Judgment and Decision Making, 7*(1), 25.

Fagerlin, A., Zikmund-Fisher, B.J., Ubel, P.A., Jankovic, A., Derry, H.A., & Smith, D.M. (2007). Measuring numeracy without a math test: Development of the Subjective Numeracy Scale (SNS). *Medical Decision Making*, *27*, 672-680, doi: 10.1177/0272989X07304449.

Forsythe Appraisals (2015). What is effective age? *Forsythe Appraisals, LLC Blog*, Retrieved on 10 August 2016 from: http://www.forsytheappraisals.com/2015/july/285-what-is-effective-age

Gibbs, N. (2015). Reading: the next steps. Supporting higher standards in schools. *Department of Education*, Retrieved on 10th August 2016 from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/409409/Reading_the_next_steps.pdf

Gigerenzer, G. (2009). Making sense of health statistics. *Bulletin of the World Health Organization, 87(8)*, 567. doi: 10.2471/BLT.09.069872

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L.M., & Woloshin, S. (2008). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, 8(2)*, 53-96.

Joireman, J., Shaffer, M., Balliet, D., & Strathman, A. (2012). Promotion orientation explains why future oriented people exercise and eat healthy: Evidence from the two-factor consideration of future consequences 14 scale. *Personality and Social Psychology Bulletin, 38,* 1272-1287, doi: 10.1177/0146167212449362.

Knapp, P., Gardner, P. H., Carrigan, N., Raynor, D. K. and Woolf, E. (2009). Perceived risk of medicine side effects in users of a patient information website: A study of the use of verbal descriptors, percentages and natural frequencies. *British Journal of Health Psychology, 14(3),* 579–594. doi: 10.1348/135910708X375344.

Liberman, N., & Trope, Y. (2003). Construal level theory of intertemporal judgement and decision. In Loewenstein, G., Read, D., & Baumeister, R. (Eds.) *Time & Decision: Economic and Psychological Perspectives on Intertemporal Choice* (pp. 245-276). New York: Russell Sage Foundation

Lopez-Gonzalez, A. A., Aguilo, A., Frontera, M., Bennasar-Veny, M., Campos, I., & Vicente-Herrero, T., Tomas-Salva, M., de Pedro-Gomez, J., & Tauler, P. (2014). Effectiveness of the Heart Age tool for improving modifiable cardiovascular risk factors in a Southern European population: a randomized trial. *European Journal Preventative Cardiology, 22,* doi:10.1177/2047487313518479

Madden, T.J., Ellen, P.S., & Ajzen, I. (1992). A comparison of the theory of planned behaviour and the theory of reasoned action. *Personality and Social Psychology Bulletin, 18(1),* 3-9.

Matthews, A., Herrett, E., Gasparrini, A., Van Staa, T., Goldacre, B., Smeeth, L., & Bhaskaran, K. (2016). Impact of statin related media coverage on use of statins:

interrupted time series analysis with UK primary care data. *BMJ*, *353*. doi:10.1136/bmj.i3283

Loewenstein, G., Read, D., & Baumeister, R. (2003.) *Time & Decision: Economic and Psychological Perspectives on Intertemporal Choice*. New York: Russell Sage Foundation

Parkes, G., Greenhalgh, T., Griffin, M., & Dent, R. (2008). Effect on smoking quit rate of telling patients their lung age: the step2quit randomised controlled trial. *BMJ, 336*(7644), 598-600. doi:10.1136/bmj.39503.582396.25

Patak, M., & Reynolds, B. (2007). Question-based assessments of delay discounting: do respondents spontaneously incorporate uncertainty into their valuations for delayed rewards? *Additive Behaviours*, *32(2)*, 351-357. doi:10.1016/j.addbeh.2006.03.034.

Peters, E. (2008). Numeracy and the perception and communication of risk. *Annals of the New York Academy of Sciences. 1128*, 1-7. doi: 10.1196/annals.1399.001.

Rakow, T., Heard, C. L., & Newell, B. R. (2015). Meeting Three Challenges in Risk Communication: Phenomena, Numbers, and Emotions. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 147-156. doi: 10.1177/2372732215601442

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black, & Prokasy, W. F. (Eds.), Classical conditioning II: Current research and theory (pp. 64-99). New York: Appleton-Century-Crofts.

Reyna, V.F., Nelson, W.L., Han, P.K., & Dieckmann, N.F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135(6)*, 943-973. doi: 10.1037/a0017327

Slovic, P. (1987). Perception of Risk. *Science*, *236(4799)*, 280-285. doi: 10.1126/science.3563507

Spiegelhalter, D. (2011). Microlives. *Understanding Uncertainty,* Retrieved on 22 October 2014 from: http://understandinguncertainty.org/microlives

Spiegelhalter, D. (2012a). Using speed of ageing and "microlives" to communicate the effects of lifetime habits and environment. *BMJ, 345*. doi:10.1136/bmj.e8223

Spiegelhalter, D. (2012b). What does a 13% increased risk of death mean? *Understanding Uncertainty,* Retrieved on 8th March 2015 from http://understandinguncertainty.org/what-does-13-increased-risk-death-mean

Spiegelhalter, D. (2015). What is your 'effective age'? *Understanding Uncertainty,* Retrieved on 10 August 2016 from: https://understandinguncertainty.org/what-your-effective-age

Spiegelhalter, D. (2016). How old are you, really? Communicating chronic risk through 'effective age' of your body and organs. *BMC Medical Informatics and Decision Making, 16(*1), 1-6. doi:10.1186/s12911-016-0342-z

Stone, E.R., Choi, Y., Bruine de Bruin, W., & Mandel, D.R. (2013). I can take the risk, but you should be safe: Self-other differences in situations involving physical safety. *Judgement and Decision Making, 8(3),* 250-267.

Story, G. W., Vlaev, I., Seymour, B., Darzi, A., & Dolan, R. J. (2014). Does temporal discounting explain unhealthy behavior? A systematic review and reinforcement learning perspective. *Frontiers in Behavioral Neuroscience, 8,* 76. doi:10.3389/fnbeh.2014.00076

Strathman, A., Gleicher, F., Boninger, D.S., & Edwards, C.S. (1994). The consideration of future consequences; weighting immediate and distant outcomes of behaviour. *Journal of Personality and Social Psychology, 66(4)*, 742-752.

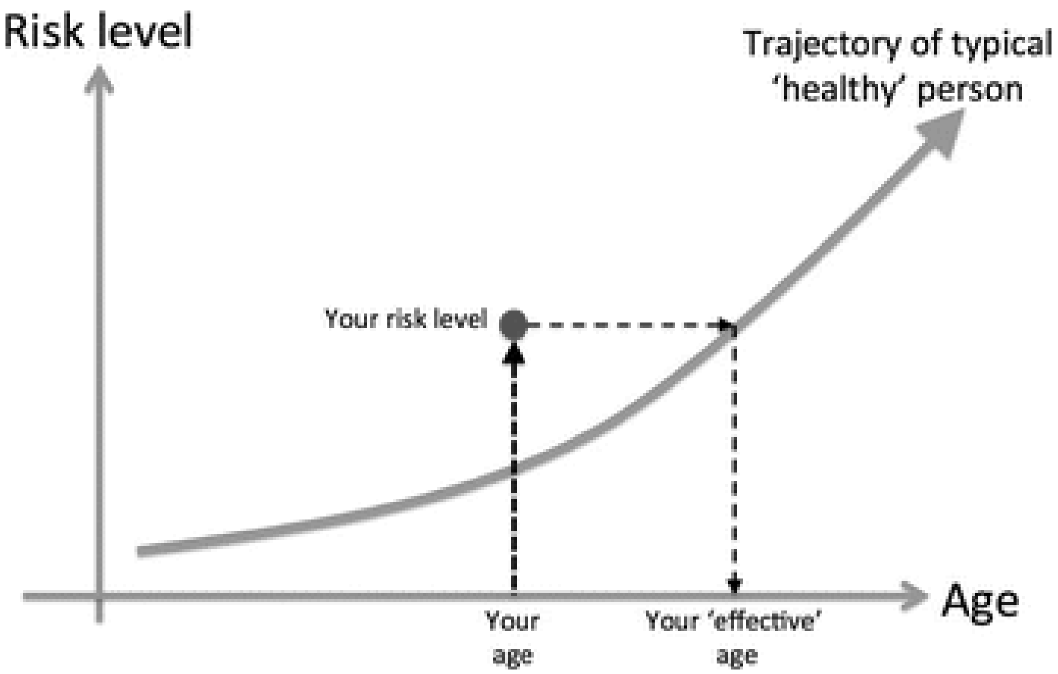
Telegraph (2015). Fizzy drinks linked to increased risk of heart failure. *The Telegraph.* Retrieved on 29th November 2016 from http://www.telegraph.co.uk/news/health/news/11971296/Fizzy-drinks-linked-to-increased-risk-of-heart-failure.html

Willey, J. (2012). Less meat, more veg is the secret for long life. *Daily Express.* Retrieved on 29th November 2016 from http://www.express.co.uk/news/uk/307781/Less-meat-more-veg-is-the-secret-for-longer-life

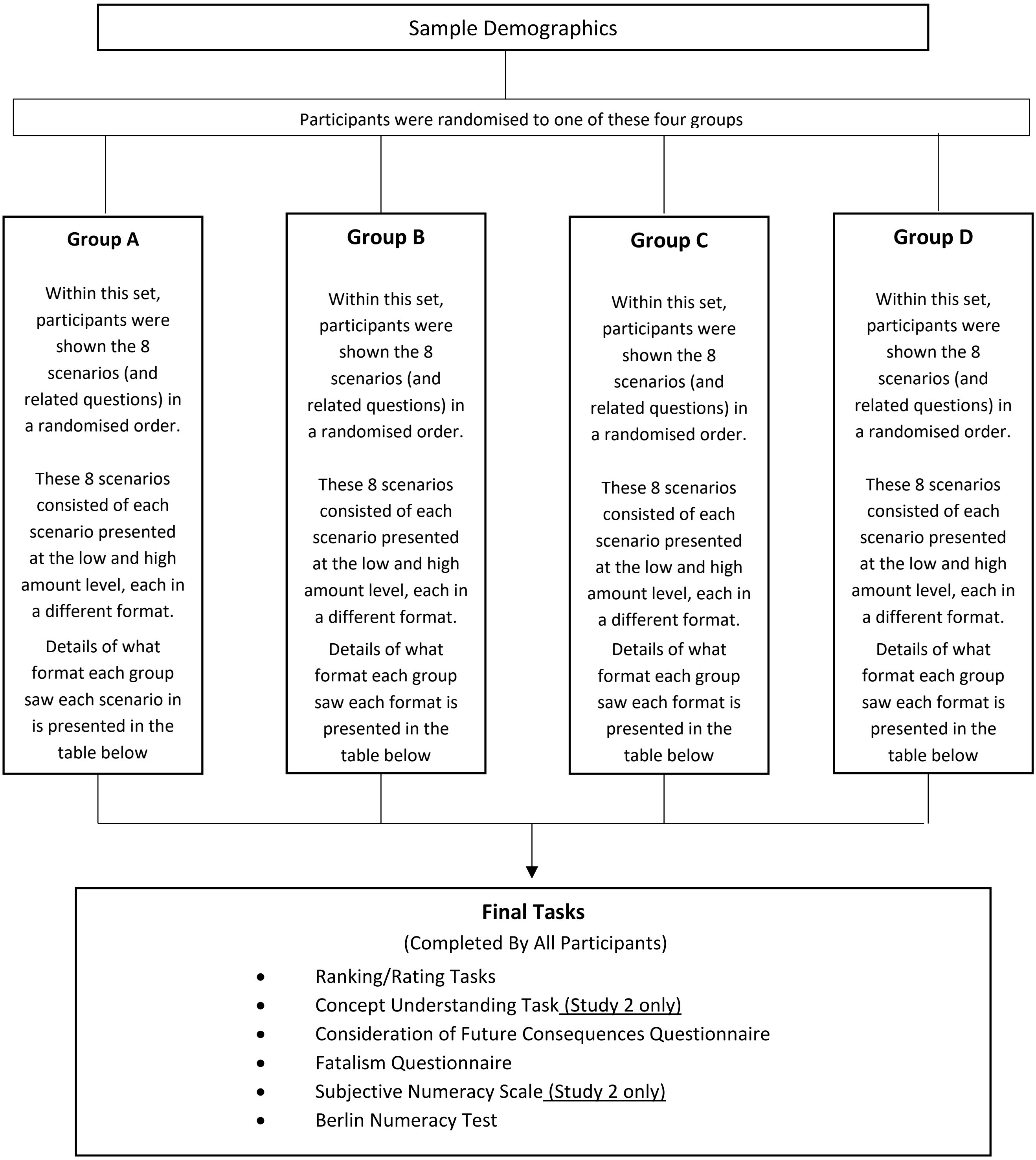
Woodcock, J., Franco, O.H., Orsini, N., Roberts, I. (2010). Non-vigorous physical activity and all-cause mortality: systematic review and meta-analysis of cohort studies. *International Journal of Epidemiology, 40,* 121-138. doi:10.1093/ije/dyq104

APPENDICES

Appendix A: Visual representation of how the concept of an effective age works from Speigelhalter (2016) (with author's permission)



Appendix B: Experimental Procedure and Latin Square Design Details:



Latin-Square Group Details:

STUDY 1	**Meat Scenario**	**TV Scenario**	**Fruit Scenario**	**Exercise**
Group A	**Hazard Ratio (%)**	**Over-life**	**Day**	**EA Change**
Group B	**EA (Change)**	**Hazard Ratio (%)**	**Over-life**	**Day**
Group C	**Over-life**	**Day**	**EA (Change)**	**Hazard Ratio (%)**
Group D	**Day**	**EA Change**	**Hazard Ratio (%)**	**Over-life**
STUDY 2	**Meat Scenario**	**TV Scenario**	**Fruit Scenario**	**Exercise**
Group A	**Hazard Ratio (%)**	**EA Age-matched**	**EA Change**	**EA Personal**
Group B	**EA Change**	**EA Personal**	**EA Age-matched**	**Hazard Ratio (%)**
Group C	**EA Age-Matched**	**Hazard Ratio (%)**	**EA Personal**	**EA Change**
Group D	**EA Personal**	**EA Change**	**Hazard Ratio (%)**	**EA Age-Matched**

EA=Effective Age

Appendix C: Worked example of each format for 50g of processed meat.

Below is an example of one of the scenarios presented. In this case, this scenario is about eating processed meat (at a low amount level) and is presented in the hazard ratio (percentage) format.

Other behaviours presented to participants were: Eating Processed Meat (100g), TV Watching (2 hours, 4 hours), Eating Fruit and Vegetables (1 portion, 5 portions) and Moderate Exercise (20 minutes, 1 hour).

All formats are illustrated below (for the 50g of processed meat scenario). Where an X is placed in the format, this replaces the code which inputted participant's age in the statement.

Hazard ratio (percentage) format:

Imagine two people who are identical in every way, except that one eats around 50g of processed meat a day and the other eats none. [50g of processed meat=1 large sausage or 3 rashers of bacon]

For the person eating **50g of processed meat a day,** this is expected to **increase their annual chance of death by 18%.**

Life expectancy (over-life format):

Imagine two people who are identical in every way, except that one eats around 50g of processed meat a day and the other eats none. [50g of processed meat=1 large sausage or 3 rashers of bacon]

For the person eating **50g of processed meat a day,** this is expected to **take around 2 years off the length of their life.**

Day format:

Imagine two people who are identical in every way, except that one eats around 50g of processed meat a day and the other eats none. [50g of processed meat=1 large sausage or 3 rashers of bacon]

For the person eating **50g of processed meat a day,** this is expected to **take around 1 hour off their life each day.**

Effective age (change) format:

Imagine two people who are identical in every way, except that one eats around 50g of processed meat a day and the other eats none. [50g of processed meat=1 large sausage or 3 rashers of bacon]

For the person eating **50g of processed meat a day,** this is expected to **add around 2 years onto their "effective age" (i.e. give them the annual chance of death of someone 2 years older).**

Effective age (age-matched) format:

Imagine two people who are identical in every way, except that one eats around 50g of processed meat a day and the other eats none. [50g of processed meat=1 large sausage or 3 rashers of bacon]

For the person eating **50g of processed meat a day, if they were your age X years old),** this is expected to **increase their "effective age" (their body's working age) to X+2 years old.**

Effective age (personal) format:

Imagine two people who are identical in every way, except that one eats around 50g of processed meat a day and the other eats none. [50g of processed meat=1 large sausage or 3 rashers of bacon]

If you were the person eating **50g of processed meat a day, with your age of X years old,** this is expected to **increase your "effective age" (your body's working age) to around X+2 years old.**

Figure Captions:

Figure Number	Caption Text
Figure 1	Figure 1: Descriptive and inferential statistics for each analysis (study 1)
Figure 2	Figure 2: Descriptive and inferential statistics for each analysis (study 2)
Figure 3	Figure 3: Number of participants rated at each level of understanding.

Separate editable documents for each figure are attached to this submission.

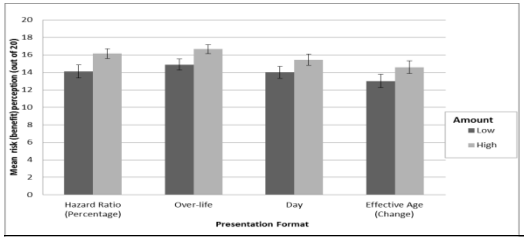
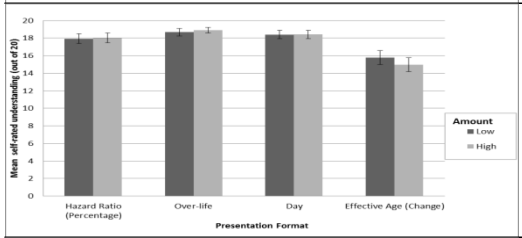
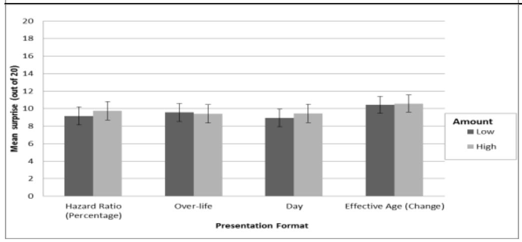
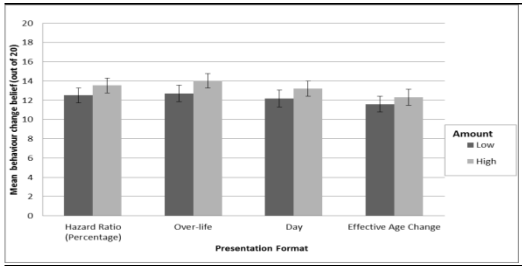
Mean ratings in each of the formats for the low and high amount levels.	**Analysis**	**Main Effect Amount**			**Main Effect Format**			**Interaction**		
	N=160	*F*	*p*	η^2	*F*	*p*	η^2	*F*	*p*	η^2
	**Risk(Benefit) Perception**	**158.33**	**<.001**	**.499**	**8.66**	**<.001***	**.052**	0.87	.459	.005
	**Understanding**	0.12	.730	.001	**49.93**	**<.001***	**.239**	.275	810*	.002
	**Surprise**	1.29	.257	.008	**2.66**	**.047**	**.016**	0.56	.641	.004
	**Behaviour Change Belief**	**32.86**	**<.001**	**.171**	**4.90**	**.003***	**.030**	0.51	.673	.003

Figure 1: Descriptive and inferential statistics for each analysis (Study 1)

Note: Greenhouse-Geisser corrected p-values are marked with ‘*’ in the tables/figure. Values in bold on the figure denote significant effects.

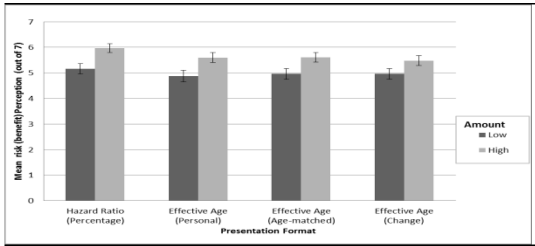
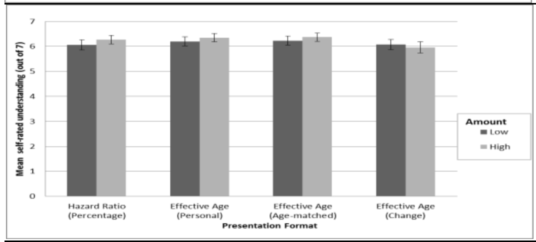
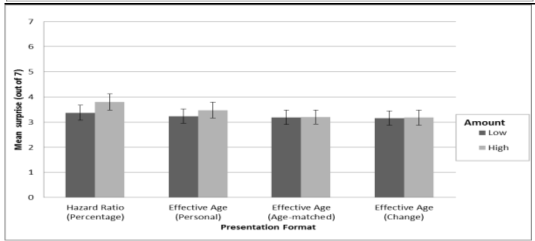
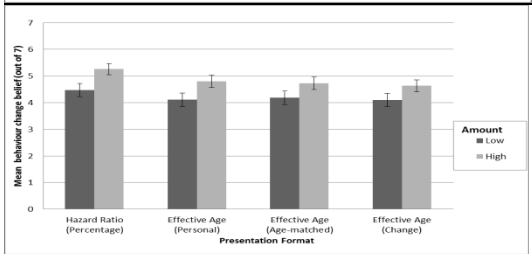
Mean ratings in each of the formats for the low and high amount levels.	**Analysis**	**Main Effect Amount**			**Main Effect Format**			**Interaction**		
	N=171	*F*	*p*	η^2	*F*	*p*	η^2	*F*	*p*	η^2
	**Risk(Benefit) Perception**	**245.42**	**<.001**	**.591**	**3.85**	**.017***	**.022**	**3.00**	**.033***	**.017**
	**Understanding**	**5.901**	**.016**	**.034**	**4.847**	**.005***	**.028**	**3.770**	**.013***	**.022**
	**Surprise**	**7.12**	**.008**	**.040**	**3.45**	**.019***	**.020**	2.21	086	.013
	**Behaviour Change Belief**	**108.53**	**<.001**	**.390**	**8.93**	**<.001**	**.050**	1.375	.249	.008

Figure 2: Descriptive and inferential statistics for each analysis (Study 2)

Note: Greenhouse-Geisser corrected p-values marked with '*' in the tables/figure. Values in bold on the figure denote significant effects.

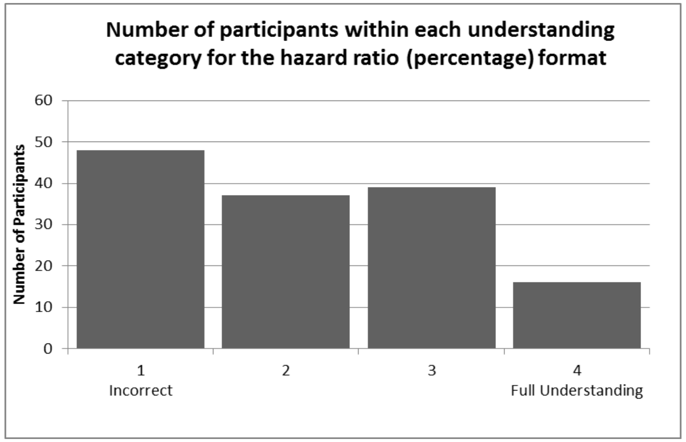


Figure 3a

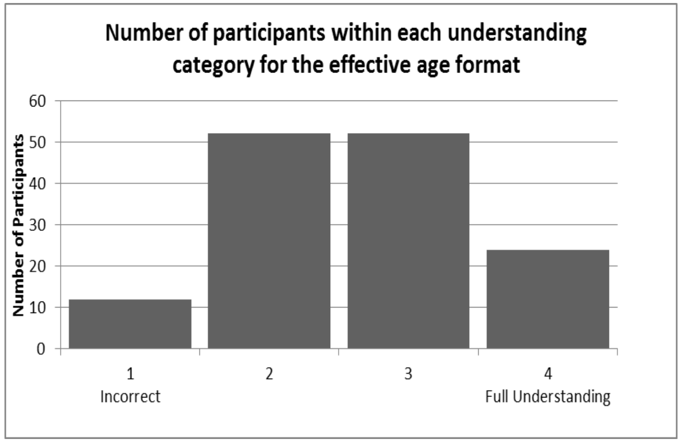


Figure 3b

Figure 3: Number of participants rated at each level of understanding.