


An Exploratory Application of Eye-Tracking Methods in a Discrete Choice Experiment

Medical Decision Making
2018, Vol. 38(6) 658–672
© The Author(s) 2018



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0272989X18782197
journals.sagepub.com/home/mdm



Caroline Vass , Dan Rigby, Kelly Tate, Andrew Stewart, and Katherine Payne

Abstract

Background. Discrete choice experiments (DCEs) are increasingly used to elicit preferences for benefit-risk tradeoffs. The primary aim of this study was to explore how eye-tracking methods can be used to understand DCE respondents' decision-making strategies. A secondary aim was to explore if the presentation and communication of risk affected respondents' choices. **Method.** Two versions of a DCE were designed to understand the preferences of female members of the public for breast screening that varied in how risk attributes were presented. Risk was communicated as either 1) percentages or 2) icon arrays and percentages. Eye-tracking equipment recorded eye movements 1000 times a second. A debriefing survey collected sociodemographics and self-reported attribute nonattendance (ANA) data. A heteroskedastic conditional logit model analyzed DCE data. Eye-tracking data on pupil size, direction of motion, and total visual attention (dwell time) to predefined areas of interest were analyzed using ordinary least squares regressions. **Results.** Forty women completed the DCE with eye-tracking. There was no statistically significant difference in attention (fixations) to attributes between the risk communication formats. Respondents completing either version of the DCE with the alternatives presented in columns made more horizontal (left-right) saccades than vertical (up-down). Eye-tracking data confirmed self-reported ANA to the risk attributes with a 40% reduction in mean dwell time to the "probability of detecting a cancer" ($P = 0.001$) and a 25% reduction to the "risk of unnecessary follow-up" ($P = 0.008$). **Conclusion.** This study is one of the first to show how eye-tracking can be used to understand responses to a health care DCE and highlighted the potential impact of risk communication on respondents' decision-making strategies. The results suggested self-reported ANA to cost attributes may not be reliable.

Keywords

breast screening, discrete choice experiment, eye tracking, risk, stated preference

Discrete choice experiments (DCEs) are a commonly used survey-based method to quantify preferences for the characteristics (termed *attributes*) of health care interventions.^{1–3} DCEs are underpinned by 2 key economic theories: random utility theory (RUT) and Lancaster's theory of consumer demand.^{4,5} In a DCE, respondents are asked to choose their preferred option from a series of hypothetical scenarios (choice tasks) defined by attributes and levels combined according to a mathematical design. Using respondents' stated choices, it is possible to infer how changes in the attributes' levels affect satisfaction or "utility" derived from an option and the probability of an option being chosen.

Systematic reviews have shown DCEs in the health care context are increasingly used to elicit preferences for benefits and risks.^{6,7} The use of DCEs to inform regulatory decision making by identifying the levels of risk consumers of health care interventions will tolerate for an associated benefit has also been mooted.^{8–10} In these contexts, risk and uncertain outcomes often feature in

Corresponding Author

Katherine Payne, Manchester Centre for Health Economics, University of Manchester, Oxford Road, Manchester M13 9PL, UK; Phone: +44 (161) 306-7906; Fax: +44 (161) 275-5205 (Katherine.payne@manchester.ac.uk).

DCE designs, yet numerical probabilistic information is a notoriously difficult concept to present and communicate clearly.^{6,11,12}

In health risk communication literature, it has been suggested that pictorial presentations of probabilistic information may help individuals' understanding.^{13–15} In contrast, health care DCEs most commonly present risk information using percentages.⁶ If DCE respondents do not understand the choice task, they may employ simplifying heuristics such as ignoring confusing attributes, violating the axiom of continuity in preferences, resulting in biased preference estimates.

Positivist qualitative research methods, such as “think-aloud,” have been used to illuminate the underlying decision strategies employed and understand whether respondents answer consistently in line with a priori expectations underpinning the DCE.¹⁶ Think-aloud is a type of cognitive interview^{17,18} purposefully developed to understand approaches to problem solving.^{19,20} Such qualitative research methods have been criticized for their susceptibility to researcher biases.²¹ It has also been highlighted in concurrent think-aloud that the burden of speaking while simultaneously completing a task may change how the task is completed.²² If the act of thinking aloud disturbs natural decision-making behavior, generated data may be of limited use and not reflect the decision-making processes under examination. Alternative retrospective think-aloud relieves the task burden but remains susceptible to other biases associated with recalling information.²³

As an alternative to cognitive interviews, psychologists have studied eye movements to understand how visual information is being processed.²⁴ In its most basic form, this has involved researcher-individual observation of participants' eyes and manual notes on pupil dilation.²⁵

More sophisticated methods have developed in line with changes to, and availability of, technology. The “eye-mind hypothesis,” proposed by Just and Carpenter,²⁶ underpins most psychological analyses of eye-tracking data. The logic of using eye-tracking data is based on the assumption that where people look is indicative of their cognitive processing. Just and Carpenter state, “There is no appreciable lag between what is fixated and what is processed.”^{26(p331)} A substantial body of research supports this proposal (see Rayner's comprehensive review²⁷). Therefore, attention, measured in “fixations” (where an individual's attention falls), can be thought of as quantification of individuals' information processing, although it is still possible that what is processed may still be ignored.

This study aimed to explore eye-tracking as a method to understand how individuals attend and process information in a DCE. A secondary aim was to investigate whether alternative approaches to presenting and communicating risk affected attention and processing strategies to understand whether pictorial presentation reduced cognitive burden or the need for simplifying heuristics.

Methods

Eye-tracking was used to investigate how respondents complete a selected example of a DCE. This study used a DCE survey designed to elicit preferences from female members of the public for a breast screening program described by 3 attributes (probability of detecting a cancer, risk of unnecessary follow-up, and out-of-pocket screening costs). Attribute labels, definitions, and levels are presented in Table 1.

This survey has been published in detail elsewhere.²⁸ In brief, respondents chose between 2 hypothetical breast screening programs and a “no screening” opt-out. Estimates of breast screening in the United Kingdom suggest uptake to the national program is approximately 75%,²⁹ implying some women who are invited do not attend. The option of “no screening” allowed women to express this preference. The levels for the “no screening” alternative were written with text to avoid confusion with icons that could represent both people screened or people not screened. Participants completed 11 choice tasks generated by Ngene³⁰ with a design minimizing the D-error with priors obtained from a pilot study. Survey respondents were randomized to 1 of 2 surveys, presenting risk as either percentages only (“percentages”) or percentages with icon arrays (“icon arrays”).

Manchester Centre for Health Economics, University of Manchester, Manchester, UK (CV, KP); Department of Economics, University of Manchester, Manchester, UK (DR); Division of Neuroscience and Experimental Psychology, School of Biological Sciences, University of Manchester, Manchester, UK (KT, AS). The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. Caroline Vass was in receipt of a National Institute for Health Research (NIHR) School for Primary Care Research (SPCR) PhD Studentship between October 2011 and 2014. The views expressed in this article are those of the authors and not of the funding body. Preparation of this manuscript was made possible by Mind the Risk, a project funded by the Riksbankens Jubileumsfond. Financial support for this study was provided by an NIHR SPCR PhD Studentship for Caroline Vass and a grant from the Riksbankens Jubileumsfond. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Table 1 Attributes and Levels Used in the Discrete Choice Experiment

Label	Attribute	Definition	Levels for Programs	Levels for Opt-Out of "No Screening"
Detect	Probability of detecting a cancer	The chance of detecting a cancer from screening over a 20-year period	3%; 7%; 10%; 14%	None: no cancers detected (0%)
Risk	Risk of unnecessary follow-up	The probability of being recalled for a procedure or procedures when no harm existed	0%; 1%; 5%; 10%; 20%	None: no unnecessary follow-ups (0%)
Cost	Out-of-pocket cost of screening over a lifetime	The costs of attending the program, including original screens and recalls. These could include transport, time off work, and carer costs.	£100 (£20 per screen); £250 (£50 per screen); £750 (£150 per screen); £1,000 (£200 per screen)	No cost to you (£0)

Eye-Tracking Metrics

Eye movement measures include those associated with fixations (where the eye is fixating on a particular aspect of the visual input) and those associated with saccades (eye movements between fixations).³¹ Saccadic patterns in eye-tracking data have been used to explain visual responsiveness, such as how quickly an individual can locate a target.³² In the context of choices, saccades have been used to understand how individuals seek information to make a decision, with research suggesting vertical movements (when alternatives are presented in columns) align with expected utility theory (EUT).³³

Pupillometry (measuring pupil size) can also reveal cognitive load.³⁴ When individuals think hard about a difficult task or one requiring significant memory load, the pupils dilate.³⁵ The exact mechanism by which this occurs is not entirely known, but pupil size is regulated (like heart rate, perspiration, and goosebumps) by the sympathetic division of the autonomic nervous system, which stimulates the body's response to stress (either fight or flight).^{36,37} There are now many data-driven studies that empirically demonstrate a relationship between task complexity and pupil dilation.^{38–40} As early as the 1960s, Hess and Polt³⁷ used a mirror and a camera to monitor academics' pupil sizes as they solved multiplication problems, noting a correlation between dilation and question complexity.

Study Population and Study Sample

The relevant study population was women who were, or would be, eligible for the national breast screening program. Although breast cancer does occur in men, national screening programs are targeted at women

because of the considerably higher prevalence.⁴¹ The sample was limited to women fluent in English and aged between 18 and 70 years (the current cutoff for routine screening in England).⁴² There were no other exclusions. The study took place in a laboratory setting requiring travel to the university's campus, and women were recruited through local advertising on public noticeboards (post offices, coffee shops) and the university's public website. A pilot study informed the experimental setup using eye-tracking conducted with employees of the University of Manchester ($n = 15$). The main exploratory study aimed for a sample size of 40 women on the basis of a similar study,⁴³ which also used a sample of 40, and feasibility of data collection.

Eye-Tracking Apparatus

A desktop-mounted EyeLink 1000 was used.⁴⁴ The eye-tracking device calculates the participant's gaze position using a camera to detect the corneal reflection from an infrared illuminator. The device recorded the eye position a thousand times a second, every millisecond. The head rest was positioned 43 cm from the screen, as per the manufacturer's recommendations, and this distance was remeasured for every participant. While the machine had a capacity for binocular recording, monocular recording of the dominant eye, with the best calibration, was conducted. The laboratory was a dark, windowless room with minimal luminosity. Choices in the DCE survey were made via a handheld games controller.

The survey containing the DCE was programmed for the eye tracker with assistance from EyeLink Experiment Builder software.⁴⁴ The survey included 3 warm-up questions, allowing participants to become familiar with the

handheld controller buttons. All choice sets and images of the DCE in both risk formats were identically sized to the nearest pixel.

Data Collection

Ethical research approval was obtained from the University of Manchester's Research Ethics Committee (AJ/ethics/1809/13/ref13178). Following consent, women were randomly (50:50) assigned to receive risk attributes framed either as percentages only or with an icon array. The women were asked to place their chin on the head rest, make themselves comfortable, and refrain from speaking. Calibration then began. If the calibration was "good" (corneal reflection was consistently recorded), then it was validated through additional calibration, ensuring all screen corners were recordable before the survey began. If either calibration or validation failed, the procedure restarted using the other eye. In the event that neither eye could be calibrated, the experiment ended.

When respondents were completing the DCE, a between-choice set calibration occurred called "drift correction." This procedure corrected for any movement of the participant's head, improving the accuracy of the collected data.

Background questions, including self-reported measures of attribute attendance, were completed by the participant on an iPad after participants had finished the DCE with the eye tracker (see online Appendix A).

Analysis

Analysis, in this exploratory study, was descriptive and focused on the recorded measures that may reveal information about individuals' decision-making processes when completing a discrete choice set. Analysis was conducted to explore the effect of risk communication format on choices made in the DCE, fixations to predefined areas of interest, saccades, and pupil size (as a proxy for task difficulty). All analyses were conducted with STATA software.⁴⁵

The Choice Data

The choice data from the DCE were analyzed using a heteroskedastic conditional logit model based on this utility function:

$$U_{nj} = \beta_{none} + \lambda_n \beta_1 Detect_{nj} + \lambda_n \beta_2 Risk_{nj} + \lambda_n \beta_3 Cost_{nj} + \lambda_n \beta_4 Detect_{nj} ICON_n + \lambda_n \beta_5 Risk_{nj} ICON_n + \varepsilon_{nj}, \quad (1)$$

where U represents individual n 's indirect utility from alternative j . β_{none} is an alternative-specific constant (ASC) term for the "no screening" option capturing differences in the mean of the distribution of the unobserved effects in the random component, ε_{nj} , between the opt-out and the other alternatives. $ICON_n$ is a dummy variable that is set equal to 1 if individual n received icon arrays in the DCE. λ is a scale parameter, inversely proportional to the variance of the error process, σ_ε^2 :

$$\lambda_n = \frac{\pi}{\sqrt{6\sigma_\varepsilon^2}}. \quad (2)$$

Icon arrays have been proposed as a means of reducing the cognitive load associated with understanding probabilities and hence improving choice consistency and reducing error variance.^{46,47} To investigate this, the scale parameter was modeled as a function of the $ICON$ dummy variable, allowing error variance to systematically vary by risk communication format:

$$\lambda_n = \exp(\gamma ICON_n). \quad (3)$$

Areas of Interest

Eye-tracking data provide highly detailed records of all locations the user has looked at, and reducing these data to a level that can be easily analyzed, taking into account computational limitations, is challenging. Data coordinates were segmented into defined areas of interest (AOIs), reducing the data file size to a more manageable level.⁴⁸ AOI quadrants were defined for each choice set, based on the sections of the task that might be stimulating for the participant, for example, attribute titles, levels presented, and response options. AOIs for attribute levels were of particular interest, which differed depending on the DCE version received (percentages only or icon arrays). Areas outside the AOIs were used to measure the amount of gaze in the "white space," which would suggest "daydreaming" and perhaps disinterest.

Fixations

Fixations were measured in milliseconds and conservatively defined as less than 1 degree of movement for 75 ms, in line with a previous study.⁴³ If a fixation was under this threshold, and another fixation occurred within 1 degree of the original fixation, then the fixations were merged. Merging adjacent fixations allowed identification of fixations that may have been missed due to measurement errors.

Collecting fixation data allowed analysis of information processing via respondents' attention to the choice task. Fixation data were analyzed in terms of 1) number of fixations, 2) average length of a fixation, and 3) total dwell time (sum of all fixation durations) to an AOI. Self-reported attribute nonattendance (ANA) was recorded in the iPad-based survey.

Saccades

Saccades were measured by their direction (angle of movement). The EyeLink 1000 tracker records degrees of vertical direction (between -180° and 180°), with zero degrees being a perfectly horizontal movement to the right. A rightward saccade was defined as movement between -45° and 45° , downward as movement between -135° and -45° , leftward as movement between -135° and 135° , and an upward saccade as a movement between 45° and 135° . Blinks were identified by EyeLink's in-built software⁴⁹ and differentiated from "normal" saccades by immediate loss of pupil image on the camera as the eyelid closed. These blink data were acknowledged but disregarded from the analysis.

The saccade data of interest were the number of saccades made and the direction of these saccades. No specific hypothesis existed about the number of saccades a respondent may make. There is evidence that more up-down movements could suggest the choices were made in line with EUT³³ and Lancaster's theory as a respondent would be considering the alternative as a whole.

Pupillometry

Pupil size was calculated by the eye tracker counting black pixels on the camera image of the eye to identify and measure pupil diameter. Pupil dilation was calculated as the difference between minimum and maximum pupil size, retained as an absolute measure rather than a percentage that can be inflated when baseline pupil size is small.⁵⁰

With the EyeLink 1000, pupil size data were not calibrated, and units of pupil measurement typically vary between studies. Pupil size was recorded as an integer number, based on the number of pixels but measured in arbitrary units, meaning that results cannot be compared across studies or even within studies if there was inconsistent luminosity or stimuli appeared at different locations on the screen, as these can affect pupil size measurement. However, in this experiment, the choice set stimuli, either icon arrays and/or percentages, occurred in precisely the same location and the head-mount, monitor, and camera

were identically located for each subject. In addition, all experiments took place in 1 laboratory with identical equipment setup and light sources, keeping consistent luminosity.

Pupil size data were described as 1) average pupil size per individual fixating to an attribute in a choice set and 2) average change in pupil size per individual fixating to an attribute in a choice set. Pupil size data were used as a measure of cognitive burden.

Influence of the Risk Presentation Format

Eye-tracking data were recorded for each choice set of the 2 alternative approaches to risk presentation completed by each participant. Data were analyzed using ordinary least squares (OLS) in a linear regression model that fitted the observed data by minimizing the sum of the squared residuals.⁵¹ This method was chosen to investigate the effect of the risk communication method on each of the outcomes of interest. The relevant outcomes summarized using count data for everyone in each choice set were number of fixations and number of saccades. In addition, pupil size and response duration level were summarized for each individual in each choice set. The aggregate outcome summarized was the dwell time. Direction of saccade was a binary outcome and calculated as the proportion of times at the saccade level (in a choice set) where the binary variable equaled unity.

For all outcomes, the regression specification in equation (4) was estimated separately for each of the 3 attributes in the DCE (detection, risk, and cost):

$$y_{i,c} = \alpha + \beta ICON_i + \varepsilon_{i,c}, \quad (4)$$

where $y_{i,c}$ denotes the outcome for individual i in choice set c . $ICON_i$ is a binary variable that equals zero if individual i completes the DCE where risk was framed as a percentage only and equal to unity if individual i completes the DCE where risk was formatted with an additional icon array. $\varepsilon_{i,c}$ is a zero-mean error term, assumed uncorrelated with the method of risk formatting. Clustering of standard errors at an individual level was included to allow observations by the same individual to be correlated. The regression of equation (4) produced the estimated coefficients shown in equations (5) and (6):

$$\hat{\alpha} = \bar{y}^{PO}. \quad (5)$$

$$\hat{\beta} = \bar{y}^{ICON} - \bar{y}^{PO}. \quad (6)$$

The estimated constant, $\hat{\alpha}$, is the average of individual choice set-level outcomes for individuals who completed the DCE with risk framed as a percentage only, and \bar{y}^{PO}

and $\hat{\beta}$ is the average difference in average outcomes between individuals who completed the DCE with risk attributes formatted as both a percentage and icon array (\bar{y}^{IAP}) and individuals who completed the DCE with risk communicated as a percentage only (\bar{y}^{PO}). Glass's delta effect sizes,⁵² using the standard deviation of the percentages only (control) group, were estimated because of unequal standard deviations across the 2 risk communication formats.

The results from the survey's debriefing questions showing self-reported attribute nonattendance were compared with the mean dwell time for each attribute using the regression specification in equation (7) estimated separately for the 3 attributes in the DCE (detection, risk, and cost):

$$y_{i,c} = \alpha + \beta ANA_i + \varepsilon_{i,c}. \quad (7)$$

$y_{i,c}$ denotes the total fixation duration for individual i in trial c . ANA_i is a binary variable that equals zero if individual i self-reported that he or she attended to an attribute and equals unity if individual i reported not attending to an attribute. $\varepsilon_{i,c}$ is a zero-mean error term, assumed uncorrelated with self-reported ANA. The regression of equation (7) produced the estimated coefficients of 8 and 9:

$$\hat{\alpha} = \bar{y}^A. \quad (8)$$

$$\hat{\beta} = \bar{y}^{ANA} - \bar{y}^A. \quad (9)$$

The estimated constant, $\hat{\alpha}$, is the mean dwell time per attribute per choice set for individuals who reported attending to an attribute (\bar{y}^A), and $\hat{\beta}$ is the difference in average (mean) dwell time between individuals who reported not attending to an attribute (\bar{y}^{ANA}) and individuals who reported attending to an attribute (\bar{y}^A).

Table 2 summarizes the eye-tracking data recorded for each outcome. Table 2 also presents the definition of \bar{y} and $\hat{\beta}$ for each OLS regression and the interpretation of the regression results. Glass's delta effect sizes⁵² were estimated using the standard deviation of the attending group.

Results

Forty-two women were recruited, but results from 2 women were excluded because their eye movements could not be accurately recorded due to calibration failure. Data from 40 women aged between 18 and 63 years were included in the exploratory analysis. These 40 women

were randomly assigned to a risk format within the DCE ($n = 20$ for percentages; $n = 20$ for icon arrays).

The results of the heteroskedastic conditional logit model analyzing the choice data are shown in Table 3. The results showed all estimated coefficients had the expected sign (negative coefficients on the attributes "risk of unnecessary follow-up" and "cost of screening program" and a positive coefficient on the attribute "probability of detecting a cancer"). Two of the 3 estimated coefficients were statistically significant (at $P < 0.001$), but the coefficient on the cost attribute was not statistically significant.

Figure 1 shows examples of the tracked eye movements for 2 respondents each completing 1 of the 2 versions of the DCE. Video examples of eye-tracking (collected in pretesting) can be found in online Appendices B and C. Visual examination of the scan paths of respondents indicated the eye-tracking experiment had face validity, with almost all fixations to the predefined AOI. On average, each participant made only 2 fixations (2.38 for the percentages version; 2.40 for the icon arrays version) to areas of white space (with no information) per choice set question. There was no significant difference ($P = 0.197$) between the 2 risk communication formats.

Visual Attention: Fixation and Dwell Times

Table 4 summarizes data for fixation and dwell times. The attributes, risk and detection, had the highest number of fixations, attracting almost twice as many fixations as the cost attribute. Fixations to these attributes were also longer in duration. The mean number of fixations made when looking at a choice set was not statistically significantly different between each version of the DCE (detect, $P = 0.587$; risk, $P = 0.366$; cost attribute, $P = 0.735$). There was no statistically significant difference in fixation duration between the risk communication formats to either the risk ($P = 0.716$), detect ($P = 0.243$), or cost attribute ($P = 0.674$).

When the number of fixations and the fixation duration were aggregated to measure the complete dwell time to an attribute, results were maintained. Eye-tracking participants spent a total of 1.8 seconds, 1.9 seconds, and 0.8 seconds on each attribute in the DCE (detect, risk, and cost), respectively. Between the 2 risk communication formats, there was no statistically significant difference in the time spent processing attributes (detect, $P = 0.487$; risk, $P = 0.702$; cost attribute, $P = 0.774$).

Table 2 Definitions and Interpretation of Eye-Tracking Outcomes Data

Outcome	Definition	Outcome Label	Type	Unit	\bar{y} in OLS Regression	β in OLS Regression	Interpretation
NOF	Number of separate occasions an individual fixates on information	Fixation	Count	Number	Mean NOF by an individual to an attribute in a trial	Difference in mean NOF when risk is presented using icon array compared with percentage only	More fixations mean respondent was processing more of the information
White space fixations	Number of separate occasions an individual fixates on areas containing no information	Fixation	Count	Number	Mean NOF by an individual to a white space in a trial	Difference in mean NOF when risk is presented using icon array compared with percentage only	More fixations to white space means confusion/no interest
Average fixation duration	Mean time of a fixation	Fixation	Mean	Milliseconds	Mean duration of a fixation to an attribute by an individual in a trial	Difference in mean duration when risk is presented using icon array compared with percentage only	Longer fixations mean more information processing
Total DT	Total time of all fixations	Fixation	Sum	Milliseconds	Mean DT to an attribute by an individual in a trial	Difference in mean DT when risk is presented using icon array compared with percentage only	Longer DT means more information processing
ANA dwell time	Total time of all fixations	Fixation and ANA	Sum	Milliseconds	Mean DT to an attribute by an individual in a trial	Difference in mean DT when respondents self-reported ANA compared with no self-reported ANA	Self-reported ANA should mean shorter DT as attributes ignored
NOS	Number of separate occasions an individual searches for information	Saccade	Count	Number	Mean NOS by an individual in a trial	Difference in mean NOS when risk is presented using icon array compared with percentage only	No hypothesis
Saccade direction	Direction of eye movements in information seeking	Saccade	Count	Binary count ^a	Proportion of saccades to an attribute in a trial, which are left-right	Difference in mean proportions when risk is presented using icon array compared with percentage only	More up-down movements mean more whole-good calculations
Average pupil size	Mean pupil size in a fixation	Pupil	Mean	Pixel units	Mean pupil size of an individual when fixating to an attribute in a trial	Difference in mean pupil size when risk is presented using icon array compared with percentage only	Larger pupil size indicates more cognitive burden
Pupil dilation	Change in pupil size while fixating (largest-smallest)	Pupil	Mean	Pixel units	Mean difference in pupil size of an individual when fixating to an attribute in a trial	Difference in mean pupil size change when risk is presented using icon array compared with percentage only	Larger changes in pupil size indicate more cognitive burden

ANA, attribute nonattendance; DT, dwell time; NOF, number of fixations; NOS, number of saccades.

^a0 = up-down; 1 = left-right.

Table 3 Heteroskedastic Conditional Logit Results^a

Heteroskedastic Conditional Logit Model (Linear Risk, Pooled Sample)			
	Attribute	Estimated Coefficient	Standard Errors
Utility parameters	Alternative specific constant (none)	-1.565***	0.41
	Detect ^b	0.141***	0.03
	Risk ^c	-0.076***	0.02
	Cost ^d	-0.031	0.03
	ICON*detect	1.789	4.29
	ICON*risk	-0.901	2.07
Scale parameter	ICON	-2.152	2.12

^aLog likelihood = -329.87321. $n = 42$. Observations = 1386.

^bProbability of detecting a cancer.

^cRisk of unnecessary follow-up.

^dOut-of-pocket cost of screening over a lifetime; ICON identifies respondents who received icon arrays and percentages.

* $P < 0.005$. *** $P < 0.001$.

Attribute Nonattendance

Comparing eye-tracking dwell time data with self-reporting from debriefing questions, it was found that mean dwell time was statistically significantly lower for participants reporting nonattendance to these attributes (“detect” and “risk”). The results in Table 4 show the difference was substantial in magnitude, with nonattendance to risk resulting in a 25% lower dwell time to the risk attributes ($P = 0.008$) and nonattendance to detect resulting in over 40% shorter dwell times ($P = 0.001$). There was no statistically significant difference ($P = 0.834$) between dwell times to cost between those who reported looking at the attribute compared with those who self-reported not attending to the attribute.

Respondents who received risk communicated as a percentage only seemed to give complete visual attention to the attribute “detect,” attending to this attribute in every choice set. Table 4 shows that when risk was framed using percentages only, complete nonattendance to the risk attribute occurred in 7 choice sets compared with only 1 choice set when risk was presented using icon arrays. Cost was the attribute most neglected by the respondents (24 choice sets completed with no apparent visual attention to this attribute).

Saccades

Information searching by respondents was evaluated through analysis of saccade data. These data showed respondents had more upward-downward eye movements, which has been suggested to be in line with EUT³³ and Lancaster’s theory. Table 5 shows there was no statistically significant difference ($P = 0.932$) between the 2

DCE versions in terms of the mean number of saccades per respondent in each choice set (around 48 movements). Respondents completing the DCE in either version made more horizontal (left-right) saccades than vertical (up-down).

Pupil Dilation

Data reporting the mean degree of pupil dilation when completing the DCE and for each of the individual attributes are presented in Table 6. The number of observations varied, reflecting that some attributes had incomplete visual ANA. For each of the 3 attributes, the mean pupil size was smaller for respondents completing the DCE with icon arrays ($P = 0.701$). None of these differences were statistically significant.

Discussion

This exploratory study suggested that eye-tracking could be used as a method to identify how respondents complete a DCE and reveal information on their choice strategies. The eye-tracking experiment appeared to be well calibrated with very few fixations to white-space areas on survey pages. This finding also suggests respondents engaged in the task rather than “daydreaming.” The study also explored if the type of risk presentation format affected DCE completion. Descriptive data on the number of fixations indicated the risk format had no effect on participants’ visual attention to relevant attributes. Results indicated there was no difference in information processing. Attributes “risk of unnecessary follow-up” and “probability of detecting a cancer” both attracted more attention

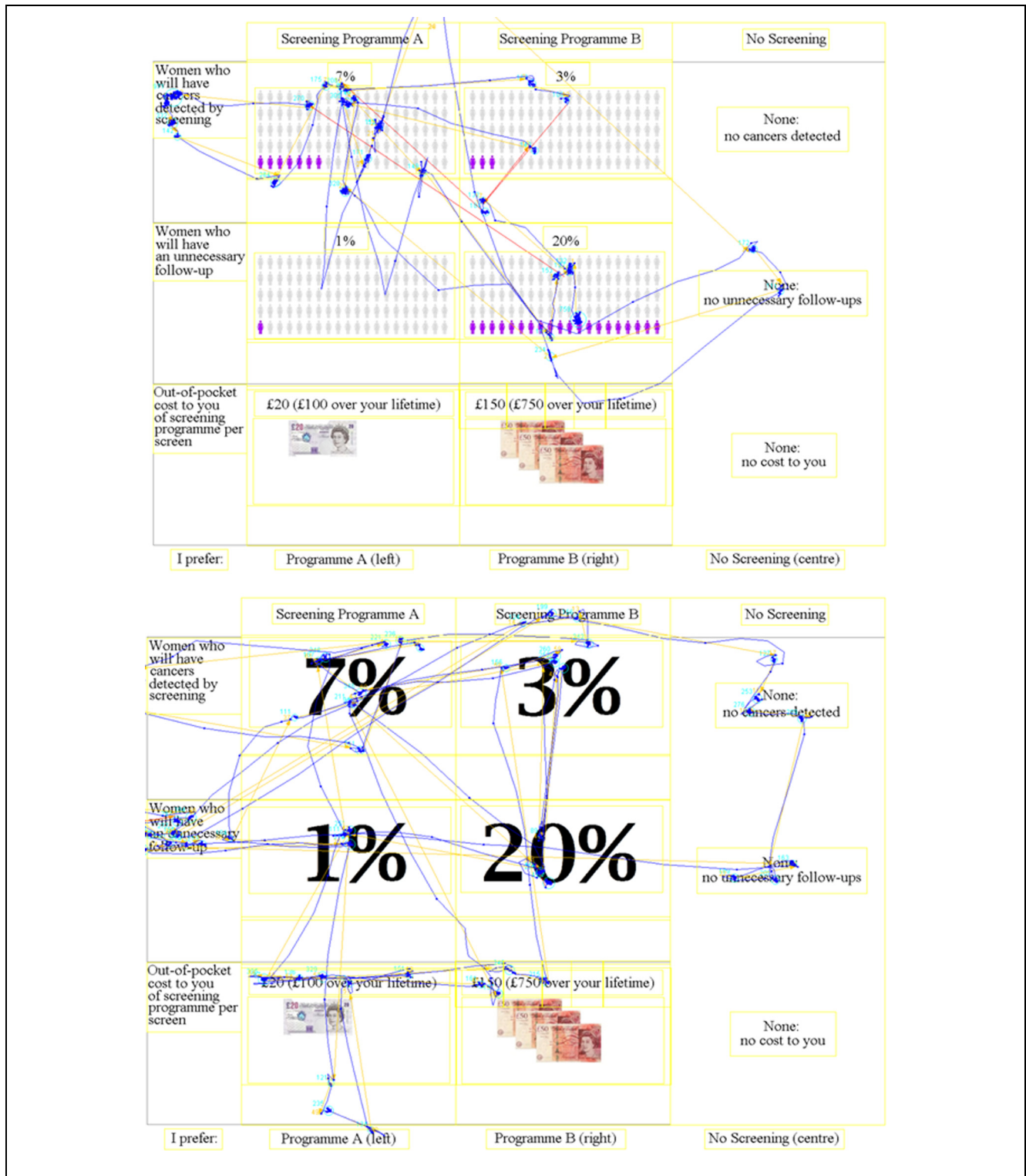


Figure 1 Example of scan paths from eye-tracking data for respondent completing the discrete choice experiment.

Table 4 Visual Attention to Each Attribute in the Discrete Choice Experiment (DCE)^a

Risk Format	Attribute in the DCE		
	Probability of Detecting a Cancer	Risk of Unnecessary Follow-up	Lifetime Cost of Screening Program
Mean number of fixations (standard errors)			
Percentages only % [α]	7.605 (0.71)	8.500 (0.84)	4.045 (0.68)
Icon arrays and percentages †	8.137	9.832	4.400
Difference due to icon array [β]	0.532 (0.97)	1.332 (1.46)	0.355 (1.04)
	<-0.103	<-0.198	<-0.066
Mean duration (milliseconds) of a fixation to each attribute in a choice set and differences between risk formats			
Percentages only % [α]	237.267 (14.95)	205.846 (9.08)	157.733 (12.16)
Icon arrays and percentages †	261.727	201.465	164.731
Difference due to icon array [β]	24.460 (20.61)	-4.38 (11.94)	6.998 (16.54)
	<-0.131	<0.064	<-0.083
Mean dwell time (total duration of fixations; milliseconds) to each attribute in a choice set and differences between risk formats			
Percentages only % [α]	1789.427 (219.59)	1919.718 (292.12)	829.427 (162.30)
Icon arrays and percentages †	1982.659	2072.573	897.536
Difference due to icon array [β]	193.232 (275.11)	152.855 (396.19)	68.109 (235.14)
	<-0.121	<-0.063	<-0.055
Effect of self-reported ANA on total dwell time (milliseconds) per trial			
No self-reported ANA [α]	1970.293 (145.70)	2010.305 (203.04)	881.676 (113.22)
Self-reported ANA	1127.796	1443.909	808.900
Difference [β]	-842.497*** (231.37)	-566.396** (203.04)	-72.776 (344.48)
	<0.573	<0.267	<0.064
Number of observations	440 ^b	440 ^b	440 ^b
Number [%] of choice sets with no attention to an attribute for each risk format			
Percentages only	0 [0.00]	7 [1.59]	9 [2.05]
Icon arrays and percentages †	1 [0.23]	1 [0.23]	15 [3.41]
Total	1	8	24

ANA, attribute nonattendance.

^aStandard errors in curved parentheses and effect sizes in angled brackets; square brackets reflect proportion of total choice sets.

^bForty participants, each completing 11 choice sets.

P* < 0.01. *P* < 0.001.

Table 5 Number and Direction of Saccades in a Choice Set^a

Risk Format	Mean Number of Saccades (SE)	Percentage of Saccades Moving Vertically (SE)	Percentage of Saccades Moving Horizontally (SE)
Percentages only [α]	47.955 (2.24)	43.5 (0.01)	56.5 (0.01)
Icon arrays and percentages †	48.228	48.9	51.1
Difference due to icon array [β]	0.273 (3.17)	5.4* (0.02)	-5.4* (0.02)
	<-0.008	<-0.510	<0.510
Observations	440 ^b	440 ^b	440 ^b

^aStandard errors in parentheses; effect sizes in angled brackets.

^bForty participants, each completing 11 choice sets.

**P* < 0.05.

than the cost attribute. This could suggest that the attributes “risk of unnecessary follow-up” and “probability of detecting a cancer” required more information processing or that this information was more important for study participants. This hypothesis was partially

confirmed by results from the heteroskedastic conditional logit model.

Other eye-tracking studies have investigated the effect of risk communication format on individuals’ information processing.^{53–55} Keller et al.⁵⁵ found no statistically

Table 6 Pupil Size and Dilation by Risk Communication Format^a

Risk Format	All	Attribute		
		Probability of Detecting a Cancer	Risk of Unnecessary Follow-up	Lifetime Cost of Screening Program
Mean pupil size per choice set (standard error)				
Percentages only [α]	1228.90 (92.24)	1248.89 (94.44)	1186.24 (90.46)	1343.14 (99.75)
Icon arrays and percentages †	1184.18	1195.84	1157.78	1264.57
Difference due to icon array [β]	-44.72 (115.55) (0.109)	-53.05 (118.59) (0.126)	-28.46 (115.63) (0.071)	-78.57 (126.33) (0.183)
Change in pupil size per trial over all fixations and fixations to each attribute and differences between risk formats				
Percentages only [α]	390.33 (33.14)	216.87 (21.30)	207.69 (25.20)	130.30 (19.23)
Icon arrays and percentages †	375.93	203.9	210.35	108.41
Difference due to icon array [β]	-14.40 (45.05) (0.077)	-12.97 (28.16) (0.088)	2.66 (33.10) (-0.018)	-21.89 (27.61) (0.172)
Observations	440 ^b	439 ^c	432 ^c	376 ^c

^aStandard errors in parentheses; effect sizes in angled brackets.

^bForty participants, each completing 11 choice sets.

^cFewer than 440 observations attributable to participants not attending to these attributes in a given choice set.

significant difference between attention to icon arrays or percentages when comparing high and low numerate samples. Measuring numeracy using 3 standardized questions,⁵⁶ split-sample analysis of our data corroborated these findings with no statistically significant differences between attention to the different formats by numeracy.

Self-reported ANA was confirmed by eye-tracking data on visual attention to the “risk” and “detect” attributes. When respondents reported that “risk” and “detect” were unimportant in choice making, data reflected that they had paid significantly less attention to these attributes. There was no difference in attention between participants reporting attending to the cost attribute and those who did not. The results of this study suggest that self-reported ANA to cost attributes, as measured through supplementary questions to the DCE, may not be robust. Balcombe et al.⁵⁷ also found in their food-based eye-tracking study that an individual reported attending to the cost attribute but never fixated on this attribute in the experiment. Qualitative research has found cost to be a delicate attribute in health care DCEs, with interviewees reporting that “I wouldn’t really be concerned about the cost.”¹⁷ This could be because it is genuinely not important or because the presence of a researcher induces some “social desirability bias” similar to the Hawthorne effect where participants align to perceived social norms.^{58,59}

Respondents in both versions of the DCE exhibited more horizontal movements than vertical, and those completing the DCE with icon arrays spent more time considering each alternative separately to make their choice. The “vertical” and “horizontal” movements are

likely related to the choice set design where alternatives were presented in columns; if alternatives were presented in rows, participants may have made even more horizontal movements. The difference may also be driven by the location of the colored icons in the array. In a study involving a simple choice set experiment, with 1 risk attribute and 1 cost attribute, Arieli et al.³³ concluded that a high proportion of vertical eye movements was indicative of respondents conducting expected payoff calculations. In this study, more vertical movements could indicate that the respondent was weighing up each alternative as a whole to decide which would offer the most utility given the uncertainty in the “detect” and “risk” attributes, suggesting EUT-type calculations. This behavior could also be seen as aligning with Lancaster’s theory,⁴ which suggests individuals value the attributes of a good or service, as individuals are taking account of each alternative to make a choice. In contrast, a high proportion of horizontal eye movements would indicate that attributes were being considered independently, such as risk and probability in this DCE. This study provided an exploratory investigation of eye-tracking to understand whether risk presentation in a DCE affected decision strategies. There was only scope to infer that different decision-making strategies were being employed between the 2 risk presentation versions. Further research is needed to understand the nature of the drivers of the difference in saccade patterns by, for example, changing the attribute order and introducing nonrisk attributes.

Results from this study also indicated that pupil dilation when completing the DCE was smaller for the icon array respondents, which implies the task was potentially

less cognitively burdensome and easier for respondents to complete. This finding is only indicative due to the small sample size, and a larger sample size would be required to investigate this preliminary finding further.

Results of the heteroskedastic conditional logit model suggested an insignificant effect of cost attribute on participants' choice. As with many forms of health care in the United Kingdom, national screening programs are free at the point of use, and in other countries or other domains, this may differ.

While increasingly used in psychology,²⁴ there are few examples of the use of eye-tracking in economics.⁶⁰ Exceptions include an eye-tracking study eliciting preferences for food, to investigate respondents' attendance, or lack of, to levels in a choice set.⁵⁷ Lack of attention, ANA, was measured by the number of times respondents reattended attribute levels. Another study recorded saccades to understand how individuals seek information in simple choice sets of financial lotteries.³³ In health care, the authors are aware of only 2 examples of eye-tracking used to identify ANA⁶¹ and for modeling of choice data.^{62,63} There are several parallels in these choice-based eye-tracking studies, with most focusing on visual attention to attributes. None have presented data on pupil dilation or compared decision strategies or attention across different attribute frames.

Limitations

Eye-tracking involves tradeoffs between precision equipment and apparatus that participants find comfortable. Equipment used in this study required participants maintain a fixed head position. Clearly, this is an unnatural reading position and could have reminded the respondents that their eye movements were being monitored, causing them to pay more attention to the screen than they would have otherwise. However, it is unlikely that the effects of this would differ between respondents randomized to the different risk formats.

Eye-tracking data also collect visual attention at the pupil's point of gaze. Although evidence from eye movement research suggests that little detailed information is extracted outside of the fovea and parafovea during written language comprehension,^{64,65} it is possible that some information about the attribute levels may have been picked up in the participant's peripheral vision without observable attention.

The sample size of 20 women in each condition could ideally have been larger and may possibly have led to

more statistical significance for some findings. However, with no priors from existing studies, power calculations could not be performed, and therefore a pragmatic approach to selecting the sample size was taken in this exploratory study. However, the results could be used to inform sample size calculations for future studies.

This study tested multiple hypotheses relating to the eye-tracking data. Although corrections for multiple hypothesis testing exist,⁶⁶ these were not applied as in tests that could have been correlated (i.e., the effect of risk communication method on mean dwell time, mean fixation time, number of fixations, mean pupil size, or change in pupil size)⁶⁷ the null hypotheses were never rejected. Applying the Bonferroni correction would therefore not have changed the results or study conclusions.

Pupil size depends on the location of the stimuli on the screen and laboratory luminosity. These 2 dependent factors influencing pupil size mean that the cognitive burden of studies cannot be compared on either an absolute or a relative scale using this metric alone. There are various causes of pupillary responses, and care must be taken in using this metric to distinguish the exact response activator. Extraneous variables, such as time of day and participants' previous activities, were not collected or controlled for in this experiment. Studies have shown that alcohol, lack of sleep, and reading may affect participants' eye movements. As respondents were randomized to different risk formats, we can assume these are unlikely to have affected the comparative results, but a larger sample size is ideally required to account for these potential sources of confounding in the experimental design.

Conclusion

This exploratory study found that eye-tracking is a promising method to understand more about how DCE respondents view the choice task and identify different decision-making strategies when different formats are used to present the survey. Differences were identified in how respondents searched for information dependent on risk presentation in the DCE. The study provided indicative findings that visual attention to attributes was greater, and cognitive burden lower, when risk was presented using an icon array. Eye-tracking data also suggested that using self-report to identify ANA may not be completely reliable. Using eye-tracking in the context of completing stated preference surveys, in general, and DCEs, specifically, is still in its infancy. This study is one of the first in the context of health care DCEs and has

demonstrated the potential of eye-tracking to answer many further research questions by providing quantitative data on the decision-making strategies employed by survey respondents.


Acknowledgments

We are grateful for feedback received at the Society for Medical Decision Making's 36th and 37th annual meetings, as well as the International Academy of Health Preference Researchers inaugural meeting. We are also grateful to experts Professor Gareth Evans, Professor Tony Howell, Dr. Michelle Harvie, and Ms. Paula Stavrinou of the Nightingale Centre at Wythenshawe Hospital, as well as Professor Stephen Campbell of the Centre for Primary Care at the University of Manchester for providing their thoughts and comments the study, including draft versions of the survey.

Supplementary Material

Supplementary material for this article is available on the *Medical Decision Making* Web site at <http://journals.sagepub.com/home/mdm>.

ORCID iD

Caroline Vass  <https://orcid.org/0000-0002-6385-2812>

References

- Ryan M. Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. *Soc Sci Med*. 1999;48:535–46.
- Ryan M. Discrete choice experiments in health care. *BMJ*. 2004;328:360–1.
- Louviere J, Lancsar E. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. *Health Econ Policy Law*. 2009;4:527–46.
- Lancaster KJ. A new approach to consumer theory. *J Polit Econ*. 1966;74:132–57.
- McFadden D. Conditional logit analysis of qualitative choice behaviour. In: Zarembka P, ed. *Frontiers in Econometrics*. New York: Academic Press; 1974. p 105–142.
- Harrison M, Rigby D, Vass CM, et al. Risk as an attribute in discrete choice experiments: a critical review. *Patient*. 2014;7:151–70.
- de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ*. 2012;21:145–72.
- Hauber A, Fairchild AO, Johnson F. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. *Appl Health Econ Health Policy*. 2013;11:319–29.
- Ho MP, Gonzalez JM, Lerner HP, et al. Incorporating patient-preference evidence into regulatory decision making. *Surg Endosc Other Interv Tech*. 2015;29:2984–93.
- Vass CM, Payne K. Using discrete choice experiments to inform the benefit-risk assessment of medicines: are we ready yet? *Pharmacoeconomics*. 2017;35:859–66.
- Lipkus I. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Med Decis Mak*. 2007;27:696–713.
- Watson V, Ryan M, Watson E. Valuing experience factors in the provision of Chlamydia screening: an application to women attending the family planning clinic. *Value Health*. 2009;12:621–3.
- Galesic M, Garcia-Retamero R, Gigerenzer G. Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychol*. 2009;28:210–6.
- Wong ST, Pérez-Stable EJ, Kim SE, et al. Using visual displays to communicate risk of cancer to women from diverse race/ethnic backgrounds. *Patient Educ Couns*. 2012;87:327–35.
- Garcia-Retamero R, Galesic M, Gigerenzer G. Do icon arrays help reduce denominator neglect? *Med Decis Mak*. 2010;30:672–84.
- Vass CM, Rigby D, Payne K. The role of qualitative research methods in discrete choice experiments: a systematic review and survey of authors. *Med Decis Mak*. 2017;37:298–313.
- Ryan M, Watson V, Entwistle V. Rationalising the 'irrational': a think aloud study of a discrete choice experiment responses. *Health Econ*. 2009;18:321–36.
- Whitty J, Walker R, Golenko X, et al. A think aloud study comparing the validity and acceptability of discrete choice and best worst scaling methods. *PLoS One*. 2014;9:e90635.
- Boren T, Ramey J. Thinking aloud: reconciling theory and practice. *IEEE Trans Prof Commun*. 2000;43:261–78.
- Ericson K, Simon H. Protocol analysis: verbal reports as data (revised edition). Available from: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:PROTOCOL+ANALYSIS+VERBAL+REPORTS+AS+DATA+REVISED+EDITION#0>
- Chenail RJ. Interviewing the investigator: strategies for addressing instrumentation and researcher bias concerns in qualitative research. *Qual Rep*. 2011;16:255–62.
- Taylor K, Dionne J. Accessing problem-solving strategy knowledge: the complementary use of concurrent verbal protocols and retrospective debriefing. *J Educ Psychol*. 2000;92:413.
- Kuusela H, Paul P. A comparison of concurrent and retrospective verbal protocol analysis. *Am J Psychol*. 2000;113:387–404.
- Rayner K. Eye movements in reading and information processing: 20 years of research. *Psychol Bull*. 1998;124:372–422.
- Kahneman D. Thinking, fast and slow. Available from: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:thinking+fast+and+slow#0>
- Just M, Carpenter P. A theory of reading: from eye fixations to comprehension. *Psychol Rev*. 1980;87:329–54.

27. Rayner K. Eye movements and attention in reading, scene perception, and visual search. *Q J Exp Psychol.* 2009;62:1457–506.
28. Vass CM, Rigby D, Payne K. Investigating the heterogeneity in women's preferences for breast screening: does the communication of risk matter? *Value Health.* 2018;21:219–28.
29. NHS Information Centre Screening and Immunisations. Breast Screening Programme 2010–2011. Available from: www.ic.nhs.uk
30. Choice Metrics. *Ngene 1.1 User Manual and Reference Guide.* Sydney, Australia: Choice Metrics; 2012.
31. Kowler E, Anderson E, Doshier B, et al. The role of attention in the programming of saccades. *Vision Res.* 1995;35:1897–916.
32. Findlay JM. Saccadic eye movement programming: sensory and attentional factors. *Psychol Res.* 2009;73:127–35.
33. Arieli A, Ben-Ami Y, Rubinstein A. Tracking decision makers under uncertainty. *Am Econ J Microecon.* 2011;3:68–76.
34. Hartmann M, Fischer MH. Pupillometry: the eyes shed fresh light on the mind. *Curr Biol.* 2014;24:R281–2.
35. Laeng B, Sirois S, Gredeback G. Pupillometry: a window to the preconscious? *Perspect Psychol Sci.* 2012;7:18–27.
36. Brodal PGA. *The Central Nervous System: Structure and Function.* 5th ed. Oxford (UK): Oxford University Press; 2015.
37. Hess EH, Polt JM. Pupil size in relation to mental activity during simple problem-solving. *Science.* 1964;143:1190–2.
38. Binda P, Pereverzeva M, Murray SO. Pupil size reflects the focus of feature-based attention. *J Neurophysiol.* 2014;112:3046–52.
39. Beatty J, Wagoner B. Pupillometric signs of brain activation vary with level of cognitive processing. *Science.* 1978;199:1216–8.
40. McGarrigle R, Dawes P, Stewart AJ, et al. Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology.* 2017;54:193–203.
41. Anderson WF, Jatoui I, Tse J, et al. Male breast cancer: A population-based comparison with female breast cancer. *J Clin Oncol.* 2010;28:232–9.
42. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet.* 2012;380:1778–86.
43. Bialkova S, van Trijp HCM. An efficient methodology for assessing attention to and effect of nutrition information displayed front-of-pack. *Food Qual Prefer.* 2011;22:592–601.
44. SR Research. *Experiment Builder Software.* Ottawa, Canada: SR Research; 2012.
45. StataCorp. *Stata Statistical Software: Release 12.* College Station, TX: StataCorp LP; 2011.
46. Vass CM, Wright S, Burton M, et al. Scale heterogeneity in healthcare discrete choice experiments: a primer. *Patient.* 2018;11:167–73.
47. Wright SJ, Vass CM, Sim G, et al. Accounting for scale heterogeneity in healthcare-related discrete choice experiments when comparing stated preferences: a systematic review. *Patient.* 2018 Feb 28. [Epub ahead of print]
48. Vansteenkiste P, Cardon G, Philippaerts R, et al. Measuring dwell time percentage from head-mounted eye-tracking data: comparison of a frame-by-frame and a fixation-by-fixation analysis. *Ergonomics.* 2015;58:712–21.
49. SR Research. *Data Viewer Software.* Ottawa, Canada: SR Research; 2013.
50. Wang JT. Pupil dilation and eye-tracking. In: Schulte-Mecklenbeck M, Kuhberger A, Ranyard R, eds. *A Handbook of Process Tracing Methods for Decision Research: A Critical Review and User's Guide.* Hove (UK): Psychology Press; 2010. p 185–204.
51. Wooldridge J. *Introductory Econometrics.* 4th ed. Available from: <http://link.springer.com/content/pdf/10.1007/978-1-4612-6292-3.pdf>
52. McGaw B, Glass GV. Choice of the metric for effect size in meta-analysis. *Am Educ Res J.* 1980;17:325–37.
53. Kreuzmair C, Siegrist M, Keller C. Does iconicity in pictographs matter? The influence of iconicity and numeracy on information processing, decision making, and liking in an eye-tracking study. *Risk Anal.* 2017;37:546–56.
54. Keller C. Using a familiar risk comparison within a risk ladder to improve risk understanding by low numerates: a study of visual attention. *Risk Anal.* 2011;31:1043–54.
55. Keller C, Kreuzmair C, Leins-Hess R, et al. Numeric and graphic risk information processing of high and low numerates in the intuitive and deliberative decision modes: an eye-tracker study. *Judgm Decis Mak.* 2014;9:420–32.
56. Schwartz L, Woloshin S, Black W, et al. The role of numeracy in understanding the benefit of screening mammography. *Ann Intern Med.* 1997;127:1023–8.
57. Balcombe K, Fraser I, McSorley E. Visual attention and attribute attendance in multi-attribute choice experiments. *J Appl Econ.* 2014;30:1–27.
58. Silverman D. *Interpreting Qualitative Data.* Thousand Oaks (CA): Sage; 1993.
59. Wickstrom G, Bendix T. The “Hawthorne effect”: what did the original Hawthorne studies actually show? *Scand J Work Environ Health.* 2000;26:363–7.
60. Orquin JL, Mueller Loose S. Attention and choice: a review on eye movements in decision making. *Acta Psychol (Amst).* 2013;144:190–206.
61. Spinks J, Mortimer D. Lost in the crowd? Using eye-tracking to investigate the effect of complexity on attribute non-attendance in discrete choice experiments. *BMC Med Inform Decis Mak.* 2015;16:14.
62. Krucien N, Ryan M, Hermens F. Visual attention in multi-attributes choices: what can eye-tracking tell us? *J Econ Behav Organ.* 2017;135:251–67.
63. Ryan M, Krucien N, Hermens F. The eyes have it: using eye tracking to inform information processing strategies in multi-attributes choices. *Health Econ.* 2018;27:709–21.

-
64. Rayner K, Slattery TJ, Bélanger NN. Eye movements, the perceptual span, and reading speed. *Psychon Bull Rev.* 2010;17:834–9.
 65. Rayner K. The perceptual span and peripheral cues in reading. *Cogn Psychol.* 1975;7:65–81.
 66. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ Br Med J.* 1995;310:170.
 67. Mayo DG, Cox DR. Frequentist statistics as a theory of inductive inference. *IMS Lecture Notes.* 2006;49:77–97.