

Graphical Perception of Icon Arrays versus Bar Charts for Value Comparisons in Health Risk Communication

Jade Kandel*, Jiayi Liu*, Arran Zeyu Wang, Chin Tseng, and Danielle Albers Szafir

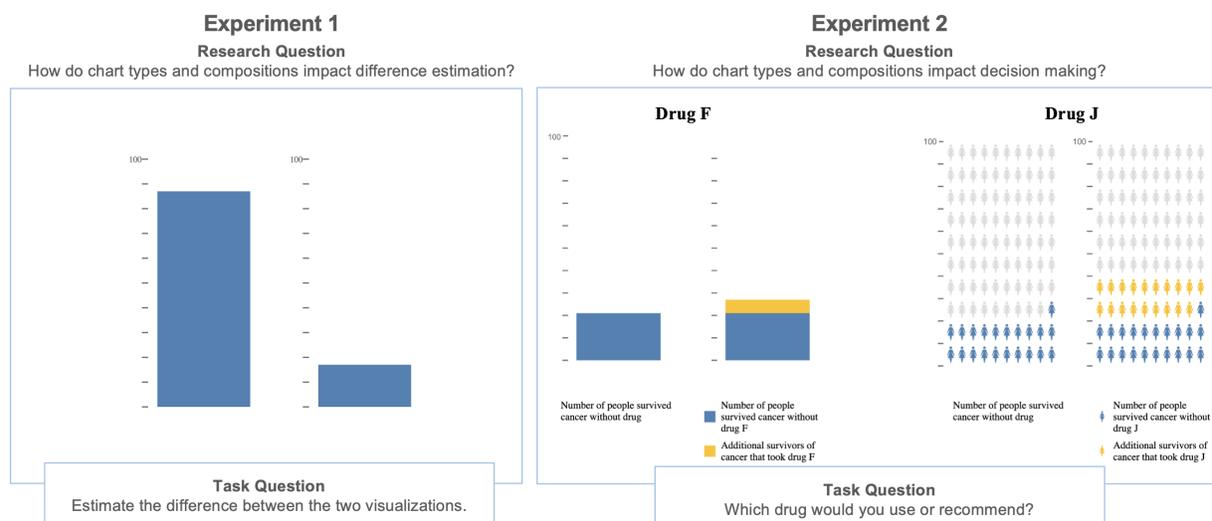


Fig. 1: We conducted two experiments examining the graphical perception of visualization *chart types* and *compositions* in health risk communication. Study 1 investigates how these design variables impact estimation of numerical difference. Study 2 investigates how these design variables influence medical decision-making through a scenario comparing survival rates between Drug F and Drug J.

Abstract—

Visualizations support critical decision making in domains like health risk communication. This is particularly important for those at higher health risks and their care providers, allowing for better risk interpretation which may lead to more informed decisions. However, the kinds of visualizations used to represent data may impart biases that influence data interpretation and decision making. Both continuous representations using bar charts and discrete representations using icon arrays are pervasive in health risk communication, but express the same quantities using fundamentally different visual paradigms. We conducted a series of studies to investigate how bar charts, icon arrays, and their layout (juxtaposed, explicit encoding, explicit encoding plus juxtaposition) affect the perception of value comparison and subsequent decision-making in health risk communication. Our results suggest that icon arrays and explicit encoding combined with juxtaposition can optimize for both accurate difference estimation and perceptual biases in decision making. We also found misalignment between estimation accuracy and decision making, as well as between low and high literacy groups, emphasizing the importance of tailoring visualization approaches to specific audiences and evaluating visualizations beyond perceptual accuracy alone. This research contributes empirically-grounded design recommendations to improve comparison in health risk communication and support more informed decision-making across domains.

Index Terms—Information Visualization, Graphical Perception, Health Risk Communication, Bar Chart, Icon Array, Composition

1 INTRODUCTION

Comparing values is fundamental for data interpretation [24] and essential for making informed decisions. People perform visual comparisons in many domains, including health risk communication, where patients often need to compare different risk factors to determine appropriate actions and treatments [9, 51]. For instance, individuals with specific genetic mutations may compare their personal cancer risk to the average population’s risk [18]. Patients may also examine risk levels before and after treatment [34] or how risk changes with age [32] to inform

cancer screenings and preventative surgery.

Visualizations can provide patients with improved access to health-related data, facilitating comprehension of differences between compared values that might be difficult to grasp from numerical data alone [19, 23, 24]. However, the presentation of healthcare data may influence the patient’s drawn conclusions [12]. For example, common visualization designs for comparative analysis in health risk communication employ both continuous and discrete data representations. *Continuous* data representations map data to visual properties in using a single, unbroken mark (e.g., a bar or line). *Discrete* methods represent data as separate, distinct units (e.g., icons in an isotype) [2]. Bar charts are one of the most common continuous representations for health risk communication, providing a familiar paradigm that supports high accuracy in value comparisons [21]. Icon arrays are discrete representations that illustrate proportions or probabilities through a collection of symbols arranged in a grid [3], making part-to-whole relationships concrete and easily countable [3, 21, 61]. While both representations have been studied in isolation [50, 55], we lack understanding in their relative effectiveness, especially in scenarios like health risk commu-

- All authors are with the University of North Carolina at Chapel Hill. E-mail: {kandelj, liujiayi, zeyuwang, chint, danielle.szafir}@cs.unc.edu
- Jade Kandel and Jiayi Liu equally contributed to the work

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxx/TVCG.201x.xxxxxx

nication. This paper examines how people compare values in health risk communication using bar charts and icon arrays for two key tasks: estimating differences and supporting risk-based decision making.

While icon arrays offer clear benefits in health risk and probability communication, including making part-to-whole relationships concrete and easily countable, particularly for individuals with limited numeracy skills [3, 21, 61], bar charts provide high accuracy for general quantitative judgments [10, 50]. In visualization research, bar charts are more common than icon arrays for making comparisons using different compositions, such as juxtaposition and explicit encoding [24, 56]. However, given the importance of understanding probabilities and making comparisons in healthcare decision-making, we must understand potential trade-offs between these visualization types to design visualizations that help people accurately understand their risks and make informed decisions [36, 37]. This requires examining not only their individual effectiveness but also how different visual compositions enhance analysis across demographic groups.

To fill these gaps, we conducted two crowdsourced experiments investigating how chart type and composition affect difference perception between values. Chart types included bar charts versus icon arrays, and compositions included juxtaposition (two charts side-by-side), explicitly encoded differences (a single chart with the difference highlighted), and their combination (two charts side-by-side with the difference highlighted). In Experiment 1, we conducted a 2×3 experiment, where participants estimated the numerical difference between two values ranging from 0-100, assessing how precisely people estimate relevant differences in risk. In Experiment 2, we conducted a 2×3 between-subject experiment where participants made treatment preference decisions between two drugs based on pre- and post-treatment risks to understand the impact of these designs on decision making.

Our results showed icon arrays and the combined composition (explicit encoding plus juxtaposition) consistently supported both accurate estimation and correct decisions, suggesting designs that support both estimation accuracy and decision making. We also identified a misalignment between estimation accuracy and decision preferences among literacy groups, as low-literacy participants selected juxtaposition formats incorrectly for decision-making despite these formats yielding lower accuracy. These findings enable us to create empirically grounded design recommendations that consider accuracy, decision-making, and literacy levels, ultimately improving health risk communication and empowering patients to make more informed medical decisions.

Contributions: We conducted two crowdsourced studies investigating value comparison (Study 1) and basic comparative decision making (Study 2) in risk communication using bar charts and icon arrays across three comparative layouts. Our paper provides the following contributions:

- An evaluation of how chart type (bar chart vs. icon array) and composition (juxtaposed, explicit encoding, or combination) affects people’s ability to accurately estimate differences between two visualized values.
- An evaluation of how chart type (bar chart vs. icon array) and composition (juxtaposed, explicit encoding, or combination) influence decision making.
- Empirically-grounded design recommendations derived from our experimental results for improving comparisons in health-risk communication and informed decision making.

2 RELATED WORK

Understanding perceptual bias in visualization is crucial for effective data communication and decision making. Our research builds upon existing knowledge of the perception of icon arrays and bar charts, particularly by examining their efficacy in health risk communication.

2.1 Health Risk Communication

Visualizations impact healthcare decisions [25], with accurate interpretation of conditional probabilities crucial for both healthcare providers and patients [14]. Misinterpretation of statistical information can lead to serious consequences in clinical settings [53]. Prior research has shown

that visualization techniques can significantly enhance probabilistic reasoning and decision-making [8]. Ottley et al. demonstrated how different visualization methods affect individual performance in Bayesian reasoning tasks, particularly in medical communication contexts where probabilistic reasoning is essential [41].

Visualizations in healthcare are common. Medical research papers use visualizations to summarize findings related to risk over time [33, 51]. Online interactive tools use stacked bar charts [34] and icon arrays [62] to assist in communicating risks to the public. However, significant challenges exist when using these charts, such as people having difficulty comprehending or misinterpreting charts, which impacts the decisions patients make about their health [1, 37]. Fagerlin et al. [15] surveyed empirical studies to provide recommendations for how healthcare providers and health educators can best communicate complex medical information to patients. Recommendations include highlighting how treatment changes risks from baseline risk, using frequencies instead of percentages, considering how age and timeline impacts risk, and comparing risks.

Comparison tasks in healthcare may involve patients comparing their personal risk with an “average” risk [18], across time [32], and before and after treatment [34]. Comparing risks may be helpful to patients because it allows personal risk statistics to be compared against a meaningful standard [59]. Without this type of information, it may be difficult for people to know how to feel and react to statistical information, making comparative risk a better indicator of worry and behavior compared to absolute risk [57]. Given the prevalence of icon arrays and bar charts in existing risk communication, understanding how people compare risk using these idioms can lead to more effective risk comprehension.

2.2 Graphical Perception in Bar Chart and Icon Array

Graphical perception systematically investigates how individuals extract and estimate various properties from visual representations [10], creating evidence-based design recommendations to mitigate biases and prevent misinterpretations [49]. Graphical perception research has investigated how people interpret data across a wide range of tasks and representations (see Quadri & Rosen [42] and Zeng & Battle [58] for surveys). Our research examines graphical perception and decision making in icon arrays and bar charts.

Bar charts are one of the most common visualization types, introduced in primary education and widely implemented across a range of domains. Previous work shows that position-based encodings are susceptible to perceptual biases including systematic overestimation and “perceptual pull” effects [54], and effectiveness varying based on horizontal separation and the presence of distractors [50]. Still, bar charts have significant advantages over other chart types. Bar chart’s simple nature can be favored by people who are less numerate [3, 16, 44]. They may also enable more accurate quantitative comparisons than alternative visual encodings such as area, angle, or color [10], especially for tasks like distributional analyses [45].

Icon arrays use discrete units to communicate part-to-whole relationships effectively. Research demonstrates their advantages in health risk communication over pie charts and random icons by portraying probabilities as countable frequencies, making them more intuitive for individuals with lower numeracy skills [3, 21, 60]. Previous work found icon arrays improve estimation compared to pie charts, bar charts, and random icons [17], though this focused on single values rather than comparisons. We lack understanding of how icon arrays impact comparison tasks, as previous research examined design elements affecting single-value perception, such as how spatial arrangements of icon arrays significantly influence probability estimation [55] and how highlighting elements within icon arrays can lead to probability overestimation [35]. Based on research on discrete representations, notably the analysis of icon types in visualizations and pictographs [26, 61], we anticipate people may achieve faster, more accurate interpretation from icon arrays due to subitizing—instantly recognizing small quantities without counting [31].

Given healthcare risk communication’s critical impact on patient outcomes, we conducted controlled experiments comparing visualization

approaches to provide evidence-based recommendations for different scenarios.

2.3 Visualization Compositions

Our experiments investigate how different visualization compositions affect comparison in perception and decision-making. Gleicher et al. [24] defined the chart designs to support visual comparison in information visualizations, abbreviated as **composition** in this paper, that support comparison tasks generally. Juxtaposition presents values separately, helping people move their attention between representations to see patterns between elements, and usually place visualizations next to each other, as in side-by-side or small multiples displays. While presenting raw values facilitates direct comparisons, viewers may still overlook or misinterpret relationships [39, 43], particularly when juxtaposed visualizations require them to rely on working memory. For example, in examining five different layouts—stacked, adjacent(juxtaposition), mirrored, overlaid, and animated—for low-level perceptual comparison tasks, Ondov et al. demonstrated that reflecting across a central axis yielded more precise value estimations than simple side-by-side juxtaposition. [39]. Alternatively, explicit encoding shows the relevant differences explicitly with a visual encoding. For example, a chart might show the increased risk associated with a particular genetic condition by first showing a bar whose height corresponds to a baseline value and then adding a highlighted extension to that bar showing the increased risk, as demonstrated in Study 1’s bar chart explicit encoding (line 1, column b) (Figure 2). Explicit encoding excels at highlighting specific relationships, but often sacrifices contextual information about the highlighted differences.

Combining it with other compositions may help preserve the broader analytical context of the comparison and allow for more versatile exploration and decision making [56]. For example, hybrid methods can combine explicit encoding and juxtaposition techniques with linked highlighting, visual overlays, and data abstraction to effectively show relationships between elements while maintaining contextual information [23, 24]. Further research by Jardine et al. showed that stacked charts provided the highest precision for comparing both means and ranges, but visualization effectiveness varies based on the visual features that best support specific comparison tasks [29]. Also, their findings suggest that people rely more on heuristic perceptual proxies than mathematical operations when extracting data values from visualizations. These insights motivate our investigation of understanding compositions’ perceptual trade-offs between icon arrays and bar charts for comparing probability pairs in health risk communication.

3 MOTIVATION

Our study is motivated by our collaboration with BRCA specialists who rely heavily on visualizations to help patients with BRCA mutations understand their elevated cancer risks and make informed decisions about risk management. Providers typically give patients resources that communicate through visualizations, including comparative mortality rates through stacked charts [34], temporal risk progression via line graphs [33], and bar charts displaying cancer risk percentages [18]. Our collaborators, who regularly use these materials to educate patients, noted that certain visualization types were more intuitive than others. One nurse practitioner specifically employed icon arrays (Figure 3.1), finding that patients better comprehended and personally connected with the colored representations of risk.

Effective visualization in these scenarios requires communication that is both intuitive and accurate. Accurate perception of risk visualizations directly impacts healthcare decisions. Misinterpretation may lead to serious consequences: risk overestimation causes unnecessary anxiety and excessive intervention, while underestimation results in neglected treatment and missed prevention opportunities.

We grounded our investigation into health risk communication in the real-world challenges faced by our collaborators. To develop empirically grounded design recommendations for health risk communication comparisons, we conducted two experiments: one focused on participants’ ability to accurately estimate differences between two visualized

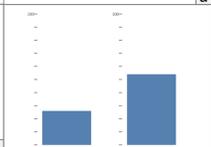
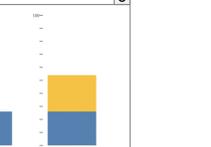
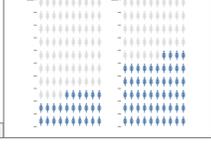
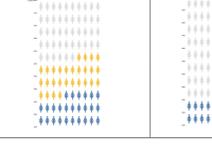
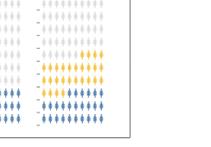
		Composition		
		Jux. a	Exp. Enc. b	Exp. Enc.+ Jux. c
Chart Type	1 Bar Chart			
	2 Icon Array			

Fig. 2: Design Variables: A table showing the chart types (1,2) and compositions (a,b,c) studied in our experiments. Abbreviations: Jux. = Juxtaposition, Exp. Enc. = Explicit Encoding.

values, and the other focused on simple decision-making through treatment selection. To systematically address these research questions, we organize our experimental motivation around three key components: the values being compared, the visualization design variables, and the specific estimation task participants perform.

Compared Values: This investigation centers on how people make comparisons between two values (corresponding to comparisons between two treatments) that range from 0 to 100, which directly correspond to risk percentages relevant to health communication contexts.

Design Variables: In both of these experiments, we investigate the effects of two **chart types** and three **compositions** (Figure 2). While healthcare communication commonly uses a range of visualization idioms, such as line charts [32], pie charts [17], tables with percentages [18], previous work and recommendations from our collaborators suggest that bar charts and icon arrays specifically are notable for their popularity and effectiveness in conveying health risks [3, 16, 19, 21, 60]. We focus on bar charts (Figure 2.1) and icon arrays (Figure 2.2) as they are pervasive and considered best practices within health risk communication [3], yet employ fundamentally different visual encoding strategies with distinct cognitive trade-offs. For composition, we focus on **juxtaposition** alone (Figure 2.a), **explicit encoding** alone (Figure 2.b), and combining **explicit encoding plus juxtaposition** (Figure 2.c). While previous work has demonstrated key tradeoffs in comparative compositions [56], the relative influence of these trade-offs in bar charts and icon arrays has not been directly compared, which we examine in our experiments.

Estimation Task: We focus on the effects of chart type and composition on value estimation (as when assessing one’s risk compared to a baseline) and decision making (as when using risk data to choose between treatments). Our experiments focus on the ability to *estimate* differences, rather than accurately calculate them, to understand working memory from brief visual exposure. While we are confident most participants could accurately calculate the difference with any chart type and composition given sufficient time, and make informed decisions from those calculated differences, in the real world not everyone takes the time to closely study visualizations [4]. Given that working memory is limited and can be misleading [26], our studies examined how visualization designs affect difference estimation during these brief exposures and the subsequent impact on decision-making processes.

4 EXPERIMENT 1: DIFFERENCE ESTIMATION

In Experiment 1, we examined whether chart type and composition affected estimation accuracy between two percentage values.

4.1 Hypotheses

Drawing on past work and our collaborators’ experiences, we hypothesized:

H1: Icon arrays yield higher estimation accuracy for small or large differences. Previous work shows discrete items improve recall for small values [27, 52]. Icon arrays enable precise value estimation

for small differences and allow counting negative space for large differences (i.e., when there are few icons that are not filled in), reducing proportional errors and utilizing subitizing to rapidly estimate small amounts [31]. Hence, we expect values less than 20 or greater than 80 to be more accurate. We do not expect to see this difference when the value differences are more moderate, between 20 and 80, making the difference in icon numbers too large for rapid estimation.

H2: Icon arrays will lead to underestimation. Previous work indicates that estimating length is more accurate than area [10], and icon arrays often times force non-symmetrical area-based comparisons rather than length judgments like bar charts. Given that people are likely creating area groupings with icon arrays [55], we predict people might group rows without including edge icons or accidentally overlooking rows, leading to a smaller estimation.

H3: Juxtaposition without explicit encoding will have more underestimation compared to designs leveraging explicit encoding. We predict the additional task of comparing two values, rather than estimating one value in a stacked graph, will lead to less accuracy. Highlighting differences, alternatively, may emphasize this value, leading to either more accuracy or overestimation.

H4: Bar charts yield higher estimation accuracy for people with lower numeracy and data literacy scores. While icon arrays were found to be very beneficial for part-to-whole interpretation for people with lower numeric and data literacy, these findings may not generalize for when two charts are shown, significantly increasing the complexity of the visualization. Previous work has indicated that people may perform better with simpler and more common charts like bar charts compared to icon arrays [20, 38].

4.2 Methods

We analyzed the effect of chart type and composition in a pair of 2×2 experiments. The first experiment compared juxtaposition alone versus explicit encoding plus juxtaposition combined. The second experiment compared explicit encoding alone versus explicit encoding plus juxtaposition combined. This two-experiment structure isolates the effects of different visual composition strategies commonly employed in health risk communication. The first experiment examines two common methods for supporting comparison in health risk communication. The second explores whether highlighting the differences compared to a baseline may encourage alternative strategies for interpretation, such as only attending the highlighted values.

In each experiment, participants were grouped into one of two composition conditions, resulting in four groups. Each participant completing trials with both chart types within their assigned composition.

4.2.1 Stimuli

We generated our stimuli was generated using HTML and JavaScript, which consisted of bar chart and icon array visualizations, generated at 460×460 pixels using D3.js [5]. Bar charts consisted of 150 pixel wide bars. Each icon array consisted of a 10×10 grid with individual icons measuring 20×40 pixels. The icons were encoded in rows from the bottom-up, as previous studies in icon array graphical perception found that people are more accurate in estimating probabilities when viewing rows compared to other arrangements such as centric, random, or around the edge [55]. Bottom-up arrangements also reflect best-practice designs in health risk communication [35]. In this study, we used a female icon, which aligns with our collaboration with BRCA healthcare providers.

Both visualization types indicated the value with steel blue (RGB: 70,130,180), and non-highlighted icons were light grey (RGB: 221,221,221), and for explicit encoding charts, the highlighted difference was colored in yellow (RGB: 255, 194, 10). As we did not screen for CVD, we ensured accessibility by selecting colors that maintain strong lightness contrast for people with CVD. The visualized values ranged from 0-100 on a y-axis with a maximum of 100, enabling interpretation as proportions, percentages, or probabilities. All visualizations had 10 equally-spaced grey tick marks (2 pixel stroke, 10 pixels in length) along the y-axis with only the maximum value labeled as "100". We used tick marks to preserve real-world chart guidelines

essential for interpretation, but omitted all y-axis labels except "100" at the maximum value to encourage visual estimation over calculation.

Participants saw bar charts and visualizations either juxtaposition, explicit encoding only, or explicit encoding plus juxtaposition. For the juxtaposition visualizations, each trial consisted of two bar charts or two icon arrays side by side with a 60-pixel gap separating the 200×460 -pixel charts. For the explicit encoding only visualizations, each trial contained one chart in the center. Figure 2 summarizes the set of tested visualizations.

For each participant, we first generated a list of 30 values for each chart type using stratified random sampling, resulting in 60 trials total. We then added three engagement checks and randomized the order of the 63 stimuli. We then rendered the stimuli sequentially during the study. We generated our dataset with stratified random sampling to ensure balanced difference magnitudes across 60 trials. We stratified the differences between values into ten equal-sized bins (1-10, 11-20, ..., 91-100). We randomly generated the smaller value and added the predetermined difference to create the larger value, ensuring neither exceeded 100. For each chart type, we selected three datasets from each of the ten different strata to guarantee balanced representation across possible differences, resulting in 30 bar chart trials and 30 icon array trials, presented in random order. Each stimulus visualized two values between 0 and 100, with the left-right order of the stimuli randomized to mitigate potential confounds.

4.2.2 Task & Prompts

Experiment 1 was separated into two different tasks, where the first experiment focuses on *comparing values*, and the second focuses on *estimating explicit differences* to compare the potential strategies used by participants in interpreting differences in risk. In both experiments, participants used a horizontal slider (640×15 pixels) with a corresponding numerical input field. The first subgroup was asked to "Estimate numerical differences between displayed visualizations." For the second experiment, to better understand how people estimate the value differences when differences were explicitly encoded in either a single or double chart, participants were asked to "Estimate numerical differences in the yellow segment." We use a numerical description in our tasks rather than a specific healthcare framing to maintain task objectivity using a neutral unit of measure.

4.2.3 Procedure

We conducted a pair of crowdsourced experiments consisting of 4 phases: (1) informed consent, (2) numeracy/graph literacy assessment and demographics, (3) task description and tutorial, and (4) formal study. Both studies followed the same procedure. At the beginning, participants provided informed consent in accordance with our IRB protocol. Participants then completed a seven-question literacy assessment, including Schwartz's validated three-question numeracy test and a four-question graph literacy test [46, 47], followed by a brief demographic questionnaire. We opted to use these measures as literacy assessments as they offered validated measures from the health risk communication literature, reflecting literacy in the context of the target application area.

Next, participants completed a tutorial to ensure task understanding, which described the study and included two example timed trials identical to those in the formal study: a bar chart with a difference of 30, and icon array with a difference of 40 (see Figure 3 for an example). The tutorial replicates what participants experience during the actual experiment, with the addition of revealing the solution. This solution is not shown to participants during the experiment itself, but is included in the tutorial to ensure understanding.

The formal study consisted of 63 trials (60 experimental trials and 3 engagement checks) ($\Delta = 100$ and $\Delta = 2$), embedded at random positions within the sequence to verify participants understood the task and response mechanisms, had sufficient visual acuity to understand the stimuli, and were not randomly guessing.

We displayed each visualization for 2 seconds to permit estimation while preventing exhaustive counting, similar to Haroz et al. [26]. A trial counter in the top-left corner displayed the number of trials completed out of the total number of trials. After the two seconds,

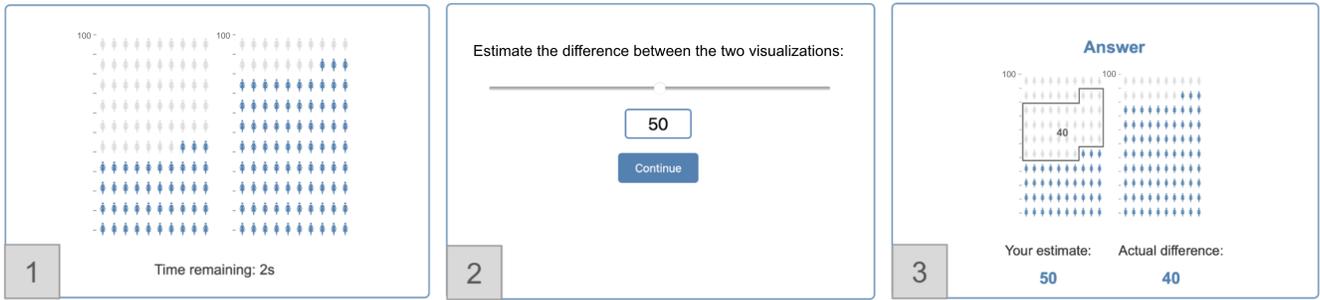


Fig. 3: An example tutorial trial for icon array juxtaposition conditions. First, participants saw the visualization for two seconds, with a timer counting down (1). Next, they entered their estimate of the difference (2). Finally, they compared their estimate of the actual difference, with the difference outlined (3).

stimuli were hidden, but participants could still input a response. Once participants pushed “Continue” to advance to the next trial.

The full experimental infrastructure and data are available on [OSF](#).

4.2.4 Participants

We recruited a total of 361 participants from the US and Canada with at least a 95% approval rating on Amazon Mechanical Turk (MTurk). We used MTurk as participants approximately reflect the general population [6], increasing likely generalizability and avoiding potential confounds and ethical concerns with actual patients [48]. The first experiment included 102 participants in the juxtaposition condition, and 100 participants in the explicit encoding plus juxtaposition condition with “comparing values” prompt. The second included 80 participants in the explicit encoding plus juxtaposition condition with the “estimate explicit encoding” prompt, and 90 in the explicit encoding condition. Due to failing more than one engagement check, we excluded 4 participants from the juxtaposition condition, 6 participants from the explicit encoding plus juxtaposition in the first experiment, and 10 from the explicit encoding condition. We analyzed data from the remaining 341 participants, aged 18-65 years old ($\mu=42.57$, $\sigma=9.19$), with an average literacy score of 5.58 out of 7 ($\sigma=1.27$). Participants took approximately 15 minutes to complete the study, and were compensated for their time. Participants who did not complete the study were not compensated or included in our analysis.

4.2.5 Measures and Analysis

We evaluated performance using signed error, unsigned error, and accuracy. Signed error (calculated as participant estimate minus actual difference) captures both direction and magnitude of estimation bias, with negative values indicating underestimation, positive values indicating overestimation, and zero representing perfect estimation. Unsigned error (the absolute value of signed error) measures estimation precision regardless of direction, focusing solely on the magnitude of deviation from the true difference. Accuracy reflects whether or not the estimate was perfect (error equal to zero).

In addition to our target variable, we also analyzed the effects of numeric and data literacy scores, categorizing individuals scoring below average as ‘low literacy’ and above average as having ‘high literacy’. Rather than using a perfect score to establish literacy [46], we include 6 as a high literacy score to provide a more inclusive definition of literacy.

For each experiment, we conducted a two-factor (chart-type \times composition) ANOVA to analyze our dependent measures. To understand how the magnitude of the actual difference influenced estimation performance (H1), we performed separate analyses across three data ranges: the full set of responses (0-100), differences between 20-80, and differences less than 20 and greater than 80 (<20 or >80). Given that statistical patterns were identical between full responses and differences between 20-80, we focused our analysis on comparing the ranges 20-80 and (<20 or >80). All post-hoc comparisons employed Tukey’s Honest Significant Difference (HSD) test with a significance threshold of $\alpha = 0.05$. In addition to removing data from participants who missed

more than one question on the engagement test, we removed responses where the error was greater than three standard deviation from the mean (error > 30). Since participants submitted this set of responses with the stimulus hidden, responses more than three times the standard deviation were highly likely to be random guesses.

4.3 Experiment 1 Results

For significant findings, we provide means and 95% confidence intervals to ensure transparent statistical communication [13]. Our analysis revealed significant differences in participant estimation errors across both chart type and chart composition. Participants’ numeric and data literacy scores correlated with both signed and unsigned error, with lower literacy scores resulting in higher error. However, we found that chart type and composition effectiveness did not vary significantly based on literacy scores. Therefore, we present focus on the impact of chart type and composition across all participants. Table 1 summarizes our ANOVA results, analyzed and grouped by difference ranges (Figure 5).

4.3.1 Chart Type:

Icon arrays generally outperformed bar charts in accuracy in both experiments across various difference ranges, contradicting our hypothesis that icon arrays would lead to underestimation. Icon arrays demonstrated significantly lower unsigned error when comparing values across juxtaposed charts (first experiment) for differences of 20-80 ($\mu_I = 5.46 \pm 0.19$; $\mu_B = 5.99 \pm 0.18$), and differences <20 and >80 ($\mu_I = 2.37 \pm 0.16$; $\mu_B = 3.64 \pm 0.16$), and when estimating the explicitly encoded difference (second experiment) for differences 20-80 ($\mu_I = 3.45 \pm 0.18$; $\mu_B = 5.43 \pm 0.19$), and differences <20 and >80 ($\mu_I = 1.53 \pm 0.18$; $\mu_B = 3.11 \pm 0.17$). For comparing juxtaposed values, with signed error, icon arrays were significantly closer to 0 than bar charts for difference >20 and <80, ($\mu_I = 0.07 \pm 0.21$; $\mu_B = 0.449 \pm 0.2$), but no statistical significance for 20-80. We found the same statistical relationship with estimating explicit differences ($\mu_I = 0 \pm 0.24$; $\mu_B = 0.34 \pm 0.25$).

4.3.2 Composition:

Participants demonstrated significantly higher unsigned errors with juxtaposition (J) compared to explicit encoding plus juxtaposition (EJ) compositions across all difference ranges, indicating lower accuracy for juxtaposition. For comparing values, juxtaposition had significantly higher unsigned error compared to explicit encoding plus juxtaposition in the difference range 20-80 ($\mu_J = 6.584 \pm 0.197$; $\mu_{EJ} = 5.068 \pm 0.199$) and differences <20 and >80 ($\mu_J = 3.359 \pm 0.173$; $\mu_{EJ} = 2.723 \pm 0.174$). Juxtaposition showed significant underestimation compared to explicit encoding plus juxtaposition compositions for difference range 20-80 ($\mu_J = -1.955 \pm 0.27$; $\mu_{EJ} = 0.267 \pm 0.272$). These results confirmed our hypothesis that juxtaposition would lead to more underestimation and less accuracy compared to the other compositions. However, we found no significant differences when the difference was <20 and >80.

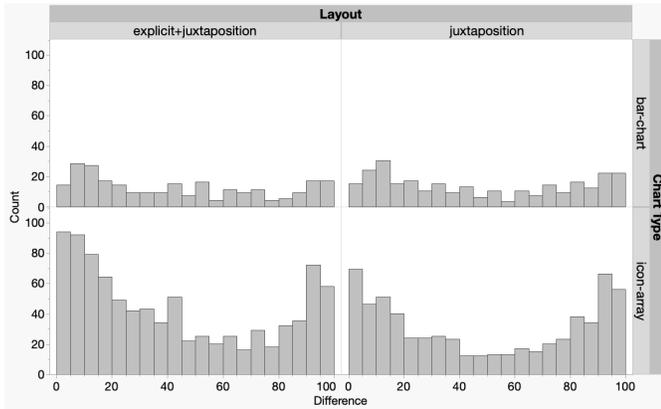


Fig. 4: Two histograms showing the number of times participants got error=0, grouped by chart type and difference. This shows that participants were more accurate when the difference was >20 and <80.

Reduced error for icon arrays with differences <20 and >80 may stem from more instances of zero error. To better understand why this phenomenon may occur and if it is connected to counting, Figure 4 shows the frequency of perfect accurate estimates (error=0) based on composition and chart type, grouped by actual difference. This demonstrates that while icon arrays consistently outperformed bar charts in accuracy, they exhibited substantially higher perfect accuracy when differences fell outside the 20-80 range. Despite general lower performance, juxtaposition also shows higher perfect accuracy for differences <20 and >80, which may explain why juxtaposition has less underestimation for <20 and >80. These findings confirm our hypothesis that icon arrays would be more accurate for differences <20 and >80 and less advantageous for differences between 20-80, and were especially influential for the juxtaposition condition.

When estimating explicitly encoded differences, explicit encoding plus juxtaposition composition was more accurate compared to explicit encoding alone. Explicit encoding had significantly higher error compared to explicit encoding plus juxtaposition in the difference range 20-80 ($\mu_E = 5.05 \pm 0.2$; $\mu_{EJ} = 3.83 \pm 0.16$) and differences <20 and >80 ($\mu_E = 2.56 \pm 0.17$; $\mu_{EJ} = 2.09 \pm 0.18$). Explicit encoding plus juxtaposition signed error was significant closer to 0 than explicit encoding for difference range 20-80 ($\mu_{EJ} = 0.049 \pm 0.16$; $\mu_E = -1.28 \pm 0.184$).

4.3.3 Chart Type \times Composition:

The analysis of Chart Type \times Composition for comparing values reveals icon arrays with explicit encoding plus juxtaposition was significantly more accurate than other visualizations. For comparing values, icon arrays with explicit encoding plus juxtaposition had a significantly lower unsigned error ($\mu = 3.45 \pm 0.19$) than bar chart with explicit and juxtaposition ($\mu = 4.75 \pm 0.2$), icon array with juxtaposition ($\mu = 5.05 \pm 0.19$), and bar chart with juxtaposition ($\mu = 5.39 \pm 0.19$). For estimating explicit differences, icon arrays with explicit encoding plus juxtaposition had a significantly lower unsigned error ($\mu = 2.87 \pm 0.21$) than icon array with explicit encoding ($\mu = 4.03 \pm 0.26$), bar chart with explicit and juxtaposition ($\mu = 4.79 \pm 0.26$), and bar chart with explicit encoding ($\mu = 6.07 \pm 0.26$).

4.3.4 Numeric and Data Literacy:

We found participant's numeric and data literacy scores correlated with unsigned error, with low literacy scores (L) resulting in significantly higher error compared to high literacy scores (H) for making comparisons ($\mu_L = 5.69 \pm 0.15$, $\mu_H = 3.97 \pm 0.11$) ($F(1,8181) = 297.72$; $p < 0.001$) and estimating explicit differences ($\mu_L = 7.68 \pm 0.33$; $\mu_H = 3.48 \pm 0.19$) ($F(1,9538) = 478.33$; $p < 0.001$). All of our findings generalized across low and high literacy groups. All participants being more accurate with icon arrays contradicts our hypothesis that bar charts would be better for people with low literacy.

4.4 Experiment 1 Discussion

Icon Arrays and Explicit Encoding Plus Juxtaposition are More Accurate: Contrary to our hypothesis, icon arrays consistently outperformed bar charts in estimation accuracy across various difference ranges. While bar charts remain standard for comparisons across domains, our findings suggest that the discrete representation of icon arrays provides perceptual benefits that can improve difference estimation. While icon arrays were most accurate in ranges >20 and <80, we found icon arrays benefits were noticeable across all difference values, especially with explicit encoding methods. We also found the explicit encoding plus juxtaposition combination was more accurate than either juxtaposition or explicit encoding alone, showing benefits of hybrid compositions for accurate comparisons.

Difference and Chart Type Impacts Juxtaposition Effectiveness: In support of our hypothesis, the effectiveness of visualization type and composition varies substantially based on the magnitude of differences being depicted. For differences between 20-80 (the middle range), juxtaposition consistently performed worse than other compositions regardless of chart type, producing higher unsigned errors and significant underestimation. However, this pattern changed dramatically for extreme differences (<20 or >80), where juxtaposition with icon arrays showed significantly lower error rates compared to all bar chart visualizations. This finding suggests that the discrete nature of icon arrays provides particular advantages for perceiving very small or very large differences—participants could effectively count individual icons or empty spaces even without explicit encoding. The frequency of perfect estimates (error=0) for icon arrays with extreme differences further supports this interpretation. Hence, designers should consider the expected magnitude of differences, and that icon arrays effectively communicate both small and large differences, even in simple juxtaposed formats.

5 EXPERIMENT 2: DECISION MAKING FROM DIFFERENCES

Our first experiment examined how individuals estimate differences between two values, similar to comparing outcomes with and without a given treatment. While accurate perception of these differences is fundamental for comparative analysis, people's decisions may not always align with perceived data differences [11]. For instance, someone might overestimate a risk—an error in Experiment 1—yet this very overestimation could drive their subsequent decision-making behavior.

In this study, we model decision making through a selection task: we ask people to choose between two different drugs based on pre- and post-treatment survival rates. While this simple decision-making paradigm ultimately relies on a comparative estimate [28], framing analytical tasks as decision scenarios can influence behavior [11]. We chose this task as it reflects real-world applications of importance to our collaborators (choosing between treatments based on binary outcome estimates) and provides an objective correct measure reflective of past studies [11, 30]. Differences between perceptual precision and selection from simple decision making in health risk communication indicates that visualization design may influence decision making behavior.

5.1 Hypotheses

Experiment 2 explores how chart type and composition impact decision-making. Drawing on the results of Experiment 1, we hypothesize:

Juxtaposition decreases drug selection: In Experiment 1, we found significant underestimation with juxtaposition compared to other compositions. Given that the difference may be perceived as smaller with juxtaposition, we anticipate explicit encoding as well as explicit encoding plus juxtaposition will increase drug selection.

Bar charts increase drug selection: Although icon arrays yielded higher accuracy in estimation tasks, this advantage may not translate to treatment selection decisions. In Experiment 1, icon arrays showed significant underestimation for the 20-80 difference range. This pattern suggests that bar charts may create a perception of greater magnitude due to their continuous visual representation, potentially amplifying the perceived effectiveness of the treatments they display.

Explicit encoding plus juxtaposition increases drug selection accuracy: Explicit encoding plus juxtaposition enhanced performance

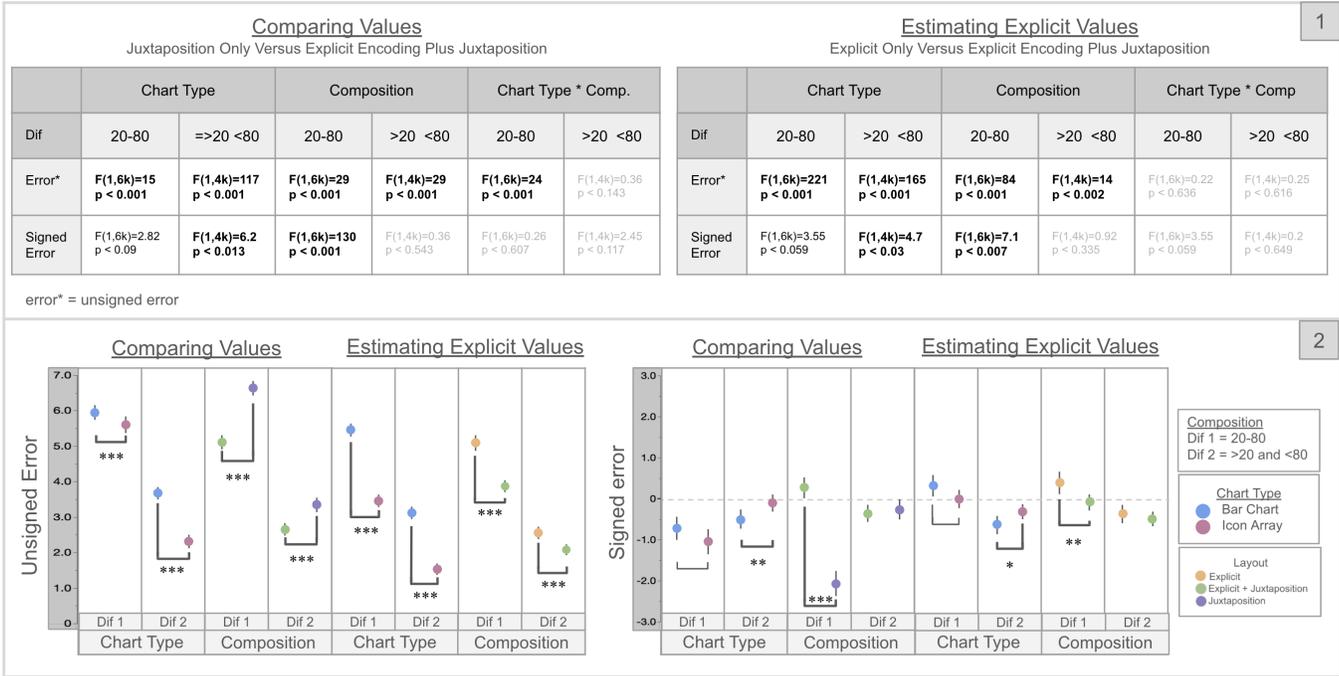


Fig. 5: 1) Results from our ANOVA tests. Bolded text indicates significant findings where ($p < 0.05$), standard text indicates marginally significant findings where ($p < 0.1$), and grey text indicates non-significant effects. 2) Means and error bars with 95% confidence intervals for unsigned error (on a scale from 0 to 7) and signed error (on a scale from -3 to 3, with a dotted line at 0). Unmarked values indicate marginally significant results ($p < 0.1$). The symbol * indicates significance with $p < 0.05$, ** indicates $p < 0.01$, and *** indicates $p < 0.001$.

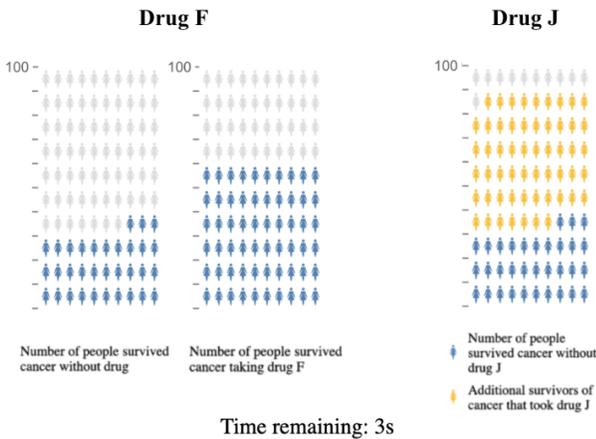


Fig. 6: An example trial for icon array juxtaposition versus icon array explicit encoding showing the efficacy of Drug F and Drug J in Experiment 2. Participants saw the visualizations for 3s.

for participants in Experiment 1. This finding indicates that redundant information may help mitigate overestimation.

5.2 Methods

Our second experiment used a 2 (chart type) x 3 (composition) between-subjects design.

5.2.1 Stimuli

Experiment 2 used the same visualization implementation as Experiment 1, with each stimulus presenting two comparative displays each using one of the three compositions. The displays were separated by 168 pixels and labeled as "Drug F" (left) and "Drug J" (right), corresponding to the input keys used to provide responses (Fig. 6).

Each visualization shows cancer survival before and after treatment. We divided participants into three groups, each seeing two compositions across two chart types. Group A saw juxtaposition and juxtaposition plus explicit encoding, Group B saw juxtaposition and explicit encoding, and Group C explicit encoding plus juxtaposition and explicit encoding. This created four possible visualization combinations within each group (2 chart types x 2 compositions). We generated visualization pairs that differed in at least one design variable. We allocated 10 trials for each of the four visualization pairs showing different compositions, and 5 trials for the two visualization pairs showing the same composition to prevent overrepresentation from shared conditions across the three studies. Each pair was initialized using a random base value from [0, 89]. We then added drug efficacy differences distributed across conditions as follows: 20% of each chart type x composition showed drugs with equal effectiveness, and 20% each at levels with survival differences of 1-5 patients (i.e., 5 more patients survived cancer with one treatment versus the other), 6-15 patients, 15-30 patients, and 30-50 patients (80% total). We balanced the left-right positioning of chart types and compositions, and distributed the more effective drug (higher survival value) equally across all chart types and compositions.

5.2.2 Task

To mirror a typical health risk decision making scenario, participants were asked "Which drug would you choose or recommend?" This task simulates real-world scenarios where people must compare visualizations across different sources or within the same document when making important health-related decisions. Participants selected their preferred drug by pressing either "F" or "J" on their keyboard or clicking corresponding on-screen buttons.

5.2.3 Procedure

Experiment 2 followed the same general protocol as Experiment 1. We adapted tutorial explanations and examples to reflect the stimuli and task, while maintaining the same overall methodology to demonstrate visualization differences and explain the task.

Participants were assigned to one of the three groups to keep the study duration manageable. Each participant completed 50 formal trials presented in a random order plus 3 engagement checks (i.e., efficacy difference of 90) positioned at the 17th, 33rd, and 50th positions in the sequence. In each trial, participants saw two visualizations for 3 seconds before the visualizations were hidden. Deliberate variation in efficacy differences (see Section 4.2.1) in the formal trials included both challenging comparisons (to test design variables impact of perception) and straightforward cases (to reinforce the task objective of selecting based on perceived effectiveness).

5.2.4 Participants:

We recruited a total of 166 participants with the same credentials as Experiment 1 from MTurk, resulting in 56 in group 1, 56 in group 2, and 52 in group 3 that passed all engagement checks. Participants were aged 18-65 years old ($\mu = 40 \pm 9.33$), had an average literacy score of $\mu = 5.56 \pm 1.12$, took about 15 minutes on average, and were compensated for their time.

5.2.5 Measures and Analysis

We analyzed data in three groups, when participants: correctly selected the more effective drug, 2) incorrectly selected the less effective drug, and 3) were shown drugs equally effective. We also considered the same literacy thresholds from Experiment 1. We performed Chi-Squared tests on the distribution of selected chart types and compositions to understand the design's influence on decision making.

5.3 Experiment 2 Results

We found balanced selection patterns when participants answered correctly, but significant imbalances ($\chi < 0.01$) when participants were incorrect or encountered equivalent values. We show imbalanced composition selection in Figure 7.

Incorrect Responses: High-literacy participants more frequently selected the incorrect drug when it was visualized with bar charts (61%). Across all participants, 43% of incorrect responses occurred when the more effective drug was visualized with juxtaposition, compared to only 26% when the effective drug was shown with explicit or explicit-plus-juxtaposition composition. High-literacy participants incorrectly selected the less effective drug more often when it was visualized with explicit encoding (36%) or explicit encoding plus juxtaposition (37%) compared to when it was visualized with juxtaposition alone (27%). Low-literacy participants, on the other hand, frequently misidentified drugs visualized with juxtaposition (42%) as more effective when the truly effective drug was actually visualized using either explicit encoding or explicit encoding plus juxtaposition.

Equivalent Values: When forced to choose between equally effective drugs, participants with low literacy were most likely to choose the drug that used explicit encoding plus juxtaposition (37% of responses). Similar to the incorrect responses, high-literacy participants chose bar charts more frequently than icon arrays (55%), and chose explicit encoding (36%) and explicit plus juxtaposition (37%) more frequently than juxtaposition (27%).

5.4 Experiment 2 Discussion

Our findings reveal important insights into how visualization design influences the interpretation of drug effectiveness data across health literacy levels. We organized our discussion around three key findings.

Bar Chart Perceived as More Effective: In partial support of our hypothesis, bar charts increased drug selection among participants with high data literacy across both incorrect responses and equivalent values. When value comparisons were difficult to estimate quickly, bar charts may have created overestimation or preference bias, potentially due to their familiar and space-filling continuous visual representation, while icons' discrete units created more empty space and a less familiar representation. While designers could strategically leverage bar charts' perceptual amplification to help patients take health risks more seriously, it should be applied with ethical caution, as exaggerating effectiveness differences could potentially trigger unwarranted anxiety or lead to disproportionate risk responses in already vulnerable populations.

Explicit Encoding plus Juxtaposition Perceived as More Effective: In partial support of our hypothesis, we found a bias towards explicit encoding plus juxtaposition, leading people to choose the drug associated with the combined representation disproportionately often both in incorrect decisions (high-literacy) and in circumstances where there was no correct decision. As with bar charts, while designers may leverage this bias to clarify relative differences, they must do so carefully to not be misleading or cause unnecessary anxiety.

Health Literacy Mediates Interpretation of Juxtaposition Encodings: Contrary to our hypothesis, lower literacy participants incorrectly favored juxtaposition unlike higher literacy participants who selected explicit encoding more. This literacy-based difference reveals how visualization interpretation is not universal but mediated by people's experience and abilities, a sentiment expressed by our healthcare collaborators. This finding highlights the need for design approaches grounded in audience characteristics, such as visualization novices [7]. While lower literacy audiences may bias towards juxtaposition, it may not represent the most effective design choice, leading to a potential conflict between preference and performance.

6 OVERALL DISCUSSION

We found that visualization chart type, composition, and literacy background impact people's data comprehension and decision making. Drawing across the two sets of findings, we offer design guidance for risk-based communication in healthcare and related domains. We structure our discussion around crosscutting findings and discuss their ramifications for design.

Guideline 1—Estimation and Decision Making Are Not the Same: Our findings indicate that accuracy in estimating differences in values does not always align with decision-making preferences. The results of Experiment 1 show that people tend to underestimate differences with juxtaposition. However, in Experiment 2, people with lower data literacy were more likely to choose the juxtaposed treatment as being more effective (i.e., having a larger difference) in ways that led to incorrect decisions. A bias towards underestimation in reading a visualization became an overestimation in effectiveness when making decisions between two options. Similarly, we saw that for high literacy, while icon arrays lead to more accurate comparisons, bar charts were more frequently selected for incorrect decisions and in cases where there was no difference in treatment.

This contrast contributes to the existing discussions that judgments and decision-making should be treated differently in visualization research and design [11, 40]. The observed dissociation between estimation accuracy and decision-making challenges the assumption that greater perceptual accuracy automatically translates to better decision outcomes. This tension between what people can accurately perceive versus what they prefer to use when making consequential choices suggests that visualization effectiveness must be evaluated through multiple lenses—not just perceptual precision [43] and that designers should keep task in mind when designing for health communication.

Guideline 2—Icon Arrays Increase Accuracy: Icon arrays consistently outperformed bar charts in estimation accuracy. This suggests that while bar charts are more commonly used across domains due to their familiarity [56], icon arrays can offer advantages in accurate interpretation, whereas bar charts introduced potential selection bias leading to inaccurate responses. Our findings demonstrate that discrete representations like icon arrays can improve comparison accuracy. While we situate our analyses in health risk communication, the lack of units in the data in Experiment 1 suggest these benefits may generalize to broader public communication such as environmental risk assessment, financial decision-making, earth science, and educational outcomes reporting. Designers should consider implementing icon arrays to improve comparison accuracy and avoid potential overestimation bias in bar charts.

Guideline 3—Combining Compositions Improves Accuracy and Decision Making: Work in visual comparison has proposed theoretical advantages combined comparative designs [23, 24]. We found that explicit encoding plus juxtaposition enhanced estimation accuracy over either composition alone and were preferred in selection tasks, even

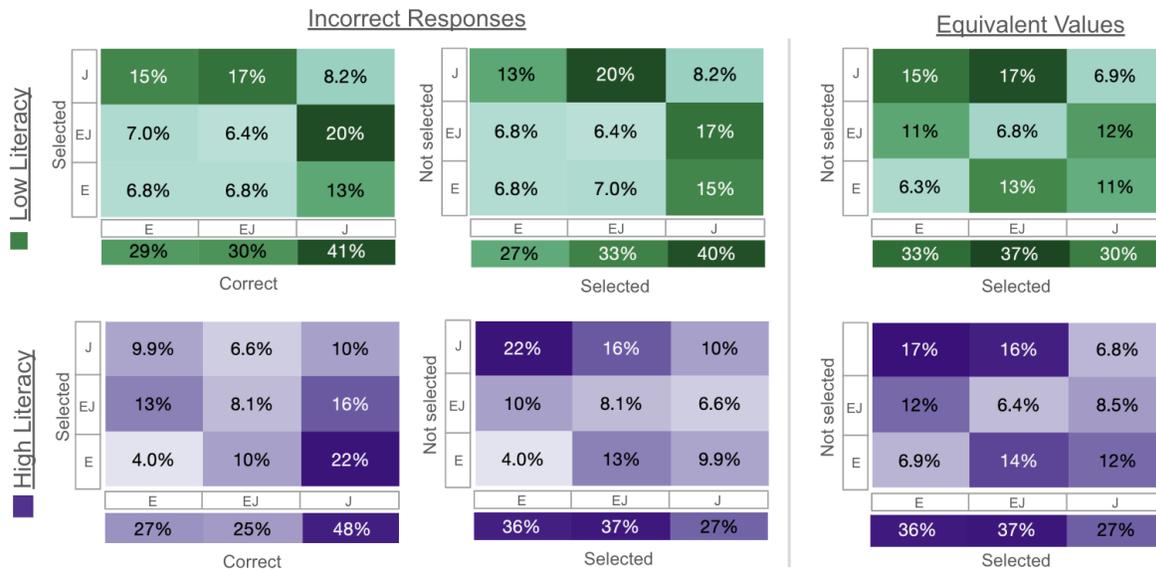


Fig. 7: Composition preferences matrix showing what the correct layout was when participants were incorrect compared to what they selected, and the percentage of cases where participants selected column compositions over row compositions for incorrect and equivalent values responses. Totals below indicate overall selection rates for each composition type across conditions.

though the combined design offers no additional information. Juxtaposition alone is more common in health risk communication, yet it led to overall underestimation and inaccurate drug selection for the lower literacy group. Given that all participants were less accurate when the more effective drug was encoded with juxtaposition, designers should consider communicating with the combined explicit encoding plus juxtaposition over just juxtaposition alone for both precise understanding and effective decision-making are essential. However, such choices may impart exaggeration biases that could shift patient comprehension and decision making.

Guideline 4—Literacy Impacts Decision Making: Numeric and graph literacy levels influenced how participants made health risk decisions with different visualization compositions. Unlike participants with high literacy, participants with lower literacy scores incorrectly chose the juxtaposition composition more frequently. In combination with the previous guidelines, this indicates that individuals with lower data visualization literacy may be drawn to simpler, side-by-side comparisons, but those comparisons may be more error prone. These results reflect past work demonstrating novices in visualization may have different behaviors [7], indicating that designers should tailor visualizations to the needs and abilities of the target audience.

7 LIMITATIONS & FUTURE WORK

Our experiments provide preliminary design recommendations for comparative visualization design in health risk communication. We selected parameters for bar charts and icon arrays to maintain consistency across experimental conditions. However, these design choices represent only a subset of possible visual configurations and may impact perceptual outcomes [55]. Future research could investigate how alternative chart types, such as raw numbers or line graphs, or additional design variables might modulate difference estimation, including layout parameters like visualization spacing, icon choice (e.g., female versus male representation), and the number of icons per row. Other design techniques in addition to chart type and arrangement, such as color, annotation, or interactive elements, may guide attention to key relationships within multi-dimensional datasets.

We focused on comparing two probabilities, a common task in health risk communication that allows for strong experimental controls while reflecting real-world needs. However, this represents only a subset of the comparative tasks healthcare communicators regularly use. Future studies should investigate how our findings translate to comparisons

across greater numbers of probabilities, multi-dimensional comparisons, such as tracking risk trajectories across multiple years or comparing risk profiles across a spectrum of cancer types, and a greater range of decision-making paradigms [28]. These complex comparison scenarios may require additional design considerations. Alternative visualization techniques, like line charts for temporal data or small multiples for categorical comparisons, may offer distinct advantages for these more complex comparative tasks. While we found benefits with using icon arrays, these benefits may not scale to multiple comparisons, given the potential for visual clutter. Further, we only tested comparisons between different chart types to manage the scope of our study and directly compare representations. Future work should consider comparing the same chart type and composition to see if certain designs elicit more accurate drug selection within the same paradigm. Beyond design variables, future research should consider systematically examining experimental parameters that affect generalizability, such as task timing, task framing, tutorial framing, and number of overall comparisons, which we tuned based on prior work and piloting.

Participants to made simple decisions in a hypothetical scenario with no real-world consequences for incorrect decisions. This approach allowed us to assess decision making without potential confound from immediate personal risk [48] but provides a reductionist view on decision making that may limit its generalizability to some scenarios. Patients must consider a range of factors in their decisions and their risk tolerance may vary [22]. While directly evaluating communication bias in scenarios with significant consequences raises ethical challenges, future work should explore how these results generalize to real-world decisions in collaboration with healthcare professionals.

8 CONCLUSION

We examined how chart types (icon arrays vs. bar charts) and compositions (juxtaposition, explicit encoding, and their combination) affect perception and decision-making in health risk communication. Our findings revealed that icon arrays and compositions using explicit encoding plus juxtaposition supported both accurate estimation and decision making, while juxtaposition was less accurate in estimation and led to more incorrect treatment selection for people with lower literacy. These insights provide healthcare communicators with design guidance for effective risk visualizations that may impact patient comprehension, decision quality, and ultimately health outcomes.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health Award 1R01HD111074-01 and NSF-IIS #2320920.

REFERENCES

- [1] S. D. Anker, I. A. Gillespie, K.-U. Eckardt, F. Kronenberg, S. Richards, T. B. Drueke, P. Stenvinkel, R. L. Pisoni, B. M. Robinson, D. Marcelli, et al. Development and validation of cardiovascular risk scores for haemodialysis patients. *International Journal of Cardiology*, 216:68–77, 2016. doi: [10.1016/j.ijcard.2016.04.151](https://doi.org/10.1016/j.ijcard.2016.04.151) 2
- [2] J. Bertin. Sémiologie graphique, 1967. 1
- [3] C. Bonner, L. J. Trevena, W. Gaissmaier, P. K. J. Han, Y. Okan, E. Ozanne, E. Peters, D. Timmermans, and B. J. Zikmund-Fisher. Current Best Practice for Presenting Probabilities in Patient Decision Aids: Fundamental Principles. *Medical Decision Making*, 41(7):821–833, 2021. doi: [10.1177/0272989X21996328](https://doi.org/10.1177/0272989X21996328) 1, 2, 3
- [4] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What Makes a Visualization Memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2306–2315, 2013. doi: [10.1109/TVCG.2013.234](https://doi.org/10.1109/TVCG.2013.234) 3
- [5] M. Bostock, V. Ogievetsky, and J. Heer. D³ Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011. doi: [10.1109/TVCG.2011.185](https://doi.org/10.1109/TVCG.2011.185) 4
- [6] M. J. Burnham, Y. K. Le, and R. L. Piedmont. Who is mturk? personal characteristics and sample consistency of these online workers. *Mental Health, Religion & Culture*, 21(9-10):934–944, 2018. doi: [10.1080/13674676.2018.1554199](https://doi.org/10.1080/13674676.2018.1554199) 5
- [7] A. Burns, C. Lee, R. Chawla, E. Peck, and N. Mahyar. Who do we mean when we talk about visualization novices? In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, 2023. article no. 819, 16 pages. doi: [10.1145/3544548.3581524](https://doi.org/10.1145/3544548.3581524) 8, 9
- [8] K. Böcherer-Linder and A. Eichler. How to Improve Performance in Bayesian Inference Tasks: A Comparison of Five Visualizations. *Frontiers in Psychology*, 10, 2019. doi: [10.3389/fpsyg.2019.00267](https://doi.org/10.3389/fpsyg.2019.00267) 2
- [9] J. J. Caban and D. Gotz. Visual analytics in healthcare—opportunities and research challenges. *Journal of the American Medical Informatics Association*, 22(2):260–262, 2015. doi: [10.1093/jamia/ocv006](https://doi.org/10.1093/jamia/ocv006) 1
- [10] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984. doi: [10.1080/01621459.1984.10478080](https://doi.org/10.1080/01621459.1984.10478080) 2, 4
- [11] E. Dimara, A. Bezerianos, and P. Dragicevic. Narratives in crowdsourced evaluation of visualizations: A double-edged sword? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 5475–5484. ACM, 2017. doi: [10.1145/3025453.3025870](https://doi.org/10.1145/3025453.3025870) 6, 8
- [12] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic. A task-based taxonomy of cognitive biases for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 26(2):1413–1432, 2020. doi: [10.1109/TVCG.2018.2872577](https://doi.org/10.1109/TVCG.2018.2872577) 1
- [13] P. Dragicevic. Fair statistical communication in hci. In *Modern Statistical Methods for HCI*, pp. 291–330. Springer, 2016. doi: [10.1007/978-3-319-26633-6_13](https://doi.org/10.1007/978-3-319-26633-6_13) 5
- [14] D. M. Eddy. Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, and A. Tversky, eds., *Judgment under Uncertainty: Heuristics and Biases*, pp. 249–267. Cambridge University Press, 1982. doi: [10.1017/CBO9780511809477.019](https://doi.org/10.1017/CBO9780511809477.019) 2
- [15] A. Fagerlin, B. J. Zikmund-Fisher, and P. A. Ubel. Helping patients decide: ten steps to better risk communication. *Journal of the National Cancer Institute*, 103(19):1436–1443, 2011. doi: [10.1093/jnci/djr318](https://doi.org/10.1093/jnci/djr318) 2
- [16] D. Feldman-Stewart, M. D. Brundage, and V. Zotov. Further Insight into the Perception of Quantitative Information: Judgments of Gist in Treatment Decisions. *Medical Decision Making*, 27(1):34–43, 2007. doi: [10.1177/0272989X06297101](https://doi.org/10.1177/0272989X06297101) 2, 3
- [17] D. Feldman-Stewart, N. Kocovski, B. A. McConnell, M. D. Brundage, and W. J. Mackillop. Perception of quantitative information for treatment decisions. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, 20(2):228–238, 2000. doi: [10.1177/0272989X0002000208](https://doi.org/10.1177/0272989X0002000208) 2, 3
- [18] FORCE. Cancer Risk for People with a BRCA1 Mutation — facingourrisk.org. <https://www.facingourrisk.org/info/hereditary-cancer-and-genetic-testing/>
- [19] S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, and J. Hullman. The science of visual data communication: What works. *Psychological Science in the Public Interest*, 22(3):110–161, 2021. doi: [10.1177/15291006211051956](https://doi.org/10.1177/15291006211051956) 1, 3
- [20] M. Galesic and R. Garcia-Retamero. Graph Literacy: A Cross-Cultural Comparison. *Medical Decision Making*, 31(3):444–457, 2011. doi: [10.1177/0272989X10373805](https://doi.org/10.1177/0272989X10373805) 4
- [21] R. Garcia-Retamero and E. T. Cokely. Communicating Health Risks With Visual Aids. *Current Directions in Psychological Science*, 22(5):392–399, 2013. doi: [10.1177/0963721413491570](https://doi.org/10.1177/0963721413491570) 1, 2, 3
- [22] M. Gerrard, F. X. Gibbons, and M. Reis-Bergan. The effect of risk communication on risk perceptions: the significance of individual differences. *JNCI Monographs*, 1999(25):94–100, 1999. doi: [10.1093/oxfordjournals.jncimonographs.a024217](https://doi.org/10.1093/oxfordjournals.jncimonographs.a024217) 9
- [23] M. Gleicher. Considerations for visualizing comparison. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):413–423, 2018. doi: [10.1109/TVCG.2017.2743961](https://doi.org/10.1109/TVCG.2017.2743961) 1, 3, 8
- [24] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011. doi: [10.1177/1473871611416549](https://doi.org/10.1177/1473871611416549) 1, 2, 3, 8
- [25] D. Gotz and D. Borland. Data-driven healthcare: challenges and opportunities for interactive visualization. *IEEE Computer Graphics and Applications*, 36(3):90–96, 2016. doi: [10.1109/MCG.2016.59](https://doi.org/10.1109/MCG.2016.59) 2
- [26] S. Haroz, R. Kosara, and S. L. Franconeri. Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1191–1200. ACM, 2015. doi: [10.1145/2702123.2702275](https://doi.org/10.1145/2702123.2702275) 2, 3, 4
- [27] S. Haroz and D. Whitney. How capacity limits of attention influence information visualization effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2422–2431, 2012. doi: [10.1109/TVCG.2012.213](https://doi.org/10.1109/TVCG.2012.213) 3
- [28] J. Hullman, A. Kale, and J. Hartline. Decision theoretic foundations for experiments evaluating human decisions. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2025. article no. 254, 15 pages. doi: [10.1145/3706598.3714063](https://doi.org/10.1145/3706598.3714063) 6, 9
- [29] N. Jardine, B. Ondov, N. Elmqvist, and S. Franconeri. The perceptual proxies of visual comparison. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):839–849, 2020. doi: [10.1109/TVCG.2019.2934786](https://doi.org/10.1109/TVCG.2019.2934786) 3
- [30] A. Kale, Y. Wu, and J. Hullman. Causal support: modeling causal inferences with visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1150–1160, 2022. doi: [10.1109/TVCG.2021.3114820](https://doi.org/10.1109/TVCG.2021.3114820) 6
- [31] E. L. Kaufman, M. W. Lord, T. W. Reese, and J. Volkman. The discrimination of visual number. *The American Journal of Psychology*, 62(4):498–525, 1949. doi: [10.2307/1418556](https://doi.org/10.2307/1418556) 2, 4
- [32] K. B. Kuchenbaecker, J. L. Hopper, D. R. Barnes, K.-A. Phillips, T. M. Mooij, M.-J. Roos-Blom, S. Jervis, F. E. Van Leeuwen, R. L. Milne, N. Andrieu, et al. Risks of breast, ovarian, and contralateral breast cancer for brca1 and brca2 mutation carriers. *JAMA*, 317(23):2402–2416, 2017. doi: [10.1001/jama.2017.7112](https://doi.org/10.1001/jama.2017.7112) 1, 2, 3
- [33] K. B. Kuchenbaecker, L. McGuffog, D. Barrowdale, A. Lee, P. Soucy, J. Dennis, S. M. Domchek, M. Robson, A. B. Spurdle, S. J. Ramus, et al. Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in brca1 and brca2 mutation carriers. *JNCI: Journal of the National Cancer Institute*, 109(7):djw302, 2017. doi: [10.1093/jnci/djw302](https://doi.org/10.1093/jnci/djw302) 2, 3
- [34] A. W. Kurian, D. F. Munoz, P. Rust, E. A. Schackmann, M. Smith, L. Clarke, M. A. Mills, and S. K. Plevritis. Online tool to guide decisions for BRCA1/2 mutation carriers. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 30(5):497–506, 2012. doi: [10.1200/JCO.2011.38.6060](https://doi.org/10.1200/JCO.2011.38.6060) 1, 2, 3
- [35] K. Nagaya and H. Shimizu. Probability overestimation induced by icon arrays. *Acta Psychologica*, 248, 2024. article no. 104352. doi: [10.1016/j.actpsy.2024.104352](https://doi.org/10.1016/j.actpsy.2024.104352) 2, 4
- [36] E. M. Nazzal, A. M. Deal, B. Borgert, H. Heiling, A. V. Bennett, S. Blalock, W. Meeks, R. Fang, R. Teal, M. B. Vu, et al. Use and usefulness of risk prediction tools in urologic surgery: Current state and path forward. *Urology Practice*, 12(4):459–468, 2025. doi: [10.1097/UJP.0000000000000808](https://doi.org/10.1097/UJP.0000000000000808) 2

- [37] E. M. Nazzal, D. Gotz, H. Heiling, A. M. Deal, K. Giannone, D. Usinger, S. Blalock, A. V. Bennett, M. E. Nielsen, D. S. Reuland, et al. Pd40-05 gist overrides numeric risk perception in surgical decision-making. *The Journal of Urology*, 211(5S):article no. e821, 2024. doi: 10.1097/01.JU.0001009356.04608.d5.05 2
- [38] Y. Okan, M. Galesic, and R. Garcia-Retamero. How People with Low and High Graph Literacy Process Health Graphs: Evidence from Eye-tracking. *Journal of Behavioral Decision Making*, 29(2-3):271–294, 2016. doi: 10.1002/bdm.1891 4
- [39] B. Ondov, N. Jardine, N. Elmqvist, and S. Franconeri. Face to face: Evaluating visual comparison. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):424–434, 2019. doi: 10.1109/TVCG.2018.2864884 3
- [40] B. Oral, P. Dragicevic, A. Telea, and E. Dimara. Decoupling judgment and decision making: A tale of two tails. *IEEE Transactions on Visualization and Computer Graphics*, 30(10):6928–6940, 2024. doi: 10.1109/TVCG.2023.3346640 8
- [41] A. Otley, E. M. Peck, L. T. Harrison, D. Afergan, C. Ziemkiewicz, H. A. Taylor, P. K. J. Han, and R. Chang. Improving bayesian reasoning: The effects of phrasing, visualization, and spatial ability. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):529–538, 2016. doi: 10.1109/TVCG.2015.2467758 2
- [42] G. J. Quadri and P. Rosen. A survey of perception-based visualization studies by task. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4893–4911, 2022. doi: 10.1109/TVCG.2021.3098240 2
- [43] G. J. Quadri, A. Z. Wang, Z. Wang, J. Adorno, P. Rosen, and D. A. Szafr. Do you see what i see? a qualitative study eliciting high-level visualization comprehension. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. ACM, 2024. article no. 204, 16 pages. doi: 10.1145/3613904.3642813 3, 8
- [44] R. Royak-Schaler, D. E. Blocker, A. M. Yali, M. Bynoe, K. J. Briant, and S. Smith. Breast and colorectal cancer risk communication approaches with low-income African-American and Hispanic women: implications for healthcare providers. *Journal of the National Medical Association*, 96(5):598–608, May 2004. 2
- [45] B. Saket, A. Endert, and Ç. Demiralp. Task-based effectiveness of basic visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(7):2505–2512, 2019. doi: 10.1109/TVCG.2018.2829750 2
- [46] P. Scalia, D. C. Schubbe, E. S. Lu, M.-A. Durand, J. Frascara, G. Noel, A. J. O’Malley, and G. Elwyn. Comparing the impact of an icon array versus a bar graph on preference and understanding of risk information: Results from an online, randomized study. *PLOS ONE*, 16(7):e0253644, July 2021. Publisher: Public Library of Science. doi: 10.1371/journal.pone.0253644 4, 5
- [47] M. Schwartz and P. C. Dixon. The effect of subject measurement error on joint kinematics in the conventional gait model: Insights from the open-source pyCGM tool using high performance computing methods. *PLOS ONE*, 13(1):e0189984, 2018. doi: 10.1371/journal.pone.0189984 4
- [48] P. Slovic, E. Peters, M. L. Finucane, and D. G. MacGregor. Affect, risk, and decision making. *Health Psychology*, 24(4S):S35–S40, 2005. doi: 10.1037/0278-6133.24.4.S35 5, 9
- [49] D. A. Szafr. The good, the bad, and the biased: Five ways visualizations can mislead (and how to fix them). *Interactions*, 25(4):26–33, 2018. doi: 10.1145/3231772 2
- [50] J. Talbot, V. Setlur, and A. Anand. Four experiments on the perception of bar charts. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2152–2160, 2014. doi: 10.1109/TVCG.2014.2346320 1, 2
- [51] H.-J. Tan, D. Gotz, H. Heiling, A. Deal, K. Giannone, D. Usinger, S. Blalock, A. Bennett, M. Nielsen, D. Reuland, et al. Visual and numeric risk information reduces variation in surgeon risk perception: results from a randomized, clinical vignette experiment. *Journal of Urology*, 209(Supplement 4):e911, 2023. doi: 10.1097/JU.0000000000003325.12 1, 2
- [52] C. Tseng, G. J. Quadri, Z. Wang, and D. A. Szafr. Measuring categorical perception in color-coded scatterplots. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, 2023. article no. 204, 20 pages. doi: 10.1145/3544548.3581416 3
- [53] H. G. Welch and W. C. Black. Overdiagnosis in Cancer. *JNCI: Journal of the National Cancer Institute*, 102(9):605–613, May 2010. doi: 10.1093/jnci/djq099 2
- [54] C. Xiong, C. Ceja, C. Ludwig, and S. Franconeri. Biased average position estimates in line and bar graphs: Underestimation, overestimation, and perceptual pull. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):875–885, 2020. doi: 10.1109/TVCG.2019.2934400 2
- [55] C. Xiong, A. Sarvghad, D. G. Goldstein, J. M. Hofman, and Ç. Demiralp. Investigating perceptual biases in icon arrays. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. ACM, 2022. article no. 137, 12 pages. doi: 10.1145/3491102.3501874 1, 2, 4, 9
- [56] S. Yi, J. Jo, and J. Seo. Comparative layouts revisited: Design space, guidelines, and future directions. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1525–1535, 2021. doi: 10.1109/TVCG.2020.3030419 2, 3, 8
- [57] L. E. Zajac, W. M. Klein, and K. D. McCaul. Absolute and comparative risk perceptions as predictors of cancer worry: Moderating effects of gender and psychological distress. *Journal of Health Communication*, 11(S1):37–49, 2006. doi: 10.1080/10810730600637301 2
- [58] Z. Zeng and L. Battle. A review and collation of graphical perception knowledge for visualization recommendation. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, 2023. article no. 820, 16 pages. doi: 10.1145/3544548.3581349 2
- [59] B. J. Zikmund-Fisher, A. Fagerlin, and P. A. Ubel. “is 28% good or bad?” evaluability and preference reversals in health care decisions. *Medical Decision Making*, 24(2):142–148, 2004. doi: 10.1177/0272989X04263154 2
- [60] B. J. Zikmund-Fisher, P. A. Ubel, D. M. Smith, H. A. Derry, J. B. McClure, A. Stark, R. K. Pitsch, and A. Fagerlin. Communicating Side Effect Risks in a Tamoxifen Prophylaxis Decision Aid: The Debiasing Influence of Pictographs. *Patient education and counseling*, 73(2):209–214, Nov. 2008. doi: 10.1016/j.pec.2008.05.010 2, 3
- [61] B. J. Zikmund-Fisher, H. O. Witteman, M. Dickson, A. Fuhrel-Forbis, V. C. Kahn, N. L. Exe, M. Valerio, L. G. Holtzman, L. D. Scherer, and A. Fagerlin. Blocks, ovals, or people? icon type affects risk perceptions and recall of pictographs. *Medical Decision Making*, 34(4):443–453, 2014. doi: 10.1177/0272989X13511706 1, 2
- [62] B. J. Zikmund-Fisher and H. O. Witteman. Icon Array — iconarray.com. <https://www.iconarray.com/>. 2