American Psychological Association

Journal of Experimental Psychology: Applied

Manuscript version of

Number Lines Can Be More Effective at Facilitating Adults' Performance on Health-Related Ratio Problems Than Risk Ladders and Icon Arrays

Marta K. Mielicki, Charles J. Fitzsimmons, Lauren K. Schiller, Dan Scheibe, Jennifer M. Taber, Pooja G. Sidney, Percival G. Matthews, Erika A. Waters, Karin G. Coifman, Clarissa A. Thompson

Funded by:
- US Department of Education, Institute of Education Sciences

Number Lines Can be More Effective at Facilitating Adults' Performance on Health-Related Ratio Problems than Risk Ladders and Icon Arrays

Marta K. Mielicki[a]

Charles J. Fitzsimmons[a]

Lauren K. Schiller[b]

Dan Scheibe[a]

Jennifer M. Taber[a]

Pooja G. Sidney[c]

Percival G. Matthews[d]

Erika A. Waters[e]

Karin G. Coifman[a]

Clarissa A. Thompson[a]

[a]Department of Psychological Sciences, Kent State University

[b]Department of Human Development, Teachers College, Columbia University

[c]Department of Psychology, University of Kentucky

[d]Department of Educational Psychology, University of Wisconsin – Madison

[e]Division of Public Health Sciences, Washington University School of Medicine in St. Louis

Author Note

This research was funded by the Department of Education Institute of Education Sciences (R305U200004). Correspondence concerning this article should be addressed to Marta K. Mielicki, Rutgers University Center for Cognitive Science, 152 Frelinghuysen Road Piscataway, NJ 08854-8020. Email: marta.mielicki@rutgers.edu The study design, hypotheses,

and analytic plan were preregistered on OSF (https://osf.io/7bfmy). The data, analytic code, and materials are available on OSF (https://osf.io/93m74/).

Abstract

Visual displays, such as icon arrays and risk ladders, are often used to communicate numerical health information. Number lines improve reasoning with rational numbers but are seldom used in health contexts. College students solved ratio problems related to COVID-19 (e.g., number of deaths and number of cases) in one of four randomly-assigned conditions: icon arrays, risk ladders, number lines, or no accompanying visual display. As predicted, number lines facilitated performance on these problems – the number line condition outperformed the other visual display conditions, which did not perform any better than the no visual display condition. In addition, higher performance on the health-related ratio problems was associated with higher COVID-19 worry for oneself and others, higher perceptions of COVID-19 severity, and higher endorsement of intentions to engage in preventive health behaviors, even when controlling for baseline math skills. These findings have important implications for effectively presenting health statistics.

Keywords: number lines, visual displays, numerical cognition, health decision making

Public Significance Statement

In Fall 2020, college students who had just returned to campus after the COVID-19 shutdown were better able to solve health-related ratio problems (e.g., comparing the number of deaths and number of cases of COVID-19) when they were paired with number lines than when they were paired with other visual displays commonly used to communicate health statistics (i.e., icon arrays and risk ladders). Number lines are seldom used in health contexts but may be a valuable tool for communicating health-related ratios.

Number Lines Can be More Effective at Facilitating Adults' Performance on Health-Related Ratio Problems than Risk Ladders and Icon Arrays

In March-April 2020, many in the U.S. believed that COVID-19 was less fatal than the seasonal flu because fewer people had died from COVID-19 (Thompson et al., 2020). This belief may have stemmed from failing to consider deaths *relative* to infections. From a public health perspective, believing that COVID-19 was less fatal could have been detrimental at the time if it caused people to question recommendations for COVID-19 prevention (e.g., mask wearing).

As of March 24th, 2020, 227,743 people worldwide were infected with COVID-19, and 9,318 of those had died (Johns Hopkins University and Medicine, 2020). In contrast, 36,000,000 people had been infected with the flu, and 22,000 of those had died (Centers for Disease Control and Prevention, 2020). To interpret which disease is more fatal, one must consider the magnitude of the ratios[1] between the numerators (deaths) and denominators (cases) for each disease. The ratio of COVID-19 deaths to infections was much larger (9,318/227,743 = .04) than for the flu (22,000/36,000,000 = .0006).

Reasoning about rational numbers – e.g., fractions, percentages, decimals, and whole-number frequencies – is often impeded by *whole number bias* (WNB: Alibali & Sidney, 2015; DeWolf & Vosniadou, 2015; Ni & Zhou, 2005). WNB occurs when people incorrectly overextend their whole-number knowledge to all rational numbers. This is the type of reasoning that likely contributed to the prevalent belief early in the pandemic that the flu was more fatal than COVID-19 (e.g., Faust, 2020; Faust & del Rio, 2020a; Faust & del Rio, 2020b; Rettner, 2020; Walker, 2020; Yan, 2020). At this time, information about COVID-19 deaths and cases was presented in the media and compared to flu deaths and cases, though these quantities were

[1] In this paper, a ratio a:b and its fraction or magnitude representation a/b are used interchangeably, thus the two quantities a and b are sometimes referred to as numerator and denominator, respectively.

generally not discussed as ratios (Thompson et al., 2020). The primary goals of the current study were 1) to investigate whether various visual displays (number lines, risk ladders, or icon arrays) facilitate more accurate performance on health-related ratio problems relative to no visual display at all, and 2) to investigate the relation between accuracy on health-related ratio problems, use of whole number bias strategies, and health cognitions and behaviors.

Nearly 30 years of research has shown that well-designed visual displays, including bar graphs, icon arrays, and risk ladders can be effective tools for communicating health risk information to patients and physicians (Lipkus, 2007; Lipkus & Hollands, 1999; Mazur & Hickam, 1993; Weinstein et al., 1994). Indeed, visual displays can often foster comprehension of risk information among people from a variety of socio-demographic characteristics, including those with limited numeracy (Garcia-Retamero & Cokely, 2013), reduce effects of gain/loss framing, and prevent denominator neglect (Garcia-Retamero & Cokely, 2017). Visual displays are increasingly being used to provide patients with personal health information, including information intended to be used to facilitate self-management of chronic disease (Turchioe et al., 2019). Several recent articles provide concrete recommendations about how and when to include visual displays to support health decision making, such as in patient decision aids or in personalized health risk calculators (Bonner et al, 2021; Trevena et al., 2021; Waters et al., 2020).

In non-health contexts, minor perceptual differences – such as whether displays include countable objects or continuous magnitudes – have been shown to impact how people reason about ratios (Begolli et al., 2020; Boyer et al., 2008). Number lines are a critical instructional tool (Gunderson et al., 2019; Hamdan & Gunderson, 2017; Moss & Case, 1999; Siegler et al., 2011); yet, they have rarely been implemented in the field of health-decision making (cf., Hawley et al., 2008).

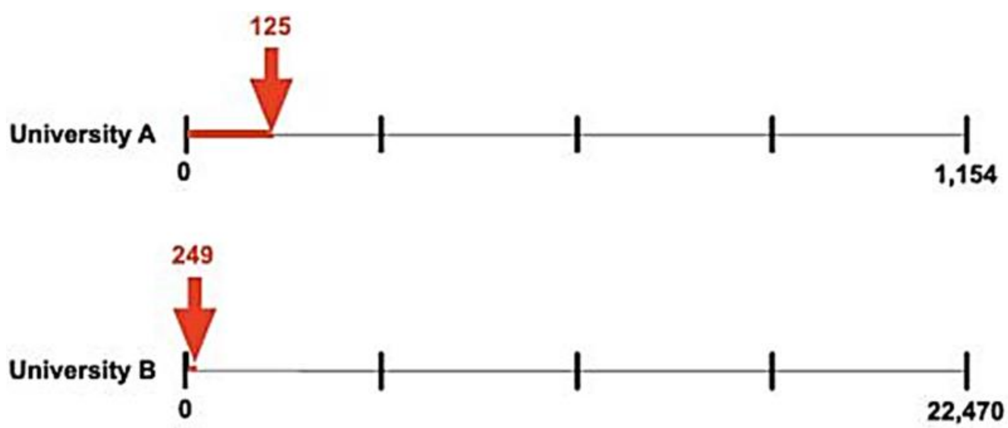
Figure 1

Examples of Each Visual Display

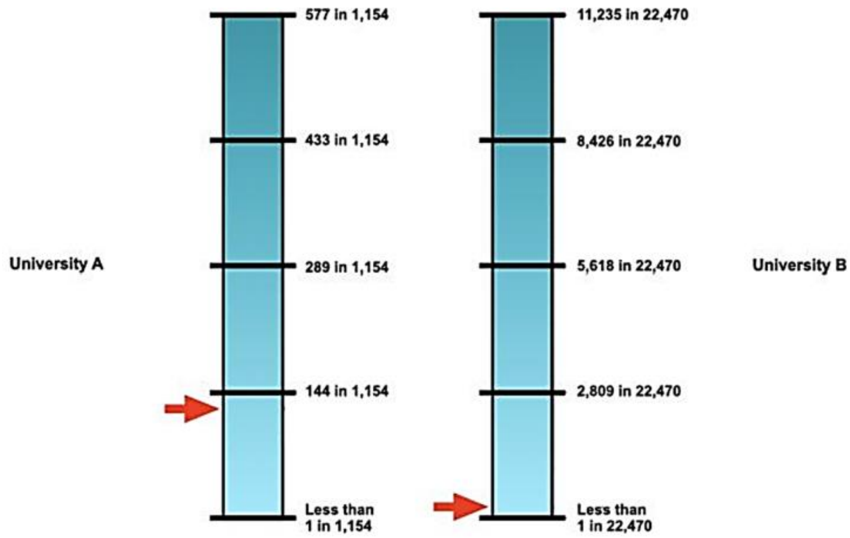
Panel A

As of early September, University A in the U.S. had tested 1,154 total students for COVID-19. At University A, 125 students had lab-confirmed cases of COVID-19 or tests that came back "positive." Another university in another part of the U.S., University B, had tested 22, 470 total students for COVID-19. At University B, 249 students had lab confirmed cases of COVID-19, or tests that came back "positive."

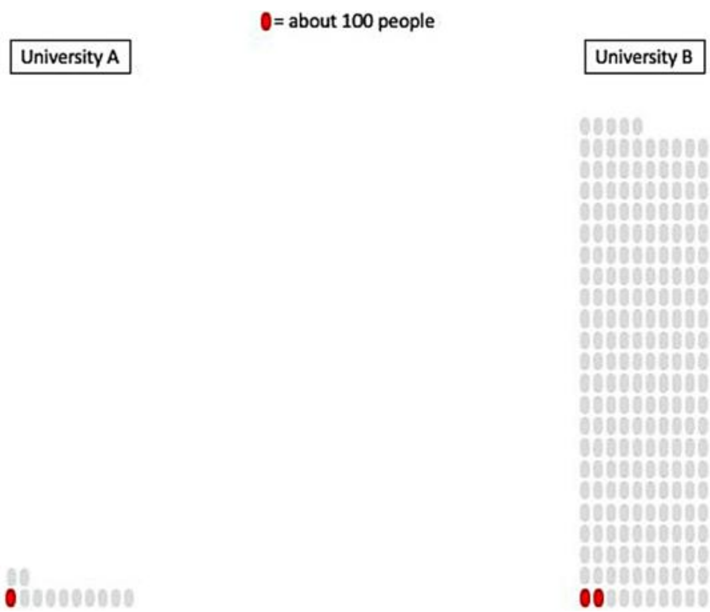
Panel B



Panel C



Panel D



Note. Participants were randomly assigned to a number line (Panels A and B), risk ladder (Panels A and C), icon array condition (Panels A and D), or to a no visual display (only Panel A) condition for each problem in Table 1. The visual displays in this figure depicted the rates of positive COVID-19 tests across two universities.

Number line interventions have helped people reason with numbers (Fazio, Kennedy, & Siegler, 2016; Fuchs et al., 2013; 2014; Moss & Case, 1999; Rittle-Johnson et al., 2001; Saxe et al., 2013; Schneider et al., 2009), and fractions specifically (Gunderson et al., 2019; Hamdan & Gunderson, 2017; Sidney, Thompson, & Rivera, 2019). Number lines are a central conceptual structure in mathematics (Moss & Case, 1999; Okamoto & Case, 1996). According to the integrated theory of whole number and fraction development (Siegler et al., 2011), what unites whole numbers and fractions is that their magnitudes can be represented on a continuous number line. Several features of number lines may contribute to their effectiveness as an instructional tool. Number lines capitalize on spatial-numeric associations because smaller numbers are represented on the left and larger numbers are represented on the right side of the mental number line (Dehaene, 2011). Even preschoolers demonstrate sensitivity to these spatial-numeric associations (Opfer et al., 2010). Number lines bridge perceptually privileged non-symbolic representations of rational-number magnitudes (Rau and Matthews, 2017) and symbolic fractions expressed as Arabic numerals (Sidney et al., 2017). People of all ages (Bhatia et al. 2020; Jacob et al., 2012; Kalra et al., 2020; Matthews et al., 2015; Park et al., 2021) and non-human primates (Vallentin & Nieder, 2008) can compare non-symbolic line ratios faster than symbolic ratios, suggesting that accurately perceiving ratios represented by line segments may be a universal, innately supported capacity. For the reasons outlined above, number lines could be a promising tool for communicating health statistics, particularly in cases where two ratios with different denominators are being compared (as was the case early in the pandemic, when ratios of deaths to cases were commonly displayed). It remains an open question whether number lines can effectively communicate numeric health-risk information, given that other types of visuals – risk ladders and icon arrays – are more commonly used in these contexts (Bonner et al., 2021; Janssen et al., 2018; Trevena et al., 2021; Waters et al., 2021).

Risk ladders (see Figure 1, Panel C) are vertically-oriented visual displays of disease risk. Although, like number lines, they represent magnitude as a linear quantity, risk ladders may be less effective at conveying COVID-19 risk magnitudes (e.g., case-fatality rates, infection rates, etc.) than number lines for several reasons. First, vertically-oriented risk ladders may not capitalize on the horizontally-oriented mental number line (Dehaene, 2011). Second, risk ladders feature an additional area dimension, which may impede rational number reasoning relative to unidimensional number lines (Gunderson et al., 2019). Third, the "rungs" of risk ladders are sometimes labeled with numeric information (e.g., Janssen et al., 2018; Sandman et al., 1993), which could impede reasoning with rational numbers (Siegler & Thompson, 2014). Finally, risk ladders sometimes spatially represent intervals non-linearly (e.g., Janssen et al. 2018), with smaller differences in risk appearing equivalent to larger differences in risk.

Another common visual display in health communications is the icon array (Figure 1, Panel D) which includes discrete, countable icons (Galesic et al., 2009; Hess et al., 2011; Zikmund-Fisher et al., 2010). One potential advantage of icon arrays is that they may facilitate translation to percentages, which are likely more familiar (e.g., a file loading from 0% to 100%) and intuitive than other rational numbers (Fitzsimmons, Woodbury, et al., 2022; Moss & Case, 1999; Schiller, 2020). However, icon arrays led to a higher prevalence of WNB than Arabic numerals or number lines for parents estimating the likelihood of side effects from taking a hypothetical drug (Fitzsimmons, Woodbury, et al., 2022). The countable units in icon arrays may elicit WNB if people are more likely to use incorrect whole-number strategies when a visual display is segmented into countable parts as compared to a continuous representation (Boyer et al., 2008).

In the current study, college students were randomly assigned to complete several health-related ratio problems (based on those used by Thompson et al., 2021, see Figure 1) with either number lines, icon arrays, risk ladders, or no accompanying visual display. We explored whether number lines, which are not commonly used to communicate health statistics, could help people reason about the magnitude of health risks more effectively than other visual displays or no display at all.

Given that a number of factors likely influence performance on health-related ratio problems, we sought to test whether the effects of visual displays on health-related ratio problem performance would be present even when controlling for objective and subjective math skills. Objective math ability is a relevant covariate for the current study because college students with high math ability will likely also be able to solve health-related ratio problems more effectively than students with lower math ability. For objective math ability, we included measures to specifically account for individual differences in ability to reason with ratios. Based on prior work linking objective math skills with health-related problem solving (Thompson et al., 2021), we included a number line estimation task, a fraction equivalence task, and a non-symbolic magnitude comparison task as measures of objective math ability. We also included an objective measure of graph literacy (Garcia-Retamero et al., 2016) to account for individual differences in participants' ability to reason with visual displays, which may relate to the effects of visual format condition in the present study. Subjective math ability is also relevant to the current study because it can explain additional variance in performance on health-related ratio problems over and above objective math skills. For instance, anxiety and negative attitudes about math could impact performance on health-related ratio problems by leading people to avoid thinking deeply about numbers (e.g., Sidney, Thalluri, et al., 2019; Thompson et al., in press). Moreover, math anxiety has been found to predict health-related math problem-solving performance (Thompson et al., 2021) as well as people's confidence in their own health-related problem-solving performance (Scheibe et al., 2022) even when accounting for objective math skills. To account for subjective math ability, we included measures of math attitudes, math anxiety, and subjective graph literacy. Finally, we included a measure to account for individual differences in relational reasoning because accurately reasoning with rational numbers requires relational reasoning (e.g., considering the *relation* between the number of deaths and the number of cases).

Does Reasoning With Ratios Predict Health Cognitions?

A secondary goal of the current study was to examine whether college students' performance on health-related ratio problems related to their health cognitions, such as perceptions of the likelihood of contracting COVID-19 or intentions to engage in preventative health behaviors. COVID-19 has high fatality and transmission rates (Piroth et al., 2021), therefore being able to accurately assess for example, the case-fatality or infection rate of COVID-19 relative to other diseases, may relate to health cognitions. People with an accurate understanding of the magnitude of risks may perceive the likelihood and seriousness of COVID-19 as higher, which could lead to greater engagement in preventative behaviors to curb the spread of COVID-19. This relationship has been supported by empirical studies (de Bruin & Bennett, 2020; Pederson & Favero, 2020; Fansher et al. 2022; Magnan et al., 2021). More generally, perceived likelihood of disease predicts engagement in health behaviors (Sheeran et al., 2014), and indicators of how people appraise their risk, including perceived likelihood and severity, are included in multiple health-behavior theories (Klein & Ferrer, 2018; Taber & Klein, 2016). Worry is another component of risk appraisal that may motivate health behaviors (Hay et al., 2006), including COVID-19 prevention behaviors (Magnan et al., 2021).

We chose to conduct this study in a sample of college students in the Fall of 2020, during the first semester on campus after the COVID-19 lockdown. The likelihood of death and severe symptoms from COVID-related complications is higher for older than younger adults (CDC, 2021). At the time these data were collected, if younger adults did not view COVID-19 as risky, they may have failed to adhere to health-safety measures and been more likely to transmit the virus to others (Boehmer et al., 2020). Alternatively, since college students tend to be more liberal than other U.S. adults, they may have been more likely to engage in preventative behaviors because political orientation predicts engagement in such behaviors (Alcott et al., 2020; Baradaran Motie & Biolsi, 2020; Painter & Qiu, 2020). Either way, it is important to probe health cognitions in this specific population.

Current Study

Near the beginning of the pandemic (March/April 2020), Thompson and colleagues (2021) developed a mathematical intervention involving several worked examples to teach adults how to calculate and compare case-fatality rates for COVID-19 versus the flu. The worked examples explicitly indicated that to appropriately compare case-fatality rates, people must consider the ratios; not just the numerators or just the denominators. The aim of the intervention was to prevent WNB errors by emphasizing the necessity of relational reasoning, and participants in the intervention condition also saw the relevant ratios presented with visual displays similar to number lines. Participants randomly assigned to the intervention relative to the control were more accurate on health-related math problems involving COVID-19 case-fatality rates. However, the “kitchen sink” nature of the intervention precluded a test of whether the number-line visual was a critical component contributing to participants' improved performance.

In the current experiment, we extend this work by testing the effectiveness of representing health-related ratios with different visual models without explicit training. In addition to our preregistered goals, investigating the effectiveness of several visual displays on performance on health-related ratio problems and the relations between performance on these problems and health cognitions and behaviors, we also tested the following two exploratory hypotheses:

(1) Participants randomly assigned to solve health-related ratio problems with number lines would be more accurate than those randomly assigned to solve problems with no visual display, even when accounting for individual differences in subjective and objective math ability, general relational ability, and ability to reason with graphs.

(2) Regardless of visual display condition, participants who were more, relative to less, accurate on the health-related ratio problems would report greater perceived COVID-19 likelihood, worry, and severity, even when accounting for individual differences in subjective and objective math ability, general relational ability, and ability to reason with graphs.

We also explored whether visual displays impacted participants' reported problem-solving strategies and responses to health cognition measures (see Supplemental Materials).

Method

Transparency and Openness

The survey flow of all measures is included in the Supplemental Materials, and question wording for all survey items is available on the OSF page for the project (see Author's Note). The data and analytic code are also available on the OSF page for the project.

Participants

We planned to collect data from 250 college student participants for this online study based on an a priori power analysis. We used an estimated effect size of $OR = 2.667$ which corresponds to a contrast of 60% accuracy (a little better than chance) vs. 80% accuracy on the health-related math problems. Using power of 80%, an alpha rate of 1% (a conservative alpha level because we planned to conduct multiple related tests), and assuming that condition would be orthogonal to all other variables, we needed $N = 214$ + number of variables in the model for 80% power for any 1 degree of freedom test in a logistic regression model.

We recruited participants ($N = 270$) from two psychology department subject pools[2], one housed in a midwestern university ($n = 143$, $M_{age} = 19.30$, $SD_{age} = 2.29$; 85% female; 77% White) and the other in a southern university ($n = 127$, $M_{age} = 18.80$, $SD_{age} = 1.41$; 88% female; 74% White). Participants were recruited through online SONA systems and were given course credit for participation in the study. This project received IRB approval from both institutions where data was collected. The data were collected during the Fall 2020 semester. This was the first semester that students were back on campus after the Spring 2020 shutdown and prior to vaccines being available. Twenty-nine percent of participants reported using a calculator to solve

[2] Significant differences in performance by university were not found for pretest, $\chi^2 = 3.52$, $p = .061$, or for health-related ratio problems, $t(235) = 0.04$, $p = .965$. Including university in the logistic mixed-effects models for item-level accuracy reported below did not change the pattern of results, and university did not emerge as a significant predictor in the model.

the health-related ratio problems, but calculator use did not vary by condition (see Supplemental materials for more details).

Design and Procedure

Participants were randomly assigned to one of four conditions: No Model (*n* = 66), Number Line (*n* = 71), Risk Ladder (*n* = 66), or Icon Array (*n* = 67). We will only describe those measures that are of central importance to testing the hypotheses of the current study. First, participants completed a pretest health-related problem about a hypothetical disease. Then participants completed the math attitudes and anxiety measures in a counterbalanced order. We chose to administer these measures before participants completed the target health-related problems, because prior work has shown that these measures are affected by completing math tasks first (Sidney et al., 2021). Then, participants completed all seven health-related math problems and one non-health math problem accompanied by the same type of visual display based on randomly-assigned condition. We included a non-health problem adapted from Denes-Raj and Epstein (1994) to assess whether the visual displays also affected decision-making accuracy on a non-health ratio problem. Note that for all problems (Table 1), participants first responded to a multiple-choice question (e.g., was Disease A or B more fatal, or were they equally fatal), then they provided an open-ended strategy report about how they solved the problem, and finally, they rated whether they thought their decision was accurate (i.e., yes/no) and indicated their confidence on a 0-100% slider. These metacognitive judgments were not central to the current study and are not discussed further. Participants rated their health cognitions and behavioral intentions immediately after completing the math problems, which allows us to investigate whether accuracy on the problems influenced responses on these measures. Participants then completed the other focal measures of individual differences in cognition (i.e., number-line estimation, non-symbolic comparisons, fraction magnitude equivalence, subjective and objective graph literacy, and the test of relational reasoning) in a randomized order. These measures were included as predictors of accurate problem solving and use of strategies consistent with WNB. Participants then completed measures not central to our current hypotheses: medical information, information about political ideology (Jost, 2006), media trust, cultural worldview (Kahan et al., 2012), and trait anxiety (Spielberger et al, 1970). Participants reported demographic information at the very end of the survey.

Materials

Health-Related Math Problems

Pretest Hypothetical Disease Problem. As in Thompson et al. (2021), participants first completed a problem in a hypothetical Disease A vs. B. context, with no accompanying visual model. Performance was scored as correct vs. incorrect. See Table 1 for the wording of this hypothetical problem.

Health-Related Ratio Problems. Participants completed eight problems (see Table 1), each of which involved determining which of two ratios was larger, a task on which participants often incorrectly apply whole-number strategies to solve (i.e., WNB). Before beginning the problems, participants read that, "All of the statistics in the problems are ***real***, accurate numbers drawn from reliable sources, such as Johns Hopkins University." [3] We designed these items based on those administered by Thompson et al. (2021). Our goal was to create items that would be relevant to undergraduates and would illustrate the severity of COVID-19. We piloted these items with some undergrads in our labs to ensure variability in performance, and to probe

[3] The order in which participants completed the problems was randomized, although the "October" question was always presented last. After the "October" question, all participants were asked an open-ended question to assess the cues that they used to make their confidence rating, "You just rated your confidence about your decision. What information did you use to decide on your confidence judgment?" Confidence ratings and cue use were unrelated to our current hypotheses and are not discussed further.

whether the scenarios in the questions seemed personally relevant to them. We attempted to word the items in a way that would not explicitly direct people to divide (i.e., "which disease is more fatal" instead of "which has the higher fatality rate").

Performance on these problems was the dependent variable of interest. Additionally, the experimental manipulation – random assignment of visual display that accompanied each decision-making problem – was instantiated in these problems (see Figure 1). The visualization for each problem included two visual displays in the same format (e.g., number lines): one representing the larger ratio and one representing the smaller ratio. Sometimes the larger ratio was represented by the left (or top) visualization, and sometimes it was represented by the right (or bottom) visualization.

We made several decisions concerning how to create the visual displays. Because the use of concrete representations can make it difficult to transfer conceptual knowledge to future problems (Kaminski et al., 2009) that differ in surface-level features, yet are conceptually similar (i.e., calculation of infection rates), we represented the icons as ovals in the current study. The horizontally-oriented number lines were bounded by "0" on the left to the denominator of the ratio on the right. Landmarks at the quartiles were included, but they were not labeled (Siegler & Thompson, 2014). A red arrow pointed to the location of the numerator of the ratio, and it was labeled with the appropriate whole number. The vertically-oriented risk ladders ranged from "Less than 1 in [denominator of the ratio]" on the bottom to [half of the denominator] to [denominator of the ratio] on the top. Labeled landmarks at the quartiles were included. These design choices were made to make the risk ladders similar to the number line visualizations (i.e., equal-sized intervals between hatch marks), but also so that the risk ladders looked similar to ones used in some health decision-making studies (i.e., top hatch mark represented 50%).

Health Cognition Measures

Perceived COVID-19 Likelihood for Self and Others. Participants completed eight items assessing their perceived likelihood of contracting COVID-19 for themselves and for people they care about. For both oneself and others, participants reported absolute and comparative cognitive and affective likelihood (adapted from Janssen et al., 2018). The self items were: "How likely do you THINK it is that you will get COVID-19 in the next 30 days?" (1 = not at all likely, 5 = extremely likely; absolute cognitive), "Compared to other women or men of your age in the United States, how likely do you THINK it is that you will get COVID-19 in the next 30 days?" (1 = much less likely, 5 = much more likely; comparative cognitive), "I FEEL like I could easily get COVID-19 in the next 30 days." (1 = strongly disagree, 5 = strongly agree; absolute affective), "Compared to other women or men of your age in the United States, how easily do you FEEL you could get COVID-19 in the next 30 days?" (1 = much less easily, 5 = much more easily; comparative affective). Items assessing perceived risk for others replaced the word "you" or "I" with "someone you care about such as a family member or friend." All items also included a "don't know" response option, and these responses were coded as missing data. Mean ratings were calculated separately for the four items relating to likelihood for self (Cronbach's $\alpha = .79$) and for the four items relating to likelihood for others (Cronbach's $\alpha = .90$).

COVID-19 Worry for Self and Others. Participants completed ratings for COVID-19 worry for themselves and for people they care about (adapted from Taber et al., 2019 and Weinstein et al., 2007) in a randomized order. The self items were: "How much do you worry about being infected with COVID-19?", "How anxious are you about being infected with COVID-19?" The items assessing worry for others referred to "someone you care about, such as a family member or friend." All items were rated on a scale of 1 = "not at all" to 5 = "a lot".

Mean ratings were calculated separately for the two items relating to worry for self (r = .86) and for the two items relating to worry about others (r = .84).

Perceived COVID-19 Severity. Participants completed ratings for three items asking about perceptions of COVID-19 severity (adapted from Lipkus et al., 2003; Cronbach's α = .92). The items were: "COVID-19 is a serious condition", "COVID-19 is dangerous", and "I believe COVID-19 is life-threatening." All items were rated on a scale of 1 = "strongly disagree" to 4 = "strongly agree."

Preventive Behavioral Intentions. This scale was adapted from Toussaint and colleagues (2020). Participants were told, "There are currently several recommendations on how to prevent the spread of the novel coronavirus (COVID-19). Indicate the extent to which you plan to do each of the following activities in the next 30 days." Participants made ratings on five-point Likert scales ranging from "never" to "all of the time" for nine preventive measures (e.g., "Use a hand sanitizer that contains at least 60% alcohol, if soap and water are not readily available."). We averaged the ratings across the nine intentions, which had good reliability (Cronbach's α = .82).

Individual Difference Measures

Math Attitudes and Math Anxiety. Participants completed the math attitudes questionnaire (MAQ; Sidney, Thompson, et al., 2019) and math anxiety questions about math in general and math involving specific types of numbers (see Thompson et al., 2021). Ratings on the MAQ were aggregated across 20 questions, and math anxiety ratings were aggregated across 6 questions. The order in which participants completed the MAQ and math anxiety measures was randomized.

Symbolic and Non-Symbolic Math Skills. Participants completed number-line estimation tasks, a fraction magnitude equivalence task, and a non-symbolic ratio task. TTo assess symbolic magnitude understanding, participants estimated the location of 10 numbers in the 0-100 range (Siegler & Opfer, 2003) and 22 numbers in the 0-1 billion range (Landy et al., 2017) in separate blocks. Estimation precision was measured as percent absolute error (PAE) and was averaged across both numerical ranges for the analyses reported below. PAE is calculated by taking the absolute value of the difference between the provided estimate and true magnitude and dividing by the numerical range. We also assessed participants' fraction magnitude knowledge with a six-item abbreviated version of a fraction magnitude equivalence task (Fitzsimmons et al., 2020) in which performance was calculated as the proportion of correct responses across the problems. Non-symbolic ratio comparison (Matthews et al., 2015) was measured across 12 trials after participants received two practice trials with feedback to explain the rules of the task (i.e., participants judged which pair of line segments had the larger ratio of white to black segment lengths). Participants did not receive feedback on the 12 target trials, and performance was calculated as the proportion of correct responses across the problems.

Subjective and Objective Graph Literacy. We adopted five items from a subjective graph literacy assessment (Garcia-Retamero et al., 2016) in which participants rated themselves on their ability to use graphs and charts on seven-point Likert scales from "Extremely bad" to "Extremely good." We averaged ratings across the five items (Cronbach's $\alpha = .94$). We adopted four problem-solving items from an objective graph literacy assessment (Galesic & Garcia-Retamero, 2011) in which participants solved problems represented in graph and chart form. Performance was calculated as the total number of correct responses across the four problems.

Test of Relational Reasoning. We adopted four items from Alexander et al. (2016) to assess participants' ability to solve four visual analogy problems. For example, one problem asked participants to complete a pattern involving shapes with 1 to 3 dots in the center of the shapes. Performance was calculated as the proportion of correct responses across the four problems.

Trait Anxiety. We included a measure of trait anxiety as a covariate in our health cognition analyses (Spielberger et al, 1970). Participants indicated how often they felt the way 20 statements described (e.g., "I feel nervous and restless"), with response scales ranging from 1 = "almost never" to 4 = "almost always". Nine items were reverse-scored such that higher values indicated higher anxiety, and the sum of all responses was included in the health cognition models reported below (Cronbach's α = .91).

Demographics. Participants responded to several items in a demographic questionnaire (see Thompson et al., 2021 and OSF (https://osf.io/7bfmy/?view_only=1916f603d76648ac986b4925f38417f2) for the full list of demographics questions that were included), and reported their gender, race, year in school, and number of math courses taken in college (see Table 2 for participant characteristics). The mean number of math courses taken in the final sample was 4.70 (SD = 1.80).

Overview of Analyses

Data Exclusion

We used the same exclusion criteria as reported in the preprint in a prior study (Thompson et al., 2021): fast responders, poor strategy reports, consistently selecting one side in the fraction magnitude equivalence task, and consistently clicking in the same region for their number line estimates. We also included two attention check items: One in the "behavioral willingness" scale and the other in the "cultural worldview" scale. If participants missed 2 of 2 attention checks, they were excluded from analyses. With these exclusions, the final sample[4] was *N* = 237 (Control: 57, Number Line: 63, Risk Ladder: 58, Icon: 59).

Analytic Plan

As indicated in our pre-registration (https://osf.io/7bfmy/?view_only=1916f603d76648ac986b4925f38417f2), we used all available data for each separate analysis after the exclusions reported above. All continuous variables were rescaled for ease of interpretation, such that *M* = 0, *SD* = 1. We conducted several preliminary analyses to determine that random assignment was successful, and those can be found in the Supplemental Materials. First, we report our pre-registration plan, then we report our deviations from that plan and the analyses we ultimately conducted because we believed they were the most appropriate. We preregistered a series of logistic regression models to assess whether condition was systematically related to decision-making accuracy and usage of strategies consistent with WNB *on each item separately* (1) to be consistent with the approach taken in the original study (Thompson et al., 2021), and (2) because pilot data from undergraduates indicated a range of accuracy on some of the problems in the final set tested here. The full regression models proposed in our preregistered analytic plan included gender, math anxiety, several proxies for magnitude knowledge (number line estimation, fraction magnitude equivalence, and non-symbolic magnitude comparison), age, and hypothetical decision making at pretest as covariates. We included gender in our preregistered analyses because it predicted health-related math problem solving in Thompson et al. (2021) and because of gender differences in magnitude estimation precision (Rivers et al., 2020), and math anxiety (Hembree, 1990). We also

[4] The pattern of results for the primary analysis of interest, examining the impact of visual displays on health-related ratio problem accuracy, did not change when the full sample was analyzed.

preregistered separate logistic regression analyses for each decision-making item to explore whether subjective graph literacy, objective graph literacy, and relational reasoning skills predict COVID-related math problem-solving accuracy and WNB strategy use. The preregistered analyses described above can be found in the Supplemental Materials.

We ultimately report analyses that differ from those we preregistered in several ways. First, we opted to conduct a logistic mixed-effects model for item-level accuracy with item-level and subject-level random intercepts instead of separate logistic regressions for each problem. We made this decision in the interest of parsimony, and to reduce the number of separate statistical tests, which would inflate the probability of Type I error. The original study (Thompson et al., 2021) that was the basis for our preregistered analysis only had three health-related problems, whereas here we had eight problems. Including item-level intercepts allowed us to account for random effects at the item level, which is important because we had relatively few items that varied in terms of health-related context. We also chose to include accuracy at pretest, math attitudes, math anxiety, number-line estimation performance, equivalence knowledge, non-symbolic magnitude comparison, subjective graph literacy, objective graph literacy, and relational reasoning skills as covariates in this model to control for individual differences in the sample, thus increasing our power to test the focal effects.

The second change to our preregistered analyses was removing age and gender as covariates in our models. We chose to do this to reduce the total number of predictors in our models to avoid overfitting (Babyak, 2004). In addition, because undergraduates comprised the sample in our analyses, participants were generally quite similar in age. Although participants in the no model control condition were slightly older than participants in the other three visual display conditions, there were no condition differences in the number of math courses participants reported having taken, nor were there any condition differences in the other math measures. The distribution of gender did not vary by condition. Also, males were underrepresented in our current sample, so gender analyses may not have been appropriate. Finally, neither gender nor age emerged as significant in our preregistered analyses (see the Supplemental Materials).

As preregistered, we ran a series of five linear regressions to test whether decision-making accuracy significantly predicted perceived COVID-19 likelihood and worry for self or for other people, and perceived COVID-19 severity. We included the same covariates in these regression models as those in the accuracy analyses. Given that the analyses we ultimately report here feature fewer statistical significance tests than the analysis plan we preregistered, we are reporting alpha values lower than 5% as significant.

Results

Problem-Solving Performance

Descriptive statistics and correlations for all measures are displayed in Table 3. The percentage of correct responses by item and condition can be found in Table 4. We fit a logistic mixed-effects model for item-level accuracy on the eight health-related problems[5] with condition, pretest decision-making accuracy, fraction magnitude equivalence performance, number line estimation precision, non-symbolic ratio performance, objective graph literacy, math anxiety, math attitudes, and subjective graph literacy as fixed factors along with item-level and subject-level random intercepts. Condition was dummy coded with the number line condition as the reference group. We fit the model using restricted maximum likelihood estimation (REML)

---

[5] We fit the same model without the bowls problem (because this problem was not health-related). The same pattern of results was obtained for the effects of visual model condition – the number line condition outperformed the other conditions, and the icon array and risk ladder conditions did not differ from the no model condition. Pretest accuracy again emerged as a significant predictor in this model, however objective graph literacy and math attitudes went from significant in the full model with all items to marginal in this model (p = .076 and p = .056, respectively).

with the *lme4* package (Bates et al., 2012) in *R* (version 4.1.1; R Core Team, 2020). For this model, VIF values were < 2, which indicates acceptable multicollinearity.

As can be seen in Table 5, the number line condition outperformed all other conditions on the health-related math problems. Moreover, recoding the conditions such that the no model condition was the reference group revealed that participants in the the risk ladder condition, $b = -0.20$, $SE = 0.22$, Wald $\chi^2 = 0.82$, $p = .364$, OR = 0.82, 95% CI[-0.64, 0.24], and the icon array condition, $b = -0.10$, $SE = 0.23$, Wald $\chi^2 = 0.20$, $p = .654$, OR = 0.90, 95% CI[-0.55, 0.34], performed no better than those in the condition where no visual display was provided. Answering the pretest decision-making item correctly was associated with higher performance on the set of decision-making items, as was higher objective graph literacy, and more positive math attitudes. None of the other covariates emerged as significant (see Table 5).

Health Cognitions and Behavioral Intentions

To test whether performance on health-related ratio problems, operationalized as the sum of the scores on six of the eight health-related problems[6] regardless of randomly-assigned condition, predicted health-cognition measures, we ran separate linear regressions for each health-related outcome. We included all covariates[7] from the full logistic mixed-effects model for health-related ratio problem performance described above. In addition, we included university as an additional covariate, because preliminary analyses indicated that participants at the two universities differed in their reported perceived risk for others, $t(230) = 2.35$, $p = .020$, $d = 0.31$, worry for self, $t(235) = 2.03$, $p = .044$, $d = 0.26$, and severity of COVID-19, $t(224.38) =$

[6] Two of the eight decision-making problems (World War 1 and Age Ranges) were excluded from the analysis here due to the rates in the problems being very similar; if participants had used rounding on these problems, two response options could be considered correct. Excluding these two items improved the reliability from .65 to .73.
[7] When we ran the models for each health outcome measure with gender, gender did not emerge as a significant predictor for any outcome. In addition, the pattern of results did not change for any health outcome, except for the severity proportion measure, for which number line estimation precision went from significant to marginal ($p = .060$).

2.50, $p = .013$, $d = 0.33$ (degrees of freedom obtained from a Welch Two Sample *t*-test due to unequal variances). We also included trait anxiety as an additional covariate in the models, because trait anxiety was significantly correlated with several health cognition measures (perceived COVID-19 likelihood for others: $r = .20$, $p < .001$; COVID-19-related worry for self: $r = .20$, $p < .001$; COVID-19-related worry for others: $r = .18$, $p < .001$). For all models, VIF values were < 2, which indicates acceptable multicollinearity. As can be seen in Table 6, performance on health-related ratio problems was a significant predictor of greater worry for self and others, greater perceived severity of COVID-19, and greater self-reported intentions to engage in preventative behaviors such as mask wearing. Performance on health-related ratio problems did not predict perceptions of the likelihood of contracting COVID-19 for self or others.

Although higher performance on health-related ratio problems was associated with higher ratings for several health cognitions, other objective math measures showed the opposite pattern. Specifically, higher performance on a fraction magnitude equivalence task was associated with lower ratings of worry for self and others, lower ratings of perceived COVID-19 severity, and lower endorsement of intentions to engage in preventative behaviors. Similarly, lower percent absolute error (i.e., higher accuracy) on the number line estimation task was also associated with lower perceived likelihood of others contracting COVID-19.

Math anxiety and attitudes towards math also predicted several health cognition measures. Higher math anxiety was associated with greater worry for self and others, greater perceived severity of COVID-19, and higher endorsement of intentions to engage in preventative behaviors. More positive attitudes towards math were also associated with higher perceived likelihood of others contracting COVID-19, greater COVID-19 worry for self, greater perceived severity of COVID-19, and higher endorsement of intentions to engage in preventative behaviors.

Discussion

Number lines facilitated college students' mathematically-accurate reasoning on health-related ratio problems, relative to risk ladders and icon arrays and no visual model. Accuracy on the health-related ratio problems accounted for significant variance in COVID-19 worry for self and others, perceptions of COVID-19 severity, and preventive health behavior intentions, even when controlling for math-related individual differences. This overlap between health-related ratio problem solving performance and health cognitions could be due to shared reliance on understanding risk magnitude. Health decision-making researchers could consider adding number lines to the repertoire of visual models commonly used to present numerical health information to help people reason about rational number magnitudes. Health researchers could also consider including measures of objective and subjective math skills (Choi et al., 2020) to better understand *who* will accurately interpret health statistics.

In the current study, the two commonly-used visuals in health-related research, risk ladders and icon arrays, did not lead to better performance than no visual display at all. The goal of including visuals in the current study and other work (Mielicki et al., 2021; Sidney, Thompson, & Rivera, 2019; Thompson et al., 2021; Fitzsimmons, Woodbury et al., 2022) is to visually communicate how the part relates to the whole, and should promote a portable and durable conceptual understanding of rational number magnitudes when people encounter them in various contexts. Though substantial research in health decision making has examined the use of visuals to help people, especially those low in numeracy, better understand risk magnitudes (Ancker et al., 2006; Bonner et al., 2021; Garcia-Retamero & Cokely, 2013; 2017; Garcia-

Retamero et al., 2012; Trevena et al., 2021), the goal of these visuals is to ease the mathematical burden such that people can more easily comprehend risk magnitudes perceptually (Waters et al., 2016). When the goal is helping people "read off" a given risk magnitude, icon arrays and risk ladders may be helpful, but our study shows that these commonly-used visuals may not facilitate relational reasoning as well as number lines, especially when the health-related statistics do not include rounded values (e.g., out of 100 or 1,000) or have different denominators, which is how COVID-19 statistics were presented early in the pandemic.

Performance on health-related ratio problems was also related to several health cognition measures and explained unique variance over and above additional covariates such as subjective and objective math ability, relational reasoning, and trait anxiety. Specifically, higher performance on these problems related to higher reported COVID-19 worry for self and others, higher perceived severity of COVID-19, and greater endorsement of intentions to engage in preventative behavior, such as mask wearing and social distancing. These findings suggest that college students' ability to reason with health-related rational numbers may contribute to their cognitions and behaviors related to COVID-19.

In addition, several covariates in our models emerged as significantly related to health cognition measures. Specifically, several measures of objective math ability were related to health cognitions, though not in the expected direction (see discussion in limitations section below), as were several measures of subjective math ability. Higher math anxiety and more positive math attitudes were related to higher ratings of COVID-19 worry for self and others, higher perceptions of COVID-19 severity, and higher endorsement of intentions to engage in preventative behaviors. The direction of the effects is particularly surprising given that math anxiety and attitudes were strongly negatively correlated. Positive math attitudes were also related to higher perceptions of COVID-19 risk for others, and higher subjective graph ability was related to higher perceptions of COVID-19 severity and higher endorsement of intentions to engage in preventative behaviors. These findings suggest that the way people feel about numerical information may play a role in how they interpret health-related numerical information when reasoning about health risks, though future research in health contexts is needed to better understand these relationships.

Limitations and Future Directions

Despite researchers commonly relying on college students as a convenience sample (Cooper et al., 2010), we focused on college students in the Fall of 2020 during their first semester back on campus after the Spring 2020 COVID-19 lockdown because we were interested in their unique perspective on COVID-19 risks. Although college students were our target population, we did ultimately recruit a convenience sample from two universities. Thus, it is not clear how our findings might generalize to all U.S. college students.

Higher performance on health-related ratio problems was associated with higher ratings for several health cognitions in this study; however, higher performance on a fraction magnitude equivalence task was associated with *lower* ratings for health cognitions (i.e., lower reported worry for self and others, lower ratings of perceived COVID-19 severity, and lower endorsement of intentions to engage in preventative behaviors). In addition, higher accuracy on the number line estimation task was also associated with lower perceived likelihood of others becoming infected with COVID-19. These results are surprising and appear to suggest that in this sample of college students, those with more precise mathematical understandings were, in general, less worried about the COVID-19 pandemic. However, we believe these findings should be interpreted with caution. The statistical models also included performance on health-related ratio problems, a measure more proximal to the health cognition outcomes of interest than mathematical measures devoid of health-context. In other words, it is unclear exactly what psychological construct the remaining variability in other math ability measures (number line estimation and fraction equivalence performance) may represent, once variability from the more proximal health-related ratio problem performance is parsed out.

Although other work has found that improving understanding of COVID-19 data can influence engagement in preventative behaviors (e.g., Lammers et al., 2020; Fansher et al., 2022), we did not observe an effect of visual display condition on health cognition measures in the current study. Since we did not train participants to reason with these COVID-19 related ratios, this could potentially explain why we did not observe an effect on health cognitions. Of note, the items assessing perceived likelihood of COVID-19 for self and others and preventative behavior intentions referred to the next 30 days and using a different time frame might have produced different responses. It is also possible that the results we observed in the current study could be specific to college students. Other work that found effects of interventions aimed at improving understanding COVID-19 data on health cognitions and behaviors (Lammers et al., 2020; Fansher et al., 2022; Thompson et al., 2021) relied on adult samples of participants. In Thompson et al. (2021), there were no reliable effects of age on the likelihood of answering health decision-making problems correctly, though the authors did not evaluate whether the effects of the worked-example training differed by age. Age was positively associated with perceived severity of COVID, though it was unrelated to baseline perceived worry or risk perceptions. Future work should explore the extent to which our findings generalize to populations with more variability in age.

This work suggests that a well-chosen and well-designed visual display can highlight mathematical relations and promote mathematically-accurate reasoning – the mere presentation of a number line without instructional intervention was sufficient to improve performance. Number lines may derive their advantage from alignment with the mental number line (Dehaene, 2011; Siegler et al., 2011; Case & Okamoto, 1996) and from implicitly drawing connections between the fractional components and percentages (Moss & Case, 1999; Schiller, 2020). This research bridges prior work in education contexts that has demonstrated a number line advantage over other types of visual displays for helping learners reason with ratios (Fazio et al., 2016; Fuchs et al., 2013; 2014; Gunderson et al., 2019; Hamdan & Gunderson, 2017; Moss & Case, 1999; Rittle-Johnson et al., 2001; Saxe et al., 2013; Schneider et al., 2009; Sidney et al., 2019; Siegler et al., 2011). The number lines in this study were specifically designed to maximize their effectiveness in illustrating health-related ratios (e.g., including unlabeled quartile landmarks, including an arrow to indicate the location of the numerator of the ratio, etc.). Future work could explore which specific features of number lines are critical for effectively communicating health statistics. Health statistics are often conveyed as rational numbers (Thompson et al., 2022), and all rational numbers, including fractions, whole-number frequencies, percentages, and decimals, have magnitudes and can be placed on number lines (Siegler et al., 2011). Future research might also investigate interventions that support drawing connections among number lines – which are infrequently used to display health statistics – and icon arrays and risk ladders, which are more commonly used to display health statistics.

More work is needed to better understand the conditions which might be best suited for using number lines to communicate health-related statistics. Recent work (Fitzsimmons et al., 2022) has indicated that an intervention combining a worked example (adapted from Thompson et al., 2021), in which ratio magnitudes were represented on number lines at each step and asking participants to solve health-related math problems accompanied by number lines may improve understanding of health-related ratios. Importantly, those randomly assigned to the intervention group also outperformed those in the control group on an immediate transfer problem that did not include number-line visual support. Perhaps incorporating number lines in other effective interventions could also impact health cognitions and behavior.

Though health statistics often involve rational numbers (Thompson et al., 2022), calculations with numbers embedded in health contexts may be different than those in purely numeric contexts because of the personal relevance, prior experience, and affective associations people have with health contexts. Moreover, being able to accurately interpret health statistics is not the same as using those statistics to make medically sound health decisions. Thus, more research is needed to determine the boundary conditions of number line benefits in health contexts. Drawing from both math cognition and health decision-making research may lead to novel inroads on research that impacts people in real-world health settings.

Concluding Remarks

The COVID-19 pandemic has highlighted the urgency of effectively communicating rational numbers in health contexts. This study bridges the gap between research in numerical cognition and health decision making by demonstrating that for health-related problems that require reasoning with rational numbers, number lines facilitate mathematically-accurate reasoning relative to icon arrays and risk ladders, which are more commonly used to present health statistics.

References

Alexander, P. A., Dumas, D., Grossnickle, E. M., List, A., & Firetto, C. M. (2016). Measuring relational reasoning. *The Journal of Experimental Education*, *84*(1), 119-151. https://doi.org/10.1080/00220973.2014.963216

Alibali, M. W., & Sidney, P. G. (2015). Variability in the natural number bias: Who, when, how, and why. *Learning and Instruction, 37,* 56-61. https://doi.org/10.1016/j.learninstruc.2015.01.003

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association, 13*(6), 608-618. https://doi.org/10.1197/jamia.M2115

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine, 66*(3), 411-421. https://doi.org/10.1097/01.psy.0000127692.23278.a9

Baradaran Motie, G., & Biolsi, C. (2021). County-Level Determinants of Social Distancing (or Lack Thereof) during the COVID-19 Pandemic. Contemporary Economic Policy, 39(2), 264-279. https://doi.org/10.1111/coep.12499

Begolli, K. N., Booth, J. L., Holmes, C. A., & Newcombe, N. S. (2020). How many apples make a quarter? The challenge of discrete proportional formats. *Journal of Experimental Child Psychology, 192,* 1-20. https://doi.org/10.1016/j.jecp.2019.104774

Bhatia, P., Delem, M., Leone, J., Boisin, E., Cheylus, A., Gardes, M-L., & Prado, J. (2020). The ratio processing system and its role in fraction understanding: Evidence from a match-to-sample task in children and adults with and without dyscalculia. *Quarterly Journal of*

Experimental Psychology, *73*(12), 2158-2176. https://doi.org/10.1177/1747021820940631

Boehmer, T. K., DeVies, J., Caruso, E., van Santen, K. L., Tang, S., Black, C. L., Hartnett, K. P., Kite-Powell, A., Dietz, S., Lozier, M., & Gundlapalli, A. V. (2020, October 2). Changing age distribution of the COVID-19 pandemic -United States, May-August 2020. *Morbidity and Mortality Weekly Report*, *69*(39), 1404-1409. https://doi.org/10.15585/mmwr.mm6939e1

Bonner, C., Trevena, L. J., Gaissmaier, W., Han, P. K., Okan, Y., Ozanne, E., Peters, E., Timmermans, D., & Zikmund-Fisher, B. J. (2021). Current Best Practice for Presenting Probabilities in Patient Decision Aids: Fundamental Principles. *Medical Decision Making*. Advance online publication. https://doi.org/10.1177/0272989X21996328

Boyer, T. W., Levine, S. C., Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology*, *44*(5), 1478-1490. https://doi.org/10.1037/a0013110

Centers for Disease Control and Prevention (CDC). (2020, March 24). *2019-2020 U.S. Flu season: Preliminary burden estimates*. https://www.cdc.gov/flu/about/burden/preliminary-in-season-estimates.htm

Centers for Disease Control and Prevention (CDC). (2021, April 16). Older adults at greater risk of requiring hospitalization or dying if diagnosed with COVID-19. *CDC*. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html

Choi, S. S., Taber, J. M., Thompson, C. A., & Sidney, P. G. (2020). Math anxiety, but not induced stress, is associated with objective numeracy. *Journal of Experimental Psychology: Applied*. Advance online publication. https://doi.org/10.1037/xap0000268

Cooper, C. A., McCord, D. M., & Socha, A. (2010). Evaluating the college sophomore problem: The case of personality and politics. *The Journal of Psychology*, *145*(1), 23-37. https://doi.org/10.1080/00223980.2010.528074

de Bruin, W. B., & Bennett, D. (2020). Relationships between initial COVID-19 risk perceptions and protective health behaviors: A national survey. *American Journal of Preventive Medicine*, *59*(2), 157-167. https://doi.org/10.1016/j.amepre.2020.05.001

Dehaene, S. (2011). *The Number Sense* (2nd ed.). New York, NY: Oxford University Press.

Denes-Raj, V. & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology*, *66*(5), 819-829. https://doi.korg/10.1037/0022-3514.66.5.8199

DeWolf, M., & Vosniadou, S. (2015). The representation of fraction magnitudes and the whole number bias reconsidered. *Learning and Instruction*, *37*, 39-49. https://doi.org/10.1016/j.learninstruc.2014.07.002

Fansher, M., Adkins, T., Lewis, R., Boduroglu, A., Lalwani, P., Quirk, M., Shah, P., & Jonides, J. (2022). How well do ordinary Americans forecast the growth of COVID-19?. *Memory & Cognition*, Advanced Online Publication. https://doi.org/10.3758/s13421-022-01288-0

Faust, J. M. (2020, April 28). Comparing COVID-19 deaths to Flu deaths is like comparing apples to oranges. *Scientific American's Newsletters.* https://blogs.scientificamerican.com/observations/comparing-covid-19-deaths-to-flu-deaths-is-like-comparing-apples-to-oranges

Faust, J. S., & del Rio, C. (2020a) Assessment of deaths from COVID-19 and from seasonal influenza. *JAMA Internal Medicine.* https://doi.org/10.1001/jamainternmed.2020.2306

Faust, J., & del Rio, C. (2020b). Relative disease burdens of COVID-19 and seasonal influenza in New York City, February 1-April 18, 2020. *medRxiv*. https://doi.org/10.1101/2020.04.22.20073551

Fazio, L. K., Kennedy, C. A., & Siegler, R. S. (2016). Improving children's knowledge of fraction magnitudes. *PloS one*, *11*(10), e0165243.

Fitzsimmons, C. J., Thompson, C. A., & Sidney, P. G. (2020). Do adults treat equivalent fractions equally? Adults' strategies and errors during fraction reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(11), 2049-2074. https://doi.org/10.1037/xlm0000839

Fitzsimmons, C. J., Sidney, P. G., Schiller, L. K., Scheibe, D., Taber, J. M., Matthews, P. G., Waters, E. A., Coifman, K. G. Thompson, C. A. (2022). *Worked examples and number lines improve adults' understanding of health risks as ratios.* [Manuscript submitted for publication]. Department of Psychological Sciences, Kent State University.

Fitzsimmons, C. J., Woodbury, L., Taber, J. M., Schiller, L., Sidney, P. G., Coifman, K. G., & Thompson, C. A. (2022). *How do visual displays impact health-risk estimates? It depends on display size, shape, and prior knowledge.* [Manuscript submitted for publication]. Department of Psychological Sciences, Kent State University.

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., Jordan, N. C., Siegler, R., Gersten, R., & Changas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology*, *105*(3), 683-700. https://doi.org/10.1037/a0032446

Fuchs, L. S., Schumacher, R. F., Sterba, S. K., Long, J., Namkung, J., Malone, A., Hamlett, C.

L., Jordan, N. C., Gersten, R., Siegler, R. S., & Changas, P. (2014). Does working memory moderate the effects of fraction intervention? An aptitude-treatment interaction. *Journal of Educational Psychology, 106*(2), 499-514. https://doi.org/10.1037/a0034341

Galesic, M., & Garcia-Retamero, R. (2011). Graph literacy: A cross-cultural comparison. *Medical Decision Making, 31*(3), 444-457. https://doi.org/10.1177/0272989X10373805

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology, 28*(2), 210-216. https://doi.org/10.1037/a0014474

Garcia-Retamero, R., & Cokely, E. T. (2013). Communicating health risks with visual aids. *Current Directions in Psychological Science, 22*(5), 392-399. https://doi.org/10.1177/0963721413491570

Garcia-Retamero, R., & Cokely, E. T. (2017). Designing visual aids that promote risk literacy: A systematic review of health research and evidence-based design heuristics. *Human Factors, 59*(4), 582-627. https://doi.org/10.1177/0018720817690634

Garcia-Retamero, R., Okan, Y., & Cokely, E. T. (2012). Using visual aids to improve communication of risks about health: A review. *The Scientific World Journal*. https://doi.org/10.1100/2012/562637

Garcia-Retamero, R., Cokely, E. T., Ghazal, S., & Joeris, A. (2016). Measuring graph literacy without a test: A brief subjective assessment. *Medical Decision Making, 36*(7), 854-867. https://doi.org/10.1177/0272989X16655334

Gunderson, E. A., Hamdan, N., Hildebrand, L., & Bartek, V. (2019). Number line unidimensionality is a critical feature for promoting fraction magnitude concepts. *Journal of Experimental Child Psychology, 187*, 1-29. https://doi.org/10.1016/j.jecp.2019.06.010

Hamdan, N., & Gunderson, E. A. (2017). The number line is a critical spatial-numerical representation: Evidence from a fraction intervention. *Developmental Psychology, 53*(3), 587-596. https://doi.org/10.1037/dev0000252

Hawley, S. T., Zikmund-Fisher, B., Ubel, P., Jancovic, A., Lucas, T., & Fagerlin, A. (2008). The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Education and Counseling*, *73*(3), 448-455. https://doi.org/10.1016/j.pec.2008.07.023

Hay, J. L., McCaul, K. D., & Magnan, R. E. (2006). Does worry about breast cancer predict screening behaviors? A meta-analysis of the prospective evidence. *Preventive Medicine: An International Journal Devoted to Practice and Theory, 42*(6), 401-408. https://doi.org/10.1016/j.ypmed.2006.03.002

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, *21*(1), 33-46. https://doi.org/10.2307/749455

Hess, R., Visschers, V. H. M., & Siegrist, M. (2011). Risk communication with pictographs: The role of numeracy and graph processing. *Judgment and Decision Making, 6*(3), 263-274.

Jacob, S. N., Vallentin, D., & Nieder, A. (2012). Relating magnitudes: The brain's code for proportions. *Trends in Cognitive Sciences*, *16*(3), 157-166. https://doi.org/10.1016/j.tics.2012.02.002

Janssen, E., Ruiter, R. A., & Waters, E. A. (2018). Combining risk communication strategies to simultaneously convey the risks of four diseases associated with physical inactivity to socio-demographically diverse populations. *Journal of Behavioral Medicine, 41*(3), 318-332. https://doi.org/10.1007/s10865-017-9894-3

Johns Hopkins University and Medicine (2020, March 24). *COVID-19 United States cases by county*. https://coronavirus.jhu.edu/us-map

Jost, J. T. (2006). The end of the end of ideology. *American Psychologist*, *61*(7), 651-670. https://doi.org/10.1037/0003-066X.61.7.651

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*(10), 732-735. https://doi.org/10.1038/nclimate1547

Kalra, P. B., Binzak, J. V., Matthews, P. G., & Hubbard, E. M. (2020). Symbolic fractions elicit an analog magnitude representation in school-age children. *Journal of Experimental Child Psychology*, *195*, https://doi.org/10.1016/j.jecp.2020.104844

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. (2009). Transfer of mathematical knowledge: The portability of generic instantiations. *Child Development Perspectives*, *3*(3), 151-155. https://doi.org/10.1111/j.1750-8606.2009.00096.x

Klein, W. M. P., & Ferrer, R. A. (2018). On being more amenable to threatening risk messages concerning close others (vis-á-vis the self). *Personality and Social Psychology Bulletin*, *44*(10), 1411-1423. https://doi.org/10.1177/0146167218769064

Lammers, J., Crusius, J., & Gast, A. (2020). Correcting misperceptions of exponential coronavirus growth increases support for social distancing. *Proceedings of the National Academy of Sciences*, *117*(28), 16264-16266. https://doi.org/10.1073/pnas.2006048117

Landy, D., Charlesworth, A., & Ottmar, E. (2017). Categories of large numbers in line estimation. *Cognitive Science*, *41*(2), 326-353. https://doi.org/10.1111/cogs.12342

Lipkus, I. M. (2007). Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Medical Decision Making*, *27*(5), 696-713. https://doi.org/10.1177/0272989X07307271

Lipkus, I. M., Green, L. G., & Marcus, A. (2003). Manipulating perceptions of colorectal cancer threat: Implications for screening intentions and behaviors. *Journal of Health Communication*, *8*, 213–228. https://doi.org/10.1080/10810730305684

Lipkus, I. M., & Hollands, J. G. (1999). The visual communication of risk. *JNCI Monographs*, *1999*(25), 149-163. https://doi.org/10.1093/oxfordjournals.jncimonographs.a024191

Magnan, R.E., Gibson, L.P. & Bryan, A.D. (2021). Cognitive and Affective Risk Beliefs and their Association with Protective Health Behavior in Response to the Novel Health Threat of COVID-19. *Journal of Behavioral Medicine*, https://doi.org/10.1007/s10865-021-00202-4

Matthews, P. G., Lewis, M. R., Hubbard, E. M. (2015). Individual differences in nonsymbolic ratio processing predict symbolic math performance. *Psychological Science*, 1-12. https://doi.org/10.1177/0956797615617799

Mazur, D. J., & Hickam, D. H. (1993). Patients' and physicians' interpretations of graphic data displays. *Medical Decision Making*, *13*(1), 59-63. https://doi.org/10.1177/0272989X9301300108

Mielicki, M., Fitzsimmons, C., Thompson, C. A., Taber, J. M., Coifman, K., Sidney, P. G., … Waters, E. A. (2022, July 8). Number Lines Can be More Effective at Facilitating Adults' Performance on Health-Related Ratio Problems than Risk Ladders and Icon Arrays. Retrieved from osf.io/93m74

Mielicki, M., Fitzsimmons, C., Woodbury, H., Marshal, H., Zhang, D., Rivera, F., & Thompson,

C. A. (2021). Effects of figural and numerical presentation formats on growing pattern performance. *Journal of Numerical Cognition*, *7*, 125-155. https://doi.org/10.5964/jnc.6945

Moss, J., & Case, R. (1999). Developing children's understanding of the rational numbers: A new model and an experimental curriculum. *Journal for Research in Mathematics Education, 30*(2), 122-147. https://doi.org/10.2307/749607

Ni, Y., & Zhou, Y. D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist, 40*(1), 27-52. https://doi.org/10.1207/s15326985ep4001_3

Okamoto, Y., & Case, R. (1996). The role of central conceptual structures in the development of children's thought: II. Exploring the microstructure of children's central conceptual structures in the domain of number. *Monographs of the Society for Research in Child Development*, *61*(1-2), 27-58. https://doi.org/10.1111/j.1540-5834.1996.tb00536.x

Opfer, J. E., Thompson, C. A., & Furlong, E. E. (2010). Early development of spatial-numeric associations: Evidence from spatial and quantitative performance of preschoolers. *Developmental Science, 13*(5), 761-771. https://doi.org/10.1111/j1467-7687.2009.00934.x

Painter, M., & Qiu, T. (2020). Political beliefs affect compliance with covid-19 social distancing orders. *Covid Economics, 4*, 103-123.

Park, Y., Viegut, A. A., & Matthews, P. G. (2021). More than the sum of its parts: Exploring the development of ratio magnitude versus simple magnitude perception. *Developmental Science*, *24*(3), 313043. https://doi.org/10.1111/desc.13043

Pedersen, M. J., & Favero, N. (2020). Social distancing during the COVID-19 pandemic: Who are the present and future noncompliers?. *Public Administration Review*, *80*(5), 805-814. https://doi.org/10.1007/s10865-017-9894-3

Piroth, L., Cottenet, J., Mariet, A. S., Bonniaud, P., Blot, M., Tubert-Bitter, P., & Quantin, C. (2021). Comparison of the characteristics, morbidity, and mortality of COVID-19 and seasonal influenza: A nationwide, population-based retrospective cohort study. *The Lancet Respiratory Medicine, 9*(3), 251-259. https://doi.org/10.1016/S2213-2600(20)30527-0

Rau, M. A., & Matthews, P. G. (2017). How to make 'more' better? Principles for effective use of multiple representations to enhance students' learning about fractions. *ZDM*, *49*, 531-544. https://doi.org/10.1007/S11858-017-0846-8

Rettner, R. (2020, May 14). How does the new coronavirus compare with the flu? *Live Science*. https://www.livescience.com/new-coronavirus-compare-with-flu.html

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*(2), 346-362. https://doi.org/10.1037/0022-0663.93.2.346

Rivers, M. L., Fitzsimmons, C. J., Fisk, S. R., Dunlosky, J., & Thompson, C. A. (2020). Gender differences in confidence during number-line estimation. *Metacognition and Learning*, *16*, 157-178. https://doi.org/10.1007/s11409-020-09243-7

Sandman, P. M., Weinstein, N. D., & Miller, P. (1994). High risk or low: how location on a "risk ladder" affects perceived risk. *Risk Analysis*, *14*(1), 35-45. https://doi.org/10.1111/j.1539-6924.1994.tb00026.x

Saxe, G. B., Shaughnessy, M. M., Gearhart, M., & Haldar, L. C. (2013). Coordinating numeric and linear units: Elementary students' strategies for locating whole numbers on the number line. *Mathematical Thinking and Learning, 15*(4), 235-258. https://doi.org/10.1080/10986065.2013.812510

Scheibe, D. A., Fitzsimmons, C. J., Mielicki, M. K., Taber, M. J., Sidney, P. G., Coifman, K., & Thompson, C. A. (2022). Confidence in COVID problem solving: What factors predict adults' item-level metacognitive judgments on health-related math problems before and after an educational intervention? *Metacognition and Learning.* Advance online publication. DOI: 10.1007/s11409-022-09300-3

Schiller, L. K. (2020). *Recharging rational number understanding* (Publication No. 27671317) [Doctoral dissertation, Columbia University]. ProQuest Dissertations and Theses Global.

Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: Their interrelations in grades 5 and 6. *Journal of Educational Psychology, 101*(2), 359-372. https://doi.org/10.1037/a0013840

Sheeran, P., Harris, P. R., & Epton, T. (2014). Does heightening risk appraisals change people's intentions and behavior? A meta-analysis of experimental studies. *Psychological Bulletin, 140*(2), 511-543. https://doi.org/10.1037/a0033065

Sidney, P. G., Thalluri, R., Buerke, M. L., & Thompson, C. A. (2019). Who uses more strategies? Linking mathematics anxiety to adults' strategy variability and performance on fraction magnitude tasks. *Thinking and Reasoning, 25*(1), 94-131. https://doi.org/10.1080/13546782.2018.1475303

Sidney, P. G., Thompson, C. A., Fitzsimmons, C., & Taber, J. M. (2019). Children's and adults' math attitudes are differentiated by number type. *The Journal of Experimental Education,* 1-32. https://doi.org/10.1080/00220973.2019.1653815

Sidney, P. G., Thompson, C. A., Matthews, P. G., & Hubbard, E. M. (2017). From continuous magnitudes to symbolic numbers: The centrality of ratio. *Behavioral and Brain Sciences*, *40*, 42-43. https://doi.org/10.1017/S0140525X16002284

Sidney, P. G., Thompson, C. A., & Rivera, F. (2019). Number lines, but not area models, support children's accuracy and conceptual models of fraction division. *Contemporary Educational Psychology*, *58*, 288-298. https://doi.org/10.1016/j.cedpscyh.2019.03.011

Siegler, R. S., & Opfer, J. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14,* 237-243. https://doi.org/10.1111/1467-9280.02438

Siegler, R. S., & Thompson, C. A. (2014). Numerical landmarks are useful – except when they're not. *Journal of Experimental Child Psychology*, *120*, 39-58. https://doi.org/10.1016/j.jecp.2013.11.014

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62*, 273-296. https://doi.org/10.1016/j.cogpsych.2011.03.001

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *The state-trait anxiety inventory (test manual)*. Palo Alto, CA: Consulting Psychologists Press.

Taber, J. M., & Klein, W. M. (2016). The role of conviction in personal disease risk perceptions: What can we learn from research on attitude strength?. *Social and personality psychology compass, 10*(4), 202-218. https://doi.org/10.1111/spc3.12244

Taber, J. M., McQueen, A., Simonovic, N., & Waters, E. A. (2019). Adapting a self-affirmation intervention for use in a mobile application for smokers. *Journal of behavioral medicine*, *42*(6), 1050-1061. https://doi.org/10.1007/s10865-019-00028-1

Thompson, C. A., Mielicki, M. K., Rivera, F. D., Fitzsimmons, C. J., Scheibe, D., Sidney, P. G., Taber, J. M. & Waters, E. A. (2022). Leveraging mathematical cognition to combat innumeracy. *Perspectives on Psychological Science*. Advance online publication. https://doi.org/10.31234/osf.io/m8ynk

Thompson, C. A., Taber, J. M., Sidney, P. G., & Coifman, K. G. (2020). Math misconceptions may lead people to underestimate the true threat of COVID-19. The Conversation. https://theconversation.com/math-misconceptions-may-lead-people-to-underestimate-the-true-threat-of-covid-19-134520

Thompson, C. A., Taber, J. M., Sidney, P. G., Fitzsimmons, C. J., Mielicki, M. K., Matthews, P. G., Schemmel, E. A., Simonovic, N., Foust, J. L., Aurora, P., Disabato, D. J., Seah, T. H. S., & Coifman, K. G. (2021). Math matters: A novel, brief educational intervention decreases whole number bias when reasoning about COVID-19 *Journal of Experimental Psychology: Applied. 27(4),* 632–656. https://doi.org/10.1037/xap0000403

Thompson, C. A., Taber, J. M., Fitzsimmons, C., & Sidney, P. (in press). Strategy reports involving attention to math are associated with accurate responses on a numeric health decision-making problem. *Journal of Numerical Cognition*.

Toussaint, L. L., Cheadle, A. D., Fox, J., & Williams, D. R. (2020). Clean and contain: Initial development of a measure of infection prevention behaviors during the COVID-19 pandemic. *Annals of Behavioral Medicine, 54*(9), 619-625. https://doi.org/10.1093/abm/kaaa064

Trevena, L. J., Bonner, C., Okan, Y., Peters, E., Gaissmaier, W., Han, P. K. J., Ozanne, E., Timmermans, D., & Zikmund-Fisher, B. J. (2021). Current Challenges When Using Numbers in Patient Decision Aids: Advanced Concepts. *Medical Decision Making*. Advance online publication. https://doi.org/10.1177/272989X21996342

Vallentin, D., & Nieder, A. (2008). Behavioral and prefrontal representation of spatial proportions in the monkey. *Current Biology, 18*(18), 1420-1425. https://doi.org/10.1016/j.cub.2008.08.042

Turchioe, M. R., Myers, A., Isaac, S., Baik, D., Grossman, L. V., Ancker, J. S., & Creber, R. M. (2019). A systematic review of patient-facing visualizations of personal health data. *Applied clinical informatics*, *10*(4), 751-770. https://doi.org/10.1055/s-0039-1697592

Walker, M. (2020, April 27). COVID-19 no worse than the flu? Hardly: New York City data show dramatic imbalance. *MedPage Today*. https://www.medpagetoday.com/infectiousdisease/covid19/86176

Waters, E. A., Fagerlin, A., & Zikmund-Fisher, B. J. (2016). Overcoming the many pitfalls of communicating risk. In M. A. Diefenbach, S. Miller-Halegoua, & D. J. Bowen (Eds.), *Handbook of Health Decision Science,* pp. 265-277. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3486-7_19

Waters, E. A., Maki, J., Liu, Y., Ackermann, N., Carter, C. R., Dart, H., Bowen, D. J., Cameron, L. D., & Colditz, G. A. (2021). Risk ladder, table, or bulleted list? Identifying formats that effectively communicate personalized risk and risk reduction information for multiple diseases. *Medical Decision Making: An international journal of the Society for Medical Decision Making, 41*(1), 74-88. https://doi.org/10.1177/0272989X20968070

Waters, E. A., Foust, J. L., Scherer, L. D., McQueen, A., & Taber, J. M. (2021). Adherence of internet-based cancer risk assessment tools to best practices in risk communication: Content analysis. *Journal of Medical Internet Research, 23*(1): e23318. https://doi.org/10.2196/23318

Waters, E. A., Taber, J. M., McQueen, A., Housten, A. J., Studts, J. L., & Scherer, L. D. (2020). Translating cancer risk prediction models into personalized cancer risk assessment tools: stumbling blocks and strategies for success. *Cancer Epidemiology and Prevention Biomarkers*, *29*(12), 2389-2394. https://doi.org/10.1158/1055-9965.EPI-20-0861

Weinstein, N. D., Kwitel, A., McCaul, K. D., Magnan, R. E., Gerrard, M., & Gibbons, F. X. (2007). Risk perceptions: Assessment and relationship to influenza vaccination. *Health Psychology, 26*(2), 146. https://doi.org/10.1037/0278-6133.26.2.146

Weinstein, N. D., Sandman, P. M., & Hallman, W. K. (1994). Testing a visual display to explain small probabilities. *Risk Analysis*, *14*(6), 895-896.

World Health Organization (2020). Novel coronavirus (2019-nCoV): Situation report - 13. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf?sfvrsn=195f4010_6

Yan, H. (2020, May 1). Critics said the flu kills more than coronavirus. Why that's not a fair comparison – and now, it's not even true. *CNN*. https://www.cnn.com/2020/05/01/health/flu-vs-coronavirus-deaths/index.html

Zikmund-Fisher, B. J., Fagerlin, A., & Ubel, P. A. (2010). A demonstration of 'less can be more' in risk graphics. *Medical Decision Making, 30*(6), 661-671. https://doi.org/10.1177/0272989X10364244

Zikmund-Fisher, B.J., Witteman, H. O., Dickson, M., Fuhrel-Forbis, A., Kahn, V. C., Exe, N. L., Valerio, M., Holtzman, L. G., Scherer, L. D., & Fagerlin, A. (2014). Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. *Medical Decision Making*, *34*(4), 443-453. https://doi.org/10.1177/0272989X13511706

Table 1

Exact Wording for All Health-Related Math Problems

Item	Wording
Pretest (Hypothetical)	Consider the following two diseases. A total of 55,924 people have been infected by Disease A, and 2,125 of those people have died from Disease A. A total of 1,677,720 people have been infected with Disease B, and 16,777 of those people have died from Disease B. Which of the two diseases above is more fatal: Disease A, Disease B, or are they equally fatal?
Age Ranges	The Center for Disease Control and Prevention (CDC) has indicated that this year there have been 504,433 total deaths from all causes for people aged 75-84. 51,365 deaths from this age group have been attributed to COVID-19. The CDC also indicated this year that there have been 625,674 total deaths from all causes for people aged 85 and older. 59,801 deaths from this age group have been attributed to COVID-19 In which age group is the cause of death more likely to be COVID-19?
Another Country	Approximately 328,200,000 people live in the U.S. As of early September, there were 5,800,472 lab-confirmed cases of COVID-19, or tests that came back "positive." Approximately 1,353,000,000 people live in another country. As of early September in this country, there were 3,234,474 lab-confirmed cases of COVID-19, or tests that came back "positive." Based on the evidence above, did more people test positive in the U.S., the other country, or were people equally likely to test positive in both countries?
Bowls	Imagine there are two bowls of jellybeans on a table in front of you. If you draw a red jellybean from one of the two bowls described below, you will win a prize. Bowl A contains 10 red jellybeans and 100 white jellybeans. Bowl B contains 90 red jellybeans and 1,000 white jellybeans. Which of the two bowls do you want to draw a jellybean from?
Global	According to the CDC, there were 50,000,000 deaths worldwide from the 1918 flu pandemic, and 675,000 of the deaths occurred

Pandemics	in the U.S. Since March, there have been 1,053,357 deaths worldwide from the COVID-19 pandemic, and 211,694 of the deaths have occurred in the U.S. In which pandemic did the U.S. have the greater share of deaths?
March COVID-19	As of March 2020, Johns Hopkins indicated that 227,743 people worldwide were infected with COVID-19, and 9,318 had died. The Center for Disease Control (CDC) indicated that 36,000,000 people were infected with the flu, and 22,000 had died. Which disease was more fatal: the flu, COVID-19, or are they equally fatal?
October COVID-19	As of October 2020, Johns Hopkins indicated that 7,768,629 people in the U.S. were infected with COVID-19, and 214,844 had died. The Center for Disease Control (CDC) indicated that 56,000,000 people in the U.S. were infected with the flu, and 62,000 had died. Which disease was more fatal: the flu, COVID-19, or are they equally fatal?
Universities	As of early September, University A in the U.S. had tested 1,154 total students for COVID-19. At University A, 125 students had lab-confirmed cases of COVID-19 or tests that came back “positive.” Another university in another part of the U.S., University B, had tested 22,470 total students for COVID-19. At University B, 249 students had lab confirmed cases of COVID-19, or tests that came back “positive.” Based on the evidence above, are people more likely, less likely, or equally likely to test positive for COVID-19 at University A as compared to University B?
World War 1	According to the Johns Hopkins website early in September, of the 6,889,086 Americans who were diagnosed with COVID-19 in the U.S., 200,641 of them died. According to Britannica, 4,355,000 Americans fought in WW1, and 116,516 died. Based on the evidence above, were people more likely, less likely, or equally likely to die from COVID-19 or in WW1?

Note: The Age Ranges and World War 1 problems were not included in the final analyses.

Table 2

Participant Characteristics for the Final Sample

	N	Percent
Gender		
Male	31	13.08%
Female	200	84.39%
Nonbinary	0	0%
Different	1	0.42%
Did Not Report	5	2.11%
Race and Ethnicity		
White	184	77.64%
Black or African American	16	6.75%
Hispanic or Latino	6	2.53%
American Indian or Alaska Native	0	0%
Asian	7	2.95%
Native Hawaiian or Pacific Islander	0	0%
Other	2	0.84%
Did Not report	5	2.11%
Multiple	17	7.17%
Year in School		
Freshman	148	62.45%
Sophomore	43	18.14%
Junior	29	12.24%
Senior	16	6.75%
Did Not Report	1	0.42%

Table 3

Descriptive Statistics and Correlations for All Measures

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Problem-Solving Perf.	4.55	1.65															
2. PAE large	0.11	0.10	**-.29**														
3. PAE small	0.03	0.02	**-.21**	.07													
4. Equivalence	0.73	0.26	**.27**	**-.35**	**-.16**												
5. Math Anxiety	4.43	2.06	**-.28**	**.19**	.12	**-.27**											
6. Math Attitudes	3.92	0.81	**.34**	**-.33**	-.12	**.34**	**-.58**										
7. Sub. Graph Lit.	25.48	5.98	**.22**	-.10	**-.15**	**.14**	**-.24**	**.39**									
8. Obj. Graph Lit.	2.27	1.09	**.27**	**-.23**	**-.13**	**.14**	**-.15**	**.20**	**.29**								
9. Non-sym. Ratio	0.69	0.19	**.16**	-.13	-.02	**.13**	**-.18**	**.13**	.09	.08							
10. Rel. Reasoning	2.73	0.97	**.25**	**-.14**	**-.26**	**.13**	-.10	**.13**	.11	**.21**	**.19**						
11. COVID Risk Self	2.38	0.82	.10	-.01	.01	-.07	-.04	.10	-.01	.06	.02	-.01					
12. COVID Risk Other	2.67	0.86	.10	.06	-.06	.04	-.03	**.15**	.02	**.14**	-.02	.09	**.32**				
13. COVID Worry Self	2.70	1.24	.09	.10	.03	**-.21**	**.17**	.07	.05	-.00	-.09	-.01	**.14**	.10			
14. COVID Worry Other	3.80	1.12	**.22**	.05	-.05	**-.14**	**.18**	.04	.11	.06	-.07	.04	**.15**	**.30**	**.57**		
15. COVID Severity	3.36	0.76	**.16**	.05	.01	**-.14**	.11	.04	**.14**	.11	-.01	.06	**.23**	**.25**	**.52**	**.52**	
16. Behavioral Intentions	4.25	0.48	**.14**	.05	.02	**-.16**	.10	.10	**.16**	.01	-.02	-.01	.03	**.13**	**.37**	**.45**	**.44**

Note: Correlations with $p < .05$ are bolded. Problem-Solving Perf. = total problem-solving score (out of 6 possible). PAE large = large range percent absolute error on the number line estimation task. PAE small = small range percent absolute error on the number line estimation task. Equivalence = proportion correct on equivalence measure. Math Anxiety = math anxiety (1 to 10 scale). Math Attitudes = math attitudes (1 to 6 scale). Sub. Graph. Lit. = subjective graph literacy (1 to 7 scales). Obj. Graph Lit. = score on objective graph literacy measure (out of 4). Non-sym. Ratio = proportion correct on non-symbolic ratio task. Rel. Reasoning = test of relational reasoning (out of 4). COVID Risk Self = perceived likelihood of COVID-19 for self (1 to 5 scale). COVID Risk Other = Perceived likelihood of COVID-19 for others (1 to 5 scale). COVID Worry Self = COVID-19 worry for self (1 to 5 scale). COVID Worry Other = COVID-19 worry for others (1 to 5 scale). COVID Severity = Perceived COVID-19 severity (1 to 4 scale). Behavioral Intentions = preventative health behavior intention inventory (1 to 5 scale).

Table 4

Percentage of Correct Responses by Item and Condition

	No Visual Model	Icon Array	Number Line	Risk Ladder
Pretest (Hypothetical)	63%	53%	49%	56%
Age Ranges	37%	10%	28%	31%
Another Country	74%	76%	82%	66%
Bowls	81%	88%	89%	85%
Global Pandemics	61%	59%	79%	57%
March COVID-19	65%	69%	79%	59%
October COVID-19	75%	88%	93%	85%
Universities	74%	81%	89%	66%
World War 1	51%	38%	52%	31%

Note: The Age Ranges and World War 1 problems were not included in the health cognition analyses.

Table 5

Logistic Mixed Effects Model for Accuracy on Health-related Ratio Problems

Fixed Effects	OR	*b* (SE)	*b* 95% CI	Wald χ^2	Random Effects	Variance
Constant	2.97	1.09 (0.44)	[0.22, 1.96]	6.06*	Subj. (Intercept)	0.63
Number Line vs. No Model	0.45	-0.80 (0.23)	[-1.26, -0.35]	11.99***	Item (Intercept)	1.30
Number Line vs. Risk Ladder	0.37	-1.00 (0.22)	[-1.44, -0.56]	20.02***		
Number Line vs. Icon Array	0.41	-0.90 (0.23)	[-1.35, -0.45]	15.52***		
Pretest Accuracy	2.15	0.76 (0.18)	[0.42, 1.11]	19.02***		
Number Line Estimation (PAE)	0.84	-0.17 (0.11)	[-0.38, 0.04]	2.65		
Equivalence Performance	1.14	0.13 (0.09)	[-0.04, 0.30]	2.18		
Objective Graph Literacy	1.20	0.18 (0.09)	[0.005, 0.36]	4.04*		
Non-Symbolic Ratio	1.17	0.13 (0.08)	[-0.03, 0.29]	2.43		
Subjective Graph Literacy	1.05	0.05 (0.09)	[-0.13, 0.22]	0.27		
Math Anxiety	0.99	-0.01 (0.10)	[-0.20, 0.18]	0.01		
Math Attitudes	1.27	0.23 (0.11)	[0.02, 0.45]	4.60*		
Rel. Reasoning	1.19	0.17 (0.09)	[-0.01, 0.35]	3.55		

Note. These analyses were based on 1896 observations (237 participants, 8 items). Log Likelihood = -960.00, AIC = 1950.00. BIC = 2033.00. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 6

Linear Regression Models for Health Cognition Measures

	Risk Self	Risk Other	Worry Self	Worry Other	Severity	Behavioral Intentions
	b (*SE*) [95% CI]	*b* (*SE*) [95% CI]	*b* (*SE*) [95% CI]	*b* (*SE*) [95% CI]	*b* (*SE*) [95% CI]	*b* (*SE*) [95% CI]
Constant	**2.15 (0.20) [1.76, 2.54]**	**2.38 (0.20) [2.00, 2.77]**	**1.93 (0.27) [1.40, 2.46]**	**2.89 (0.25) [2.41, 3.38]**	**2.81 (0.17) [2.48, 3.14]**	**3.97 (0.11) [3.76, 4.19]**
Problem-Solving Performance	0.04 (0.04) [-0.04, 0.12]	0.01 (0.04) [-0.06, 0.09]	**0.15 (0.05) [0.04, 0.26]**	**0.19 (0.05) [0.09, 0.29]**	**0.12 (0.03) [0.05, 0.18]**	**0.06 (0.02) [0.02, 0.11]**
Pretest	0.10 (0.13) [-0.15, 0.35]	**0.27 (0.13) [0.03, 0.52]**	-0.14 (0.17) [-0.48, 0.19]	0.08 (0.16) [-0.23, 0.39]	**-0.27 (0.11) [-0.49, 0.06]**	-0.10 (0.07) [-0.23, 0.04]
Num. Line Estimation (PAE)	0.02 (0.08) [-0.13, 0.17]	**0.16 (0.08) [0.01, 0.31]**	0.14 (0.10) [-0.06, 0.34]	0.10 (0.09) [-0.08, 0.29]	0.02 (0.06) [-0.10, 0.15]	0.02 (0.04) [-0.06, 0.11]
Equivalence Performance	-0.11 (0.06) [-0.23, 0.01]	-0.01 (0.06) [-0.13, 0.12]	**-0.30 (0.08) [-0.46, -0.14]**	**-0.20 (0.08) [-0.35, 0.05]**	**-0.16 (0.05) [-0.26, -0.05]**	**-0.10 (0.03) [-0.17, -0.04]**
Objective Graph Literacy	0.05 (0.07) [-0.08, 0.18]	0.11 (0.06) [-0.02, 0.23]	-0.01 (0.09) [-0.180, 0.16]	0.02 (0.08) [-0.13, 0.18]	0.05 (0.05) [-0.06, 0.16]	-0.02 (0.03) [-0.09, 0.05]
Non-Symbolic Ratio Performance	0.04 (0.06) [-0.08, 0.15]	-0.03 (0.06) [-0.14, 0.09]	-0.09 (0.08) [-0.25, 0.06]	-0.06 (0.07) -0.2, 0.08	0.002 (0.05) [-0.10, 0.10]	-0.01 (0.03) [-0.07, 0.05]
Subjective Graph Literacy	-0.07 (0.06) [-0.20, 0.06]	-0.06 (0.06) [-0.183, 0.07]	0.04 (0.09) [-0.13, 0.20]	0.11 (0.08) [-0.05, 0.26]	**0.11 (0.05) [0.00, 0.21]**	**0.07 (0.03) [0.00, 0.14]**
Math Anxiety	0.04 (0.07) [-0.11, 0.18]	0.08 (0.07) [-0.06, 0.22]	**0.34 (0.10) [0.15, 0.53]**	**0.33 (0.09) [0.16, 0.50]**	**0.15 (0.06) [0.04, 0.27]**	**0.10 (0.04) [0.03, 0.18]**
Math Attitudes	0.13 (0.08) [-0.02, 0.28]	**0.19 (0.08) [0.04, 0.33]**	**0.40 (0.10) [0.20, 0.61]**	**0.20 (0.09) [0.01, 0.38]**	**0.13 (0.06) [0.00, 0.26]**	**0.12 (0.04) [0.03, 0.20]**
Relational Reasoning	-0.04 (0.07) [-0.17, 0.09]	0.05 (0.06) [-0.08, 0.18]	-0.03 (0.09) [-0.20, 0.14]	0.01 (0.08) [-0.15, 0.17]	0.02 (0.06) -0.09, 0.13]	-0.02 (0.04) [-0.09, 0.05]
University	-0.01 (0.12) [-0.26, 0.23]	0.15 (0.12) [-0.09, 0.38]	**0.38 (0.16) [0.05, 0.70]**	0.02 (0.15) [-0.28, 0.31]	**0.32 (0.10) [0.11, 0.52]**	0.08 (0.07) [-0.05, 0.21]
Trait Anxiety	0.00 (0.06) [-0.13, 0.13]	0.12 (0.06) [-0.00, 0.24]	0.13 (0.08) [-0.04, 0.29]	0.12 (0.08) [-0.03, 0.28]	0.02 (0.05) [-0.05, 0.12]	0.01 (0.03) [-0.06, 0.08]

R^2	0.04	0.13	0.20	0.19	0.19	0.14
Adjusted R^2	-0.01	0.08	0.16	0.15	0.14	0.10
Residual *SE*	0.82	0.82	1.13	1.04	0.71	0.46
F Test	$F(12, 210) =$ 0.80	$F(12, 218) =$ 2.68	$F(12, 223) =$ 4.75	$F(12, 223) =$ 4.33	$F(12, 223) =$ 4.22	$F(12, 223) =$ 3.09

Note: Predictors were entered simultaneously for each model. Bolded values indicate $p < .05$.