1-in-X" bias: "1-in-X" format causes overestimation of health-related risks

Miroslav Sirota[1], Marie Juanchich[1], Jean-Francois Bonnefon[2]

[1]Department of Psychology, University of Essex

[2]Toulouse School of Economics (TSM-R), CNRS, University of Toulouse Capitole

Word count (text only): 4,674

Author Note

The support of Jean-Francois Bonnefon from the ANR-LabEx IAST is gratefully acknowledged.

To be published in Journal of Experimental Psychology: Applied

Correspondence concerning this paper should be addressed to Miroslav Sirota, Department of Psychology, Wivenhoe Park, University of Essex, Colchester, CO4 3SQ, United Kingdom (email: msirota@essex.ac.uk)

The data and materials are publicly available on the Open Science Framework at https://osf.io/dsp8e/?view_only=614a2dafda45493c8f06429274110e65

Abstract

According to the “1-in-X” effect, “1-in-X” ratios (e.g., 1 in 12) trigger a higher subjective probability than numerically equivalent “N-in-X*N” ratios (e.g., 3 in 36). Here we tested: (i) the effect on objective measures, (ii) its consequences for decision-making, (iii) whether this effect is a form of bias by measuring probability accuracy, and (iv) its amplification in people with lower health literacy and numeracy. In parallel-designed experiments, 975 participants from the general adult population participated in one of five experiments following a 2(format: “1-in-X” or “N-in-X*N”) × 4(scenarios) mixed design. Participants assessed the risk of contracting a disease on either a verbal probability scale (Exp. 1), or a numerical probability/frequency scale with immediate (Exp. 2-3) or delayed presentation (Exp. 4-5). Participants also made a health-related decision and completed a health literacy and numeracy scale. The “1-in-X” ratios yielded higher probability perceptions than the “N-in-X*N” ratios and affected relevant decisions. Critically, the “1-in-X” ratios led to a larger objective overestimation of numerical probabilities than the “N-in-X*N” ratios. People with lower levels of health literacy and numeracy were not more sensitive to the bias. Health professionals should use “1-in-X” ratios with great caution when communicating to patients, because they overestimate health risks.

Keywords: “1-in-X” effect, bias, health risk communication, health literacy, numeracy

Public Significance Statement

In five experiments, we present evidence that people incorrectly interpret “1-in-X” ratios (e.g., 1 in 13) as conveying a higher probability than the numerically equivalent “N-in-X*N” ratios (e.g., 10 in 130), and that this effect is due to an overestimation of “1-in-X” ratios. Health professionals should use “1-in-X” ratios with caution since these ratios might make patients to overestimate health risks, and to make ill-informed decisions as a result.

The accurate perception and understanding of probabilistic information is the cornerstone of well-informed medical decisions and health-related behaviour change (e.g., Lipkus, 2007). To facilitate such accurate perception and understanding, prior research focused on optimising the communication of probabilistic health information (e.g., Fagerlin, Ubel, Smith, & Zikmund-Fisher, 2007a; Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007; Lipkus, Samsa, & Rimer, 2001; Sirota, Juanchich, & Hagmayer, 2014a), while taking into account individual differences such as numerical and graphical literacy (e.g., Gaissmaier et al., 2012; Galesic, Garcia-Retamero, & Gigerenzer, 2009). In the present paper, we have contributed to this research programme by documenting the probability perception of health risks when communicated using “1-in-X” (e.g., “1 chance in 13”) and “N-in-X*N” (e.g., “10 chances in 130”) ratios. We have assessed the systematic departure from correct probability estimates that these formats generate when communicated to people with different levels of health literacy and numeracy.

“1-in-X” effect or “1-in-X” bias?

According to the “1-in-X” effect, people perceive “1-in-X” ratios to be more likely than “N-in-X*N” ratios (Pighin et al., 2011). For example, 1 chance in 13 is perceived as greater than 10 chances in 130, even though the probability is numerically the same. This effect is robust as it occurs across different education levels, cultures and outcomes (Pighin et al., 2011; Pighin et al., 2015; Sirota, Juanchich, Kostopoulou, & Hanak, 2014). The “1-in-X” effect runs counter to the ratio bias effect (Yamagishi, 1997), according to which people neglect the denominator of fractions expressing probabilities (Bonner & Newell, 2008; Denesraj et al., 1995; Galesic et al., 2009; Kirkpatrick & Epstein, 1992; Okan, Garcia-Retamero, Cokely, & Maldonado, 2012). One critical difference between the two literatures is that ratio biases are typically obtained when people are presented with two ratios jointly, and asked to make a comparative judgment; whereas the “1-in-X” effect is typically obtained when people are presented with ratios separately, and asked to make an absolute magnitude judgment. Since health providers and consumers are often required to understand probability and frequency information using an absolute magnitude judgment, we have focused on such absolute judgments in this paper.

What we do not know yet about the “1-in-X” effect, although it is critical, is whether the effect is actually a form of bias, that is, an inaccurate distortion of probability perception, leading to suboptimal decisions. In other words, we do not know what the source of the probability distortion is: “1-in-X” format or the other formats to which researchers usually compared to “1-in-X”. Indeed, the fact that “1-in-X” ratios increase the magnitude of subjective probability judgments – measured in prior research exclusively on Likert scales using verbal probability expressions – does not logically entail that these judgments are inaccurate: they could actually be *better* calibrated (i.e., closer to the “true” objective probability of an event). Establishing whether or not the “1-in-X” effect is a form of bias is important for at least three related reasons.

First, there have been some recent calls to retire the “1-in-X” formats from health risk communication (Zikmund-Fisher, 2011, 2014). Even though such calls rely on the fact that the format change is cost-free and therefore easy to implement, they implicitly assume that the “1-in-X” effect is distorting probability perception. The supporting evidence for such claims is so far indirect. For instance, it seems that “1-in-X” ratios are not only subjectively higher than “N-in-X*N” ratios, but are also systematically higher compared with other formats such as percentages or pictorial representations (Pighin et al., 2011; Sirota, Juanchich, Kostopoulou, & Hanak, 2014b). Nevertheless, to fully justify the calls to withdraw “1-in-X” ratios from medical practice, we need more direct evidence of the inaccuracy of “1-in-X” probability perceptions.

Second, “1-in-X” ratios are highly prevalent in health communication (Sirota et al., 2017). For example, family physicians overwhelmingly prefer to communicate prenatal risks using “1-in-X” ratios (80.4%) when compared to alternative numerical formats (e.g., 3.8% cases of “N-in-X*N” format). Also, the UK’s National Health Service mostly uses “1-in-X” ratios to convey medical information to patients on its website: 45.7% cases conveyed in the “1-in-X” compared with 13.7% cases conveyed in the “N-in-X*N” format (Sirota et al., 2017). This high prevalence can magnify even small inaccuracy effects, and amplify the potential to bias patients’ judgment and decision-making, which underlies the urgency of the recent calls to abandon “1-in-X” ratios.

Third, we know that people with lower levels of health literacy and numeracy (i.e., the ability and disposition to understand and use numerical health information) are more prone to various cognitive biases, which impair their ability to engage in informed and shared health decision-making (Galesic & Garcia-Retamero, 2011; Galesic et al., 2009; Garcia-Retamero & Galesic, 2009; Gigerenzer et al., 2007; Kaphingst et al., 2015; Okan, Garcia-Retamero, Cokely, & Maldonado, 2012; Peters, 2012; Reyna, Nelson, Han, & Dieckmann, 2009; Reyna, 2008). Accordingly, we need to assess whether people with lower health literacy and numeracy levels might be differentially affected by a “1-in-X” bias, with amplified consequences on their health-related decision-making.

The Present Experiments

In this paper, we have reported five experiments which aim to systematically investigate the impact of the “1-in-X” format on probability perception accuracy. To do so, we used several response scales, including numerical scales which allow for an objective measure of accuracy, in contrast with the verbal probability scales used in prior research which only reflected variations in subjective probabilities. Using these numerical scales allows us to assess inaccuracy, but this comes at a price: instead of simply perceiving an event as more or less likely, as they would when they are told about medical risk, participants may try to conduct a precise mathematical calculation to translate this risk correctly onto the numerical scale. To tap into the same psychological processes as those involved in producing the "1-in-X" effect using verbal probability scales, we introduced several strategies to avoid precise calculation. First, we simply instructed participants not to precisely calculate the answers. Second, we used ratios for which numerical values are not easily calculated (e.g., "1 in 13" rather than "1 in 10"). Third, we designed variants of the experiment in which the estimation of the objective probability is delayed to a subsequent screen, instead of being presented together with the ratio. Finally, we asked participants whether they had calculated their answer, to be able to control statistically for this possibility.

In five experiments, we investigated four research questions. First, we investigated whether the "1-in-X" effect would be observed across the different scales (verbal, frequency and probability scales) and different procedures (immediate or delayed presentation of the scale), allowing for differences in the magnitude of the effect across scales and procedures (Question 1). Second, we investigated whether the "1-in-X" effect would affect decisions, in addition to subjective probability judgments (Question 2). Third, and most importantly, we tested whether judgments based on the "1-in-X" format would be more biased – i.e. show greater deviations from the objectively correct value than judgments based on the "N-in-X*N" format (Question 3). Fourth, we tested whether people with lower health literacy and numeracy levels would be more prone to the "1-in-X" bias (Question 4).

Methods

Participants

The stopping rule for the required number of participants was determined a priori to avoid possible increased false-positive rates (Bakker, van Dijk, & Wicherts, 2012). Based on a power analysis, we aimed to recruit at least 90 participants in each condition of the five experiments (i.e., at least 180 participants in each experiment, 900 participants overall). Such a sample size would allow us to detect at least a medium effect size, *Cohen's d* = 0.42 (an overall meta-analytical effect detected in the prior literature, see Sirota et al., 2014b), when assuming $\alpha = .05$, $1 - \beta = .80$, and a conservative two-sided *t*-test (Cohen, 1988). We increased the sample size by 10% to account for a possible attrition rate.

As a result, 1,003 Amazon Mechanical Turk workers completed the online questionnaire and were rewarded with $1 for their participation. The data from 28 people was removed based on a priori exclusion criteria – self-reported careless responding, when the participants indicated "no" to the question: "Lastly, it is vital to our study that we only include responses from people that devoted their full attention to this study. You will receive your reward for participation in this study no matter what, however, please tells us: In your honest opinion, should we use your data in our analyses in this study?" (Meade & Craig, 2012). Participants knew that their answer would not affect the financial compensation for their participation.

The final sample of 975 participants (54.7% males, age range from 18 to 74 years, *M* = 35.0, *SD* = 11.2) took part in one of five experiments. The sample was heterogeneous in terms of occupation: management and professionals (22.2%), unemployed, students and homemakers (19.3%), sales and office (17.8%), other categories (16.4%), service (12.4%) and some other less common occupations such as government-workers or those involved in the farming industry. Participants had different education levels: 0.5% did not complete their high school education, 30.8% achieved high school education, 56.5% achieved a college degree, 10.2% achieved a master's degree and 2.1% achieved a PhD or other professional degree.

Design

Five experiments were carried out using a parallel design (e.g., Sirota, Kostovičová, & Vallée-Tourangeau, 2015). As shown in Figure 1, participants were first allocated randomly to one of the five experiments, and then allocated randomly to one of the two experimental format conditions within that experiment ("1-in-X" vs. "N-in-X*N"). In each format condition, participants read four scenarios focusing on the risks of contracting different diseases that had different probability magnitudes; the presentation order of the scenarios was randomised for each participant. We chose to use a parallel design for three reasons. First, this design allowed us to test our questions using a variation of the response scales, whilst allowing us to analyse the results of each experiment individually. Second, we chose this design because we planned to conduct a small-scale meta-analysis and this design allowed us to avoid an arbitrary stopping rule for the number of included experiments that could inflate the chances of a type I error (Ueno, Fastrich, & Murayama, 2016). Third, this parallel design allowed us to draw direct causal claims about the variables manipulated across the experiments.

Insert please Figure 1 around here

Materials and Procedure

After giving informed consent and random allocation to conditions (Figure 1), participants were instructed to express their intuitions about the probability of some events, rather than a calculated mathematical response. Then, participants read four scenarios about the risk of contracting a disease whilst travelling abroad (with a random order of presentation for each participant). For instance, one scenario described the risk of being infected by malaria whilst travelling to Kenya: “Imagine that you have booked a trip to Kenya and now you learn that the risk of being infected by malaria during your trip to Kenya is [1 in 13 / 7 in 91]”. The remaining three scenarios described travel to Sierra Leone, Norway and Slovakia and featured risk magnitudes for Ebola, the flu and Lyme disease, respectively (see Table 1). Different probability ratios were selected with the condition that they would not be easy to calculate (as it would be, for example, if the risk was 1 in 10). The scenarios are presented in the Appendix. The order of presentation of the scenarios was randomised.

In all of the scenarios, participants assessed the probability of contracting the illness: (“In your opinion, the probability of being infected by malaria during your trip to Kenya is …”) on a scale that differed across experiments. In Exp. 1 (*verbal probability scale*), participants ticked one of the buttons of an 11-point Likert scale anchored with verbal quantifiers (1: *extremely low*, 11: *extremely high*). In Exp. 2 (*numerical probability scale*), participants selected their answer on a visual analogue probability scale ranging from 0 to 100 with increments of 1. The scale had 11 numerical labels (0, 10, 20 …) and moving the cursor on the slider showed the exact numerical value (e.g., 26). In Exp. 3 (*numerical frequency scale*), participants selected their response on a visual analogue frequency scale ranging from 0 to 286 by increments of 1. The scale had 11 labels (0, 29, 57 …) and moving the cursor on the slider enabled participants to see the exact value (e.g., 26). In Exp. 4 (*delayed numerical probability scale*), the same numerical probability scale was used as in Exp. 2, however the scale was presented on a subsequent page, hence its presentation was delayed. In Exp. 5 (*delayed frequency probability scale*), the same numerical frequency scale was used as in Exp. 3, however the scale was presented on a subsequent page, hence its presentation was delayed.

After they had assessed the probability of contracting the disease in the scenario, participants reported whether they calculated the answer ("Did you calculate the answer to the previous question?"; note this was only asked in Exp. 2-5) and decided whether they would take action to remedy the risk. ("Given this risk, how likely are you to cancel your trip to Kenya?") This was done by ticking a radio button on a 6-point Likert scale (1: *unlikely to cancel*, 6: *likely to cancel*). After answering all the scenarios, participants rated how easy it was to assess the risks. ("How difficult was it to map the risks associated with the travelling expressed as ratios on the provided response scale?") This was done using a 5-point Likert scale (ranging from 1: *very easy* to 5: *very difficult*).

Insert please Table 1 around here

Next, participants filled in a brief measure of health literacy and numeracy. The health literacy measure – *The Newest Vital Sign* (Weiss et al., 2005) – consisted of the nutritional label for an ice-cream accompanied by six open-ended questions (e.g., "If you eat the entire container, how many calories will you eat?") for which the answers were coded either as correct or incorrect. This measure was selected because it correlates highly with well-established measures of literacy and it is quick and simple to perform (Weiss et al., 2005). The internal consistency of the sum score was acceptable although low (Cronbach's α = 0.52). *The Subjective Numeracy Scale* (Fagerlin et al., 2007b; McNaughton, Cavanaugh, Kripalani, Rothman, & Wallston, 2015) consisted of three questions (e.g., "How good are you at working with fractions?") assessed on a 6-point Likert scale ranging from 1: *not good at all* to 6: *extremely good*. The average score had good internal consistency (Cronbach's α = 0.83). We used both measures separately because their correlation was relatively small, r =

0.24, indicating that they do not measure the same construct. Other authors adopted a similar approach (e.g., Rodríguez et al., 2013).

Finally, participants recorded some basic socio-demographic information (i.e., age, gender, education and occupation). The study was conducted in accordance with the ethical standards of the American Psychological Association. The project was approved by the Ethics Committee of the Department of Psychology, University of Essex. We have reported all experiments, measures, manipulations and exclusions here. The data are publicly available on the Open Science Framework at https://osf.io/dsp8e/?view_only=614a2dafda45493c8f06429274110e65

Results

Does the "1-in-X" effect alter probability perception? (Question 1)

We observed the "1-in-X" effect as a general pattern: the "1-in-X" ratios received greater magnitude than the "N-in-X*N" ratios. The effect was observed quite consistently across different scenarios and experiments (Figure 2). Clearly, the effect was more pronounced with the verbal probability scales (Exp. 1) and less so with the numerical scales (Exp. 2-5, Figure 2, panel E). To simplify the presentation of the results, we averaged the probability and frequency across the four scenarios in each experiment. (This was reasonable, because a mixed ANOVA showed a significant effect of the ratios format but no interaction between the scenarios and ratios format.)

Insert please Figure 2 around here

In Exp. 1 (verbal probability scale), the “1-in-X” ratios led to significantly higher probabilities than the “N-in-X*N” ratios, $F(1, 190) = 19.4$, $p < .001$, $\mu_p^2 = 0.09$. In Exp. 2 (numerical probability scale), the “1-in-X” ratios led to a higher probability than the “N-in-X*N” ratios, but not significantly so, $F(1, 197) = 0.4$, $p = .532$, $\mu_p^2 = 0.002$ and this did not change after controlling for the calculation covariate, $F(1,195) = 0.6$, $p = .423$. In Exp. 3 (numerical frequency scale), the “1-in-X” ratios led to a higher probability than the “N-in-X*N” ratios, but not significantly so, $F(1, 188) = 1.1$, $p = .297$, $\mu_p^2 = 0.01$ and did not change after controlling for the calculation covariate, $F(1,186) = 1.1$, $p = .305$. In Exp. 4 (delayed numerical probability scale), the “1-in-X” ratios led to a higher probability than the “N-in-X*N” ratios and the difference was statistically significant, $F(1, 194) = 5.9$, $p = .016$, $\mu_p^2 = 0.03$, and did not change after controlling for the calculation covariate, $F(1,192) = 5.7$, $p = 0.017$. Finally, in Exp. 5 (delayed numerical frequency scale), the “1-in-X” ratios led to a higher probability than the “N-in-X*N” ratios but the difference was not statistically significant, $F(1, 191) = 1.2$, $p = .284$, $\mu_p^2 = 0.01$, and did not change after controlling for the calculation covariate, $F(1,192) = 0.9$, $p = .354$.

Thus, the direction of the effect was consistently replicated, but the findings were mixed in terms of statistical significance. To better assess the overall effect of the “1-in-X” ratio, we conducted a meta-analysis on the aggregated data using a random effect model. We found that the overall “1-in-X” meta-analytical effect was $g = 0.27$, 95% CI[0.08, 0.47], $z = 2.8$, $p = .005$. The scale was a significant moderator in a random effect model, $g = -0.44$, 95% CI [-0.75, -.14], $QM(1) = 7.9$, $p = .005$. The effect was more pronounced for the verbal probability scale, $g = 0.63$, 95% CI[0.35, 0.91]), compared with the numerical scales, $g = 0.19$, 95% CI[0.05, 0.32], but the overall effect across the four numerical scale experiments was still significant, $z = 2.6$, $p = .009$. Thus, we replicated the “1-in-X” effect for the numerical as well as for the verbal scales.

Does the "1-in-X" effect alter decision-making? (Question 2)

Participants in the "1-in-X" condition were more likely to take a safety measure to reduce the risk (i.e., cancel their trip, Figure 3). In Exp. 1 (verbal probability scale), participants in the "1-in-X" condition were significantly more likely to cancel their trip, $F(1, 190) = 5.6$, $p = .019$, $\mu_p^2 = 0.03$. In Exp. 2 (numerical probability scale), 3 (numerical frequency scale) and 4 (delayed numerical probability scale), participants in the "1-in-X" condition were slightly more likely to cancel their trip, but the effect was not significant, $F(1, 198) = 0.2$, $p = .678$, $\mu_p^2 < 0.01$, $F(1, 189) = 0.1$, $p = .827$, $\mu_p^2 < 0.01$, $F(1, 196) = 0.4$, $p = .515$, $\mu_p^2 < 0.01$, respectively. In Exp. 5 (delayed numerical frequency scale), participants in the "1-in-X" condition were significantly more likely to cancel their trip, $F(1, 192) = 4.6$, $p = .034$, $\mu_p^2 = 0.02$. We summarised the effect on decision-making through a meta-analysis of the aggregated data using a random effect model. We found that the overall "1-in-X" meta-analytical effect on decision-making was small but significant, $g = 0.17$, 95% CI[0.04, 0.29], $z = 2.5$, $p = .012$ (Figure 3, panel F).

Is the "1-in-X" effect an overestimation bias? (Question 3)

To assess objective accuracy (Exp. 2 to 5) we transformed the answers given on the frequency scales into probability (0-100%) scales and then calculated the distance from the objective value for each estimate. For example, a deviation of 0 meant that the answer was perfectly aligned with the objective value, a deviation of -5 meant that a participant underestimated the objective value by a magnitude of 5 percentage points, and a deviation of + 5 meant that a participant overestimated the objective value by a magnitude of 5 percentage points.

Overall, across all the scenarios and experiments, participants were less accurate in the "1-in-X" condition, overestimating the objective value by 6.0 percentage points (*SD* =

17.1), compared to an overestimation of 3.5 points in the "N-in-X*N" condition (*SD* = 14.3). The trend to overestimate was consistent across all scenarios (Figure 4).

Insert please Figure 4 around here

For statistical inference, we used a multi-level modelling approach, which allowed us to explicitly model other relevant variables (e.g., adjusting for calculation, health literacy, numeracy) and their interactions with ratios, which would not be possible in a meta-analytical approach. Through an iterative process, the final model featured ratio and scenario as fixed factors and random intercepts within experiments, scenarios and subjects. (We removed the full random structure due to convergence problems; we also removed the interaction between the fixed effects of ratio and scenario from the final model, since comparing the model with and without interaction yielded a non-significant difference, $\chi^2(3) = 2.45$, $p = .484$). In this final model, we found a significant effect of ratio, $F(1, 777.72) = 6.1$, $p = .014$, and a significant effect of scenario, $F(3, 2341.30) = 25.6$, $p < .001$. Critically, both the "1-in-X" and "N-in-X*N" ratios were not perfectly accurate, since the intercept was 7.0 and "N-in-X*N" decreased this in accuracy by 2.2 points.

We then tested whether such a bias might have occurred due to performing actual calculations and/or subjective perception (ease) of ratios. The random intercept model reported above was adjusted for the self-reported measure of calculation and the interaction with the ratio. In this extended model, the observed effect of the ratio remained the same and the calculation did not affect the deviation, $b = 0.8$, $F(1, 2956.9) = 1.0$, $p = .317$; nor did it interact with the ratio format, $b < 0.1$, $F(1, 2965.5) < 0.1$, $p = .996$. In a similar way, we did not observe any effect of ratio format on the self-reported ease. Across the four experiments featuring a numerical scale, the subjective ease was virtually the same in the “1-in-X” condition (*M* = 2.4, *SD* = 1.0) as in the “N-in-X*N” condition (*M* = 2.4, *SD* = 1.0). We found no significant effect of ratio on the subjective ease, $b < 0.1$, $F(1, 194.99) = 0.03$, $p = .864$, in a multi-level model with ratio as the fixed factor, with a random intercept and slope within the experiments. Thus, the effect of the ratio format on inaccuracy could not be attributed to the different objective calculation requirements of the ratios nor to the subjective perception of computation difficulty.

Are people with lower health literacy and numeracy levels more prone to the “1-in-X” bias? (Question 4)

Finally, we wanted to identify the role of health literacy (*M* = 5.0, *SD* = 1.2, max = 6) and numeracy (*M* = 4.5, *SD* = 1.1, max = 6) might play in the propensity for the “1-in-X” bias. We used the same multi-level model with ratio and scenario as fixed factors and added heath literacy and numeracy and their interaction into the model. Literacy was substantially negatively skewed (-1.28) and to improve its skewness the score was reflected, logarithmically transformed and reflected back (final skewness was improved -0.41). Both variables were then mean-centered for better interpretability. According to the final random-intercept model, we found the same patterns regarding our main fixed factors: a significant intercept, $b_0 = 7.2$, $t(3.8) = 4.3$, $p = .015$ and a significant effect of ratio, $F(1, 771.68) = 6.6$, $p = .010$ and scenario, $F(3, 2341.34) = 25.6$, $p < .001$. Health literacy and subjective numeracy significantly and substantially decreased deviation by $b = -11.8$ points, $F(1, 773.32) = 33.3$, $p < .001$ and by $b = -1.76$ points, $F(1, 771.60) = -3.1$, $p = .002$, respectively. However, health literacy, $b = 1.9$, $F(1, 772.06) = 0.3$, $p = .617$, subjective numeracy, $b = 1.2$, $F(1, 772.93) = 2.2$, $p = .140$, or both, $b = -4.3$, $F(1, 772.15) = 1.7$, $p = .199$, did not interact with ratio manipulation. Thus, even though more health-literate and numerate participants were less likely to overestimate the objective probabilities, they were not any less sensitive to the “1-in-X” bias.

Discussion

Across the five experiments reported here, we found that the “1-in-X” format increased perceived probability compared with the “N-in-X*N” format, not only when participants provided their perceptions on a verbal probability scale but also on numerical probability and frequency scales. While the effect was smaller for the numerical scales, it was reliable, being directionally replicated across the four experiments featuring a numerical scale, and detected as significant in a meta-analysis. This finding thus replicates and extends the prior research on the “1-in-X” effect (Oudhoff & Timmermans, 2015; Pighin et al., 2011; Sirota et al., 2014b). It is also consistent with the findings that responses provided on numerical scales are less prone to biases than on verbal ones (Windschitl & Weber, 1999). Clearly, participants’ answers are more constrained by numerical scales, perhaps because they offer the possibility to provide *the* correct answer. Another noteworthy finding was the overall “1-in-X” effect on related decision making. Despite the fact that the “1-in-X” effect on probability was small, it was still strong enough to affect the involved decisions.

Critically, when using numerical scales, our participants systematically overestimated numerical probabilities/frequencies expressed in a “1-in-X” format (compared with their objective value). Therefore, we can conclude that the “1-in-X” effect is a form of bias, leading to higher overestimation of numerical quantities relative to the “N-in-X*N” ratios. Our findings provide direct evidence for such a bias, the existence of which was only indirectly supported by prior research showing that the “1-in-X” format resulted in a higher subjective probability when compared with other formats such as “N-in-X*N”, percentages or visual representations (Oudhoff & Timmermans, 2015; Pighin et al., 2011; Sirota et al., 2014b). Evidence that the “1-in-X” effect is a form of bias, along with the finding that the “1-in-X” ratios also affected decision-making, provides more support for the recommendation to eliminate the “1-in-X” format when communicating health risks to patients (Zikmund-Fisher, 2011, 2014). The “1-in-X” bias, however, is not dependent on levels of health literacy or numeracy. This means that people with lower numeracy/health literacy skills are not more sensitive to this particular format variation. Such a finding also points towards a different root of the bias than lack of numeracy or difficulties interpreting health information, which are traditionally assumed to be behind probability biases. For example, individuals with lower numeracy skills have been shown to be less sensitive than more numerate individuals to variations in denominators (Ghazal, Cokely, & Garcia-Retamero, 2014; Reyna & Brainerd, 2008).

A few limitations of the current studies need to be discussed. First, we need to consider the possibility that the “1-in-X” bias is a methodological artefact. One might argue, for instance, that different calculation requirements of the ratios might have caused the “1-in-X” bias. However, we do not think the “1-in-X” bias can be explained via this mechanism. First, it is hard to consider that a ratio of 1 in 7 would be much harder to compute than 10 in 70. Hence, if participants – despite our request to estimate the numbers rather than calculate them – failed in a process of calculation, then one would expect the opposite effect: to perform worse with the ratios containing higher numbers (10 in 70) rather than lower numbers (1 in 7). Such miscalculation, furthermore, fails to explain why the bias would occur with verbal probabilities, as we have seen in Exp. 1 and elsewhere (Pighin et al., 2011; Sirota et al., 2014b). In our data, self-reported attempts to calculate answers did not interact with the ratios, in fact, they did not change the accuracy of the estimates.

Second, we have measured decision-making using a continuous measure rather than an actual choice task. As such, the effect of the ratio on the decision-making scale can be considered as proof of concept rather than an estimation of clinical significance. Even though we know that the “1-in-X” ratios are very prevalent in clinical practice (Sirota et al., 2017), we are still lacking information on how important the effect is in clinical terms. Future research may, in particular, focus on establishing clinical significance using ecologically valid materials, and map the effect on actual dichotomous decisions. Finally, future research should also try to explain the mechanism underlying the effect and test the suggestions of other authors, e.g., gist interpretation (Zikmund-Fisher, 2014), severity overestimation by association (Sirota et al., 2017), or ease of imagination (Oudhoff & Timmermans, 2015). In this respect, identifying the boundary conditions through which the effect is occurring could shed more light on the mechanisms as well.

To conclude, we extended the “1-in-X” effect on medical probabilities to numerical estimation scales, which allowed us to establish that the “1-in-X” effect was really a form of bias and a source of objective mistakes, impeding informed decision-making, although people with lower levels of health literacy and numeracy did not show increased vulnerability to this effect. The evidence we have provided strongly suggests that risk communication theories should not assume different ratio formats to be able to communicate risk as effectively, and that risk communicators should be especially cautious when using “1-in-X” ratios.

References

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psycholological Science, 7*, 543-554. doi: 10.1177/1745691612459060

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd ed. ed.). Hillsdale, NJ: Lawrence Erlbaum.

Fagerlin, A., Ubel, P. A., Smith, D. M., & Zikmund-Fisher, B. J. (2007a). Making numbers matter: Present and future research in risk communication. *American Journal of Health Behavior, 31*, S47-S56.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007b). Measuring numeracy without a math test: development of the Subjective Numeracy Scale. *Med Decis Making, 27*, 672-680. doi: 10.1177/0272989x07304449

Gaissmaier, W., Wegwarth, O., Skopec, D., Muller, A. S., Broschinski, S., & Politi, M. C. (2012). Numbers can be worth a thousand pictures: Individual differences in understanding graphical and numerical representations of health-related information. *Health Psychology, 31*, 286-296. doi: 10.1037/a0024850

Galesic, M., & Garcia-Retamero, R. (2011). Do low-numeracy people avoid shared decision making? *Health Psychology, 30*, 336-341. doi: 10.1037/a0022723

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychology, 28*, 210.

Garcia-Retamero, R., & Galesic, M. (2009). Communicating treatment risk reduction to people with low numeracy skills: A cross-cultural comparison. *American Journal of Public Health, 99*, 2196-2202.

Ghazal, S., Cokely, E. T., & Garcia-Retamero, R. (2014). Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment and Decision Making, 9,* 15-34.

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

Kaphingst, K. A., Stafford, J. D., McGowan, L. D. A., Seo, J., Lachance, C. R., & Goodman, M. S. (2015). Effects of racial and ethnic group and health literacy on responses to genomic risk information in a medically underserved population. *Health Psychology, 34*, 101-110. doi: 10.1037/hea0000177

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

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21*, 37-44.

McNaughton, C. D., Cavanaugh, K. L., Kripalani, S., Rothman, R. L., & Wallston, K. A. (2015). Validation of a short, 3-item version of the Subjective Numeracy Scale. *Medical Decision Making*. doi: 10.1177/0272989x15581800

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17,* 437-455. doi: doi: 10.1037/a0028085

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2012). Individual differences in graph literacy: Overcoming denominator neglect in risk comprehension. *Journal of Behavioral Decision Making, 25*, 390-401.

Oudhoff, J. P., & Timmermans, D. R. (2015). The effect of different graphical and numerical likelihood formats on perception of likelihood and choice. *Medical Decision Making, 35*, 487-500.

Peters, E. (2012). Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science*, *21*, 31-35.

Pighin, S., Savadori, L., Barilli, E., Cremonesi, L., Ferrari, M., & Bonnefon, J. F. (2011). The 1-in-X Effect on the subjective assessment of medical probabilities. *Medical Decision Making, 31*, 721-729. doi: 10.1177/0272989x11403490

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences, 18*, 89-107. doi: 10.1016/j.lindif.2007.03.011

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

Reyna, V. F. B. C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning & Individual Differences*, *18*, 89-107. doi: 10.1016/j.lindif.2007.03.011

Rodríguez, V., Andrade, A. D., García-Retamero, R., Anam, R., Rodríguez, R., Lisigurski, M., et al. (2013). Health literacy, numeracy, and graphical literacy among veterans in primary care and their effect on shared decision making and trust in physicians. *Journal of Health Communication, 18*, 273-289. doi: 10.1080/10810730.2013.829137

Sirota, M., Juanchich, M., & Hagmayer, Y. (2014a). Ecological rationality or nested sets? Individual differences in cognitive processing predict Bayesian reasoning. *Psychonomic Bulletin & Review, 21*, 198-204. doi: 10.3758/s13423-013-0464-6

Sirota, M., Juanchich, M., Kostopoulou, O., & Hanak, R. (2014b). Decisive evidence on a smaller-than-you think phenomenon: Revisiting the "1-in-X" effect on subjective medical probabilities. *Medical Decision Making, 34*, 419-429. doi: 10.1177/0272989x13514776

Sirota, M., Juanchich, M., Petrova, D., Garcia-Retamero, R., Walasek, L., & Bhatia, S. (2017). Health professionals prefer to communicate risk-related numerical information using "1-in-X" ratios. *Medical Decision Making, Advanced Online Publication*.

Sirota, M., Kostovičová, L., & Vallée-Tourangeau, F. (2015). Now you Bayes, now you don’t: effects of set-problem and frequency-format mental representations on statistical reasoning. *Psychonomic Bulletin & Review, 22*, 1465-1473. doi: 10.3758/s13423-015-0810-y

Ueno, T., Fastrich, G. M., & Murayama, K. (2016). Meta-analysis to integrate effect sizes within an article: Possible misuse and Type I error inflation. *Journal of Experimental Psychology - General, 145*, 643-654. doi: 10.1037/xge0000159

Weiss, B. D., Mays, M. Z., Martz, W., Castro, K. M., DeWalt, D. A., Pignone, M. P., et al. (2005). Quick assessment of literacy in primary care: The Newest Vital Sign. *The Annals of Family Medicine, 3*, 514-522. doi: 10.1370/afm.405

Windschitl, P. D., & Weber, E. U. (1999). The interpretation of "likely" depends on the context, but " 70% " is 70% - Right? The influence of associative processes on perceived certainty. *Journal of Experimental Psychology-Learning Memory and Cognition, 25*, 1514-1533.

Zikmund-Fisher, B. J. (2011). Time to retire the 1-in-X risk format. *Medical Decision Making, 31*, 703-704. doi: 10.1177/0272989x11418238

Zikmund-Fisher, B. J. (2014). Continued use of 1-in-X risk communications is a systemic problem. *Medical Decision Making*, *34*, 412-413. doi: 10.1177/0272989x13516198

Table 1

Scenarios used in Exp. 1-5 with the focal disease, conveyed risk and ratio format.

Scenario	Disease	Risk	Ratio “1-in-X” vs. “N-in-X*N”
Kenya	Malaria	7.69%	1 in 13 vs. 7 in 91
Sierra Leone	Ebola	1.20%	1 in 83 vs. 2 in 166
Norway	Flu	14.29%	1 in 7 vs. 10 in 70
Slovakia	Lyme disease	33.33%	1 in 3 vs. 13 in 39

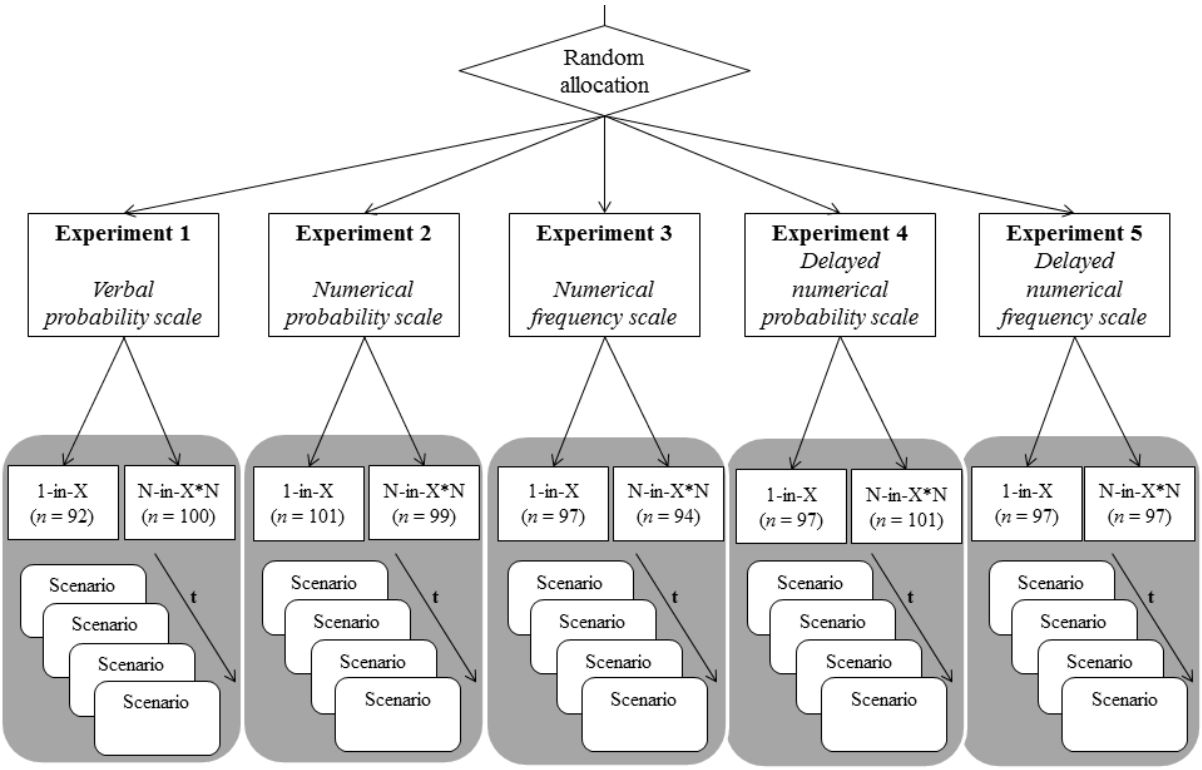

Figure 1. Design of five parallel experiments and designs within each experiment.

Note. Scenario 1 (Kenya), Scenario 2 (Sierra Leone), Scenario 3 (Norway), Scenario 4 (Slovakia). The presentation order of the scenarios was randomised.

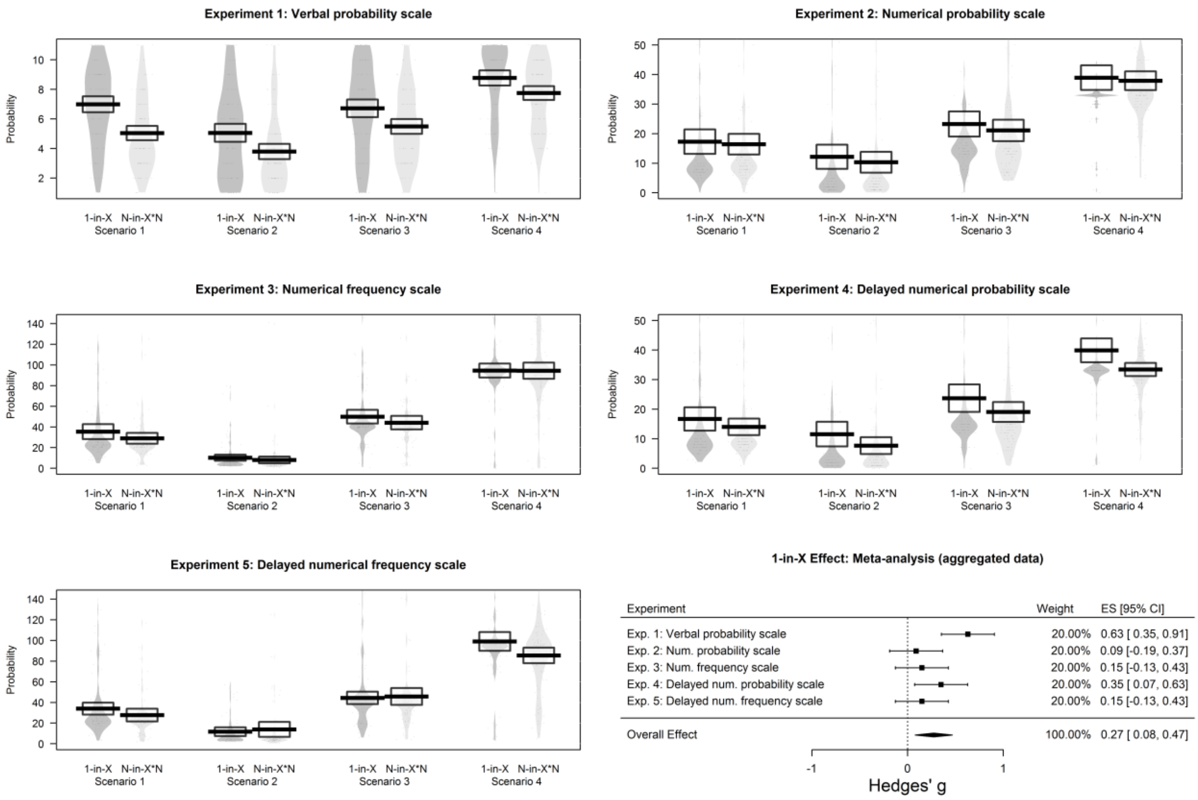

Figure 2. Effect of the ratios' formats ("1-in-X" vs. "N-in-X*N") on the perceived probability: Verbal probability scale (panel A), numerical probability scale (panel B), numerical frequency scale (panel C), delayed numerical probability scale (panel D), delayed numerical frequency scale (panel E) and the internal meta-analysis of the effect on aggregated data across all scales (panel F).

Note. The middle bars represent the mean and the boxes represent 95% CIs. Scenario 1 (Kenya), Scenario 2 (Sierra Leone), Scenario 3 (Norway) and Scenario 4 (Slovakia).

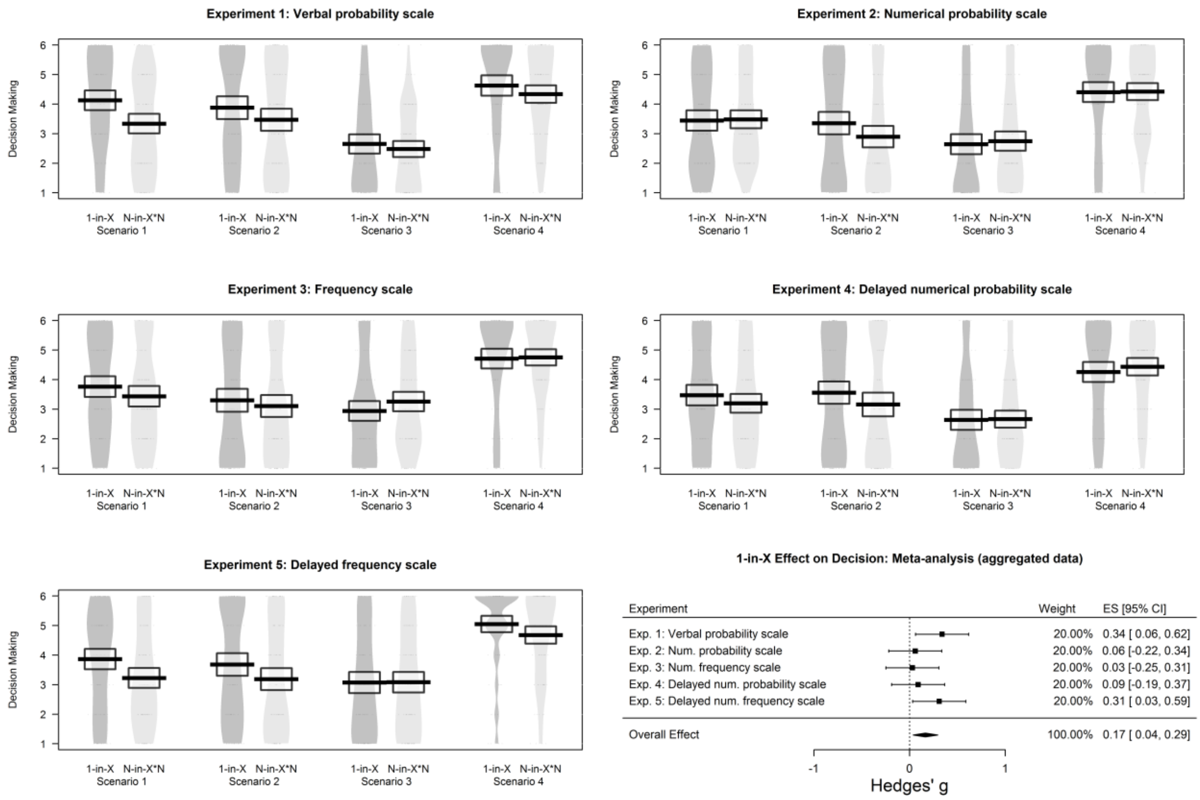

Figure 3. Effect of the ratios' formats ("1-in-X" vs. "N-in-X*N") on related decision-making: Verbal probability scale (panel A), numerical probability scale (panel B), numerical frequency scale (panel C), delayed numerical probability scale (panel D), delayed numerical frequency scale (panel E) and the internal meta-analysis of the effect on aggregated data across all scales (panel F).

Note. The middle bars represent the mean and the boxes represent 95% CIs. Scenario 1 (Kenya), Scenario 2 (Sierra Leone), Scenario 3 (Norway) and Scenario 4 (Slovakia).

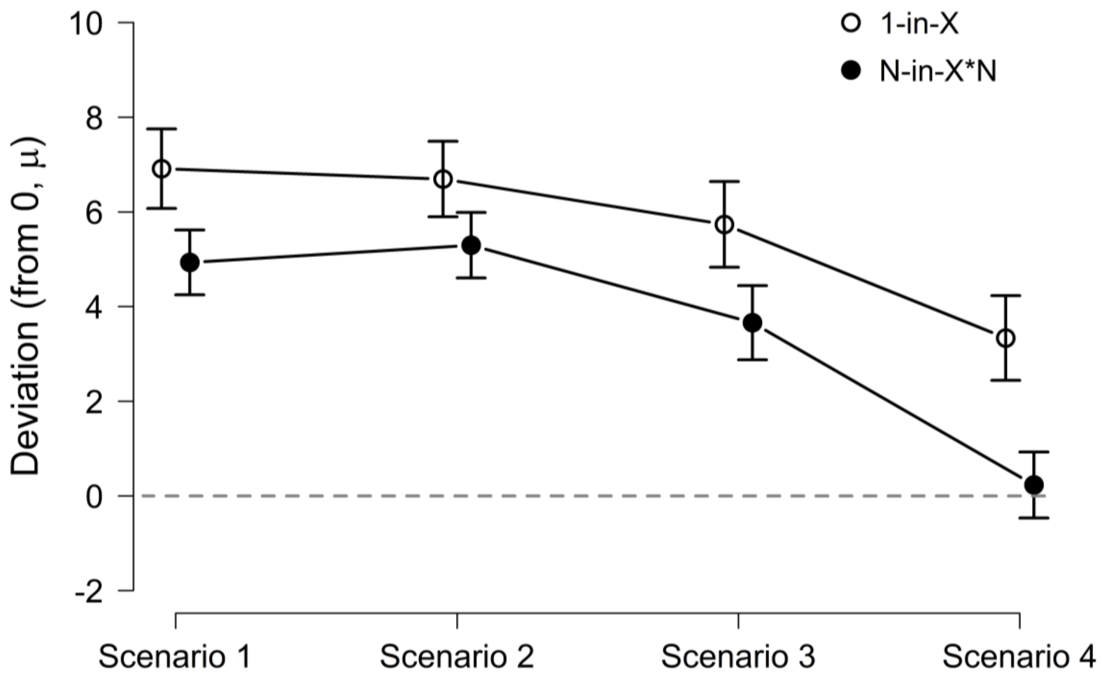

Figure 4. Effect of the ratios' formats ("1-in-X" vs. "N-in-X*N") on the accuracy (deviation on a probability scale; 0 represents perfect accuracy) across four scenarios (aggregated across Exp. 2-5).

Note. The middle points represent the mean and the bars represent ±1 SE. Scenario 1 (Kenya), Scenario 2 (Sierra Leone), Scenario 3 (Norway) and Scenario 4 (Slovakia).