Decisive Evidence on a Smaller-Than-You-Think Phenomenon: Revisiting the "1-in-X" Effect on Subjective Medical Probabilities

Miroslav Sirota, PhD, Marie Juanchich, PhD, Olga Kostopoulou, PhD, Robert Hanak, PhD

*Accurate perception of medical probabilities communicated to patients is a cornerstone of informed decision making. People, however, are prone to biases in probability perception. Recently, Pighin and others extended the list of such biases with evidence that "1-in-X" ratios (e.g., "1 in 12") led to greater perceived probability and worry about health outcomes than "N-in-X*N" ratios (e.g., "10 in 120"). Subsequently, the recommendation was to avoid using "1-in-X" ratios when communicating probabilistic information to patients. To warrant such a recommendation, we conducted 5 well-powered replications and synthesized the available data. We found that 3 out of the 5 replications yielded statistically nonsignificant findings. In addition, our results showed that the "1-in-X" effect was not moderated by numeracy, cognitive reflection, age, or gender. To quantify the evidence for the effect, we conducted a Bayes factor meta-analysis and a traditional meta-analysis of our 5 studies and those of Pighin and others (11 comparisons, N = 1131). The meta-analytical Bayes factor, which allowed assessment of the evidence for the null hypothesis, was very low, providing decisive evidence to support the existence of the "1-in-X" effect. The traditional meta-analysis showed that the overall effect was significant (Hedges'* g *= 0.42, 95% CI 0.29–0.54). Overall, we provide decisive evidence for the existence of the "1-in-X" effect but suggest that it is smaller than previously estimated. Theoretical and practical implications are discussed.* ***Key words:*** *"1-in-X" effect; subjective probability; probability perception; meta-analysis; Bayes factor meta-analysis.* ***(Med Decis Making 2014;34:419–429)***

Given a steadily increasing number of screening tests and treatment options, patients have to make complex decisions that involve probability estimation. For example, pregnant women may have to decide whether to undergo amniocentesis, which will indicate their probability of having a Down syndrome baby but will also increase their risk of miscarriage. As the majority of these decisions rely on probabilistic information, accurate perception and processing of probabilities of health-related outcomes are crucial for informed decision making. One source of bias in the perception of probability is related to the format that is used to communicate the information. It has frequently been shown that mathematically or logically equivalent information, presented in different formats, may be perceived differently.[1] A new instance of the format-related biases was recently demonstrated in the domain of medical risk communication. Pighin and others[2] found that when probability information was presented as "1-in-X" ratios, it led to higher perceived

Received 12 February 2013 from School of Medicine, King's College London, UK (MS, OK); Kingston Business School, Kingston University London, UK (MJ); and Faculty of Business Management, University of Economics in Bratislava, Bratislava, Slovakia (RH). Financial support for this study was provided in part by a grant awarded to Marie Juanchich by Kingston Business School. The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report. Revision accepted for publication November 2013.

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

Address correspondence to Miroslav Sirota, Medical Decision Making and Informatics Research Group, Department of Primary Care & Public Health Sciences, School of Medicine, King's College London, Capital House, 7th floor, 42 Weston Street, SE1 3QD, London, UK; telephone: +44 (0)20 7848 6652; fax: +44 (0)20 7848 6620; e-mail: miroslav.sirota@kcl.ac.uk.

Reprints and permission:
http://www.sagepub.com/journalsPermissions.nav
DOI: 10.1177/0272989X13514776

probabilities than when it was presented as ''N-in-X*N'' ratios. The authors asked participants to assess the probability of a 45-year-old woman having a child affected by Down syndrome and gave them the probability in 1 of 2 formats:

Format A: ''1 in 12''
Format B: ''10 in 120''

Although the 2 formats present the same statistical information, format A, the ''1-in-X'' format, triggered higher probability estimates than format B, the ''N-in-X*N'' format.[2] The effect was found to be robust, as it was demonstrated across different samples and in scenarios involving different medical conditions and different ratios. Based on this evidence, some researchers recommended avoiding the ''1-in-X'' format when communicating statistical information to patients since such a format may create confusion and biased risk perception.[2,3]

Despite its apparent robustness, the ''1-in-X'' effect does not appear entirely consistent with 2 other ratio-based effects reported in the literature, namely ''group diffusion''[4] and ''denominator neglect.''[5] According to the group-diffusion effect, the greater the reference class, the smaller the perceived probability. For example, ''10% of 1000 individuals'' led to smaller probability estimates than ''10% of 10 individuals.''[4,6] As with the ''1-in-X'' effect, the group-diffusion effect would predict that probability in format A (''1 in 12'') would be perceived as higher than in format B (''10 in 120''), but because the denominator is larger and not because of the value 1 in the numerator. Consequently, according to the group-diffusion effect, any format with a smaller denominator than ''10 in 120'' (e.g., ''5 in 60'') and not only ''1-in-X'' format (e.g., ''1 in 12'') would result in a higher perceived probability.

The ''1-in-X'' effect is inconsistent with the denominator-neglect effect, which postulates that ''1-in-X'' ratios lead to lower probability estimates than ''N-in-X*N'' ratios because comparisons are based primarily on the numerators. For example, participants perceived that they were less likely to get a chocolate chip cookie from a jar that contained 1 chocolate chip cookie and 9 plain cookies than from a jar containing 10 chocolate chip cookies and 90 plain cookies.[5] The denominator-neglect effect is a robust phenomenon that applies to many probability magnitudes, to various denominators, and to positive as well as negative outcomes.[7–9] In our earlier medical example, denominator neglect predicts that probability conveyed by format A (''1 in 12'') would be perceived as lower than the same probability conveyed by format B (''10 in 120''), because the numerator is lower. This prediction is exactly the opposite of the ''1-in-X'' effect (as well as the group-diffusion effect).

We consider the denominator-neglect and group-diffusion effects to be inconsistent with the ''1-in-X'' effect, although different research paradigms studying the ratio effects may moderate the outlined inconsistencies. Indeed, the research paradigms differ to some extent, for example, in terms of a possibility to compare the ratios (i.e., a joint v. separate evaluation paradigm). A pure separate presentation of quantities enabling only absolute judgments yielded the ''1-in-X'' and the group-diffusion effects,[2,4] whereas some involvement of a joint presentation of quantities enabling comparative judgments yielded the denominator-neglect effect. Prior research has demonstrated that the extent of involvement of a joint presentation causes different preferences and preference reversals as people can use more contextual knowledge.[10] Yet current limited evidence indicates that the group-diffusion effect occurs in both modes[4,6] and that the denominator-neglect effect occurs to some extent in a separate evaluation mode,[5,7] particularly if the numerator is salient.[6]

We propose that given the current available evidence, further investigation of the ''1-in-X'' effect is needed before rejecting the ''1-in-X'' format from health-related communication.[3] From a more general perspective, this research may be seen as a response to recent calls for replications in psychology.[11,12] In the 5 experiments presented in this paper, we accumulated more evidence about the ''1-in-X'' effect. Experiment 1 is a replication testing the robustness of the effect in the context of the alternative group-diffusion and denominator-neglect effects studied in a pure separate evaluation mode (i.e., separate evaluation of the ratios in a between-subjects design). Experiment 2 is a replication, focusing on the role of individual differences, namely, numeracy and cognitive reflection. Experiments 3, 4, and 5 are replications testing the occurrence of the ''1-in-X'' effect in different adult populations (Slovak and British), using different scenarios and measurement scales. Finally, we synthesized the available data on the ''1-in-X'' effect using a Bayes factor meta-analysis and a traditional meta-analysis: Both methods reliably asses robustness and overall size of effects.[13,14]

EXPERIMENT 1

This experiment aimed to replicate the ''1-in-X'' effect and to establish its existence using somewhat stricter criteria than those used by Pighin and others,[2] who used only 1 ratio contrast per experiment (e.g., ‘‘1 in 12’’ versus ‘‘10 in 120’’). Specifically, to establish the ‘‘1-in-X’’ effect 2 conditions need to be met. First, the ‘‘1-in-X’’ ratio (‘‘1 in 12’’) should elicit higher probability estimates than 2 different ‘‘N-in-N*X’’ ratios (‘‘5 in 60’’ and ‘‘10 in 120,’’ Hypothesis 1). Second, the two ‘‘N-in-N*X’’ ratios should not yield different probability estimates, otherwise the ‘‘1-in-X’’ effect could simply be categorized as a case of group diffusion (‘‘5 in 60’’ > ‘‘10 in 120,’’ Hypothesis 2); the reversed pattern would indicate the denominator neglect (‘‘5 in 60’’ < ‘‘10 in 120,’’ Hypothesis 3). We were also interested in the extent to which ratio formats depart from other frequently used formats in health communication, and therefore we included a percentage condition in our design (‘‘8.33%’’; Question 1). We used the same dependent measures as Pighin and others (i.e., subjective probability, worry, severity) and added a measure of subjective frequency. The rationale for adding a frequency measure draws on research showing that frequency judgments are more accurate than judgments of probability.[15] Therefore, we assumed that inducing a frequency interpretation of the focal ratio, instead of a probabilistic one, might provide a simple remedy to the ‘‘1-in-X’’ effect.

Method

Given the overall effect size found by Pighin and others,[2] $g = 0.59$ (95% CI 0.39 to 0.78) and given $\alpha = 0.05$ and power $1 - \beta = 0.80$, the sample size was determined to be at least 46 per group in all of the experiments reported here.[16] All datasets presented in this study were analyzed using a 2-tailed test of significance $\alpha = 0.05$, and all confidence intervals reported are 95%. Gender and age did not interact with the format in any of the experiments presented here. Therefore, they are not included in the analysis.

In this experiment, 226 undergraduate Slovak students of management (148 females, age range 18–45 years, $\bar{x} = 21.29$, $s = 1.84$) participated without incentive. In a randomized between-subjects design, participants read a medical scenario featuring a probability in 1 of the following 4 formats: ‘‘1 in 12,’’ ‘‘5 in 60,’’ ‘‘10 in 120,’’ or ‘‘8.33%.’’

Participants completed a short web questionnaire in a classroom. After giving their informed consent, participants read the short Down syndrome scenario used in Experiment 4 by Pighin and others (see online appendix).[2] Participants first assessed the subjective probability (‘‘In your opinion, the probability that a 45-year-old woman will have a child affected by Down syndrome is. . . .’’). Then, they reported the degree of severity of Down syndrome, the degree of worry they would feel if they were a 45-year-old pregnant woman, and the frequency of occurrence of Down syndrome among fetuses carried by 45-year-old women. Probability, severity, worry, and frequency were judged on 7-point Likert scales (1: *extremely low, not at all severe, not at all worried, not at all frequent* and 7: *extremely high, extremely severe, extremely worried, extremely frequent*). Finally, participants provided sociodemographic information.

Results and Discussion

Table 1 presents the average subjective probability, severity, worry, and frequency of Down syndrome given the 4 formats. Table 1 shows that as predicted by the ‘‘1-in-X’’ effect, the highest probability, worry and frequency perceptions were found in the ‘‘1 in 12’’ format compared with the ‘‘10 in 120,’’ ‘‘5 in 60,’’ and ‘‘8.33%’’ formats. As expected, the formats did not yield different severity estimates.[2]

We conducted an analysis of orthogonal planned contrasts to test our 3 hypotheses and 1 research question with a maximal power. First, the 3 ratios were contrasted with the percentage condition, then the ‘‘1 in 12’’ ratio was contrasted with the average of the ‘‘5 in 60’’ and ‘‘10 in 120’’ ratios, and finally ‘‘5 in 60’’ was contrasted with the ‘‘10 in 120’’ ratio (Table 2). As shown in Table 2, the ‘‘1 in 12’’ ratio had a higher subjective probability and frequency than the ‘‘5 in 60’’ and ‘‘10 in 120’’ ratios, in line with the ‘‘1-in-X’’ effect, predicted by Hypothesis 1. Yet, the ‘‘1 in 12’’ ratio did not lead to greater severity or greater worry. Furthermore, the ‘‘5 in 60’’ ratio yielded approximately the same probability, severity, worry, and frequency as the ‘‘10 in 120’’ ratio, which does not support Hypothesis 2 (the existence of the group-diffusion effect) nor the opposite direction of such effect (i.e., the denominator-neglect effect, Hypothesis 3). Finally, the percentage condition yielded a lower probability estimate than the ratio formats (Question 1).

To conclude, findings showed a statistically significant ‘‘1-in-X’’ effect, but this effect was considerably smaller than that found by Pighin and others ($g = 0.39$ v. $g = 0.59$, respectively).[2] Moreover, findings did not lend support to either the prediction that people infer subjective probability from the size of the denominator while neglecting the numerator (group-diffusion effect) or the prediction that they focus on the numerator while neglecting the denominator of a ratio

Table 1 Average Subjective Probability, Perceived Severity of the Medical Condition, Worry, and Frequency of Occurrence of Having a Child Affected by Down Syndrome as a Function of Numerical Formats

Numerical Formats	n	Probability $\bar{x}$ (s)	Severity $\bar{x}$ (s)	Worry $\bar{x}$ (s)	Frequency $\bar{x}$ (s)
1 in 12	55	3.96 (1.66)	5.85 (0.80)	5.60 (1.41)	4.20 (1.51)
5 in 60	57	3.46 (1.45)	5.93 (1.12)	5.18 (1.44)	3.53 (1.36)
10 in 120	56	3.32 (1.38)	5.75 (1.12)	5.41 (1.50)	3.73 (1.39)
8.33%	58	3.34 (1.40)	5.82 (1.06)	5.21 (1.60)	3.40 (1.26)
Overall	226	3.52 (1.48)	5.84 (1.03)	5.35 (1.49)	3.71 (1.41)

Table 2 Differences in Subjective Probability, Perceived Severity of the Medical Condition, Worry, and Frequency of Occurrence of Having a Child Affected by Down Syndrome as a Function of Numerical Formats

	Mdiff	g	Confidence Interval of g	t	P
Hypothesis 1: "1 in 12" > ("5 in 60," "10 in 120")					
Probability	1.15	0.39	0.06 to 0.71	2.38	0.02
Severity	0.03	0.01	−0.31 to 0.33	0.09	0.93
Worry	0.61	0.20	−0.12 to 0.52	1.25	0.21
Frequency	1.14	0.41	0.08 to 0.73	2.51	0.01
Hypothesis 2: "5 in 60" > "10 in 120"					
Hypothesis 3: "5 in 60" < "10 in 120"					
Probability	0.13	0.09	−0.28 to 0.46	0.49	0.63
Severity	0.18	0.17	−0.20 to 0.54	0.92	0.36
Worry	−0.24	−0.16	−0.53 to 0.21	−0.84	0.40
Frequency	−0.21	−0.15	−0.52 to 0.22	−0.79	0.43
Question 1: ("1 in 12," "5 in 60," "10 in 120") > "8.33%"					
Probability	0.71	0.16	−0.14 to 0.46	1.05	0.29
Severity	0.05	0.01	−0.31 to 0.33	0.11	0.92
Worry	0.57	0.13	−0.17 to 0.43	0.83	0.41
Frequency	1.27	0.31	0.01 to 0.61	2.01	0.05

Note: *Mdiff* = point estimate of mean difference for a given contrast; g = standardized effect size Hedges' g converted from the t tests; CI = 95% confidence intervals for g; t = 2 independent-samples t test; P = 2-way statistical significance.
Contrasts are defined orthogonally as follows: Contrast 1 compares ("1 in 12," "5 in 60," "10 in 120") with "8.33%" (tests Question 1); Contrast 2 compares "1 in 12" with ("5 in 60," "10 in 120") (tests Hypothesis 1); Contrast 3 compares "5 in 60" with "10 in 120" (tests Hypotheses 2 and 3).

(denominator-neglect effect). The ratio format affected frequency perception consistently with the "1-in-X" effect but not with the other 2 effects. These results are in line with those of Pighin and others on subjective probability judgments and may weaken suggestions that frequency judgments are more accurate than probabilistic ones. Frequency judgments thus cannot provide a simple remedy to the "1-in-X" effect.

One possible explanation for the difference in magnitude of the effect between our findings and those of Pighin and others is that Pighin and others[2] recruited samples with different characteristics (female patients in the maternity ward of an Italian hospital and employees of local offices and companies). Although a general adult population differs from a student population in many respects, the level of numeracy and cognitive reflection are individual differences especially relevant to the ratio biases documented in the literature.[17–21] The following experiment aimed to assess the effect of these individual differences on the magnitude of the "1-in-X" effect.

EXPERIMENT 2

This experiment was tailored to examine the potential interaction effects of numeracy and cognitive reflection with the "1-in-X" effect. There is ample evidence that the occurrence of biases based on different formats of presentation of probabilities relates to the individual's level of numeracy.[18,20,22–24] For example, individuals with low numeracy were more

Table 3 Average Subjective Probability of Having a Child Affected by Down Syndrome According to the Numerical Formats, Numeracy (Median Split), and Cognitive Reflection (Median Split)

			Numeracy		Cognitive Reflection	
Numerical Formats	N	Overall	Low (n = 114)	High (n = 132)	Low (n = 128)	High (n = 118)
1 in 10	128	5.41 (2.72)	5.73 (2.70)	5.08 (2.72)	5.28 (2.83)	5.62 (2.55)
10 in 100	118	4.92 (2.50)	5.31 (2.62)	4.66 (2.39)	4.86 (2.67)	4.97 (2.39)
Total	246	5.18 (2.62)	5.55 (2.66)	4.86 (2.55)	5.12 (2.76)	5.24 (2.47)

Note: Values are given as $\bar{x}$ (s).

prone to the denominator-neglect effect than were individuals with high numeracy. Moreover, people with lower cognitive reflection ability compared with those with higher ability were more prone to contextual effects (e.g., framing effect) or other biases in decision making.[21,25–27] As a result, we hypothesized that the "1-in-X" effect will occur for individuals with a low level of numeracy and cognitive reflection, whereas the effect will be smaller for individuals with higher levels of numeracy and cognitive reflection.

Method

A total of 246 Slovak undergraduate students of management (166 females, age 18–33 years, $\bar{x}$ = 21.34, s = 1.31; 2 participants failed to report their sociodemographics), different from those in Experiment 1, participated in the study without incentive.

Participants assessed the subjective probability of a medical condition in a 2 (ratio format: "1 in 10" v. "10 in 100") × 2 numeracy (median split numeracy: low v. high) × 2 cognitive reflection (median split cognitive reflection: low v. high) between-subjects design.

Participants completed a short online questionnaire in a classroom. After reading the informed consent, they read a long version of the Down syndrome scenario as used in Experiment 2b by Pighin and others[2] (see online appendix at http://mdm.sagepub.com/supplemental). Subsequently, participants rated the probability of having a child affected by Down syndrome (the same wording as in Experiment 1). Participants provided judgments on an 11-point Likert scale (1: *extremely low* and 11: *extremely high*). Participants then responded to the 11-item Lipkus Numeracy Scale[21,28] and the 3-item Cognitive Reflection Test.[21,25] The order of these 14 items was randomized. The internal consistency for both measures was satisfactory (respectively, $\alpha = 0.74$; $\alpha = 0.60$). Numeracy scores ranged from 1 to 11 ($\bar{x} = 8.2$, $s = 2.4$, $Mdn = 9$) and cognitive reflection scores ranged from 0 to 3 ($\bar{x} = 0.8$, $s = 1.0$, $Mdn = 0$), which are similar to values in the studies where they had a significant effect on probability perception.[17,21] Both scores were skewed (respectively, $g = -1.2$; $g = 1.1$), and therefore we entered the variables into analysis as median splits. Finally, participants provided sociodemographic information.

Results and Discussion

Mean and standard deviations of subjective probability as a function of the ratio formats, levels of numeracy, and cognitive reflection are summarized in Table 3. The table shows that participants perceived a higher probability when reading the "1 in 10" ratio compared with the "10 in 100" ratio. Table 3 also indicates that individuals with low numeracy perceived a greater probability than individuals with high numeracy across the format conditions. Further, individuals with low cognitive reflection perceived approximately the same probability as those with higher cognitive reflection.

A 3-way analysis of variance was conducted with ratio formats, level of numeracy, and level of cognitive reflection as independent variables and subjective probability as a dependent variable. Format did not have a significant effect on perceived probability, $F(1, 238) = 1.27$, $P = 0.26$, $\eta^2_{par} = 0.01$, nor did cognitive reflection, $F(1, 238) = 1.47$, $P = 0.23$, $\eta^2_{par} = 0.01$. Numeracy had a statistically significant main effect on perceived probability, $F(1, 238) = 4.00$, $P = 0.047$, $\eta^2_{par} = 0.02$. More important, the analyses revealed no significant interaction effects of numeracy, cognitive reflection, and ratio formats, all Fs < 1.

The results of the present experiment show that although participants perceived a higher probability when reading a "1 in 10" ratio than when reading a "10 in 100" ratio, this effect was small and not statistically significant ($g = 0.19$, CI -0.06 to 0.44). The "1-in-X" effect did not occur more among individuals with low numeracy or low cognitive reflection than among individuals with high numeracy or high cognitive reflection, and thus it is not attributable to an insufficient numerical ability or to intuitive information processing.[29]

The lack of any interaction effects minimizes the possibility of a small ''1-in-X'' effect owing to participants' characteristics but does not eliminate it completely. This finding does not discard the role of other factors such as material characteristics (i.e., using Down syndrome scenario) or specific ratios (e.g., using ''1 in 10'' v. ''1 in 100'' comparison). In fact, even if students were somewhat immune to the effect when assessing a Down syndrome probability, the inconclusiveness of our findings does not rule out the existence of the ''1-in-X'' effect in a general population and for other medical scenarios. Thus, in the following experiments, we test the occurrence of this effect on a general adult population in Slovakia (Experiments 3 and 4) and the United Kingdom (Experiment 5) using different scenarios and ratios.

EXPERIMENT 3

Method

A total of 94 people (55 females, age range 18–66 years, $\bar{x}$ = 32.16, s = 11.34; 1 participant failed to report the sociodemographics) were recruited by convenience sampling from the Slovak general adult population. Workers of 3 local medium-sized offices received an e-mail invitation sent by a contact person to join an electronically administered questionnaire with a chance to win a voucher worth €50. In a between-subjects design, participants read the short version of the Down syndrome scenario as used in Experiment 1 and were randomly presented with 1 of the ratios ''1 in 12'' (n = 46) or ''10 in 120'' (n = 48). Participants assessed the subjective probability of having a child affected by Down syndrome as in Experiment 1 using a 7-point Likert scale (1: *extremely low*, 7: *extremely high*). Finally, participants provided sociodemographic information.

Results and Discussion

The mean subjective probability perception was higher in the ''1 in 12'' compared with the ''10 in 120'' condition (respectively, $\bar{x}$ = 4.17, s = 1.68; $\bar{x}$ = 3.31, s = 1.69). The effect was medium-sized (g = 0.51, CI 0.10 to 0.92) and statistically significant, $t(92)$ = 2.48, P = 0.02. Thus, Experiment 3 supports the existence of the ''1-in-X'' effect in a general adult population. To further test the robustness of the ''1-in-X'' effect, we intended to replicate the effect in a general adult population with another scenario.

EXPERIMENT 4

Method

A total of 96 individuals (55 males, age range 19–62 years, $\bar{x}$ = 32.74, s = 11.27; 3 participants did not report their sociodemographics), different from those in Experiment 3, were recruited by convenience sampling (the same recruitment method as in Experiment 3) from the Slovak general adult population. In a between-subjects design, participants read a short version of the malaria scenario as used in Experiment 1 by Pighin and others[2] (see online appendix at http://mdm.sagepub.com/supplemental) and were randomly presented with either ''1 in 200'' (n = 48) or ''5 in 1000'' (n = 48). Participants then assessed the subjective probability of being affected by malaria while traveling to Kenya (''In your opinion, the probability of being affected by malaria while travelling to Kenya is. . . .'') on a 7-point Likert scale (1: *extremely low*, 7: *extremely high*). Finally, participants provided sociodemographic information.

Result and Discussion

The mean subjective probability perception was again higher in the ''1 in 200'' condition compared with the ''5 in 1000'' condition (respectively, $\bar{x}$ = 3.40, s = 1.73; $\bar{x}$ = 2.79, s = 1.74). The effect size of this difference was small (g = 0.34, CI −0.06 to 0.74) and not statistically significant, $t(94)$ = 1.71, P = 0.09.

It is also possible that the underpinning mechanism of the smaller ''1-in-X'' effect is culturally or linguistically specific as some authors suggested for the mechanisms of some other ratio biases.[6,30] Indeed, Pighin and others[2] conducted their studies on Italian samples, whereas the studies reported here were conducted on Slovak participants. To test whether the effect is more pronounced in another culture or using a different language, the next experiment tested the ''1-in-X'' effect on a sample from the British general adult population.

EXPERIMENT 5

Method

A total of 105 participants (57 males, age range 19–75 years, $\bar{x}$ = 48.73, s = 11.95) from the British adult

Table 4 Average Subjective Probability, Worry of Having a Child Affected by Down Syndrome, and Intention to Advise Amniocentesis as a Function of Numerical Formats

Numerical Formats	*n*	Probability	Worry	Advice
1 in 12	52	5.69 (2.97)	7.25 (2.27)	6.69 (2.70)
10 in 120	53	4.87 (2.92)	6.98 (2.54)	7.11 (2.95)
Total	105	5.28 (2.96)	7.11 (2.40)	6.90 (2.82)
Hedges' *g* (95% confidence interval of *g*)		0.27 (−0.11 to 0.65)	0.10 (−0.28 to 0.48)	−0.15 (−0.53 to 0.23)
t test (*P* value)		1.44 (0.15)	0.51 (0.57)	−0.76 (0.45)

population took part in this experiment. The sample varied not only in age but also in education (no formal education 2.9%; GCSEs or equivalent 21.9%; A levels or equivalent 19.0%; undergraduate 33.3%; postgraduate education 22.9%). In a between-subjects design, participants were presented either with a ''1 in 12'' or a''10 in 120'' ratio format.

Participants were contacted via e-mail by a marketing company and invited to take part in a web questionnaire investigating probability perception in exchange for a small-value voucher. After giving their informed consent, participants read the long version of the Down syndrome scenario, as used in Experiment 2. Participants assessed the probability of having a child affected by Down syndrome for Anna with a slightly changed wording (''In your opinion, what is the probability that Anna will have a child affected by Down syndrome?'') and the worry about the occurrence of such an event. Both judgments were given on an 11-point scale (1: *extremely low* or *not at all worried* and 11: *extremely high* or *extremely worried*). Because we were also interested in behavior change intentions induced by different formats, participants were asked whether they would recommend to Anna to undertake an invasive medical investigation (amniocentesis), which could reduce her uncertainty but could cause miscarriage (''Anna is offered an amniocentesis to test whether the fetus is affected by Down syndrome. The test is very reliable but quite risky as it can cause miscarriage.''). Participants provided their recommendation on an 11-point scale (1: *not at all* and 11: *completely*). Finally, participants provided sociodemographic information.

Results and Discussion

Table 4 exhibits the mean subjective probability, worry, and intention to advise amniocentesis as a function of the format: Participants perceived a slightly greater probability but roughly the same worry and willingness to advise amniocentesis when the probability was presented as a ''1 in 12'' ratio than when it was a ''10 in 120'' ratio. Table 4 also summarizes the effect sizes of these differences and their statistical significance. Results replicated the direction of the ''1-in-X'' effect, but the effect was small ($g = 0.27$) and not significant. All other differences were close to zero. The null effect on amniocentesis uptake could be explained by the fact that factors other than the perceived probability affect the uptake, such as socioeconomic status.[31]

These results support the occurrence of a small, nonsignificant ''1-in-X'' effect in a population very different from those of the prior studies, which rules out the cultural or linguistic sensitivity explanation.

REANALYSIS AND SYNTHESIS OF EVIDENCE

All 5 experiments reported here support the existence of the ''1-in-X'' effect, but 3 of them failed to reject the null hypothesis, which—logically—should not lead to dismissing the effect. Therefore, we assessed support for the null and alternative hypotheses by reanalyzing the 5 pairs of comparisons (i.e., ''1-in-X'' v. ''N-in-X*N'') reported by Pighin and others and the 6 pairs of comparisons reported here with a default Bayesian *t* test.[32]

The rationale of this reanalysis relies on the fact that this method, in contrast with traditional *P* values, allows evidence to be quantified in favor of the null or alternative hypotheses. The use of Bayes factor analysis has been strongly recommended both in medicine[33] and psychology.[32,34] The Bayesian *t* test computes a Bayes factor (BF), which is the ratio of the probability of the data given H_0 (no difference between the ratios) to the probability of the data given H_1 (''1-in-X'' generates higher probability than ''N-in-X*N''). Thus, the Bayes factor expresses the ratio of marginal likelihood of the data under the models H_0

Table 5 Bayes Factor Reanalysis of the Experiments Testing the ''1-in-X'' Effect

Experiment	df	t	P	g	BF_{01}	Evidence Category
Pighin and others, 1	61	2.5	0.01	0.62	0.338	Anecdotal for H_1
Pighin and others, 2a	78	4.2	<0.001	0.93	0.003	Very strong for H_1
Pighin and others, 2b	98	2.4	0.02	0.48	0.461	Anecdotal for H_1
Pighin and others, 3	81	2.3	0.02	0.50	0.538	Anecdotal for H_1
Pighin and others, 4c	94	2.4	0.02	0.49	0.456	Anecdotal for H_1
Sirota and others, 1i	110	1.7	0.09	0.32	1.771	Anecdotal for H_0
Sirota and others, 1ii	109	2.2	0.03	0.41	0.721	Anecdotal for H_1
Sirota and others, 2	244	1.5	0.14	0.19	3.351	Substantial for H_0
Sirota and others, 3	92	2.5	0.02	0.51	0.365	Anecdotal for H_1
Sirota and others, 4	94	1.7	0.09	0.34	1.669	Anecdotal for H_0
Sirota and others, 5	103	1.4	0.16	0.27	2.656	Anecdotal for H_0

Note: df = degrees of freedom; t = 2 independent-samples t-test; P = 2-way statistical significance; g = Hedges' g converted from t tests; BF_{01} = Bayes' factor. Evidence category for BF_{01} = evidence to support H_0: decisive evidence (>100), very strong evidence (100 – 30), strong evidence (30 – 10), substantial evidence (10 – 3), anecdotal evidence (3 – 1); evidence to support H_1: decisive evidence (<1/100), very strong evidence (1/100 – 1/30), strong evidence (1/30 – 1/10), substantial evidence (1/10 – 1/3), anecdotal evidence (1/3 – 1). t tests were computed post hoc in Experiment 1; otherwise, t tests in 2a, 2b, and4 are those reported by Pighin and others.[2]

and H_1: in other words, how many times the data are more likely to occur under H_0 compared with H_1. For example, if the BF_{01} value is 3, then the data are 3 times more likely to occur under the null hypothesis. A Bayes factor with a value lower than 1 indicates that the alternative hypothesis is more likely, whereas with a value greater than 1 indicates that the null hypothesis is more likely. Furthermore, Bayes factor values may also be interpreted as evidence categories; for example, values between 1 and 3 indicate anecdotal evidence to support the null hypothesis, whereas values greater than 100 indicate decisive evidence to support the null hypothesis (see the note of Table 5 for a complete categorization).[35,36]

To compute the default Bayesian test, we used the web-based applet designed by Rouder and others.[32] Table 5 presents traditional t test information (with associated degrees of freedom and P values), Hedges' g, Bayes factor, and the evidence category for each comparison. Results showed that only 1 comparison produced very strong evidence supporting the ''1-in-X'' effect (Experiment 2a, Pighin and others[2]) and only 1 case of substantial evidence favoring the null hypothesis (Experiment 2, reported here). The other comparisons yielded only anecdotal evidence to support either H_1 or H_0.

Clearly, neither individually reported P values nor Bayes factors allowed us to make a decisive statement about the existence of the ''1-in-X'' effect. There is, nevertheless, another option to establish the existence of an effect: data synthesis. Therefore, we conducted 2 types of synthesis for the data (11 comparisons, N = 1131): 1) a meta-analytical Bayes factor to assess the synthesized evidence to support either the null or the alternative hypothesis, and 2) a meta-analysis to estimate the magnitude of the overall effect size.[13]

First, we computed a meta-analytical Bayes factor using the R script[37] provided by Jeff Rouder.[14] The advantage of the meta-analytical Bayes factor is its cumulative function. For a single study, a small effect can be interpreted as support for the null hypothesis, but with an increasing number of replications, anecdotal evidence in favor of the null hypothesis may add up to strong evidence in favor of the alternative hypothesis.[14] Thus, systematic replication of the direction of the effect, which may have the character of anecdotal evidence supporting either the null or the alternative hypothesis, may add up to strong evidence supporting existence the ''1-in-X'' effect. Indeed, this was the case here: The 1-tailed meta-analytical Bayes factor was $BF_{01} = 5.82^{-10}$ and the 2-tailed Bayes factor was $BF_{01} = 1.16^{-09}$, which can be interpreted as decisive evidence to support the alternative hypothesis ergo the existence of the ''1-in-X'' effect.

After establishing the existence of the effect, we conducted a small-scale meta-analysis[13] of the 11 comparison pairs to estimate the actual size of the effect. The meta-analysis of Hedges' g values[13] was conducted with the metafor package[38] in the R statistical environment.[37] Individual values entered into the meta-analysis are summarized in Table 5. The population parameter of the effect size was treated in a random effect model; the selected comparisons were homogenous, $Q(\text{df} = 10) = 10.10$, $P = 0.43$. The

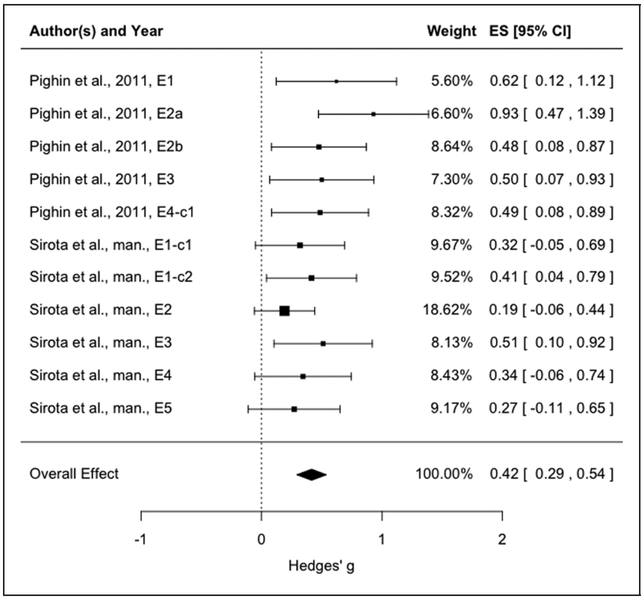

Figure 1 *Meta-analysis of the eleven ''1-in-X'' versus ''N-in-N*X'' comparisons (using Hedges'* g *derived from the studies of Pighin and others and those presented here). Data, from left to right, are the study with the number of an experiment and comparison, a study weighting, an effect size (ES) of* g*, and its 95% confidence interval (CI) in brackets.*

individual comparisons' effects and overall ''1-in-X'' effect are depicted in Figure 1. As illustrated, all of the effects are in the same (positive) direction. The overall size of effect g was 0.42 and was significantly different from zero (CI 0.29 to 0.54).

Overall, the analyses converged to produce decisive evidence of the existence of the ''1-in-X'' effect but resulted in estimation of a smaller effect size than previously offered. The effects obtained in our studies turned out to be considerably smaller (overall $g = 0.31$, CI 0.17 to 0.45) than those found by Pighin and colleagues ($g = 0.59$, CI 0.39 to 0.78), and therefore some of our replication attempts produced consistent but nonsignificant trends.

GENERAL DISCUSSION

The ''1-in-X'' effect postulates that although mathematically equivalent, ''1-in-X'' ratios trigger perceptions of higher probability than ''N-in-X*N'' ratios. We accumulated more evidence on this effect and its moderation and then synthesized it with prior studies. We replicated the direction of the ''1-in-X'' effect in all 5 experiments using different scenarios, measurement scales, populations, and cultures. However, we found statistically significant results in only 2 experiments. Numeracy, cognitive reflection, gender, and age did not moderate the ''1-in-X'' effect. A synthesis of all available evidence, conducted via a Bayes factor meta-analysis and traditional meta-analysis, provided decisive evidence supporting the existence of the ''1-in-X'' effect yet a considerably smaller effect size estimation than previously suggested.

Overall, these results are consistent with those of Pighin and others[2] supporting the existence of the ''1-in-X'' effect. They differ in the estimation of its magnitude, which explains our difficulties in replicating a statistically significant effect in the experiments powered to a medium-sized effect. We believe that such estimation manifests a more general problem of replicability of published effects.[34,39] Even established effects may decline in magnitude (or completely disappear) in replication attempts, possibly due to the statistical self-correction of initially exaggerated outcomes.[40]

Besides the process of statistical self-correction, some methodological differences between the 2 sets of experiments should be considered as a potential cause of the smaller effect. First, we collected the data via online questionnaires and not via paper-and-pencil questionnaires, as was the case in the original studies. Nevertheless, if this deflated the effect, then our results represent more realistic estimation of the effect, because patients encounter health-related statistical information more often in a web-based rather than paper-and-pencil form.[41] Second, we sampled our participants from Slovak and British adult populations and not from the Italian adult population, as was the case in the original studies, which points toward cultural sensitivity of the ''1-in-X'' effect. An ongoing discussion about the extent to which some ratio effects are (not) culturally universal[6] would support such an explanation; however, the fact that the ''1-in-X'' effect was small in 2 culturally different samples weakens such a cultural-sensitivity explanation.

Our findings also have theoretical implications. The first implication concerns the ratio effects. We failed to observe the denominator-neglect and group-diffusion effects but not the ''1-in-X'' effect in Experiment 1. This finding could indicate either different mechanisms underpinning these effects or purely methodological or statistical artifacts at work. The negative findings perhaps occurred because the effects are bounded to a specific research paradigm. The current evidence does not support such a conjuncture, since the effects were observed in different paradigms. Alternatively, the negative findings represent a normal statistical variation for the effects, which failed to be revealed in the study powered to detect medium-sized effects (particularly for the group-diffusion effect). More cumulative evidence would shed some light on the existence of these ratio effects, their occurrence in specific designs, and their boundary conditions, which should be, in turn, reflected in their conceptualization. Since the ''1-in-X'' effect was substantially more pronounced in our study than the group-diffusion effect using the same design, one could assume a different mechanism behind these two.

The second implication concerns individual differences. We found that numeracy, gender, and age affected subjective probability, although they did not interact with the ratio formats. The presence of individual differences indicates that participants in our experiments processed the numerical information to at least some extent—these differences would not occur otherwise. Therefore, the smaller ''1-in-X'' effect found here should not be attributed to the lack of motivation to process numerical information. Furthermore, lack of interaction of formats with numeracy and cognitive reflection suggests that the ''1-in-X'' effect does not occur because of a weak ability to work with numbers[20] or because of miserly cognitive processing.[26] Further research should test other possible mechanisms, for example, the explanation that the ''1-in-X'' effect is mediated by different gist interpretations of the ratios.[3] Future research should also focus on the probability calibration of the 2 formats (''1-in-X'' v. ''N-in-X*N'') since current evidence failed to address this aspect.

In summary, we provide decisive evidence confirming the existence of the ''1-in-X'' effect but estimate the effect to be smaller than initially thought. Health professionals should carefully consider the selection of ratio formats when communicating probabilistic information to their patients. Even small effects substantially bias people's probability estimates if the effects have a cumulative nature, as is the case here.[13] We also strongly advocate replication studies to cumulate reliable and high-quality evidence for guiding risk communication policies.[42] The harnessing of sophisticated statistical techniques and use of sound methodological designs are crucial in this endeavor.

ACKNOWLEDGMENT

We thank Jeff Rouder for providing the R script to compute the meta-analytical Bayes factor; Vladimira Cavojova for a back-translation of the tasks; and Karl-Halvor Teigen, Brian Zikmund-Fisher, and three anonymous reviewers for helpful suggestions on an earlier version of this manuscript.

REFERENCES

1. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science. 1981;211(4481):453–58.

2. Pighin S, Savadori L, Barilli E, Cremonesi L, Ferrari M, Bonnefon JF. The 1-in-X effect on the subjective assessment of medical probabilities. Med Decis Making. 2011;31(5):721–29.

3. Zikmund-Fisher BJ. Time to retire the 1-in-X risk format. Med Decis Making. 2011;31(5):703–4.

4. Yamaguchi S. Biased risk perceptions among Japanese: illusion of interdependence among risk companions. Asian J Soc Psychol. 1998;1(2):117–31.

5. Miller DT, Turnbull W, McFarland C. When a coincidence is suspicious—the role of mental simulation. J Pers Soc Psychol. 1989;57(4):581–89.

6. Price PC, Matthews TV. From group diffusion to ratio bias: effects of denominator and numerator salience on intuitive risk and likelihood judgments. Judgm Decis Mak. 2009;4(6):436–46.

7. Bonner C, Newell BR. How to make a risk seem riskier: the ratio bias versus construal level theory. Judgm Decis Mak. 2008;3(5): 411–16.

8. Denes-Raj V, Epstein S. Conflict between intuitive and rational processing—when people behave against their better judgment. J Pers Soc Psychol. 1994;66(5):819–29.

9. Kirkpatrick LA, Epstein S. Cognitive experiential self-theory and subjective-probability—further evidence for 2 conceptual systems. J Pers Soc Psychol. 1992;63(4):534–44.

10. Hsee CK. The evaluability hypothesis: an explanation for preference reversals between joint and separate evaluations of alternatives. Organ Behav Hum Decis Process. 1996;67(3):247–57.

11. Bakker M, van Dijk A, Wicherts JM. The rules of the game called psychological science. Perspect Psychol Sci. 2012;7(6): 543–54.

12. Francis G. Publication bias and the failure of replication in experimental psychology. Psychon Bull Rev. 2012;19(6):975–91.

13. Cumming G. Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-analysis. London: Routledge; 2012.

14. Rouder J, Morey R. A Bayes factor meta-analysis of Bem's ESP claim. Psychon Bull Rev. 2011;18(4):682–89.

15. Gigerenzer G. Why the distinction between single-event probabilities and frequencies is important for psychology (and vice-versa). In: Wright G, Ayton P, eds. Subjective Probability. Chichester (UK): Wiley; 1994. p 129–61.

16. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. ed. Hillsdale (NJ): Lawrence Erlbaum; 1988.

17. Liberali JM, Reyna VF, Furlan S, Stein LM, Pardo ST. Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. J Behav Decis Mak. 2012;25(4):361–81.

18. Peters E. Beyond comprehension: the role of numeracy in judgments and decisions. Curr Dir Psychol Sci. 2012;21(1):31–35.

19. Reyna VF, Brainerd CJ. The importance of mathematics in health and human judgment: numeracy, risk communication, and medical decision making. Learn Individ Differ. 2007;17(2): 147–59.

20. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. Psychol Bull. 2009;135(6):943–73.

21. Sirota M, Juanchich M. Role of numeracy and cognitive reflection in Bayesian reasoning with natural frequencies. Stud Psychol. 2011;53(2):151–61.

22. Peters E, Vastfjall D, Slovic P, Mertz CK, Mazzocco K, Dickert S. Numeracy and decision making. Psychol Sci. 2006;17(5): 407–13.

23. Bonner C, Newell BR. In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. Mem Cognit. 2010;38(2):186–96.

24. Reyna VF, Brainerd CJ. Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. Learn Individ Differ. 2008;18(1):89–107.

25. Frederick S. Cognitive reflection and decision making. J Econ Perspect. 2005;19(4):25–42.

26. Toplak ME, West RF, Stanovich KE. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. Mem Cognit. 2011;39(7):1275–89.

27. Sirota M, Juanchich M, Hagmayer Y. Ecological rationality or nested sets? Individual differences in cognitive processing predict Bayesian reasoning. Psychon Bull Rev. 2013 June 21 [Epub ahead of print].

28. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Med Decis Making. 2001;21(1):37–44.

29. Sirota M, Juanchich M. To what extent do politeness expectations shape risk perception? Even numerical probabilities are under their spell! Acta Psychol. 2012;141(3):391–99.

30. Yamaguchi S, Gelfand M, Ohashi MM, Zemba Y. The cultural psychology of control—illusions of personal versus collective control in the United States and Japan. J Cross Cult Psychol. 2005;36(6):750–61.

31. Pighin S, Savadori L, Barilli E, et al. Using comparison scenarios to improve prenatal risk communication. Med Decis Making. 2013;33(1):48–58.

32. Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon Bull Rev. 2009;16(2):225–37.

33. Goodman SN. Toward evidence-based medical statistics, 2: the Bayes factor. Ann Intern Med. 1999;130(12):1005–13.

34. Wagenmakers E-J, Wetzels R, Borsboom D, van der Maas HL. Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). J Pers Soc Psychol. 2011;100(3):426–32.

35. Jeffreys H. Theory of Probability. Oxford (UK): Oxford University Press; 1961.

36. Wetzels R, Matzke D, Lee MD, Rouder JN, Iverson GJ, Wagenmakers E-J. Statistical evidence in experimental psychology: an empirical comparison using 855 t-tests. Perspect Psychol Sci. 2011;6(3):291–98.

37. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing [program]. Vienna (Austria): R Foundation for Statistical Computing; 2011.

38. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36:1–48.

39. Barhillel M, Budescu DV. The elusive wishful thinking effect. Bull Psychon Soc. 1992;30(6):471–71.

40. Schooler J. Unpublished results hide the decline effect. Nature. 2011;470(7335):437–37.

41. Eysenbach G. Consumer health informatics. BMJ. 2000; 320(7251):1713–16.

42. Sirota M, Juanchich M. Risk communication on shaky ground. Science. 2012;338(6112):1286–87.