w214 Final Project

Joshua Noble, Ryan T Orton, Sandip Panesar

4/8/2021

Icons, Charts, Denominators

Joshua Noble, Ryan T. Orton, Sandip Panesar

W241 Spring 2021

Introduction

Reading and understanding ratios, particularly when presented numerically, can be a taxing task even for people who claim to be at ease with mathematics **Reference 1**. The entire field of data visualization has been built around the premise that humans count objects and see ratios better when presented with graphical representations of those quantities rather than numerical ones. We wanted to test whether this was true in a very specific context reading ratios. Many ratios seem quite easy to read and understand e.g. 1/2, 25%, '2 out of 3', but many others are more difficult to interpret. Consequently, it is more difficult to process their context, particularly when related to probability or efficacy. There are numerous examples of this effect in everyday life: Millions of people buy lottery tickets despite the odds; random sampling is foreign enough that companies still use it as a brain-teaser question in job interviews; and as we've seen in the past six months, vaccine efficacy is rather difficult to understand.

Study Rationale and Hypothesis

Vaccine efficacy is measured by calculating the risk of disease among vaccinated and unvaccinated persons and determining the percentage reduction in risk of disease among vaccinated persons relative to unvaccinated persons. The greater the percentage reduction of illness in the vaccinated group, the greater the vaccine efficacy. The basic formula is written as:

$$\text{Vaccine Efficacy} = \frac{\text{Risk among unvaccinated group} - \text{Risk among vaccinated group}}{\text{Risk among unvaccinated group}}$$

It is unclear exactly how well this is understood, however. Certainly the COVID-19 pandemic has provided us ample opportunity to witness how these efficacy results can be confusing.

According to Garcia-Retamero et al **Reference 2**, denominator neglect is the focus on the number of times a target event has happened (e.g. the number of treated and nontreated patients who die) without considering the overall number of opportunities for it to happen (e.g. the overall number of treated and nontreated patients). In their 2 studies, the authors' addressed the effect of denominator neglect in problems involving treatment risk reduction where samples of treated and non-treated patients and the relative risk reduction were of different sizes. They also tested whether using icon arrays helped people take these different sample sizes into account.

Based upon the above, our causal research question is:

Are icon arrays (or other visual aids) more effective than text-based numerical ratios for interpretation of numerical ratios?

We hypothesize, based upon the aforementioned evidence that they are. More specifically, that a group presented with icon arrays correctly interprets the numerical ratios presented in them at a greater rate, compared to a group presented with only text or numbers.

Methods

The original paper on which our study was based examined numerical ratios and icon arrays. In order to expand the test we wanted to test both a chart that required interpretation, an icon array which might require counting, and a numerical ratio which showed a vaccine efficacy. Our experiment was conducted using the Berkeley XLab to recruit survey takers.

Research Design

We adopted an alternative approach to the classical pre-test/post-test ROXO and instead used a RXRXO post-test control group design. Our survey did not involve any observation prior to treatment or control randomization. The subjects were tested via a multi-factorial approach: All users were first presented with several "stem" questions about the COVID-19 pandemic and children returning to classrooms. These questions were the same for all participants and were intended to obscure the actual purpose of our survey so that participants wouldn't immediately think of it as testing ratios or mathematical concepts. Participants were first presented with a question from the chart treatment stage: Either a numerical description or a chart. Then the same subject was presented with a question from the icon array stage: Either a text-based description or an icon array. Within each stage, subjects would then be assigned randomly (using Qualtrics' survey flow randomization algorithm) into either control or treatment groups (**Figure 1.**):
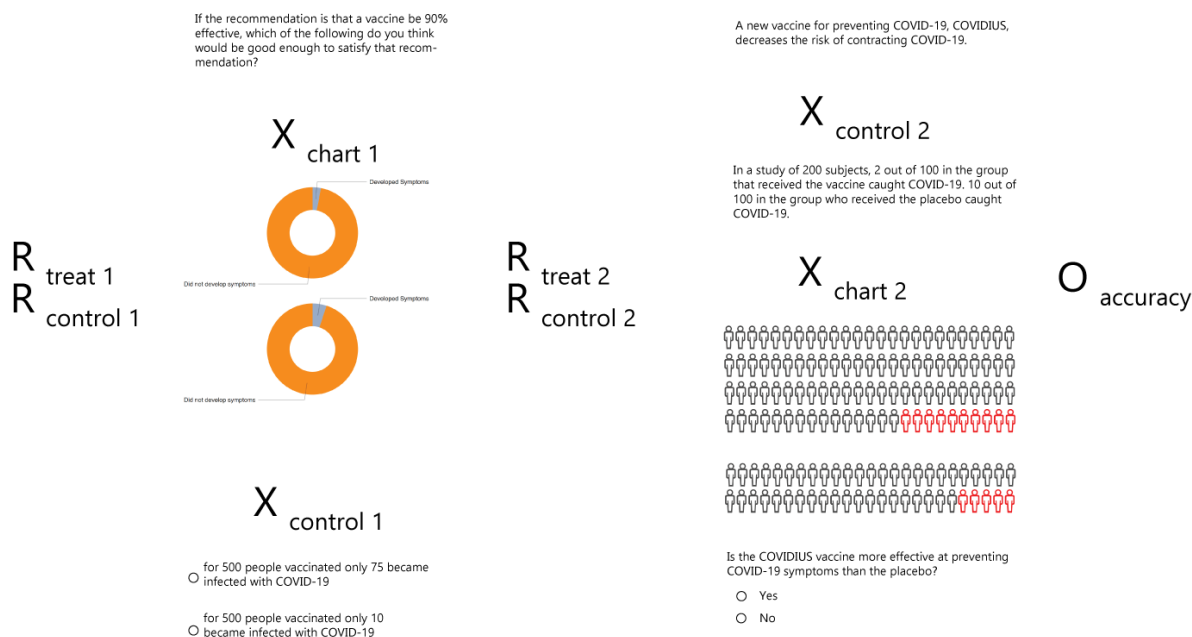

Figure 1 A diagram demonstrating the overall RXRXO study design.

Pie Chart Questions

Pie charts are a particularly ineffective form of visualization **Reference 3**. Our method involved participants differentiating between pie charts. In information visualization, when presented with several pie charts, subjects often have difficult interpreting or differentiating between them **Reference 4**. The pie charts we utilized lacked numbers and used ratios that were deliberately intended to be difficult to visually decode: 80%, 90%, 95%, and 97%. In this arm, the pie chart was either right or wrong, depending upon the ratio presented. If a pie chart without labels is more effective than a numerical ratio at expressing efficacy then this provides us with alternative evidence to test the hypothesis that visual aids were still more effective than either text or numerical ratios **Figure 2**
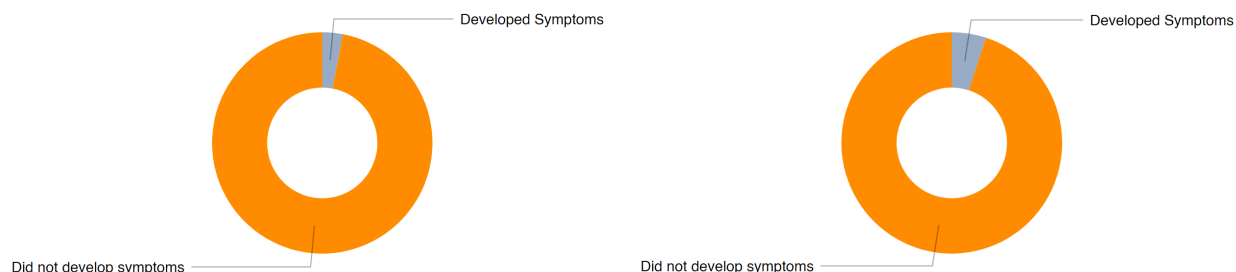

Figure 2 a. A pie chart representing 97% efficacy, b. A pie chart representing 85% efficacy.

Icon Array Questions

For the icon array questions, icon arrays were employed to test numerical ratios of efficacy against icon arrays that represented vaccine efficacy. These icon arrays differ from the pie chart pairings in that they contain countable quantities of different colored icons and thus can be more precise than the charts **Figure 3**.
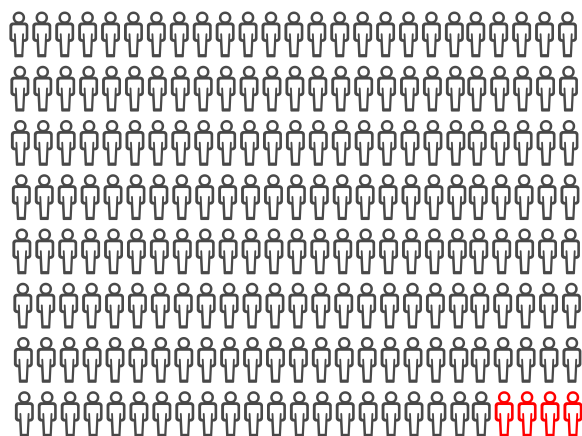

Figure 1: Figure 3a.

Figure 3 a. Icon array representing the treatment group of a hypothetical vaccine trial, whereby 4 out of 200 (represented by red and black figures, respectively) contracted COVID-19 following vaccination. b. Icon array representing the control group of a hypothetical vaccine trial, whereby 10 out of 100 (represented by red and black figures, respectively) contracted COVID-19 following vaccination. Altogether, vaccine efficacy is 80%.

The 'correct' answer to the question presented to the icon array subjects is also significantly easier to interpret compared to those in the chart arm. The correct answer was always 'yes', i.e. the vaccine was efficacious at
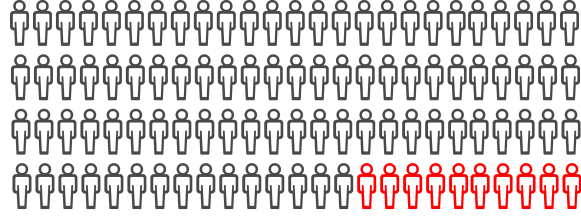
Figure 2: Figure 3b.

the 80% level.

Confounding Variables

In addition to the study-specific data, we also acquired a range of demographic data of participants. Numerous factors may affect the responses to our questions, namely how well subjects were able to understand them. These include age, level of educational attainment and English as a second language (ESL) status. We therefore kept data from the Berkeley Xlab experimental survey pertaining to these factors for use in analysis.

Analytic Plan

1. Descriptive statistics.
2. Group-based Statistical Tests (Chi-Squared).
3. Regression Analysis with and without potential control features.

Pilot Data

We ran a pilot study using Qualitrics which received 19 valid responses. This survey was open for approximately 10 days. The structure of the Pilot survey contained two arms which tested a chart against a numerator/denominator numerical representation and an icon array against a numerator/denominator numerical representation. We didn't collect demographic data from respondents. The accuracy of responses to the two questions is shown below:

q1_right/q2_right	0	1	Total
0	20% (1)	80% (4)	100% (5)
1	23% (3)	77% (10)	100% (13)
Total	22% (4)	78% (14)	100% (18)

Figure 3: Pilot Study Results

The pilot study was not large enough to provide reliable sample sizes but it did provide confirmation that our survey itself was working as intended.

In the chart treatment group, users answered correctly 85.7% of the time, while in the control they answered 83.3% of the time. **Table 1.**.

The pilot study was not large enough to provide reliable sample sizes but it did provide confirmation that our survey itself was working as intended.

q1_right/q2_right	0	1	Total
0	20% (1)	80% (4)	100% (5)
1	23% (3)	77% (10)	100% (13)
Total	22% (4)	78% (14)	100% (18)

Figure 4: Table 1.

Results

The overall breakdown of our study is provided in **Figure 4.**. In all there were 309 respondents. 45 subjects were lost to attrition (test-stage surveys, failure to complete). For the first question, 131 subjects were presented with the pie-chart (treatment group) and 133 subjects were presented with the text control question. For the second question, 128 subjects were presented with the icon array (treatment group), 128 subjects were presented with the text control question.

Figure 4. A diagram demonstrating the study flow and numbers of respondents for each respective control and treatment stage.

Stem Questions

These were asked mainly as a distractor for subjects, prior to being presented with the questions of interest. The data is presented in **Table 2.**

Demographics

Regarding relevant demographics, participant age ranged from 18 to 54, with a mean age of 23.5 (standard deviation 5.5). Relevant demographics of the test subjects are presented, grouped by educational level in **Table 3.**.

Approximately half of the respondents had completed some college (likely undergraduate students), a quarter had completed their undergraduate studies, while only 8% progressed onto further education. 16% of respondents had no college. Interestingly, the majority of respondents in all educational groups were ESL students, with over 4/5 of postgraduate respondents reporting ESL status.

Between-Group Comparison

Table 4. summarizes the responses received for the first set of questions, where the subjects were exposed to either the text or the pie chart. 83.5% of those in the control answered the question correctly, while only 68.9% in the treatment did. As these data were categorical, consisted of counts between groups, and were generated through random sampling we utilized a χ^2 test to determine if the proportion of right and wrong answers was significantly different between groups. The χ^2 value was 7.42 and had a p-value of 0.006, indicating a highly significant difference.

Table 5. summarizes the responses received for the second set of questions, where the subjects were exposed to either the text or the icon array. 84.4% of those in the control answered the question correctly, while 93.8% in the treatment did. As these data were categorical, consisted of counts between groups, and were generated through random sampling we utilized a χ^2 test to determine if the proportion of right and wrong answers was significantly different between groups. The χ^2 value was 5.80 and had a p-value of 0.0160, indicating a significant difference at the 5% level.
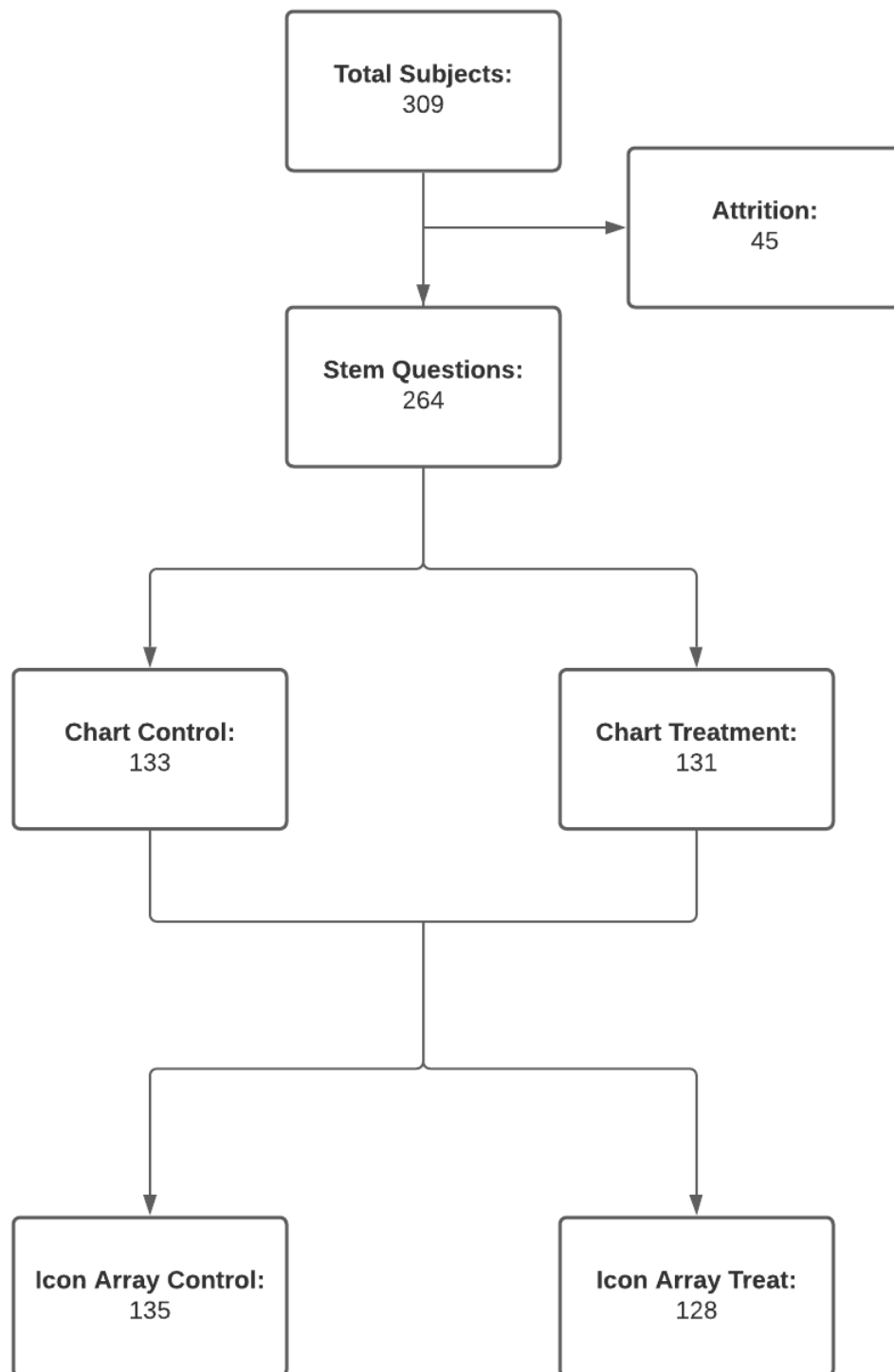
Figure 5: Figure 4.

	Question	Count
1	Should Elementary Schools be Opened?	
2	Yes	114
3	No	150
4	How Many Days Should Schools Open?	
5	None	110
6	<5 Weekdays	116
7	Mon-Fri	38
8	At What Capacity Should Schools Open?	
9	25%	95
10	50%	125
11	75%	28
12	100	16
13	How Effective Should the Vaccine Be?	
14	85%-90%	98
15	90%-95%	96
16	>95%	69

Figure 6: Table 2.

	Educational Level	Count	%Total	Mean Age	%ESL
1	Bachelor's degree	67	25.4	25.4	68.7
2	Advanced degree (Master's, Doctorate)	21	8.0	29.0	81.0
3	Some college	134	50.8	22.4	79.1
4	No college	42	15.9	21.4	71.4

Figure 7: Table 3.

	Right	Wrong	%Right
Control	111	22	83.5
Treat	82	37	68.9

Figure 8: Table 4.

	Right	Wrong	%Right
Control	114	21	84.4
Treat	120	8	93.8

Figure 9: Table 5.

To summarize, the treatment arm of the chart control group had a higher proportion of incorrect answers compared to the control arm, while the treatment arm of the icon array group had a higher proportion of correct answers compared to the controls. These differences were significant at the 5% level at least, confirming our original hypothesis. Nevertheless, the χ^2 approach cannot control for the additional demographic factors that we identified as potentially being relevant, so regression analysis may offer further insight by allowing us to control for these factors.

Regression Analysis

We created a simple model that regresses the number of correct answers against the participants being in two treatment groups:

	Dependent variable:
	all_q_right
q1_treat	−0.014
	(0.070)
q2_treat	0.150**
	(0.069)
Constant	1.551***
	(0.064)
Observations	264
R^2	0.018
Adjusted R^2	0.011
Residual Std. Error	0.557 (df = 261)
F Statistic	2.454* (df = 2; 261)
Note:	*p<0.1; **p<0.05; ***p<0.01

Our second model regresses the response for each arm of the study against being in the respective treatment group and the education level of the participant.

	Dependent variable:	
	q1_right	q2_right
	(1)	(2)
q1_treat	−0.087	
	(0.055)	
q2_treat		0.110***
		(0.039)
as.factor(Edu_Level)Bachelor's degree	0.035	−0.072
	(0.110)	(0.078)
as.factor(Edu_Level)No college	0.093	−0.204**
	(0.118)	(0.083)
as.factor(Edu_Level)Some college	0.178*	−0.136*
	(0.103)	(0.073)
Constant	0.656***	0.953***
	(0.099)	(0.070)
Observations	264	264
R^2	0.031	0.054
Adjusted R^2	0.016	0.039
Residual Std. Error (df = 259)	0.441	0.312
F Statistic (df = 4; 259)	2.087*	3.678***
Note:	*p<0.1; **p<0.05; ***p<0.01	

We anticipated that the age of the participants might be a confounding factor and so regressed the test scores against the reported age from survey takers

	Dependent variable:
	all_q_right
Birthyear	−0.002
	(0.006)
Constant	1.674***
	(0.152)
Observations	257
R^2	0.0002
Adjusted R^2	−0.004
Residual Std. Error	0.536 (df = 255)
F Statistic	0.061 (df = 1; 255)
Note:	*p<0.1; **p<0.05; ***p<0.01

Another possible confounding factor we anticipated was whether participants had English as a first language.

	Dependent variable:
	all_q_right
q1_treat	−0.012
	(0.070)
q2_treat	0.150**
	(0.069)
as.factor(ESL == "Yes")	0.038
	(0.079)
Constant	1.522***
	(0.089)
Observations	264
R^2	0.019
Adjusted R^2	0.008
Residual Std. Error	0.557 (df = 260)
F Statistic	1.707 (df = 3; 260)
Note:	*p<0.1; **p<0.05; ***p<0.01

Finally, we built a fully saturated model including all of our confounding factors:

	Dependent variable:
	all_q_right
q1_treat	−0.032
	(0.070)
q2_treat	0.097
	(0.069)
as.factor(ESL == "Yes")	−0.012
	(0.075)
Birthyear	0.00001
	(0.007)
as.factor(Edu_Level)Bachelor's degree	−0.031
	(0.141)
as.factor(Edu_Level)No college	−0.073
	(0.165)
as.factor(Edu_Level)Some college	0.063
	(0.137)
Constant	1.600***
	(0.245)
Observations	257
R^2	0.020
Adjusted R^2	−0.008
Residual Std. Error	0.537 (df = 249)
F Statistic	0.718 (df = 7; 249)
Note:	*p<0.1; **p<0.05; ***p<0.01

References

1. **Insert reference from intro paragraph here**
2. [Do Icon Arrays Help Reduce Denominator Neglect?]
3. [(William S. Cleveland, The Elements of Graphing Data, Hobart Press, 1994]
4. (Edward Tufte, The Visual Display of Quantitative Information, Graphics Press, 1983, p. 178.)