The Effect of eBook Cover Attributes on the Book's Appeal

Spring 2021 W241 Final Report

Isaac Vernon, Renjun Cheng, Jacky Wong, and Ken Opamuratawongse

Abstract

Over the past decade, and especially in recent times, eBook purchases have taken up an increasing share of the number of overall book purchases. With eBooks more popular than ever, there are incentives to investigate the potential factors that can cause an ebook to succeed and derive some insight into whether authors should invest capital into those areas. We will run an experiment to measure the effect that varying the color and image size of an eBook cover will have on the appeal or attractiveness to the customer.

The experiment was designed as a lineup design incorporating the methods of a factorial design to create the various treatments. Our experiment was conducted as a survey, with randomization giving us a between-subjects design. Overall, we found that when presented with other books that are colored and have regular sized images, there are significant gains in relative attractiveness from having a cover that is in color and contains an image of reasonable size for both the general population as well as the subpopulation of ebook readers.

Background

In the past decade, a large volume of book purchases have shifted from physical copies to ebooks. This means that the author and publisher need to take into account the different medium of the ebook store. The determinants and effects for the level of attraction of the customer towards a book need to be considered. There are many factors that could be varied -- the price, the rating, the number of reviews. However, there is another component that is more central to the experience of browsing for a new book: the cover. The cover has two main parts that can be varied: the title and the art of the cover. It would be difficult to use the title to create a rigorous experiment as there isn't that much we can vary within the title that is without controversy --

people have different ideas about what makes a good title. Changes to the title should also have smaller effects overall (unless of course a change is large or blatant enough, though that might tip our hand in our survey design). Therefore, we focus on determining what effect various other factors of a cover, such as the coloring and image size, have on the success or attractiveness of a kindle ebook. This will be helpful for all authors, but especially for self-published authors (a growing trend with the rise of ebooks), as this can help them better determine the payoff they can achieve by putting some of their limited resources into cover design.

Research Question

As part of a semester long course with limited funding we can only run an experiment within our time limits and resource constraints. This means that we can't use A/B testing on Amazon's kindle ebook website, an ideal experiment which would give us the best results. However, we can do our best to approximate and replicate the conditions from this ideal version of the experiment. We will then address the research question, what factors of a book cover affect the likelihood of a person choosing the book from the kindle ebook store?

Hypothesis

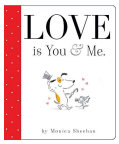
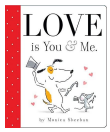
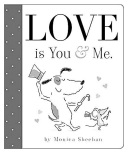
Our goal is to find out the effect of varying the color and image size on the attractiveness or draw of the book on potential consumers. We believe that there are effects to be found as the human eye is drawn to color or things that stand out (including the images on a cover) and as such covers in black in white or covers with small or no images should be less appealing to the customer.

In line with our theory above, we expect that a colored title will be more attractive to customers (both in overall terms and for those who identify as kindle ebook readers) and that a larger image will similarly improve the draw to potential readers.

Experiment Design

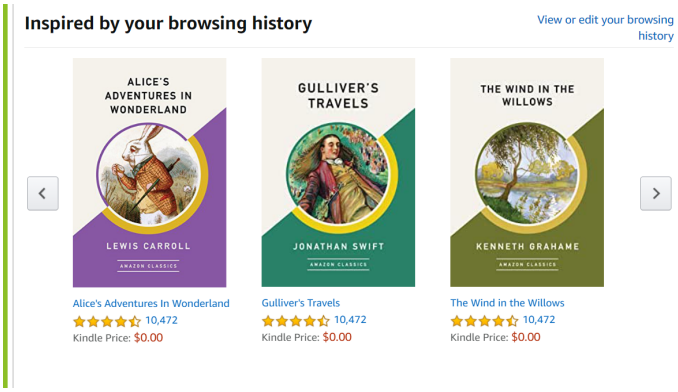
Experiment Overview (Factorial Design)

There are two main designs that we use for our experiment, the first of which is the factorial design. We decided that there were two main, independent factors that we wanted to vary -- the cover color and the image size. This led us into designing our treatments as a 2x3 factorial design, giving us two factors for color (Color or Black and White) and three factors for image size (Regular Image, Small Image, No Image). This is a valid setup for a factorial design as the two factors are independent from one another (we ensured that the ebook cover that we treated would allow for independent factors under this setup). Overall, this setup gave us six different treated covers (Regular and Color, Regular and Black and White, Small and Color, etc) for use in our survey design as shown below.

2x3 Factorial Design	No Image	Small Image	Regular Image
Color			
Black and White			

Experiment Overview (Lineup Design)

The other approach that we utilized was a lineup design. Since our goal was to determine the relative attractiveness of an ebook to a customer, we decided to proxy this with choice rates for our particular treated book. We then had to determine how we wanted to generate these rates from our survey form. To do so, we chose to run our treatments as part of a lineup. To conform with the actual conditions of the ebook store, the lineup of books was presented as they would appear on the Amazon Kindle eBook store carousel. We then randomized the treatment to the respective choice rates.



Power Calculation

We did our power calculations a couple of different ways to check our calculations. The first method was to compute power by hand using the formula for proportions from two independent samples. This uses the two-sided sample equation $n = 2(\frac{Z_{1-\alpha/2} + Z_{1-\beta}}{(\frac{|p_1 - p_2|}{\sqrt{p_1(1-p_1)}})})^2$ where n is the sample size needed in each group to achieve the given level of power. We then had to make some decisions about our parameters and estimate the effect size that we would see. Since we didn't have enough users in our pilot study for those estimates to be useful, we had to come up with our best estimates. We assumed that the base choice rate for a book in our lineups should be 33% (we tested to see if this rate was different during the pilot study) and that one of our main effects (going from color to black and white) would essentially halve the choice rate (change to 17%). Given our assumptions ($\alpha = .05$, $\beta = .8$and the effect size going from .33 to .17), we then calculated that we would need 115 samples in each group to achieve a power of .8. We estimated that the total effect of moving through our image size factors would be similar (33% - 17% over Regular Image to No Image). We then assumed that each step in the process would take an equal rate so to detect an effect of cover size would need 505 samples per for the intermediate step and 115 samples each for the overall effect (Regular to No Image). Even though it was somewhat unlikely that we would get the samples we needed for the intermediate effect we decided to stick with our design (there were at least 5 teams that we knew of participating, so with $2500 at least we assumed we could get close to a thousand respondents -- still not enough for 80% power, but still fairly good). We confirmed our calculations with an online two-sample proportion calculator (https://www.stat.ubc.ca/~rollin/stats/ssize/b2.html).

Project Timeline

The project topic was proposed during week 5 of this course. At the time we were going to attempt to measure an individual's valuation of a book based off of the cover attributes (contingent valuation). By week 9, this idea had evolved into the current concept we are presenting: a lineup experiment paired with a factorial design to get the relative choice rates of various treated covers in constant lineups. In week 11, we performed our pilot survey, attempting to identify the best genre lineups to include in the final survey. After analyzing the results of this pilot survey, we finalized our Qualtrics survey and sent it out as part of the Omnibus XLab survey during week 13. We then waited around a week for results, and used the following week to conduct our interpretations of the findings.

Proposal	Initial Design	Pilot Survey	Final Survey	Results and Analysis
Week 5	Week 9	Week 11	Week 13	Week 15

Survey Tools

We used two different survey tools throughout the project. The first was Google Forms. This was used in our pilot survey to elicit responses as we didn’t need to implement any advanced survey techniques to accomplish our goal of narrowing down our lineups to our final choices. Therefore, as Google Forms was the most familiar tool (as we had used it previously), we utilized it during our pilot survey.

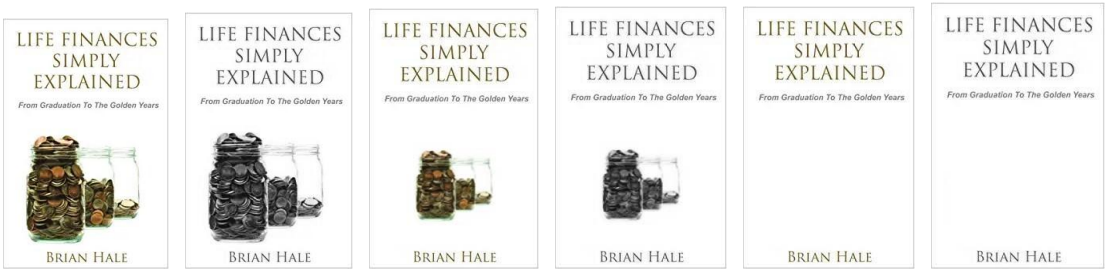
The second, and most important, survey tool that we utilized was Qualtrics. Specifically, we leveraged the advanced functions that it supports, including randomization, separating questions across different pages, and allowing for an introduction prior to our section of the survey. Qualtrics was used for the implementation of our final study, as we included our questions among a larger Omnibus survey for response.

Enrollment and Recruitment

As we had volunteered our entire budget for the project towards the Berkeley XLab Qualtrics survey platform, we decided to gather respondents for our pilot study locally and for free. This meant that the recruitment for our pilot study included family, friends, other students (contacted through Slack), and respondents who saw our post on NextDoor in San Rafael, California. This enrollment could potentially introduce selection bias into our pilot study (it's possible that those who choose to take the survey will have systematic differences when choosing books from a lineup) though any potential effect should be negligible.

Our enrollment for our final version of the survey followed a different form. As one of the Berkeley XLab groups, we had our respondents come from the Berkeley XLab survey network. They were given $20 worth of credit to complete the full omnibus survey which included our questions. The survey also had a few other respondents. These were collected from Amazon's Mechanical Turk when the survey and survey design was being tested. We have a perfect response rate for those contacted through Berkeley XLab, but there is attrition from the Mechanical Turk population.

Treatment



We have 6 different treated versions of each book, one for each combination of factors in our 2x3 factorial design. The above images are one of the treatment spreads. You can see that there are 6 different versions of a cover for the book, each a different combination of color and image (These are Regular Size and Full Color, Small Image and Color, No Image and Color, Regular Size and Black and White, Small Image and Black and White, and No Image and Black and White).

We repeated this treatment step 4 times with different covers, as we planned to run the treatment over 4 different versions of the experiment (using new books for each variation of the experiment). We also chose to run our experiment as a lineup. As shown below, this means that we provided two other book choices along with one of our treated versions. These choices were presented on a version of the Amazon kindle ebook site (with other factors such as price, rating, and the number of reviews staying constant across all choices). Users were then asked "Select your favorite of the books below". We then would be able to collect the relative choice rates for each version of the treatment.

Potential Outcomes

First, for each individual question (one of our 4 genre lineups) we will be making 13 comparisons. These correspond to the full slate of comparisons for a 2x3 factorial design (including both the four main effects and the nine interaction effects). Some examples include the overall effects of switching from a colored cover to black and white, the interaction effect from switching colors while having a regular sized image, and the overall effect of going from a regular image to a small image. We will also aggregate the data from the four versions of the experiment we ran and run the same 13 comparisons as well as running a regression on the overall results for point estimates and error estimates. Finally, we will repeat each of these comparisons for the subset of our data where the user identifies as a kindle ebook purchaser (corresponding to responses of "I buy ebooks" or "I buy both" when asked to identify the type of books that they purchase).

Randomization

We decided to randomize each question at the individual level. This means that we randomized the version of this question you would see, with the six different variants corresponding to the six different treatment options (you will randomly see one of the treated versions in the lineup that is presented to you). After we set our treatment randomization, we also set the survey to randomize choices equally, ensuring that we have close to 55 observations for each version of the question. All of this randomization was created using the tools in Qualtrics (the randomization was set in the survey flow).

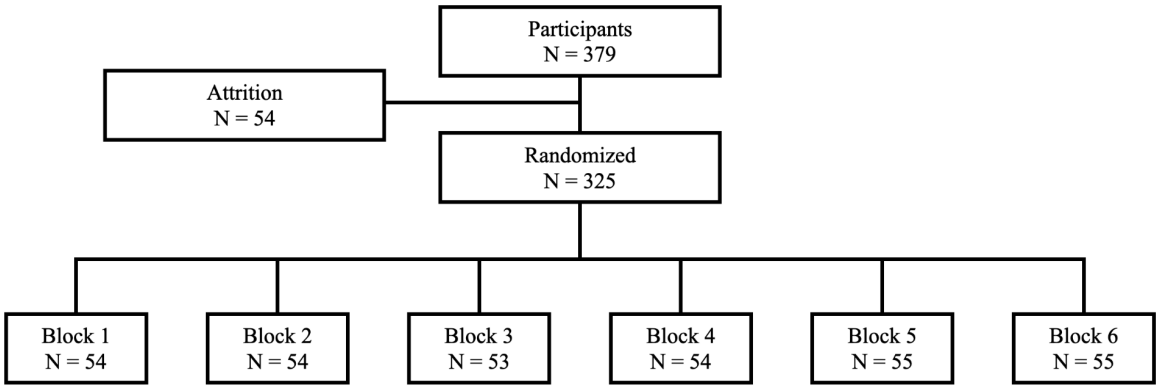
Observations and Outcome measures

The relevant data we received from Berkeley Qualtrics came from 25 multiple-choice questions. One question was a covariate question about the type of books the user purchases and the other 24 questions corresponded to the six treated options over the four different questions. More specifically, we have 4 types of lineups (Children, Finance, Classics, and Non-fiction) and 6 different treatment options so there are 4*6, 24 questions.

We chose to focus on the selection rate of the treated book cover as our key measurement. This was the intention from the beginning of this experiment design study, as this corresponds to the popularity or attractiveness of the book.

Data Completeness

The flowchart shows the survey procedure for one of our questions. There were 379 responses to the survey. Of these, 54 respondents failed to respond to any of our questions and were marked as attrited from the survey. The 325 remaining participants were then randomly assigned into one of six blocks for each question, with each block corresponding to a different version of the lineup (each block shows a different treated version of our cover of interest). The process of randomization was then repeated three more times, with the randomization being repeated at each step (being assigned to block 1 during the first question does not automatically assign you to block 1 for the remaining questions). There was no attrition during this part of the process.



Results

General Results

Children Lineup	No Image	Small Image	Regular Image
Color	.296	.4	.315
Black and White	.296	.327	.283

Financial Lineup	No Image	Small Image	Regular Image
Color	.278	.509	.377
Black and White	.259	.444	.327

NonFiction Lineup	No Image	Small Image	Regular Image
Color	.411	.55	.625
Black and White	.392	.5	.357

Classic Lineup	No Image	Small Image	Regular Image
Color	.393	.451	.531
Black and White	.327	.354	.404

We can see that the preliminary results that we get are roughly in line with the trends that we expected going into the experiment. These are the decrease in choice rates going from color covers to black and white as well as a decrease in choice rate going from an image to no image (there doesn't appear to be a consistent trend for the effect of changing from a regular image to a small image). From these preliminary results, we can then conduct our analysis two different ways; regression and significance tests (both foundational methods in factorial designs).

Regression

OLS Regression Results

Dep. Variable:	Y	R-squared:	0.781
Model:	OLS	Adj. R-squared:	0.664
Method:	Least Squares	F-statistic:	6.672
Date:	Sun, 18 Apr 2021	Prob (F-statistic):	0.000851
Time:	18:45:41	Log-Likelihood:	40.846
No. Observations:	24	AIC:	-63.69
Df Residuals:	15	BIC:	-53.09
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.2463	0.034	7.208	0.000	0.173	0.319
Color	0.0258	0.039	0.653	0.524	-0.058	0.110
Regular	0.0242	0.039	0.612	0.550	-0.060	0.108
Small	0.0877	0.039	2.223	0.042	0.004	0.172
Color_reg	0.0934	0.056	1.673	0.115	-0.026	0.212
Color_small	0.0454	0.056	0.813	0.429	-0.074	0.164
Type_Finance	0.0463	0.032	1.435	0.172	-0.022	0.115
Type_Non-Fic	0.1529	0.032	4.745	0.000	0.084	0.222
Type_Classic	0.0902	0.032	2.800	0.013	0.022	0.159

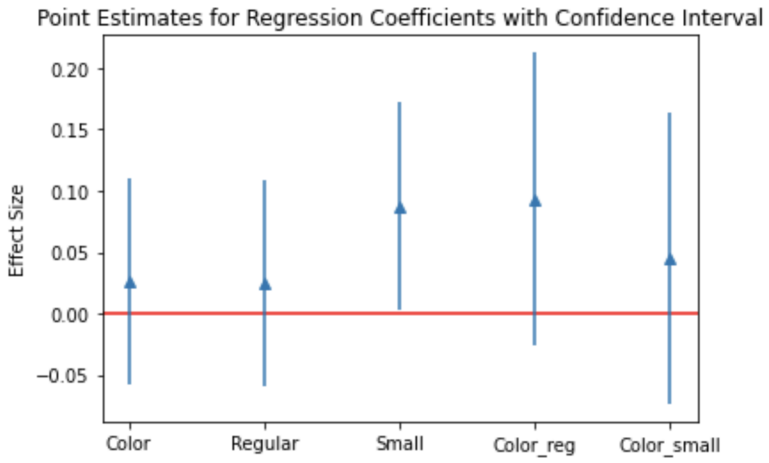
Omnibus:	1.353	Durbin-Watson:	2.207
Prob(Omnibus):	0.508	Jarque-Bera (JB):	0.899
Skew:	-0.026	Prob(JB):	0.638
Kurtosis:	2.053	Cond. No.	10.3

The first way that we analyzed our data was with a similar method to the one used by Rosen in his discrimination study. This is to use a regression with interaction terms included.

Above is our regression table along with point estimates and standard errors. The regression estimates of Color & Regular, Color & Small are the outcome with interaction while the terms Color, Regular, and Small indicate these respective changes to the base case of Black and White with No Image. The Type_ variables are one-hot encodings of the different genres (we see that the Non-Fiction and Classic genres caused significant differences in our base case percentage).

In this regression model, our base case (represented by the constant term) is a book with a Black and White cover with No Image from the children's book lineup. Constant-coefficient indicates there we expect 24% of people could choose that book among 3 books. The color coefficient indicates if the book cover is colored (a Colored cover with No Image), there will be 2.58% more people choosing that. For Regular Image size (this means a Black and White cover with a Regular Image), there will be 2.42% more, and smaller size (again a Black and White cover with a Small Image): 8.77%. Other than that, the interaction coefficient indicates that if the book cover is both Colored and has a regular-sized image, the chosen rate will go up an extra 9.34%. Similarly, Colored with Smaller size goes up an extra 4.54%.

Overall, we find one significant measure through this regression. We can see that the coefficient for Small (changing from No Image and Black and White to a Small Black and White Image) is significant, meaning that there does appear to be a relevant change (that is also practically significant in this case) in this situation. We can see this visualized in the plot below (shows point estimates with 95% confidence intervals)



Significance Tests

	Childrens	Overall	NoImage	SmallImage	RegImage	RegMinNo.Pval	RegMinSm.PVal	SmMinNo.PVal
1	Color	0.337	0.296	0.4	0.315	0.019 (1)	-0.085 (0.467)	0.104 (0.35)
2	Black and White	0.302	0.296	0.327	0.283	-0.013 (1)	-0.044 (0.772)	0.031 (0.887)
3	Overall		0.296	0.364	0.299	0.003 (1)	-0.065 (0.387)	0.067 (0.361)
4	Color-B&W (PVal)	0.035 (0.578)	0 (1)	0.073 (0.552)	0.032 (0.882)			

	Financial	Overall	NoImage	SmallImage	RegImage	RegMinNo.Pval	RegMinSm.PVal	SmMinNo.PVal
1	Color	0.389	0.278	0.509	0.377	0.1 (0.373)	-0.132 (0.237)	0.231 (0.023)
2	Black and White	0.344	0.259	0.444	0.327	0.068 (0.569)	-0.117 (0.289)	0.185 (0.07)
3	Overall		0.269	0.477	0.352	0.083 (0.239)	-0.125 (0.083)	0.209 (0.002)
4	Color-B&W (PVal)	0.045 (0.464)	0.019 (1)	0.065 (0.629)	0.05 (0.731)			

	NonFiction	Overall	NoImage	SmallImage	RegImage	RegMinNo.Pval	RegMinSm.PVal	SmMinNo.PVal
1	Color	0.524	0.411	0.55	0.625	0.214 (0.047)	0.075 (0.556)	0.139 (0.189)
2	Black and White	0.416	0.392	0.5	0.357	-0.035 (0.862)	-0.143 (0.186)	0.108 (0.361)
3	Overall		0.402	0.526	0.481	0.079 (0.31)	-0.046 (0.592)	0.124 (0.086)
4	Color-B&W (PVal)	0.108 (0.065)	0.019 (1)	0.05 (0.729)	0.268 (0.011)			

	Classic	Overall	NoImage	SmallImage	RegImage	RegMinNo.Pval	RegMinSm.PVal	SmMinNo.PVal
1	Color	0.455	0.393	0.451	0.531	0.138 (0.223)	0.08 (0.551)	0.058 (0.68)
2	Black and White	0.361	0.327	0.354	0.404	0.077 (0.541)	0.05 (0.717)	0.027 (0.913)
3	Overall		0.361	0.397	0.465	0.104 (0.165)	0.069 (0.377)	0.035 (0.683)
4	Color-B&W (PVal)	0.094 (0.106)	0.066 (0.608)	0.097 (0.384)	0.127 (0.282)			

	Combined	Overall	NoImage	SmallImage	RegImage	RegMinNo.Pval	RegMinSm.PVal	SmMinNo.PVal
1	Color	0.426	0.345	0.48	0.456	0.11 (0.026)	-0.024 (0.694)	0.134 (0.006)
2	Black and White	0.356	0.318	0.404	0.343	0.025 (0.654)	-0.061 (0.219)	0.086 (0.076)
3	Overall		0.332	0.441	0.398	0.066 (0.054)	-0.043 (0.22)	0.109 (0)
4	Color-B&W (PVal)	0.071 (0.011)	0.028 (0.608)	0.076 (0.126)	0.113 (0.023)			

We can investigate the results that we get when we use significance testing over the 13 criteria. In the individual genre tables, we find few significant results at an alpha level of .05. The results that are significant are for changing from a Small Image to No Image in the Financial lineup (both for the overall results as well as within the Colored cover subset). We have other results that are marginally significant, including the overall change from Color to Black and White for

the NonFiction and Classic lineups and more subsetted results for the change from a Small Image to No Image.

For the overall results aggregated over the tables (since all treatments/factor levels were the same this aggregation is valid) we find two main effects that are significant at an $\alpha = .05$level. These are the main (overall) effects of switching the cover from Colored to Black and White and from moving from a Small Image to No Image. We also have marginal significance in the case of moving from a Regular Image to No Image. Within these main effects, we also have significant results from the interactions. The change of Color only appears to be significant within the Regular sized images by our results, and the effect of changing the Image size is only significant for Colored covers (both in the Regular Image to No Image case and the Small Image to No Image case).

Overall, we do find some significant results that are in line with our hypothesis. It appears that changing from a Colored cover to a Black and White cover as well as changing from an Image to No Image have significant effects on the attractiveness of the book to customers. We also find that there appears to be no real effect on changing the image from a Regular size to a Small size (this could be due to the magnitude of the change that we instituted) or on changing from Colored to Black and White for covers without an image (this agrees with our logic, since even if there was an effect, it would likely be too small to detect at our current level of power). One interesting result we found was that our point estimates for moving from a Regular Image to a Small Image are negative (a book is more popular with a smaller image on the cover). This was in contrast to our intuition, though this is likely not a meaningful result given that our constituent lineup results had an equal number of positive and negative point estimates for this effect.

These results lead us to conclude that there are significant benefits to be had for an author that creates their cover in a certain way. It appears that in general, colored covers with an image will do a lot towards maximizing the appeal of the ebook, at least in the overall population.

We can also see the effects for those respondents who identify as ebook purchasers (we tracked a covariate for each respondent that categorized the type of books they purchase and read). This

narrowed down our responses to 164 individuals (of which 17 were still lost to attrition), for which our outcome tables can be found below. Within this subgroup, we also find a significant effect for the change from Small image to No image in the Financial lineup (as well as marginal significance on the Regular image to No image category). There are other marginally significant outcomes as well, though we fail to find any other convincing measures of significance within the lineups (due to the reduced power of our tests). Finally, in the overall, aggregated data we see the same significant effects that we found from the larger population as a whole (Color to Black and White as well as Regular or Small image to No image). Additionally, the other point estimates again follow the hypothesis we had at the start of the experiment.

Again, these results lead us to conclude that there are significant benefits to be had for an author that creates their cover in a certain way. It appears that within the subgroup of ebook readers, colored covers with an image (of any reasonable size) will do a lot towards maximizing the appeal of the ebook when contrasted against other colored, regular image covers.

	Childrens	Overall	NoImage	SmallImage	RegImage	RegMinNo.Pval	RegMinSm.PVal	SmMinNo.PVal
1	Color	0.347	0.25	0.381	0.423	0.173 (0.29)	0.042 (1)	0.131 (0.502)
2	Black and White	0.236	0.278	0.28	0.172	-0.105 (0.623)	-0.108 (0.535)	0.002 (1)
3	Overall		0.261	0.326	0.291	0.03 (0.91)	-0.035 (0.869)	0.065 (0.647)
4	Color-B&W (PVal)	0.111 (0.196)	-0.028 (1)	0.101 (0.68)	0.251 (0.081)			

	Financial	Overall	NoImage	SmallImage	RegImage	RegMinNo.Pval	RegMinSm.PVal	SmMinNo.PVal
1	Color	0.365	0.233	0.5	0.389	0.156 (0.412)	-0.111 (0.675)	0.267 (0.072)
2	Black and White	0.274	0.207	0.381	0.261	0.054 (0.899)	-0.12 (0.596)	0.174 (0.301)
3	Overall		0.22	0.447	0.317	0.097 (0.394)	-0.13 (0.304)	0.226 (0.023)
4	Color-B&W (PVal)	0.091 (0.315)	0.026 (1)	0.119 (0.602)	0.128 (0.592)			

	NonFiction	Overall	NoImage	SmallImage	RegImage	RegMinNo.Pval	RegMinSm.PVal	SmMinNo.PVal
1	Color	0.551	0.429	0.586	0.607	0.179 (0.342)	0.021 (1)	0.158 (0.415)
2	Black and White	0.42	0.421	0.52	0.32	-0.101 (0.709)	-0.2 (0.252)	0.099 (0.729)
3	Overall		0.425	0.556	0.472	0.047 (0.812)	-0.084 (0.5)	0.131 (0.297)
4	Color-B&W (PVal)	0.131 (0.156)	0.008 (1)	0.066 (0.831)	0.287 (0.07)			

	Classic	Overall	NoImage	SmallImage	RegImage	RegMinNo.Pval	RegMinSm.PVal	SmMinNo.PVal
1	Color	0.507	0.409	0.455	0.63	0.221 (0.211)	0.175 (0.349)	0.045 (1)
2	Black and White	0.368	0.208	0.483	0.391	0.183 (0.293)	-0.091 (0.705)	0.274 (0.074)
3	Overall		0.304	0.471	0.52	0.216 (0.053)	0.049 (0.766)	0.166 (0.143)
4	Color-B&W (PVal)	0.139 (0.127)	0.201 (0.247)	-0.028 (1)	0.238 (0.162)			

	Combined	Overall	NoImage	SmallImage	RegImage	RegMinNo.Pval	RegMinSm.PVal	SmMinNo.PVal
1	Color	0.443	0.317	0.49	0.525	0.208 (0.004)	0.035 (0.722)	0.173 (0.019)
2	Black and White	0.324	0.267	0.42	0.28	0.013 (0.966)	-0.14 (0.054)	0.153 (0.039)
3	Overall		0.293	0.455	0.402	0.109 (0.032)	-0.053 (0.339)	0.161 (0)
4	Color-B&W (PVal)	0.119 (0.004)	0.05 (0.548)	0.07 (0.399)	0.245 (0.001)			

Conclusions

Overall, we found results in line with the hypothesis that we had coming into the experiment. We found statistically and practically significant results for the effect of a colored cover and the effect of having an image. This is for the case where the cover in mention is compared to two other colored, image-containing covers of similar genre. However, as this seems to be a common occurrence on the kindle store (backed up by years of browsing, but no real numbers to confirm this), then we conclude that there are advantages that an author can get by investing in their cover design when it comes to the marketability of their ebook to both the general population as well as the subgroup of ebook readers.

Future Enhancements and Limitations

The first issue we ran into was the sample size from our final survey was less than we expected it to be. This meant that our test did not have quite enough power to detect all the effects that appeared to be there (according to our point estimates). A second issue was the counterfactual we were running the experiment against. Our results can only generalize to the effects of a change given that the treated cover is compared against other Colored, Regular size image covers. We argue that this is actually not much of an issue, as if one browses the Amazon kindle ebook store, the majority of covers that you'll see fall under this category. Finally, one more possible source of error could come in from the lineup design. Despite the decent performance of certain genre

lineups in our pilot survey, there were two lineups where our control cover was chosen over 50% of the time, which is not ideal as we want the base choice rates to be evenly split (or close to it).

There are a couple of improvements that could enhance this design in future iterations. The first would be to simply collect larger sample sizes, both for the precision of estimates and overall power. Another version would be to include more versions of lineups, to expand the counterfactual. Finally, the ideal way of testing this experiment would be through A/B testing on the Amazon website, though this is likely infeasible and as such a survey should be a good enough proxy.