Dating Experiment

Kalvin, Dan, and Jennifer

December 3, 2017

Abstract

The idea of self-presentation at the early stages of relationships has been well studied in the realms of human interactions and electronic communications. Our experiment continues this research in this domain with a focus on the impact of self-presentation of education attainment on online dating success on Tinder. Our research finds that despite anectdotal claims that females with higher education experienced greater difficulty in finding suitors, females presenting a medical degree on their Tinder profile experienced a 6% increase in match rates over those without any education listed. Women who presented a PhD on their Tinder profile experienced no statistically significant impact in match rates.

1 Introduction

Since its release in September 2012, Tinder has experienced a meteoric rise from TechCrunch's Crunchie Award for "Best New Startup of 2013" [1] to becoming a world-wide dating application with over 50 million users.[2] As of April 2015, Tinder users swipe through 1.6 billion profiles and make more than 26 million matches per day.[3] The application is most popular with young adults with 83% of the user base being under the age of 35.[4] Despite the immense popularity of the platform, there are very few published studies regarding factors that impact Tinder match rate success. Our research begins with a brief overview of the use of technology in dating and builds upon prior research regarding self-presentation in the early stages of relationships. Specifically, we study the impact of education attainment on Tinder match rate success.

To achieve and contextualize the goal of our study, our paper begins with a brief history of online dating followed by a review of prior studies in this arena. Our research design and data collection processes are discussed along with challenges associated with the aforementioned data collection on Tinder. After our results are analyzed, we propose ways in which future studies may improve upon our results and add robustness to this field of research.

1.1 Brief History of Online Dating

While many believe that online dating is a new phenomenon propelled by the launch of Match.com in 1995, the true conception of the idea goes back hundreds of years. In fact, the first known use of a person using new communication tools to find love and companionship was in 1695 when a 30-year-old British bachelor placed a personal ad seeking "some good young gentlewoman with the fortune of 3,000 pounds or thereabouts."[5] The movement began with the upper echelons of society and moved to the general populous in the mid 1800s, however it was still viewed as deviant behavior.[6] The expansions of the American frontier further popularized personal ads, however, they became mainstream as lonely World War I soldiers sought female pen pals.[7]

As technology advanced through the mid-twentieth century, dating services began to develop sophisticated methodology for connecting potential suitors. In 1965, a group of Harvard students launched Operation Match, a $3 dating service in which users would submit paper questionnaires that would be processed on an

[1] https://techcrunch.com/video/tinder-wins-best-new-startup-of-2013-crunchies-awards-2013/518118930/
[2] https://en.wikipedia.org/wiki/Tinder_(app)
[3] http://mashable.com/2015/04/15/coachella-tinder-usage-sky-rockets/#1PeiJxGLBmqU
[4] http://www.businessinsider.com/a-lot-of-people-on-tinder-arent-actually-single-2015-5
[5] https://www.huffingtonpost.com/susie-lee/timeline-online-dating-fr_b_9228040.html
[6] Id.
[7] Id.

IBM 1401 mainframe computer. The computer matched the users to five potential matches and the results were returned by mail.[8] By the fall of 1965, the business had tens of thousands of users and offices throughout the country.[9]

The 1980s brought about the rise of the VCR and the incredibly awkward era of video dating.[10] Suitors would submit a personal video to dating services describing themselves and what they were seeking in a partner. The service would then review the videos and manually match clients by common interest.[11]

The 1990s saw the development of the Internet with the world's first website and server going live at CERN on December 20, 1990.[12] Within five years, there were 25 million internet users in the United States- that number grew to over 270 million users by 2015.[13] The combination of the new communications technology with the timeless human search for relationships led to the launch of Match.com in 1995. The success of online dating was instantly obvious and in 2002, Wired Magazine prophetically stated, "Twenty years from now, the idea that someone looking for love won't look for it online will be silly, akin to skipping the card catalog to instead wander the stacks because the right books are found only by accident."[14] Since that article was written, more than 49 million Americans have tried online dating and approximately 17% of marriages in the last year were products of online dating.[15]

1.2 Prior Studies

Self-presentation and self-disclosure are well studied aspects of the relational development. In Erving Goffman's 1959 study "Presentation of Self In Everyday Life", Goffman explores the way in which individuals use strategic activities to "convey an impression to others which it is in his interest to convey."[16] These strategic activities are most important at the beginning of a relationship where an individual will "alter their self-presentational behavior in accordance with the values desired by the prospective date."[17] As Rowatt, Cunningham, & Druen noted, "men, but no women, chang[ed] their self-reported personality characteristics and physical appearance when they expected to meet a potential date. Additionally, their propensity to exaggerate these characteristics was enhanced when the method of meeting was via email."[18] Dating success, however, is not solely tied to an exaggerated presentation of the ideal self. Reis and Shaver's research indicated that the need to highlight one's positive attributes are experienced in tandem with the need to present' one's true self.[19] Interestingly, research by Gibbs, Ellison, and Heino indicates that 94% individuals disagreed that they had intentionally misrepresented themselves in their online communication and 87% felt that intentionally misrepresenting one's self was unacceptable.[20]

Somewhat contradicting prior research, 2010 research by Hitsch, Hortacsu, and Airely found no evidence for strategic behavior.[21] Their research found strong same-race preferences among users, which did not differ across age, income, or education levels. The study, however, did show gender differences in mate preferences, specifically that women had stronger preference for income and men had stronger preference for physical attributes.[22]

Our platform of interest, Tinder, has been the focus of a few published studies. Gatter & Hodkinson found that "despite common stereotypes about those who use different types of online dating. . . no differences were

[8] https://www.wired.com/2014/08/tech-time-warp-ibm-1401-dating/
[9] Id.
[10] http://www.businessinsider.com/found-footage-awkward-80s-video-dating-2015-12
[11] Id.
[12] https://ourworldindata.org/internet/
[13] Id.
[14] http://www.pbs.org/pov/xoxosms/infographic-technology-dating/
[15] https://www.statisticbrain.com/online-dating-statistics/
[16] http://clockwatching.net/~jimmy/eng101/articles/goffman_intro.pdf
[17] http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2006.00020.x/full
[18] https://pdfs.semanticscholar.org/0b3c/a535b2ca4bcb6506d8d579b61b70a41ac961.pdf
[19] http://www.affective-science.org/pubs/1998/LaurenFBPl1998.pdf
[20] http://journals.sagepub.com/doi/10.1177/0093650205285368
[21] https://link.springer.com/article/10.1007/s11129-010-9088-6
[22] Id.

found in motivations, suggesting that people may use both Online Dating Agencies and Tinder for similar reasons."[23] In specifying the motivations behind Tinder use, Sumter, Vandenbosch, and Ligtenberg found six motivating factors for use: Love, Casual Sex, Ease of Communication, Self-Worth Validation, Thrill of Excitement, and Trendiness.[24]

1.3 Research Question and Hypothesis

Building upon prior research, our study attempts to answer the question "How does success impact date-ability across the genders?" As success can be measured in a variety of ways, our study used self-presented education level attainment as a proxy. While our study is solely focused in online dating, we are encouraged by the findings of Gatter & Hodkinson that our findings may generalize to online and offline dating.

Based upon prior research indicating that women have stronger preference for income and men have stronger preference for physical attributes, we hypothesized that females would experience a smaller change in match rate for the same change in education as compared to males. Stated more directly: as the male profile's listed education changes from no education to MD, we would expect a greater change in match rate than in the female profile for the same increase in education.

$$H_0 : Male\ and\ female\ profile\ receive\ the\ same\ change\ in\ rate\ of\ matches\ with\ changes\ to\ education$$
$$H_1 : The\ female\ profile\ gets\ a\ lower\ change\ in\ match\ rate\ as\ education\ increases\ compared\ to\ the\ male\ profile$$

2 Research Design

Our experiment created two fictitious online profiles that mimicked actual user profiles on Tinder. The profiles were of a male and a female, each 29 years of age. The profiles included five pictures of each individual in various settings, the individual's name, age, level of education attainment, and a comment stating, "Moving in couple of weeks and looking forward to meeting new people!"

The treatment classes are different levels of educational attainment. Specifically, whether the individual has earned a PhD, MD, or a Bachelor's Degree. The control group displayed no educational attainment level in the profile.

Twenty nine years old was the age selected for the profiles as it gives the individual enough time to plausibly have completed any of the treatment levels of education. Further, it is an age where users were more likely to be looking for long-term relationships, rather than short-term flings. In each profile's filter settings, the age range of potential suitors was restricted to 24-34.

Based upon prior research, physical attributes tend to be highly significant in partner selection, especially for males, and therefore a single male and a single female were selected for the profile images in order to eliminate variation in match rate based upon aesthetics. Further, the images used were the same images the individuals had used in prior dating profiles to ensure validity of image selection.

The factitious profiles were displayed over a four week period in the eight largest cities in the United States: New York, Los Angeles, Chicago, Houston, Phoenix, Philadelphia, San Antonio, and San Diego.[25] It is important to note that neither individual whose pictures we used had previously used Tinder in any of the test cities, thus reducing the potential for any contamination or bias in the results. Each profile-educational level combination was displayed to 100 users in each city for a one week period of time in November-December, 2017. Test subjects were randomly selected by "swiping right" on every other profile displayed. It should be noted that Tinder profile viewership was restricted to only those who were chosen as test subjects. This eliminated the ability for a test subject to view the profile at multiple levels of education. The only potential

[23] http://www.tandfonline.com/doi/abs/10.1080/23311908.2016.1162414
[24] http://www.sciencedirect.com/science/article/pii/S0736585316301216
[25] https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population

for spillover would be by an individual having multiple profiles and having been selected twice as a test subject or by recognizing the profile while viewing Tinder on another user's account. Additonally the research team cross compared a unique identifier for profiles to remove any duplicates.

After one week, we returned to the profile to collect data on which users also "swiped right" on our test profile. This factorial design of two sexes, eight cities, and four education levels allowed for the analysis of different impacts of test conditions across sexes and locations.

Linearly, the research design was:

$$R - -0_1,\ RX_1 - 0_2,\ RX_2 - 0_3,\ RX_3 - 0_4$$

3 Randomization Engineering

Gathering a larger sample of profiles gave our analysis more power and confidence in the trends observed from the data. As the power calculations below demonstrate, if the study is conducted with individual suitor profiles as the measurement unit there would be a 95% probability of generating results that lead to the rejection of the null in the presence of a true treatment effect, an exceptionally high power. Alternatively, if the analysis considered the group level, then each city(8), each treatment(4) and each gender (2) would result in a much smaller n of 64 and the power in that experiment would be about 8%. The larger the sample size is the more robust the experiment is at identifying subtle differences in the data.

Taken from P93 in G&G

$$\beta = \Phi\left(\left(\frac{|\mu_t - \mu_c|\sqrt{N}}{2\sigma}\right) - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)$$

Applying to our data:

```
# Parameters
alpha = 0.05
mu_c = 100
mu_t = 102
sigma = 15

# calc final - individual level
n_1 = 3000
power_indiv = pnorm(((abs(mu_t - mu_c)) * sqrt(n_1))/(2 * sigma) -
    qnorm(1 - (alpha/2)))
power_indiv
```

```
## [1] 0.9546312
```

```
# calc final - city level (8 cities, M/F, 4 treatments)
n_2 = 8 * 2 * 4
power_city = pnorm(((abs(mu_t - mu_c)) * sqrt(n_2))/(2 * sigma) -
    qnorm(1 - (alpha/2)))
power_city
```

```
## [1] 0.07684319
```

The pilot demonstrated that mannually copying and pasting the details from each suitor's profile prohibited collecting a larger number of profiles. It took about three hours to gather four hundred profiles, meaning just over 2 minutes per profile. The use of automation was a very effective tool that enabled the collection of over five thousand suitor profiles in a four week period, which would have taken one hundred eight five hours by hand. Luckily, the website version of Tinder exactly matched the mobile phone version, greatly streamlining and standardizing the data collection approach.

The research team wrote scripts in AutoHotKey (PC) and Maestro (the MAC equivalent) that control the operating system of the computer. The code would open up Tinder and navigate to developers' tools where the website's Javascript rendered source code (the HTML behind the webpage) was available. The source code listed the suitors name, age, a unique photo source path, and two details about the profile. After collecting and storing the source code in the appropriate folder that designated the city and treatment level, the code would automatically right swipe on the profile collect.

Tinder restricts users from seeing repeat profiles to prevent communication between people who previously interacted unsuccessfully, making the trials independent. Each week a new randomly selected batch of ladies and gentlemen were swiped on and put into the experiment. Any user of the Tinder application is subject to the behind the scences algorithm that serves up potential profiles; both the test male and female profiles, along with the profiles of the general Tinder user.

To further enhance the level of randomness within treatment assignment, every other profile was systematically rejected from the experiment. Given more time, the treatments could be delivered in alternative orders to help determine if the algorithm modifies the suitor population based on features of the profile. As the experiment was constrained to a five week window, the collection of covariate distributions augments the research team's understanding of the mechanisms driving tinder matches. For example, to help demonstrate the balance in treatment assignment, the distribution of suitor ages for each treatment condition are displayed in figures 1.1-1.4 below.

Fig. 1.1: Suitor Ages for the No Education Treatment

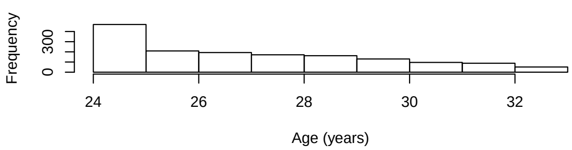


Fig. 1.2: Suitor Ages for the BS Treatment

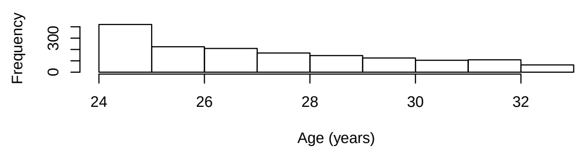


Fig. 1.3: Suitor Ages for the MD Treatment

Fig. 1.4: Suitor Ages for the PhD Treatment

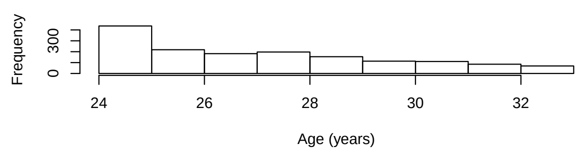


Fig. 2.1: Distribution of Male Suitor Ages

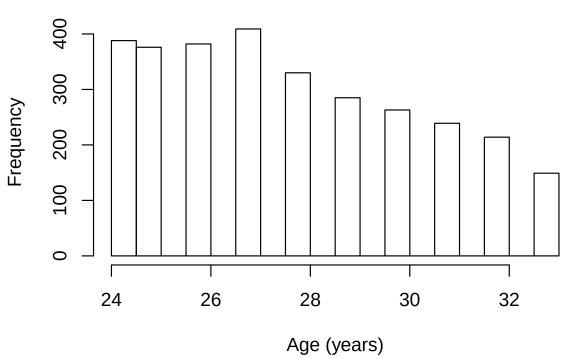
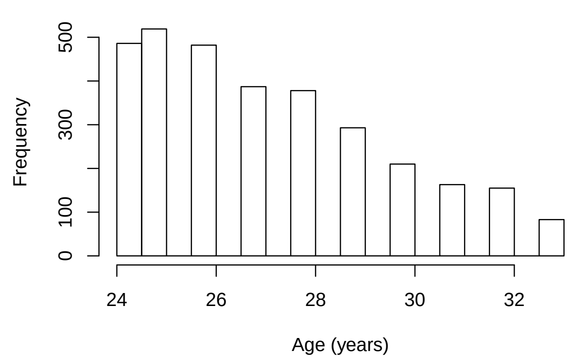


Fig. 2.2: Distribution of Female Suitor Ages



It is interesting to see that the age distribution has a right skew for all treatment conditions, and it is unclear if depending on Tinder to select suitors is truly random– it could be the case that Tinder selects suitors based on the age or age filter set in the test profile, that older suitors were more likely to artifically reduce their age (paid accounts for those age 29 and above were twice as expensive), or that the general population of Tinder users is centered around the mid-20s age range. But the consistency of this distribution across all conditions indicates that there was indeed an equal probability of receiving a given treatment condition, at least amongst the suitors selected by Tinder.

The age distribution of suitors to the Male profile (female suitors) does seem to be centered approximately 2 years younger than that of suitors to the Female profile (shown in Figures 2.1-2.2), and the skew appears to increase slightly with increasing education level. However, statistical tests for the difference in average age by test condition yields a significant difference only between the male and female conditions (shown in Table 1 below), and its effect size is small– approximately half a year. Furthermore, since the treatment assignment was randomized within each city and within each gender, this study is generally less concerned with covariate imbalances between cities and between genders.

Table 1: Comparison of Average Age

	Dependent variable:		
	Gender	Treatment	Location
	(1)	(2)	(3)
female	0.590 (0.066)***		
bs		0.171 (8.922)	
md		−0.091 (0.000)***	
phd		0.099 (0.000)***	
houston			−0.012 (0.132)
losangeles			0.143 (0.133)
newyork			0.280 (0.132)**
philadelphia			−0.009 (0.133)
phoenix			0.005 (0.131)
sanantonio			−0.072 (0.137)
sandiego			0.061 (0.134)
Constant	27.171 (0.045)***	27.414 (609.994)	27.411 (0.094)***
Note:			*p<0.1; **p<0.05; ***p<0.01

4 Experimental Materials (e.g. treatment materials)

Objectively measuring individual's success in life is ambiguous and difficult, but a reasonably good proxy for research purposes could be educational attainment, job title prestige or educational institution prestige. Below is a table of potential example degrees of variation for each of the proposed measures of success:

Educational Attainment	Job Title	Educational Prestige (nearest)
Advanced Degree (JD, PhD, MD, etc)	Doctor	IVY (Harvard)
College Degree	Teacher	State School (UMass)
Associate Degree	Social Worker	Community College (Bunker Hill CC)

As a team we considered the strengths and weakness of each respective success measure. We planned on customizing the education institution to reflect broadly known schools in the vicinity of each city in the experiment. To enhance consistency of school selection across geography, Barron's Ranking list and the US News school ranking would have been consulted. One of the drawbacks of educational prestige that could not be remedied, however, was that educational prestige is most often determined while the individual is in their early twenties. At such a young age the future of any person is extremely malleable and fertility concerns are not particularly potent. Generally, it also takes time after obtaining an undergraduate education to become particularly skilled or develop expertise or to develop a positive reputation in a domain of interest. Due to the inadequacy of using educational prestige as a proxy for success, the research team leaned away from using it as a treatment group in the study.

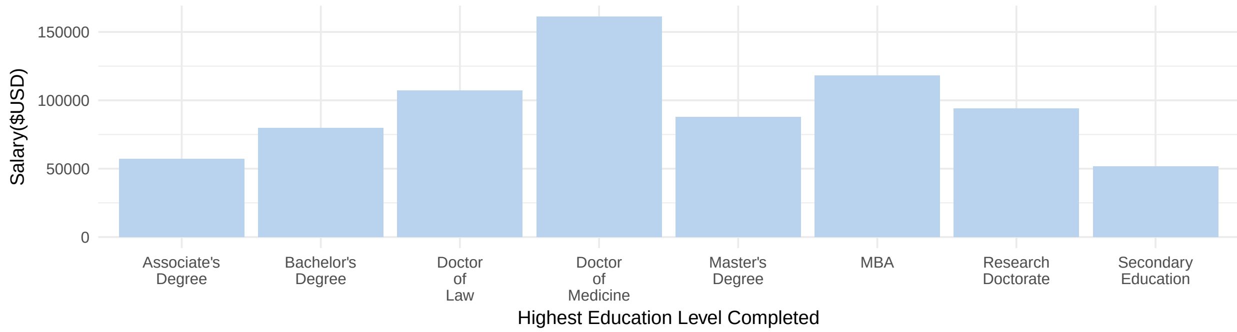
Professions then appeared as the better measure for success as individuals' career paths are more fleshed out a few years later. Most adults know that doctors make significantly more than teachers who make more than social workers, however, all three professions are in the business of helping people, and this could introduce potential bias. To elaborate on the point about bias, men might seek women with a nurturing profession (picturing the woman raising babies at home) and women might avoid men in a nurturing profession for fear of being replaced in traditional household roles. Additionally researchers at Microsoft recently published a paper demonstrating gender-based job title bias in the general public's vernacular, implying that using job title as a proxy for success potentially introduces bias into an experiment. [26] Ultimately, professions each carry their own reputation and predefined characteristics and controlling for that would be exceptionally difficult. Additionally, the novelty effect for women and men in unconventional roles also raised some concerns. Cross comparing average salary data from the Occupation Employment Statistics Survey and the Current Population Survey reveal that top paying professions are male dominated. For example, the Bureau of Labor Statistics data shows that 66% of dentists, 64% of lawyers, and 73% of CEOs are men.[27] [28] Meaning that any high paying job selected would be more common among male profiles than female profiles making an apples to apples cross gender comparison questionable. For these reasons, job title was disregarded as the the final success proxy.

The government and society have created very clear breaks in educational attainment that are known to all, uniform across geographic lines, and would be easy to administer for an experiment. Generally people perceive access to education in the United States as exceptionally balanced across gender, and bias concerns are mitigated. As the graph below indicates, on average there is a positive association between higher education and income levels, providing some validation for the positive life impact. After deciding to run the experiment with educational attainment as the measure of success, the degrees of treatment were identified as MD, BS, and PhD. The research team was particularly interested in the difference between the impact of having a Medical Degree versus having a Research Degree. Both degrees require about five years to obtain, however there is significant difference in expected earnings. Due to concerns regarding the realness of the profile and limited time, the team decided not to test the impact of the associate's degree because other profiles on Tinder did not prominently display an associate's degree. The control group would be a profile giving no information for the particular success measure. The treatments were delivered in the following non-monotonic order: MD, BS, no education, and PhD.

In order to effectively deliver the treatment variable, showing the profile to a limited audience of potential suitors was paramount. To elaborate on that point, should a potential suitor see the profile change from "MD" to "BS" that would not only ruin the data point, but also risk the injured party flagging the account to the administrators and terminating the experiment prematurely. Fortunately, the premium Tinder account enabled the research team to control who had access to the profile so only the one hundred individuals in each city assigned to that week's treatment were exposed- effectively elminating almost all potential spillover.

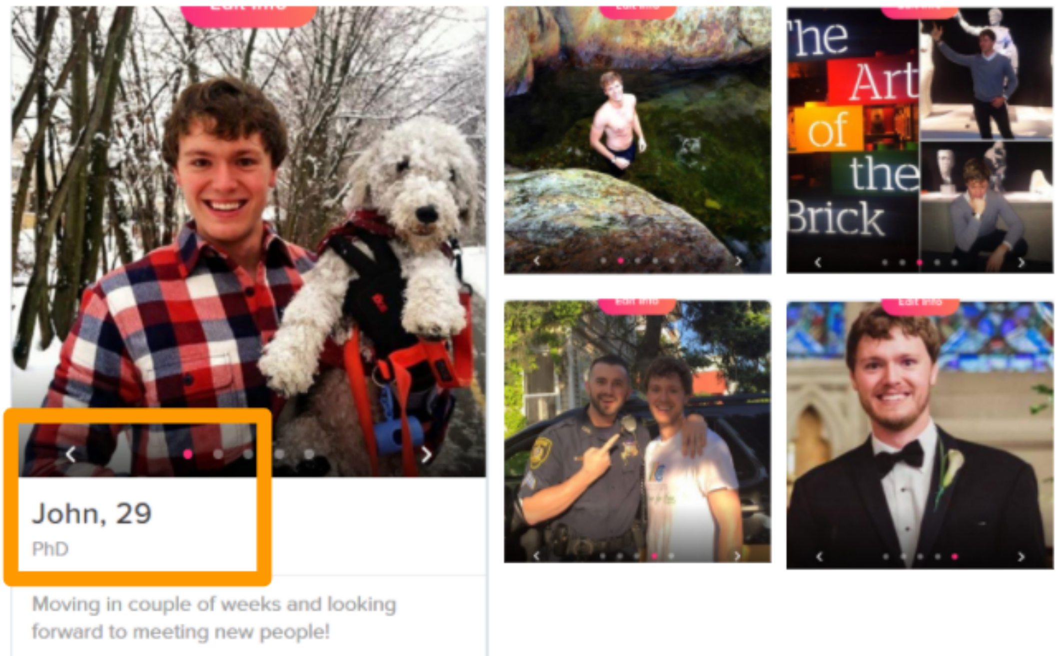
[26] Bolukbasi, 2016
[27] https://www.bls.gov/oes/tables.htm
[28] https://www.bls.gov/cps/tables.htm

Fig. 3.1: Salary by Education Level

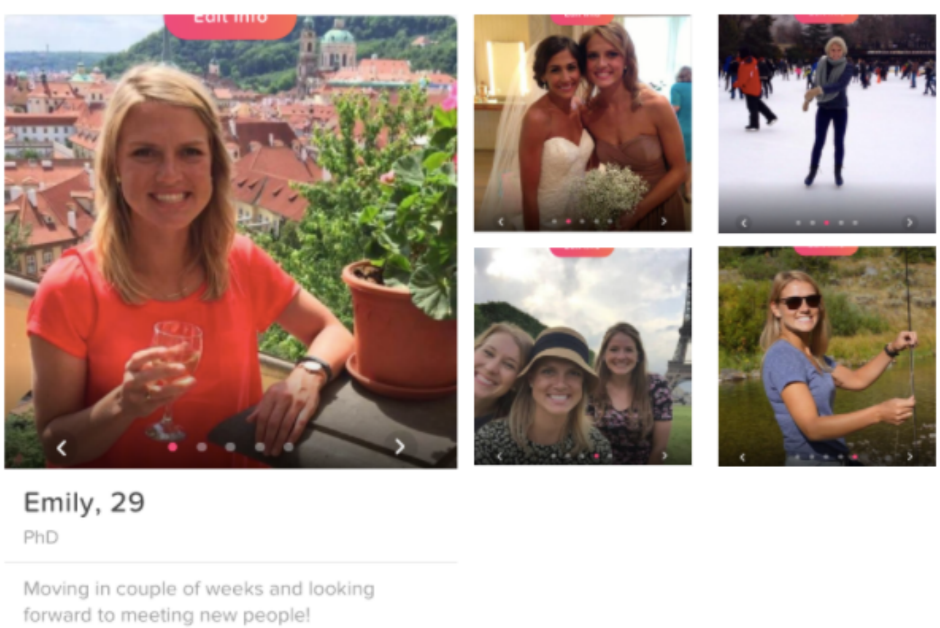


The potential suitors were shown a profile that either read: “MD”, “PhD”, “BS” or the field to provide education information was left blank. A sample real profile can be seen below; along with an overview of the male and female profiles used in the experiment.

Treatment



Female Profile



5 Measurement of Variables

As discussed previously, the treatment in this experiment is the test profile's exposure of a given education level to a potential suitor, and the outcome is whether or not the suitor matches with the test profile. When applying an education level treatment to the test profile, the measured test subject covariates include its age, sex, and location, and the measured suitor covariates include age, sex, whether or not school information is provided, whether or not job information is provided, and whether or not instagram photos are provided (and their count, if so). These variables are described in Table 3 below.

Table 3: Variable Descriptions

	Variable	Variable Type	Source	Description / Possible Values
1	test profile education level	treatment	test profile	4 possible values*: No education listed (control), Bachelor's, MD, PhD
2	suitor profile match	outcome	Tinder application	2 possible values: match or no match
3	test profile sex	covariate	test profile	2 possible values: female or male
4	test profile location	covariate	test profile	8 possible values: Chicago, Houston, Los Angeles, New York, Philadelphia, Phoenix, San Antonio, San Diego
5	suitor age	covariate	suitor source code	Due to the filters set in the test profiles, this variable can range from 24 - 34 years.
6	suitor school***	covariate	suitor source code	2 possible values: school information was detected or not
7	suitor job***	covariate	suitor source code	2 possible values: job information was detected or not
8	suitor instagram***	covariate	suitor source code	2 possible values: suitor profile included an instagram link or not
9	number of instagram photos***	covariate	suitor source code	if a suitor profile contained a link to an instagram account, this variable is the number of photos in the account

* The initial experimental plan also included the 'Associates' education level as a treatment condition (see Section 4 for further details).

** The test profile age was an originally planned covariate, but due to time constraints, only one age was used for the entire experiment.

*** These variables are believed to be only partially observed– see Section 5.2 for further details.

5.1 Issues with Measurement of Outcome

After swiping right on a suitor, the Tinder application only provides notification if that suitor likes the test profile in return, and does not provide an indication of whether or not a suitor actually saw the profile or if the suitor disliked the test profile. Our experiment therefore contains some rate of non-compliance and we are unfortunately unable to determine the compliance rate– our estimated average treatment effect is thus the effect of the intent to treat.

The intent to treat effect is in fact the result of interest. In practical terms, it describes the effect on match rate a Tinder user should expect when listing a certain education level in his or her profile. The complier average causal effect, on the other hand, would describe the likelihood that someone swipes right after seeing a certain education level, which is not the research question under examination.

5.2 Issues with Measurement of Covariates

The structure of the relevant source code in the Tinder browser interface is such that it contains a link to a suitor's profile image (which we use as a unique identifier for that suitor), the suitor's name, the suitor's age, and finally a maximum of two additional details that may include any combination of school information, job information, and instagram information.

It is known that a Tinder profile may display all three of these pieces of information, but none of the source codes collected contained more than two details. It is highly unlikely that amongst the 6270 suitors encountered in the experiment, none displayed more than two of these details, so we believe that school information, job information, and instagram information are only partially observed, and without more in-depth examination of each suitor profile at the time of experiment execution, our analysis is unable to fully determine the education and job status/level, as well as instagram information, for each suitor. Due to time and manpower constraints, thorough examination of each profile was unfortunately not an option– we were required to use an automated swiping method, which acquired covariate information from source code. It is also known that a Tinder profile may additionally display the suitor's distance and favorite spotify songs, but none of this information was present in the suitor source codes acquired during the experiment, and these two covariates were thus unobserved.

The source code indicates which suitor details are the profile image, the name, and the age, but does not indicate which details are school, job, or instagram information. Our data collection process searches the text of these details to make a best guess as to what type of detail it is (instagram information is easily identified, but not school or job information). So for some of the school and job details that were collected, it could not be determined whether the detail represented education or employment in a university setting.

6 Experiment Results

6.1 Exporatory Data Analysis

The randomization of our treatment assignment was dependent on Tinder's selection of suitors. Since Tinder's algorithm for this selection is unknown and potentially complex, the team was interested in how the available covariates were balanced across test conditions.

6.1.1 Missing Instagram Information

Like school and job information, instagram information was missing from many of our suitor profiles. However, it was easily detected when present in the source code since the format for an instagram detail was consistent across suitor profiles (of the form "X Instagram Photos"). One technical detail of the structure of suitor source codes is that school and job information took priority– an instagram detail is only present when less than two school or job details are present. Therefore, whether or not a profile's source code contains an instagram detail could be used to represent the amount of information a suitor chose to include in his or her profile– the missingness of instagram details is thus examined across experimental conditions. Table 4 below shows the count of profiles with instagram information detected between the male and female test profiles.

Table 4: Profile Instagram Data by Gender

	Gender	Total Profiles	Profiles with Instagram	% Missing Instagram Info
1	Female	3096	439	85.8204134366925
2	Male	3174	425	86.6099558916194

Table 5 below shows the count of profiles with instagram information detected between all treatment conditions.

Table 5: Profile Instagram Data by Treatment Condition

	Education Level	Total Profiles	Profiles with Instagram	% Missing Instagram Info
1	No Education	1589	262	83.5116425424795
2	BS	1595	185	88.4012539184953
3	MD	1508	213	85.8753315649867
4	PhD	1578	204	87.0722433460076

Table 6 below shows the count of profiles with instagram information detected between all testing locations.

Table 6: Profile Instagram Data by City

	City	Total Profiles	Profiles with Instagram	% Missing Instagram Info
1	Chicago	785	90	88.5350318471338
2	Houston	793	91	88.5245901639344
3	Los Angeles	793	156	80.327868852459
4	New York	793	109	86.2547288776797
5	Phildelphia	777	108	86.1003861003861
6	Phoenix	794	106	86.6498740554156
7	San Antonio	740	80	89.1891891891892
8	San Diego	795	124	84.4025157232704

Table 7 below shows the results of the tests for a difference in average missingness of instagram information, by gender, treatment condition, and location. While some differences are statistically significant, we do not consider them practically significant– the missingess of instagram information seems to be balanced across test conditions.

Table 7: Comparison of IG Missingness

	Dependent variable:		
	Gender	Treatment	Location
	(1)	(2)	(3)
female	0.008 (0.009)		
bs		-0.049 (0.012)***	
md		-0.024 (0.013)*	
phd		-0.036 (0.013)***	
houston			0.0001 (0.016)
losangeles			0.082 (0.018)***
newyork			0.023 (0.017)
philadelphia			0.024 (0.017)
phoenix			0.019 (0.017)
sanantonio			-0.007 (0.016)
sandiego			0.041 (0.017)**
Constant	0.134 (0.006)***	0.165 (0.009)***	0.115 (0.011)***

Note: *p<0.1; **p<0.05; ***p<0.01

6.1.2 Missing Age Values

Table 8 below shows the count of profiles that contain the suitor's age, between males and females.

Table 8: Profile Age Data by Gender

	Gender	Total Profiles	Profiles with Age	% Missing Age
1	Female	3096	3035	1.97028423772609
2	Male	3174	3156	0.567107750472595

Table 9 below shows the count of profiles that contain the suitor's age, between all treatment conditions.

Table 9: Profile Age Data by Education Level

	Education Level	Total Profiles	Profiles with Age	% Missing Age
1	No Education	1589	1569	1.25865324103209
2	BS	1595	1571	1.50470219435737
3	MD	1508	1487	1.39257294429708
4	PhD	1578	1564	0.887198986058302

Table 10 below shows the count of profiles that contain the suitor's age, between all testing locations.

Table 10: Profile Age Data by City

	City	Total Profiles	Profiles with Age	% Missing Age
1	Chicago	785	777	1.01910828025478
2	Houston	793	783	1.2610340479193
3	Los Angeles	793	775	2.26986128625473
4	New York	793	775	2.26986128625473
5	Phildelphia	777	772	0.643500643500639
6	Phoenix	794	791	0.377833753148615
7	San Antonio	740	735	0.67567567567568
8	San Diego	795	783	1.50943396226415

Table 11 below shows the results of the tests for a difference in average missingness of the suitor's age, by gender, treatment condition, and location. While some differences are statistically significant (between males and females, for example), we do not consider them practically significant. It appears that the missingess of age information is also balanced across test conditions.

The important result of the exploratory data analysis is that no significant difference in covariates was found between treatments, so the experiment still yields an apples-to-apples comparison.

6.2 Results

As a research team we were extremely excited to see any kind of results. In the pilot and in the first round of testing we struggled with a low overall match rate (0/400 swipes and 5/1,600 respectively) and therefore we could not perform initial calculations to evaluate the treatments' effects. The primary concern was that the typical Tinder user made decisions based solely on physical attractiveness and there would be no variance across our treatment variable. It wasn't until week four of swiping that we could compare the numbers across treatments. Unfortunately including the data points from the pilot and the second week of the project would have introduced too much bias because the profiles were modified thereafter; so that data was excluded from the analysis.

After enhancing the profiles in week three more matches finally came in. The two tables below summarize the match distribution across key variables. The first table highlights the extreme imbalance across gender

Table 11: Comparison of Age Missingness

	Dependent variable:		
	Gender	Treatment	Location
	(1)	(2)	(3)
female	−0.014 (0.003)***		
bs		−0.002 (0.004)	
md		−0.001 (0.004)	
phd		0.004 (0.004)	
houston			−0.002 (0.005)
losangeles			−0.013 (0.006)*
newyork			−0.013 (0.006)*
philadelphia			0.004 (0.005)
phoenix			0.006 (0.004)
sanantonio			0.003 (0.005)
sandiego			−0.005 (0.006)
Constant	0.994 (0.001)***	0.987 (0.003)***	0.990 (0.004)***
Note:			*p<0.1; **p<0.05; ***p<0.01

where the female profile received almost forty four times the number of matches compared to the male profile. The second table displays the match rate across the treatment groups and that the higher education levels were associated with higher match rates. The high level figures intimate a causal treatment effect, but more robust calculations must be performed before that claim can be made.

High Level Overview of Outcomes

```
##        Female Indicator
## Matches    0    1  Sum
##     0   3161 2524 5685
##     1     13  571  584
##     Sum 3174 3095 6269
```

Table 12: Profile Matches by Sex

		Female Profile	Male Profile
1	Matches	571	13
2	Total Suitors	3096	3174

```
##        Treatment
## Matches   BS Control   MD  PhD  Sum
##     0   1468    1454 1330 1433 5685
##     1    127     135  177  145  584
##     Sum 1595    1589 1507 1578 6269
```

Table 13: Profile Matches by Education

		No Education (Control)	BS	MD	PhD
1	Matches	135	127	177	145
2	Total Suitors	1589	1595	1508	1578

Before building sophisticated models, the research team paused to consider the power of the experiment.

While the delta between the treatment and control means was very small, the large sample size engenders a high powered experiment.

Actual Power of the experiment:

```
# Parameters
alpha = 0.05
mu_c = mean(df_female$Matched[df_female$noedu == 1])
mu_c
```

```
## [1] 0.160414
```

```
mu_t = mean((df_female$Matched[df_female$md == 1]), na.rm = T)
mu_t
```

```
## [1] 0.2289326
```

```
sigma = sd((df_female$Matched[df_female$md == 1]), na.rm = T)

# calc final - individual level
n_1 = 3034
power_indiv = pnorm(((abs(mu_t - mu_c)) * sqrt(n_1))/(2 * sigma) -
    qnorm(1 - (alpha/2)))
power_indiv
```

```
## [1] 0.9942697
```

7 Modeling Choices:

To better understand the predictive power and magnitude of each variable the research team exclusively leveraged regression models. This type of model is known for being highly interpretable and there are many different varieties of regression available. For analyzing our experiment we selected both linear and logistic model implementations.

7.1 Linear Regression:

The first model shown below is a standard ordinary least squares regression where the outcome variable is the linear measurement of the match variable. As the regression table shows, we regressed the match indicator against each of the treatments, the female indicator and we also tested for interactions between the treatment variables and gender (equation shown below). The highest level of education, MD, has a positive and significant ATE of 0.06 for the female account meaning that the number of matches actually increased by 6% when the treatment was applied compared to when the control with no education information was applied. Prior observational research from match.com indicated that income levels for females was not significant and it is surprising to see a different pattern. The regression shows there were no other statistically significant variables so the other treatments, BS and PhD, did not provide evidence of causal relationships.

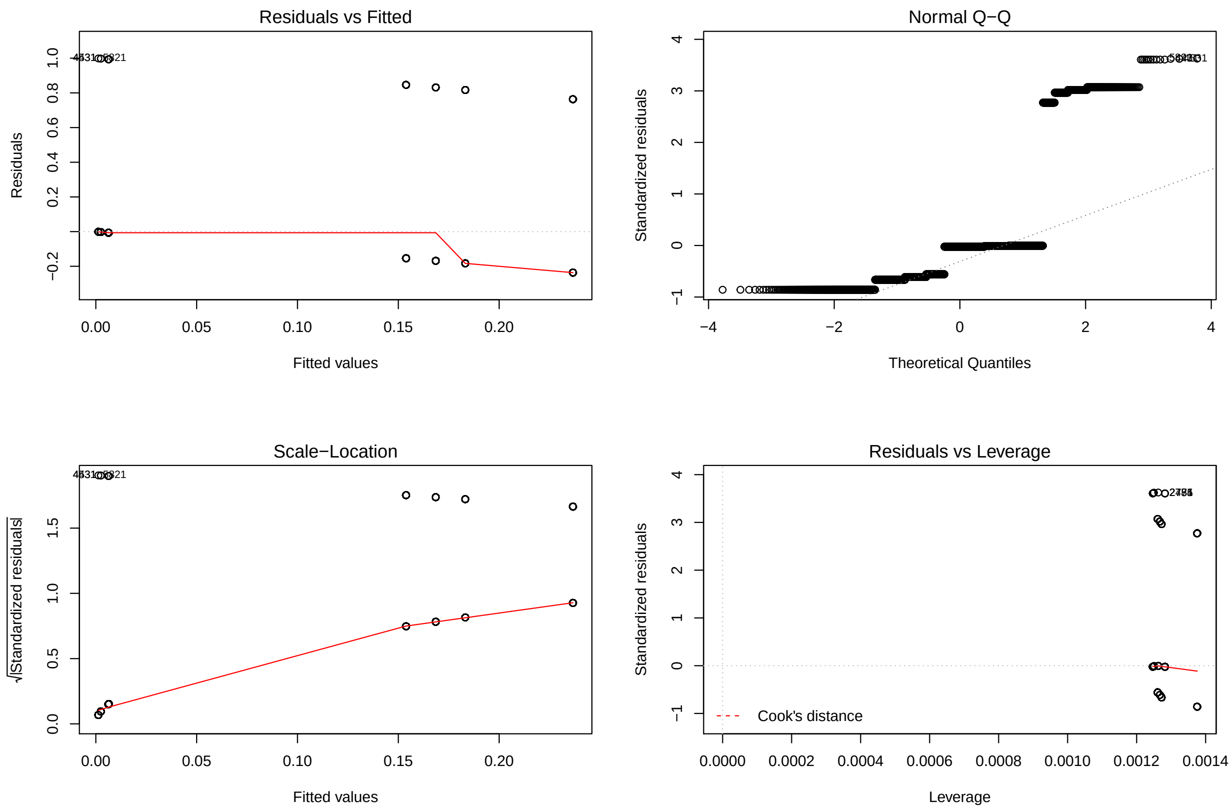
Not surprisingly, with only thirteen male matches and over five hundred female matches the data results in a highly significant effect on the female indicator variable where the female match rate is sixteen percent points higher. The research team did not want to draw too many conclusions from the male matched records because any findings would just be p-hacking. In the next series of regressions, all the male records were removed from the analysis and the sample size drops from n = 6,269 to n = 3,034. Additionally, the seventy nine suitor profiles without age information were also removed, but as discussed in section 6.1.2, there is no bias introduced with this action.

7.1.1 Models

$y_i = \beta_0 + \beta_1 Z_i + e_i Y_{Matches} = \beta_0 + \beta_1 MD + \beta_2 PhD + \beta_3 BS + \beta_4 female + \beta_5 female * MD + \beta_6 female * PhD + \beta_7 female * BS + e_i$

```
##
## Call:
## lm(formula = Matched ~ md + bs + phd + female + female * md +
##     female * bs + female * phd, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.23659 -0.16857 -0.00641 -0.00250  0.99874
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.002500   0.009745   0.257  0.79754
## md           0.003910   0.013870   0.282  0.77801
## bs           0.003734   0.013773   0.271  0.78629
## phd         -0.001237   0.013816  -0.090  0.92864
## female       0.166068   0.013830  12.008  < 2e-16 ***
## md:female    0.064111   0.019828   3.233  0.00123 **
## bs:female   -0.018456   0.019539  -0.945  0.34492
## phd:female   0.015876   0.019592   0.810  0.41779
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2756 on 6261 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1018, Adjusted R-squared:  0.1008
## F-statistic: 101.4 on 7 and 6261 DF,  p-value: < 2.2e-16
```

The research team did have some concerns regarding the validity of a linear regression with a binary outcome variable. With any regression model it is important to consider the BLUE assumptions, which is exactly what the plots below explore:



To address the violated homoskedacity seen the in residuals vs fitted and scale-location graphs above, gender model 2 uses robust standard errors, but the coefficient estimates are still the same. More importantly, the female MD treatment is still statistically significant even with the wider confidence interval generated by using robust standard errors.

```
library(sandwich)
model_gender2 = coeftest(model_gender, vcov = vcovHC)
model_gender2
```

```
##
## t test of coefficients:
##
##                 Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   0.0025000  0.0017678  1.4142  0.157349
## md            0.0039103  0.0033633  1.1626  0.245021
## bs            0.0037344  0.0032969  1.1327  0.257379
## phd          -0.0012374  0.0021728 -0.5695  0.569056
## female        0.1660678  0.0134614 12.3366 < 2.2e-16 ***
## md:female     0.0641107  0.0209408  3.0615  0.002212 **
## bs:female    -0.0184561  0.0188023 -0.9816  0.326342
## phd:female    0.0158757  0.0193307  0.8213  0.411525
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results from the prior page are included as the base model in column one, but the research team wanted to add additional covariates to the model little by little to fully understand their impact. The first variable considered was age; as the suitor profiles increase in age, there is a significant but small decrease in the match rate of -0.5% seen in the model results in column two. Column three considers a model with location covariates, but unlike in the baseline model, there is no control for location that would be the obvious choice to leave out. Consequently, the coefficients for location are compared to the values for Chicago. The female profile received significantly more matches in LA than any other city, and Chicago had slightly more matches than the remaining five cities, as indicated by the negative coefficients; however these deltas were not all statistically significant. Column four considers the influence of a suitor sharing his or her job title or school, but the regression shows the additional covariates do not add bias. Finally column five details a comprehensive model that amalgamates all the covariates. The most important takeaway from these more detailed models is that the MD treatment is highly significant across in every instance.

Table 14: Comparison of treatments

	Dependent variable:				
	Base	Age	Location	Details	All
	(1)	(2)	(3)	(4)	(5)
md	0.004 (0.003)	0.069 (0.021)***	0.064 (0.020)***	0.068 (0.021)***	0.064 (0.020)***
bs	0.004 (0.003)	−0.010 (0.018)	−0.013 (0.018)	−0.013 (0.018)	−0.011 (0.018)
phd	−0.001 (0.002)	0.026 (0.019)	0.023 (0.019)	0.024 (0.019)	0.025 (0.019)
female	0.166 (0.013)***				
md:female	0.064 (0.021)***				
bs:female	−0.018 (0.019)				
phd:female	0.016 (0.019)				
age		−0.005 (0.003)**			−0.007 (0.003)***
losangeles			0.168 (0.032)***		0.170 (0.032)***
houston			−0.067 (0.026)***		−0.068 (0.026)***
newyork			−0.012 (0.028)		−0.009 (0.028)
phoenix			−0.056 (0.026)**		−0.057 (0.026)**
sandiego			−0.001 (0.028)		−0.002 (0.028)
sanantonio			−0.051 (0.027)*		−0.052 (0.028)*
philadelphia			−0.047 (0.027)*		−0.049 (0.027)*
school				−0.009 (0.014)	−0.003 (0.014)
job				0.006 (0.017)	0.007 (0.017)
Constant	0.002 (0.002)	0.306 (0.072)***	0.170 (0.023)***	0.164 (0.015)***	0.354 (0.075)***

Note: *p<0.1; **p<0.05; ***p<0.01

7.2 Logistic Regression:

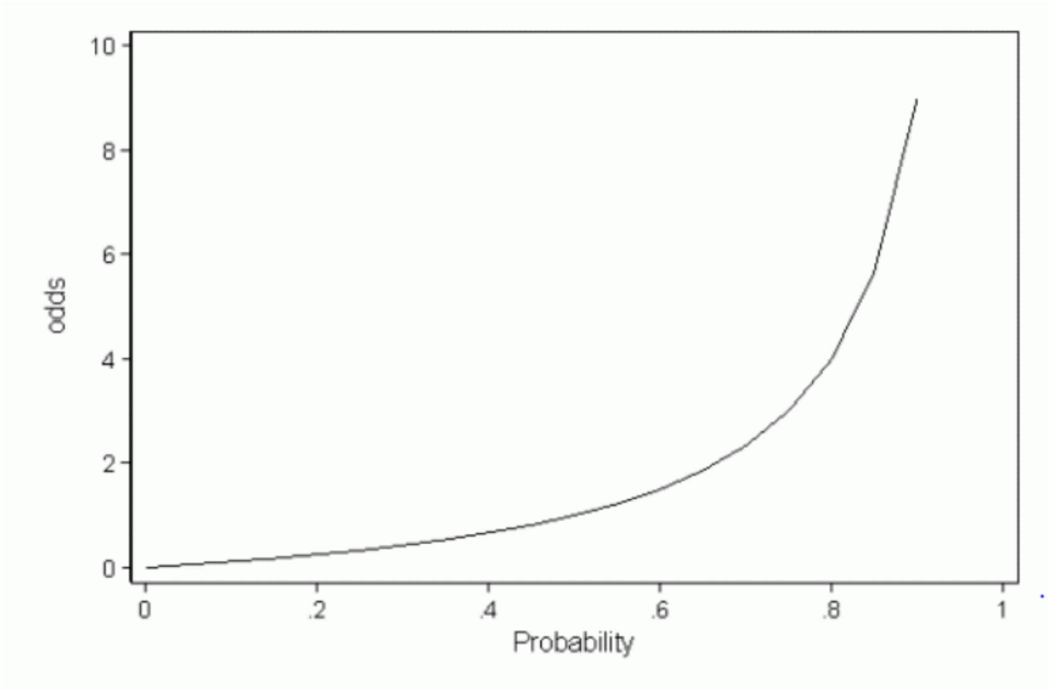
After reviewing the results of five different regression models, the research team observed that all of the coefficients were very low and we could not shake the feeling that something was off. After some research we found that models with probabilities close to 0 and 1 are prime candidates for logistic regression. What does it mean for a probability to be close to 0 or 1? If you take the results from the baseline model and plug in a test case, for this example consider a male with an MD, the probability of receiving a match is 0.0064 (calculation built off of baseline regression) or practically 0.

$$p = 0.0025 + 0.0039 * 1$$
$$p = 0.0064$$

We updated our modeling technique to logistic regression to better fit the data. The regression output below shows a consistently statistically significant female indicator, but instead of the model coefficients showing percentage point difference in probability, the logistic model outputs a log odds ratio. This measure is very difficult to interpert so we exponentiated the coefficients and interpreted them as regular odds-ratios. The coefficient 4.39 for the female indicator actually means the probability of having a match based on that variable alone (see table converting odds ratios to probability for support) is over 90% and this magnitude of influence is much more practically significant than the 16% seen in the Linear model. The other important call out is that the treatment variable interacted with the female indicator is no longer statistically significant. While the regression coefficient now looks negative, it must be converted to an odds ratio. After that transformation, the probability of a match for a female with an MD is actually a positive 35% probability. Based on our original raw numbers we know that the observed MD match rate is closer to 20% and our new model also has a very large weight on the intercept, indicating there is more analysis needed.

```
##
## Call:
## glm(formula = Matched ~ md + bs + phd + female + female * md +
##     female * bs + female * phd, family = "binomial", data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7348  -0.6076  -0.1134  -0.0708   3.6536
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -5.9890     0.7080  -8.459  < 2e-16 ***
## md            0.9455     0.8382   1.128    0.259
## bs            0.9175     0.8382   1.095    0.274
## phd          -0.6843     1.2258  -0.558    0.577
## female        4.3931     0.7144   6.150 7.76e-10 ***
## md:female    -0.5212     0.8481  -0.615    0.539
## bs:female    -1.0265     0.8493  -1.209    0.227
## phd:female    0.7854     1.2329   0.637    0.524
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3884.0  on 6268  degrees of freedom
## Residual deviance: 3105.3  on 6261  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 3121.3
##
## Number of Fisher Scoring iterations: 9
```

```
##  (Intercept)           md           bs          phd       female
##  0.002506266  2.574193548  2.503136763  0.504424779 80.894817073
##    md:female    bs:female   phd:female
##  0.593808575  0.358265452  2.193225315
```



To enable comparison with the linear regression model, we dropped the male data and explored the impact of additional covariates in the tables below. Now if we consider the impact of the MD treatment we see once again that it is significant in all of the covariate models, and the odds ratio hovers around 1.55 or ~ 60% probability. Despite adding several different coefficients, our models still have highly significant intercepts, high AIC values, and a potentially over-stated treatment effect. What the logistic regressions imply is that the model may have explanatory power, but does not have much predictive power. While we are confident in the causal relationship seen, additional endogenous variables likely exist.

```
##  (Intercept)          md          bs         phd      female
##  0.002506266  2.574193548  2.503136763  0.504424779 80.894817073
##    md:female   bs:female  phd:female
##  0.593808575  0.358265452  2.193225315
```

```
## (Intercept)          md          bs         phd         age
##   0.5209477   1.5584427   0.9218764   1.1964360   0.9642122
```

```
##  (Intercept)           md           bs          phd  losangeles
##    0.2028092    1.5307286    0.9059346    1.1861079   2.4024543
##      houston      newyork      phoenix     sandiego  sanantonio
##    0.5932723    0.9220836    0.6566167    0.9927362   0.6770610
## philadelphia
##    0.7071552
```

```
## (Intercept)          md          bs         phd      school         job
##   0.1956520   1.5516287   0.9073514   1.1819531   0.9432156   1.0401049
```

```
##  (Intercept)           md           bs          phd         age
##    0.7622943    1.5351761    0.9133814    1.2002779   0.9531272
##   losangeles      houston      newyork      phoenix     sandiego
##    2.4366636    0.5902777    0.9391539    0.6494332   0.9858438
##   sanantonio philadelphia       school          job
##    0.6707464    0.6976407    0.9779182    1.0504544
```

Table 15: Comparison of treatments

	Dependent variable:				
	Base	Age	Matched Location	Details	All
	(1)	(2)	(3)	(4)	(5)
md	0.946 (0.838)	0.444 (0.133)***	0.426 (0.135)***	0.439 (0.133)***	0.429 (0.136)***
bs	0.918 (0.838)	−0.081 (0.141)	−0.099 (0.143)	−0.097 (0.141)	−0.091 (0.144)
phd	−0.684 (1.226)	0.179 (0.135)	0.171 (0.137)	0.167 (0.135)	0.183 (0.137)
female	4.393 (0.714)***				
md:female	−0.521 (0.848)				
bs:female	−1.026 (0.849)				
phd:female	0.785 (1.233)				
age		−0.036 (0.018)**			−0.048 (0.018)***
losangeles			0.876 (0.169)***		0.891 (0.171)***
houston			−0.522 (0.203)**		−0.527 (0.204)***
newyork			−0.081 (0.187)		−0.063 (0.188)
phoenix			−0.421 (0.198)**		−0.432 (0.199)**
sandiego			−0.007 (0.185)		−0.014 (0.186)
sanantonio			−0.390 (0.210)*		−0.399 (0.211)*
philadelphia			−0.347 (0.196)*		−0.360 (0.196)*
school				−0.058 (0.096)	−0.022 (0.099)
job				0.039 (0.116)	0.049 (0.118)
Constant	−5.989 (0.708)***	−0.652 (0.500)	−1.595 (0.158)***	−1.631 (0.113)***	−0.271 (0.531)
Observations	6,269	3,034	3,034	3,034	3,034
Log Likelihood	−1,552.673	−1,417.188	−1,373.097	−1,419.059	−1,369.560
Akaike Inf. Crit.	3,121.345	2,844.376	2,768.194	2,850.119	2,767.119
Note:					*p<0.1; **p<0.05; ***p<0.01

```
## Warning in cbind(gender, age, loc, details, all): number of rows of result
## is not a multiple of vector length (arg 1)

##                  gender       age       loc   details       all
## (Intercept) 0.002506266 0.5209477 0.2028092 0.1956520 0.7622943
## md          2.574193548 1.5584427 1.5307286 1.5516287 1.5351761
## bs          2.503136763 0.9218764 0.9059346 0.9073514 0.9133814
## phd         0.504424779 1.1964360 1.1861079 1.1819531 1.2002779
```

7.3 Randomization Inference

Our results are further confirmed by randomization inference under the sharp null hypothesis that treatments have no effect on any individual– in other words, this test assumes that the education level shown to each suitor had no effect on whether or not there was a match. The experiment was blocked on gender and location, so we simulate a new randomization of the treatment assignment within each combination of gender and location, and re-calculate the ATE under that treatment assignment. For each treatment, we repeat this procedure 10,000 times and compare the aggregated ATE's under simulation with our actual estimate in order to gain a sense of the likelihood of observing our estimates if the treatment did not actually have an effect.

Figures 4.1-4.3 show the distribution of the ATE (for each treatment) under the sharp null hypothesis, with a vertical line indicating the actual ATE estimate observed in the experiment. The percentage of ATEs from simulation that are at least as extreme as the observed ATE is the p-value under randomization inference.

Figure 4.2 shows the distribution of ATE of the MD treatment under the sharp null hypothesis

Fig. 4.1: ATE Distribution Under the Sharp Null (BS)

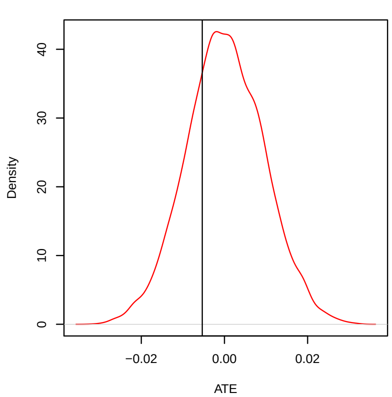


Fig. 4.2: ATE Distribution Under the Sharp Null (MD)

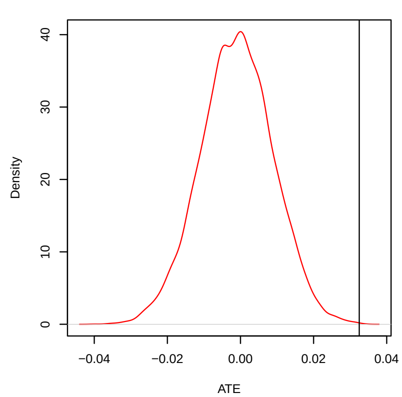
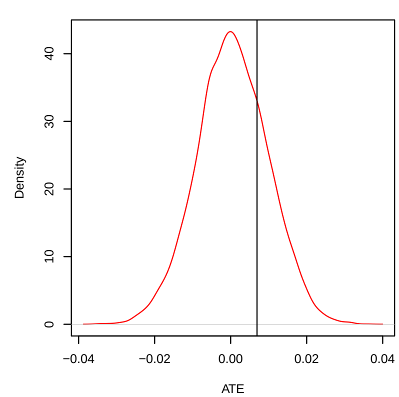


Fig. 4.3: ATE Distribution Under the Sharp Null (PhD)



For the BS treatment, randomization inference yields a p-value of 0.2512, which is not significant. The p-value for the PhD treatment is again not significant under randomization inference (0.2266). However, randomization inference shows a significant p-value of 10^{-4} for the MD treatment, which provides further confirmation of the results of our analysis.

8 Conclusion

Our results indicate that females with an MD presented in their dating profile receive a statistically significant increase in match rate of approximately 6.4% over females without an education level presented. Further, our study found that indicating a bachelor's degree or a PhD had no statistically significant impact on female match rate. These results are interesting as they contradict previous research that suggests males are less influenced by success factors in a partner than their female counterparts.

Despite these positive results, there are many changes that we would incorporate in further studies in this field. Tactically, we would put an emphasis on fully capturing all covariates in a suitor's profile. While this is not currently possible to automate, this could be performed manually. The study could also be improved by running the research in the target markets to remove any suspicion caused by test subjects being alerted that our factitious user was hundreds or thousands of miles away. Additionally, with increased technical resources, we would be interested in re-running this study for all profiles over the same time period to remove any variation caused by differences between weeks. Further studies would also be necessary to understand differences in the measuring period- i.e. would the results have changed if we allowed test subjects to match with our profiles over a two week period instead of only a one week period? Finally, we would further broaden our treatment levels to include other levels of education such as master's degrees, JDs, PharmDs, etc.

To improve the generalizability of these results, we would seek to also test this theory across different dating sites and services. Further, the scope of the study would need to broaden to include international subjects, as well as subjects in rural and suburban locations, and to incorporate demographic information regarding race, religion, and income. Testing on other platforms would also allow us to see if similar results are obtained in gay and lesbian online dating communities, which is another topic of interest to our team.

9 Works Consulted

"FAQ: How Do I Interpret Odds Ratios in Logistic Regression?" IDRE Stats, stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-interpret-odds-ratios-in-logistic-regression/.

"Linear vs. Logistic Probability Models: Which Is Better, and When?" Statistical Horizons, statisticalhorizons.com/linear-vs-logistic.

"Tables Created by BLS." U.S. Bureau of Labor Statistics, U.S. Bureau of Labor Statistics, www.bls.gov/oes/tables.htm.