How Likely Are You to Date Your Half-Sibling?

Nathan Holt
Kayleigh Wheeler
Mariya Belchikova
Erika Bliss

May 4, 2017

1 Introduction

30,000 to 60,000 children are estimated born through IVF every year in the United States. With the invention of in vitro fertilization (IVF), multiple children may have the same father and not even realize it. An anecdotal story shared about IVF regards two friends who underwent IVF realizing that their babies almost had the same father, as each's chosen sperm donor was the other's second choice for sperm donor. They later joked that if either of them had changed their minds, they could have been key players in the case of "accidental incest" [4]. In this paper, we attempt to quantify the probability of inadvertently dating your half-sibling due to the advent of in vitro fertilization.

2 Preliminary Assumptions and Simplifications

For simplicity, we look at the area of Silicon Valley (the setting of the aforementioned anecdotal story). This area has plenty of well-documented statistics regarding population, income, sperm donations, and other factors that lends itself well to determining the probability of dating your half-sibling.

We only consider half-siblings where both people are the result of in vitro fertilization. Furthermore, we only consider the half-siblings who unknowingly date each other. We do not account for fringe accounts of half-siblings dating by choice. We are also only accounting for sperm donations, not egg donations, as sperm donors are much more common than egg donors.

When discussing dating, we make a couple of assumptions. First, we assume that an eligible range for people to date is 18 to 64 years old. We believe it is reasonable, most people don't start seriously dating when they're very young or very old. Further, we assume that within that age range everyone is single and that each person is equally attracted to each other. Additionally, people date (on average) five people over the course of their lives. This is backed up by an article found on The Daily Mail [6].

Regarding women who elect for in vitro fertilization, we make the assumption that they are under 35 years old. This is supported by an article by AttainFertility [1].

Regarding sperm donation, sperm banks have certain regulations about the number of times a single donor can donate. For the sake of this paper (and to increase the overall chances of being related), we assume that each donor donates the maximum amount of sperm he's eligible to. We also make the assumption that all children born from sperm donations stay in the general location in which they were donated at. This means that none of these people are moving away and are all only eligible to date people in their area.

3 Mathematical Derivation of the Model

Our model consists of two parts. In the first part, we derive the chance that a single baby born through IVF in Silicon Valley is related to another baby born through IVF in Silicon Valley. The second part uses combinatorics to determine the probability that you would date your half sibling using the chance that you would be related, as found in the first part.

To begin to quantify the probability that you would be related to your half-sibling, we first sought to determine how many potential sperm donors there are in Silicon Valley by calculating the number of males in Silicon Valley that meet the eligibility requirements of one cyrobank. Using the number of sperm donors in UK, we scale the potential number of sperm donors in Silicon Valley to the actual number of sperm donors in Silicon Valley. By knowing the maximum number of sperm one donor can donate and the success rate of a sperm in the sperm bank becoming a fetus, we deduce the probability of being related to another baby born through IVF.

First, we obtained an estimate to the number of people in Silicon Valley. According to the Silicon Valley Institute for Regional Studies, there are three million residents in the Silicon Valley area [7]. We now attempt to quantify the number of possible sperm donors in the Silicon Valley area. According to California Cryobank, several requirements must be met in order to be a sperm donor [9]:

- The donor must be at least 5'9".
- The donor must be between 19 and 38 years old.
- The donor must be in a 4-year university or have a degree.
- The donor must not have had sex with men.
- The donor must be in good health.
- The donor must legally be allowed to work.

We will not attempt to quantify sexuality, health, or working status. Instead, we will assume that all males that satisfy height, age, and education requirements are eligible to become sperm donors.

According to the $2007-2008$ US Census, 52% of the Silicon Valley population are males [8]. This means that there are approximately $1,560,000$ males in the Silicon Valley area who we will consider as potential sperm donors.

When looking at the US Census data, we can also get a breakdown of the ages of males in California. From the statistics, we obtain that approximately 29% of citizens in California are between the ages of 20 to 39. We will assume this is roughly the same percentage of people between the ages of 19 and 38.

Next, based on Census data, we can conclude that the state of California has $39,144,818$ citizens total [2]. Thus, the percentage of people in California that live in Silicon Valley is $\frac{3000000}{39144818}$.

Multiplying the population of California, the percentage of people between 18 and 39, and density of Silicon Valley, we obtain a number of people in Silicon Valley that meet the age restriction and are male:

$$39144818 \times .29 \times \frac{3000000}{39144818} \times .52 = 452400$$

Thus, there are 452400 people in Silicon Valley that are between 18 and 39 and male.

Next, based on Census data, we obtain height demographics based on age in the Silicon Valley area. The data says that approximately 57.8% of people between 20 and 29 meet the height requirement, and 54.6% of people between 30 and 39 meet the height requirement [10]. We take the average of these two to get an approximate percentage of people that satisfy the height requirement. Approximately 56% of people between 20 and 39 meet the height requirement. Furthermore, we assume this is approximately the same percentage as people between the ages of 19 and 38.

When we multiply this by number of males between 20 and 39, we obtain that there are approximately 253344 males in Silicon Valley that meet both the height and age requirements.

According to the US Census, 32% of California citizens have a degree, and 29% of citizens are attending college currently [2]. We will assume this is approximately the same percentages as in Silicon Valley. Thus, we conclude that 61% of Silicon Valley citizens meet the education requirements required to be a sperm donor. If we multiply this by the number of males in Silicon Valley who meet the height and age requirements, we obtain the number of males in Silicon Valley who meet California Cryobank's height, age, and education restrictions:

$$253344 \times .61 = 154540$$

Thus, there are $154,540$ possible sperm donors which we will consider for the purpose of this paper.

Now that we have the number of possible sperm donors in the Silicon Valley area, we must approximate the number that actually become sperm donors.

Based on an article published in Metro, the UK was said to have 480 sperm donors in 2010 [5]. We will assume this is roughly the same amount as today. The UK has a population of 63.26 million, which we will chop this in half since we assume half of the UK population is male. Thus, this gives us a ratio of sperm donors to males in the UK:

$$\frac{480}{31630000} \approx 0.000015$$

We will use this as a rough estimate to gauge the number of donors in Silicon Valley. We will guesstimate that there are around 200 sperm donors in Silicon Valley. We obtained this guess by saying it needs to be less than 480 (the number of donors in the UK, where the population is much greater), but above 100.

Thus, we will say that Silicon Valley has approximately 200 sperm donors. This is how many we will assume moving forward.

Now we calculate, for one sperm donor, how many children that donor will father. We assume that each sperm donor is donating the maximum amount of sperm, which according to a TIME Magazine article, is 25 per $800,000$ people in the area [3]. When we resize this ratio to match Silicon Valley, we approximately obtain 93 sperm. Furthermore, according to AttainFertility, there is a 40% success rate in a given deposit in the sperm bank becoming a baby [1]. Thus, we multiply these together.

$$93 \times .4 \approx 37$$

Thus, each sperm donor will be assumed to be the father of 37 babies. Now we can obtain the total number of babies born from the Silicon Valley donors by multiplying the number of babies per donor by the number of donors in Silicon Valley.

$$37 \times 200 \approx 7400$$

Thus, we conclude that there are approximately $7,400$ babies born from Silicon Valley sperm donors.

To calculate the likelihood that a given person in Silicon Valley is the product of a sperm donation, we divide the number of in vitro babies by the total population of Silicon Valley:

$$\frac{7400}{3000000} \approx 0.00247$$

Thus, there is a .247% chance that a given person in Silicon Valley is born by in vitro fertilization. Now that we have the likelihood that a given person is a product of sperm donation, we want to calculate the likelihood that they're related. We can divide the number of successful in vitro babies per donor by the number of in vitro babies in Silicon Valley to obtain the likelihood a given baby has a specific donor. This gives us a .2% chance that two given in vitro babies have the same donor. Thus, we obtain the final likelihood by multiplying

the likelihood someone is a product of sperm donation by the likelihood that they have the same parent.

$$.002 \times .00247 \approx .00000494$$

Thus, we obtain our conclusion. There is approximately a .000494% chance that you are related to another person born through IVF in Silicon Valley.

4 Quantifying Potential Dating Partners

We can quantify the number of potential dating partners over the course of ones life using counting techniques from Combinatorics. Ignoring sexual/romantic preferences for the sake of simplicity, we can compute the binomial coefficient to calculate how many potential combinations of dating partners a person may choose over the course of their life. The binomial coefficient tells us how many ways there are of taking n things and selecting k of them. So, we will take n to be the population of Silicon Valley (3 million), and k to be the 5 dating partners someone chooses over the course of their lives.

$$\binom{3000000}{5} = 2024993250007874996250000600000 \approx 2.025 \times 10^{30}$$

The huge number is the number of possible combinations of 5 dating partners a person may select over the course of their lives, of course assuming they are willing to date anyone in (and only people from) Silicon Valley. However, since we have the percentage of the time a person would date their half-sibling, we can figure out how many combinations of dating partners there are where at least one of them is your half-sibling.

$$.0000123 \times \binom{3000000}{5} \approx 2.50 \times 10^{25}$$

This is perhaps a surprising result that such a small percentage chance of dating your half-sibling is equivalent to having as many possibilities as the order of 10^{25}. However, this is not many in comparison with the total number of possibilities, which is 5 orders of magnitude larger. Thus, it still abides by our model and isn't contradictory. It just seems that way at first glance.

5 Computational Automation

Since we've only explored the likelihood in Silicon Valley, the logical next step is to expand to other locations. We wanted to chose a location in which we thought there was a smaller chance that you would date your half-sibling, in order to validate our model. Salt Lake City, Utah was chosen as the second city. We believed that there would be a lower likelihood of dating your half-sibling in Salt Lake City compared to Silicon Valley because the median income in Salt

Lake City is much lower compared to the median income in Silicon Valley. To do this, we created a MATLAB script that will calculate the probability given certain variables about the location.

```
%% Fill in information for city
%Salt Lake City, Utah
% total_pop = 192672;
% percent_males = 0.513;
% percent_age = 0.386;
% percent_height = 0.56;
% percent_education = .61;
% median_income = 44510;

% Silicon Valley, California
total_pop = 3000000;
percent_males = 0.52;
percent_age = 0.58;
percent_height = 0.56;
percent_education = .61;
median_income = 90747;

%% Number of children produced by sperm donors

%Number of males who donate sperm
eligible_males = floor(total_pop * percent_males *
    percent_age * percent_height * percent_education)
percent_donate = 0.00168;
donating_males = floor(eligible_males * percent_donate)
%donating_males = 500

% Possible children produced by each male
baby_per_pop = 1/32000;
allowance = total_pop * baby_per_pop;
success_rate = 0.4;
baby_per_male = floor(success_rate * allowance)

% Total children produced via sperm donor
affordability = median_income/90747; % use Silicon Valley
     as ideal
total_babies = floor(affordability * baby_per_male *
    donating_males)

%% Likelihood of dating sibling
ave_num_partners = 5;
possible_matches = nchoosek(total_pop,ave_num_partners);
percent_chance_sibling = (baby_per_male / total_babies)*(
```

```
    total_babies / total_pop);
matches_with_sibling = possible_matches *
    percent_chance_sibling;
num_not_sibling = total_pop - baby_per_male;
matches_no_siblings = nchoosek(num_not_sibling,
    ave_num_partners);
%likelihood = 1 - matches_no_siblings/possible_matches;
likelihood = matches_with_sibling/possible_matches;
percent = likelihood*100;
fprintf('There is a %i percent chance of dating your half
    sibling',percent)
```

Here we have code available to run computations for the probability in Silicon Valley (as done previously step by step), and Salt Lake City. Statistics were taken from the census, as was the case with Silicon Valley.

When this script is run for the probability in Silicon Valley, we obtain 0.00000487 as the probability, whereas in our paper we claimed it was 0.00000494. This slight discrepancy comes from us rounding off certain numbers in our calculation, whereas the computer does not round. For example, earlier in the paper we claimed there were approximately 200 males eligible to donate sperm, but in the code we can get a more precise number of 232, and use this in calculating the probability. This is where the slight discrepancy comes from; it's not inconsistent.

According to our snippet, the probability in Salt Lake City is approximately 0.000453%, which is only slightly smaller than 0.000494% in Silicon Valley.

6 Conclusion

In summary, the likelihood of dating your half-sibling over the course of your life is very small. We looked at two differing locations - Salt Lake City and Silicon Valley - with very different populations, incomes, and demographics, and both had less than a 0.001% chance. We conclude that it is very unlikely a US citizen would ever date their half-sibling, even with the increasing prevalence of in vitro fertilization.

References

[1] *6 IVF Success Factors*. URL: `http://attainfertility.com/article/ivf-success-factors`.

[2] *Census profile: California*. 2015. URL: `https://censusreporter.org/profiles/04000US06-california/`.

[3] Eliana Dockterman. "'Delivery Man': 9 Sperm-Donation Questions You're Too Embarrassed to Ask". In: *TIME* (Nov. 2013). URL: http://healthland.time.com/2013/11/22/delivery-man-the-9-sperm-donation-questions-youre-too-embarrassed-to-ask/.

[4] Holly Finn. "In Vitro Veritas". In: *The Wall Street Journal* (Sept. 2011). URL: https://www.wsj.com/articles/SB10001424053111904106704576582981912838112.

[5] Haley Leaver. "Rising to the challenge: The battle against Britain's sperm shortage". In: *Metro* (July 2014). URL: http://metro.co.uk/2014/07/22/rising-to-the-challenge-the-battle-against-britains-sperm-shortage-4806109/.

[6] "No one said finding The One would be easy: The average women will kiss FIFTEEN men, enjoy TWO long-term relationships and have her heart broken TWICE". In: *Daily Mail* (Jan. 2014). URL: http://www.dailymail.co.uk/femail/article-2532213/No-one-said-finding-The-One-easy-The-average-women-kiss-FIFTEEN-men-enjoy-TWO-long-term-relationships-heart-broken-TWICE.html.

[7] *Population Growth in Silicon Valley*. May 2015. URL: https://www.jointventure.org/images/stories/pdf/population-brief-2015-05.pdf.

[8] *Population in the U.S. - Google Public Data Explorer*. Oct. 2016. URL: https://www.google.com/publicdata/explore?ctype=l&dl=en&ds=kf7tgg1uo9ude_&hl=en&met_y=population&nselm=h&rdim=sex&strail=false#!ctype=l&strail=false&bcs=d&nselm=h&met_y=population&fdim_y=age_group:5&fdim_y=county:06081&scale_y=lin&ind_y=false&rdim=sex&idim=sex:Female:Male.

[9] *Sperm Donor Requirements*. URL: http://www.spermbank.com/how-it-works/sperm-donor-requirements.

[10] *Statistical Abstract of the United States: 2011*. URL: https://www2.census.gov/library/publications/2010/compendia/statab/130ed/tables/11s0205.pdf.