The use of statistical analysis in testing some genealogical hypotheses

Bill Celmaster

Draft – June 27, 2016

Introduction

Frequently genealogists can make reasonable inferences about certain branches of family trees, despite the fact that there aren't complete or conclusive records to prove the deductions. This is possible despite the missing data, because the genealogist still has enough context and general experience to construct a *likely* inference. What I'd like to do here, is to attempt to quantify the likelihood – or probability – of the correctness of this kind of inference. In particular, I will examine, based on vital records from the early 19th century, a certain kind of hypothesis that comes up frequently when attempting to establish 18th century Jewish family connections.

A key tool that I will use is Bayes' theorem, in conjunction with what is known as *The Rule of Total Probability* (Freund, 1988).

> *If B_1, B_2,..., and B_k are mutually exclusive events of which one must occur, then*
>
> $$P(B_i|A) = \frac{P(B_i)\cdot P(A|B_i)}{\sum_{j=1}^{k} P(B_j)\cdot P(A|B_j)} \qquad (1)$$
>
> *for I = 1, 2, ... or k and where $P(B|A)$ is the conditional probability of B relative to A (i.e., the probability of B given A).*

The general situation to be studied here, has to do with a Jewish family branch whose progenitors (*C* and/or his wife *D)* and some descendants are known based on vital records or other contemporaneous documents, but whose ancestors are only hypothesized – based on anecdotal evidence. There is an Ashkenazi Jewish tradition of naming children after one's deceased ancestors, so the names of descendants are correlated to the names of ancestors. The question is whether the records of descendants support or reject certain hypotheses about the ancestors. In the language of statistical analysis, a *null hypothesis* (Freund, 1988) is proposed:

$H_\emptyset$: "*X* is the ancestor (e.g. father, grandfather, etc.) of *C* (or *D*)

with an alternative hypothesis:

H_A: "*X* is not the ancestor of *C* (or *D*)"

I will begin by setting up a fairly generic family tree with certain assumptions about naming patterns and about how records are randomly selected. This generic situation can be directly applied to a number of situations encountered in practice. However, there are many exceptions to the assumptions and forms of the family trees, and each exception needs its own separate analysis. I will first develop the methods for analyzing the generic case, starting with some simple examples and then eventually deriving equations for the most general version of the generic case. The next section of the paper will be to apply these methods – as well as exploring some of the exceptional situations mentioned above. The family to be studied is from the town of Drobin in Poland, and concerns the members of the family of Rabbi Jonatan of Drobin born in the early 18th century. Jonatan was a relative of the well-known Rabbi Jonatan Eybeschuetz and consequently a great deal of genealogical information can be extracted from rabbinic literature. As part of that information, is a claim that Jonatan of Drobin was the son of Aron, broth of Jonatan Eybeschuetz. However, after exploring many hundreds of vital records of Jonatan's family in Drobin and other towns, I have been struck by a curious absence of the name Aron. The question then is this. How likely is it that Jonatan was the son of Aaron, given the names of Jonatan's sons and grandsons? Statistical analysis is applied. Ultimately, this investigation turns out to show that even with the absence of the name Aron amongst Jonatan's descendants, it is still statistically plausible that Aron was indeed the father of Jonatan. Perhaps that will come as no surprise to anyone basing their assumptions on genealogies in rabbinical writings. Even so, it is heartening to find that this question was able to be examined systematically to arrive at a quantitative answer.

Probabilistic combinatorics as applied to family trees

Assumptions and General Method

The starting point of analysis is a hypothesized family tree starting with generation 0 – two parents A and B. The tree goes backward in time (the 'ancestor-generations') starting with generation 1, the parents of A and B, then their parents – generation 2 – and so on. The tree also goes forward in time (the 'descendant-generations') starting with generation (-1), the children of A and B, then their children – generation (-2) – and so on. The analysis will be based on existing records and anecdotes (we'll refer to these as 'known') which provide partial information about various members of the family tree. The question we will be addressing is this: given known information about the descendant-generations, what is the probability of various hypothesized ancestor-generations? This question can be converted, via the Bayes equation (1), into the question "what is the probability of known information about the descendant-generations, assuming various hypotheses about the ancestor-generations?"

A set of general assumptions[1] are proposed as the basis of analysis of this situation.

GA1. Names of sons (generation (-1)) are given only for deceased ancestors (generations 1 and earlier).

[1] Although all assumptions are stated in terms of sons, precisely the same assumptions could be made for daughters.

GA2. When sons are given single names, assume that names are chosen in order so that near-generations are exhausted before the next ancestor-generation is used – but that within a generation, names are chosen at random.

GA3. The number of sons follows a Poisson distribution with an expected value of N/5, where N is the number of child-bearing years of the couple. Here is the basis for that assumption. First of all, it seems plausible that childbirth follows a Poisson distribution (Derrida, Manrubia, & Zanette, 1999), at least during the best part of child-bearing years (approximately 25 years). Secondly, it seems plausible that during that time, the average number of children is 10, divided equally into sons and daughters (hence 5). In situations where parents are known to have had a shorter than average child-bearing period (owing to early death or late marriage of one of the parents), the expected value would be shorter[2].

GA4. Names, other than those of the ancestors whose names are known[3], are all equally probable unless explicitly noted otherwise[4].

GA5. *Random selection*: Observations (finding civil records of confirmed sons) are a random selection of the sons[5]. (Note that some records may be missing from city archives, owing to infant deaths, emigration, etc.). More precisely, for each descendant-generation, the known names have been selected entirely randomly from the collection of descendants of that generation – irrespective of chronology. (Note that other selection-assumptions could also be of interest. For example, we might have the actual birth records of one or more of the descendant-generations, and therefore rather than a random selection, we would have a chronologically-ordered selection. However, in what follows, the analysis will deal only with this *random selection* assumption.)

GA6. All ancestors (generations 1 and earlier) are deceased prior to the births of any of the sons. (This assumption is much more restrictive than we require for the analyses that follow.)

In practice, it suffices to limit the numbers of ancestor and descendant generations to be studied. Most often in what follows, I will only go out as far as the 3rd ancestor-generation and as far as the 1st descendant-generation. The general approach, based for simplicity (but it would be trivial to extend) on only one descendant-generation, is to start with the sons' names ($S_1, S_2, \dots, S_n$) that we have discovered from generation (-1). As stated in assumption GA5, we'll assume there are other sons whose names we don't know, and we'll also assume – as stated in assumption GA5 – that we don't have any information

---

[2] This assumption, as well as others, could potentially be studied empirically by surveying the Charedi communities in Israel – to obtain contemporary data regarding names, birth and death-dates of children and the people after whom they are named.

[3] Or more generally, the names of male ancestors needn't be known – but could, as in the case of Aron, be hypothesized with some probability representing the likelihood of that hypothesis.

[4] In point of fact, each town and each time-period have names that are distributed highly non-equally. If a name is rare, that often turns out to provide valuable genealogical insights about families where that name turns up I records.

[5] These assumptions turn out to be reasonably valid when studying Polish Jewish records of generations born prior to 1808, but whose names appear in records later than 1808. It will be outside the scope of this paper to account for situations where (as often happens) there is some partial birth-order information about the known sons. In the case-study considered later, results would be fairly insensitive to whatever birth-order information we have.

about the birth order of either the known or unknown sons. The sons S_i have been selected randomly from the full set of sons in generation (-1), ($V_1, V_2, \ldots V_m$), where V_i are ordered by birthdate so that $i < j$ implies that V_i is born before V_j. We then proceed by enumerating the number of ways in which the ancestor-generations can be consistent with known information about the ancestors (the hypothesis). For each of those ways, count the number of son-combinations that are consistent with the application of the naming rules GA1 – GA6. This process gives us the total of all son-selections that are consistent with the ancestor-hypothesis. This process gives us the total of all descendant-selections that are consistent with the ancestor-hypothesis. Next, count the subset of above, of all of the sons' name-combinations (V_i) that include the *known* names (S_i). Divide this by the total, to obtain the *pre-selection* probability. In the final step, recalling *the random selection* rule GA5, we multiply the pre-selection probability by the *selection factor*[6] $\frac{1}{\binom{m}{n}}$ to obtain the net probability.

Family trees have many different patterns of known and hypothesized members, so it is challenging to come up with an exhaustive, yet easy-to-use, collection of combinatorial rules that can be applied. Some examples follow. For simplicity of notation in this paper, I will use letters to represent individuals in the family tree and will also use the same letters to represent their names. Usage will easily be determined by context.

Some generic examples

Introduction – Starting with a known set of ancestors

As above, suppose we have obtained the names of sons ($S_1, S_2, \ldots, S_n$) in generation (-1) of a family tree. Those names have been selected as described in assumption GA5 above, from a chronologically ordered set of sons ($V_1, V_2, \ldots V_m$). The ancestors from generation 1 are named X_1 and X_2. The ancestors from generation 2 are named X_3, X_4, X_5 and X_6. For simplicity, we assume that none of the names X_i are the same as each other. This is all illustrated in Chart 1.

Generic Tree

X_3 + wife	X_4 + wife	X_5 + wife	X_6 + wife
X_1	wife of X_1	X_2	wife of X_2
	father	wife	

V_1	V_2	V_3	V_4

Chart 1

Now, for Chart 1, count how many ways the V_i can be named, consistent with the ancestors in the chart.

1.1. According to rules GA1 and GA2, and given assumption GA6, the first two sons (in order of birth) will be named after the two grandfathers of generation 1. Furthermore, also from rule GA3, the naming can be in either order. Therefore, we have, with equal likelihood, ($V_1 = X_1$, $V_2 = X_2$) or ($V_2 = X_1$, $V_1 = X_2$) (i.e. 2! permutations[7])

[6] We follow the notation $\binom{m}{n} \equiv \frac{m!}{n!(m-n)!}$

[7] I will use factorial signs to indicate permutations, even when they aren't necessary such as in the case of 2!

1.2. For each of the above two possibilities, V_3 and V_4 can be matched with any two of X_3, X_4, X_5, X_6, and in any order. This is computed as $\frac{4!}{2!} = 12$ possibilities.
1.3. There are therefore a total of $2! * \left(\frac{4!}{2!}\right) = 24$ possible combinations of V_i consistent with the ancestors. Furthermore, from assumption GA5, all those combinations are equally probable.

Next, to further illustrate the procedure, let's assume for the generic tree of Chart 1 that we know the names of 2 sons (S_1, S_2) and those are a subset of the names of the males in generation 1 and generation 2. In other words, we happen to have selected (but not in chronological order) sons (S_1, S_2) which are some permutation of 2 of the ancestors $(X_1, X_{2,}, X_3, X_4, X_5, X_6)$. We now count which of the above 24 combinations of V_i are consistent with the 2 known names of the sons. There are a number of interesting cases.

1.4. Suppose that the 2 observed names are X_1 and X_2. We follow 1.1 to see that all 24 combinations of V_i contain the names X_1 and X_2.
1.5. Suppose this time, that the two observed names are X_1 and X_3. As before, we follow 1.1 to see that all combinations of V_i contain the name X_1, but only 6 of those combinations also contain the name X_3.[8]

Finally, we use the above information to compute probabilities.

1.6. For the situation in 1.4, the pre-selection probability is obtained by dividing the 24 combinations obtained in 1.4, by the total obtained in 1.3. This results in a pre-selection probability $\frac{24}{24} = 1$. The after-selection "net" probability is then obtained by multiplying this result by the selection factor $\frac{1}{\binom{4}{2}}$. So finally, the probability for situation 1.4 is $p_{1.4} = \frac{1}{6}$.
1.7. For the situation in 1.5, the pre-selection probability is obtained by dividing the 6 combinations obtained in 1.5, by the total obtained in 1.3. This results in a pre-selection probability $\frac{6}{24} = \frac{1}{4}$. The after-selection "net" probability is then obtained by multiplying this result by the selection factor $\frac{1}{\binom{4}{2}}$. So finally, the probability for situation 1.4 1.5 is $p_{1.5} = \frac{1}{24}$.

Up to this point, we have started with a specific choice of ancestors. In general, we only know (or hypothesize) the names of some of those ancestors. In such cases, there are many possible configurations of ancestors which are consistent with the names we know. All of these must be considered, and for each of these, we also need to count the possible configurations of descendants. Some examples follow.

Example 1 – Names are known of all the sons

Example 1a – none of the great-grandfathers are known

In this example, assume that we know (for example, through anecdotes) that there were exactly 4 sons in generation (-1), and furthermore that we know (from records) their names. Furthermore, assume we

[8] There are 4 great-grandfathers, so if X_3 is selected as one of the names of sons, then that leaves 3 other options for the great-grandfather after which the 4th son is named. For each of those 3 options, there are 2 possible configurations corresponding to the two order-permutations of Y_1 and Y_2, thus a total of $(3 * 2)$ configurations.

know the two grandfathers in generation 1, but we don't know the names of any of the great-grandfathers in generation 2. The family tree looks like this.

Example 1a

X_1 + wife	X_2 + wife	X_3 + wife	X_4 + wife
A	wife of A	B	wife of B
	father	wife	

A	B	Y_1	Y_2

Chart 2

Here, we have replaced the names (V_1, V_2, V_3, V_4) in Chart 1 with the names of the 4 known sons (S_1, S_2, S_3, S_4). By assumption GA3, two of those sons must be named A and B, and the other two sons have been given the names Y_1 and Y_2. Again from assumption GA3, the sons A and B will appear chronologically before the other two sons. Note that in Chart 2, unlike Chart 1, the row of sons is <u>not</u> intended to be representative of chronological order and we will consider all chronological orders that are consistent with the assumptions about the ancestors and with the rules GA1 – GA7.

From assumption GA5, each X_i is chosen with equal probability from a collection of names. Assume that there is a large number, N of possible names. Then there are a total of N^4 possible name combinations[9] in generation 2. From 1.1 to 1.3**Error! Reference source not found.** above, for each of those combinations, there are $2! * \left(\frac{4!}{2!}\right)$ configurations of 4 sons consistent with that combination, hence a total of $T_{1a} = 2! * \left(\frac{4!}{2!}\right) * \ N^4$ configurations.

From assumption GA3, two of the X_i must match the two Y_i. There are exactly $c_{1a} = \frac{4!}{2!} * N^2$ such combinations, which can be more generally computed as

$$c = \left(\frac{l!}{(l-k)!}\right) * N^{\ l-k} \qquad (2)$$

where "c" is the number of combinations, "l" is the number of unknowns (X_i) in the ancestor generation being considered, and "k" is the number of (known) Y_i's in descendant generation (-1), who are not named after the grandfathers of generation 1. For each of the above ancestor-combinations, the sons A and B can appear in either order, and the sons Y_1 and Y_2 can also appear in either order, so there are $2! * 2!$ possible configurations of the 4 sons. Hence we have altogether $Q_{1a} = 2! * 2! * \frac{4!}{2!} * N^2$ possible configurations of 4 sons (in chronological order) that are consistent with the ancestors and known descendants hypothesized in Chart 2.

[9] Throughout this paper, I will ignore effects of relative order $o\left(\frac{1}{N}\right)$. So, for example, I will not consider situations where one of the great-grandfathers has the name Y and another great-grandfather, who could potentially have any one of N names, happens to have the same name Y. Since that occurs only once out of N times, it's net impact will change the overall results by a small multiple of $\frac{N-1}{N}$.

Putting all this together, and then multiplying by the selection factor $\frac{1}{\binom{4}{4}} = 1$, the probability of the known sons being consistent with the hypothesized ancestors for Example 1a, is

$$P(sons|example\ 1a) = \frac{Q_{1a}}{T_{1a}} * \frac{1}{\binom{4}{4}} = \frac{2!*2!*\frac{4!}{2!}*N^2}{2!*\left(\frac{4!}{2!}\right)*N^4} = \frac{2}{N^2} \qquad (3)$$

Example 1b – the name of one great-grandfather is known

We now consider a slight variation to Example 1a where one of the 4 grandfathers in generation 2, is hypothesized to have the name "C". Neither Y_1 nor Y_2 is the same name as C. The family tree is shown in Chart 3.

Example 1b

X_1 + wife	X_2 + wife	X_3 + wife	C + wife
A	wife of A	B	wife of B
	father	wife	

A	B	Y_1	Y_2

Chart 3

Similar to Example 1a, there are a total of N^3 possible name combinations in generation 2. There are $2! * \left(\frac{4!}{2!}\right)$ configurations of 4 sons consistent with each combination, and therefore a total of $T_{1b} = 2 * \left(\frac{4!}{2!}\right) * N^3$ configurations.

From assumption GA3, two of the X_i must match the two Y_i . Similar to the argument of Example 1a, and using equation (2), there are $c_{1b} = \frac{3!}{1!} * N$ such combinations. For each of the above ancestor-combinations, there are as before $2! * 2!$ possible configurations of the 4 sons. This leads to $Q_{1b} = 2! * 2! * \frac{3!}{1!} * N$ possible configurations of 4 sons that are consistent with the ancestor hypothesis of Chart 2.

Therefore, the probability of the known sons being consistent with the hypothesized ancestors for Example 1b, is

$$P(sons|example\ 1b) = \frac{Q_{1b}}{T_{1b}} * \frac{1}{\binom{4}{4}} = \frac{2!*2!*\frac{3!}{1!}*N^1}{2!*\left(\frac{4!}{2!}\right)*N^3} = \frac{1}{N^2} \qquad (4)$$

Generalizing Examples 1a and 1b

We next consider a generalization of the previous examples, where there are S sons ($3 \le S \le 6$), all of whose names are known, and two of whom are named A and B. The others are named Y_i. We also examine the situation where there are G known great-grandfathers, where $0 \le G \le 6 - S$ (so for instance in Example 1b, $G = 1$). The other great-grandfathers are named X_i. The Y_i's are not the names of any of the known great-grandfathers.

There are N^{4-G} possible great-grandfathers. In a way similar to 1.1 to 1.3 above, we find that for each combination of great-grandfathers, there are $2! * \left(\frac{4!}{(6-S)!}\right)$ configurations of S sons, consistent with that combination, hence a total of $T_{Ex1} = 2! * \left(\frac{4!}{(6-S)!}\right) * N^{4-G}$ configurations.

The sons A and B, match the grandfathers of generation 1. From assumption GA3, the X_i must match the Y_i. Similar to the argument of Example 1a, and using equation (2), there are $c_1 = \frac{(4-G)!}{(6-G-S)!} N^{6-G-S}$ such combinations. For each of the above ancestor-combinations, there are $2! * (S-2)!$ "randomized" configurations of S sons, consistent with that ancestor-combination. This leads to $Q_{Ex2} = 2! * (S-2)! * \frac{(4-G)!}{(6-G-S)!} * N^{6-G-S}$ possible configurations of S sons that are consistent with the known descendants and the ancestor hypothesis.

Again multiply by the selection factor $\frac{1}{\binom{S}{S}}$ to obtain the probability of the known sons being consistent with the hypothesized ancestors.

$$P(sons|example\ 1) = \frac{Q_{Ex1}}{T_{Ex1}} * \frac{1}{\binom{S}{S}} = \frac{2!*(S-2)!*\frac{(4-G)!}{(6-G-S)!}*N^{6-G-S}}{2!*\frac{4!}{(6-S)!}*N^{4-G}}$$

$$= \frac{(S-2)!*(6-S)!*(4-G)!}{(6-G-s)!*4!} * \frac{1}{N^{S-2}} = \frac{(S-2)!*\binom{6-S}{G}}{\binom{4}{G}} * \frac{1}{N^{S-2}} \qquad (5)$$

Application of Bayes theorem

This example is a hybrid of Example 1a and 1b. We make the assumption that examples 1a and 1b are the only possible configurations of ancestors. To be more precise, we assume that either all names in the second generation are unknown and of equal likelihood to one another or that one name is known and the others are unknown and of equal likelihood to one another. We also need to assign a probability to each case: $P_{1a} \equiv p(1a)$and $P_{1b} \equiv p(1b)$, with $P_{1a} + P_{1b} = 1$.

What is of ultimate interest, is the probability that the hypothesized ancestors of Example 1a are consistent with the observed descendants of Example 1a (or alternatively we can ask the same question of Example 1b). We apply the Bayes equation (1) to equations (3) and (4).

$$P(example\ 1a|sons) = \frac{P_{1a}*P(sons|example\ 1a)}{P_{1a}*P(sons|example\ 1a)+P_{1b}*P(sons|example\ 1b)}$$

$$= \frac{P_{1a}*\frac{2}{N^2}}{P_{1a}*\frac{2}{N^2}+P_{1b}*\frac{1}{N^2}} = \frac{2P_{1a}}{2P_{1a}+P_{1b}} \qquad (6)$$

Example 2 – Names are known of some of the sons

Up to now, the examples have examined family trees where we know all the sons' names (from generation (-1)). However, in most real situations, there are sons whose records haven't been found and whose names are therefore unknown.

Example 2a – Only three sons are known

We begin by considering a minor variation of Example 1b. As before, we assume there were 4 sons, but this time, we know the names only of 3 of those sons. Said differently, we have identified a set of 3 sons S_1, S_2 and S_3, but we happen to know that there were 4 sons.

Example 2a

X_1 + wife	X_2 + wife	X_3 + wife	C + wife
A	wife of A	B	wife of B
	father	wife	

A	B	Y_1	Z_1

Chart 4

The sons in Chart 4 are again not necessarily shown in chronological order. As before, Y_1 is a son whose name (not C) we know (so altogether, the 3 identified sons are A, B and Y_1). But this time, we don't know the name of the 4th son Z_1. As usual, there are N^3 possible combinations of great-grandfathers. For each, there are $2! * \left(\frac{4!}{2!}\right)$ possible combinations of 4 sons consistent with that combination, hence a total of $T_{2a} = 2! * \left(\frac{4!}{2!}\right) * N^3$ configurations.

The sons A and B, match the grandfathers of generation 1. From assumption GA3, Y_1 must match one of the X_i. There are exactly $c_{2a} = 3 * N^2$ such combinations of great-grandfathers. As was done in 1.1 through 1.3 we must count the ways in which the sons can be named, consistent with the information we have.

1.8. There are, as usual, (2!) ways of matching the names of the grandfathers
1.9. The remaining two sons include the name Y_1. The 'first' son can be named after the great-grandfather whose name is Y_1. In that case, the 'second' son (whose name we don't know) would be named after any one of the other 3 great-grandfathers. Alternatively, the 'second' son can be named after the great-grandfather whose name is Y_1 and the 'first' son can have one of the 3 other names of great-grandfathers. So there are $\frac{3!}{1!}$ ways for the remaining two sons to match the great-grandfathers.
1.10. There are therefore a total of $2! * \left(\frac{3!}{1!}\right) = 12$ possible combinations of sons consistent with the ancestors. Furthermore, from assumption GA5, all those combinations are equally probable.

In total, putting together the number of possible combinations of ancestors, with the number of possible sons (consistent with the information we have) for each such combination, we get $Q_{2a} = 3 * 2! * \frac{3!}{1!} * N^2$.

Finally, the probability of the known sons being consistent with the hypothesized ancestors for Example 2a, is obtained by multiplying $\frac{Q_{2a}}{T_{2a}}$ by the selection factor $\frac{1}{\binom{4}{3}}$.

$$P(sons|example\ 2a) = \frac{Q_{2a}}{T_{2a}} * \frac{1}{\binom{4}{3}} = \frac{3*2!*\frac{3!}{1!}*N^1}{2!*\left(\frac{4!}{2!}\right)*N^3} * \frac{1}{4} = \frac{3}{8} * \frac{1}{N} \qquad (7)$$

Generalizing Example 2a

This further generalizes Example 1. As before, there are 2 ancestor generations, and two known sons named after grandfathers. There are S sons ($3 \le S \le 6$), K of whose names we know ($3 \le K \le S$), including two of whom are named A and B. The other known sons are named Y_i and the unknown sons are named Z_i. We also examine the situation where there are G_2 known great-grandfathers, where $0 \le G_2 \le 6 - S$. The other great-grandfathers are named X_i. The Y_i's are not the names of any of the known great-grandfathers. Just as for Example 1, the total possible combinations of ancestors and S sons is $T_{Ex2} = T_{Ex1} = 2! * \left(\frac{4!}{(6-S)!}\right) * N^{4-G_2}$.

Similar to Example 1, there are $c_2 = \frac{(4-G_2)!}{(6-G_2-K)!} N^{6-G_2-K}$ possible configurations of great-grandfathers, consistent with the names of the sons. For each of these, we can follow the analysis of 1.8 through 1.12 to see that the number of possibilities are $2! * (S-2)! * \binom{6-K}{S-K}$.[10] Putting all this together, we have a total number of consistent possibilities as $Q_{Ex2} = 2! * (S-2)! * \binom{6-K}{S-K} * \frac{(4-G_2)!}{(6-G_2-K)!} N^{6-G_2-K}$.

Finally, we have

$$P(sons|example\ 2) = \frac{Q_{Ex2}}{T_{Ex2}} * \frac{1}{\binom{S}{K}} = \frac{2!*(S-2)!*\binom{6-K}{S-K}*K!*(S-K)!*\frac{(4-G_2)!}{(6-G_2-K)!}*N^{6-G_2-K}}{2!*\left(\frac{4!}{(6-S)!}\right)*S!*N^{4-G_2}}$$

$$= \frac{(S-2)!*K!}{S!} * \frac{\binom{6-K}{G_2}}{\binom{4}{G_2}} * N^{2-K} \qquad (8)$$

Example 3 – Adding generation #3

It seems intuitively obvious that if there are 6 sons or less, then generations 3 and earlier will have no effect on the results. Here we want to examine the situation where there are 7 sons or more. In particular, we want to see whether there are significant differences between an analysis based on having 8 males in generation 3 (which is the default), or based on some other number of males in generation 3 (which, although inaccurate, could simplify the calculations). The situations will be examined in increasing order of complexity.

[10] The factor $K! * (S-K)!$ comes from noting that the total number of ways that the known names can occur amongst the $S!$ randomized names is obtained by counting number of ways the K known (and therefore selected) names can be ordered, multiplied by the number of ways that the remaining names can be ordered.

Example 3a – all grandfathers and great-grandfathers but no other ancestors are known

Example 3a

X_1 + wife	X_2 + wife	X_3 + wife	X_4 + wife	X_5 + wife	X_6 + wife	X_7 + wife	X_8 + wife
C	wife of C	D	wife of D	E	wife of E	F	wife of F
	A		wife of A	B		wife of B	
			father	wife			

A	B	C	D	E	F	Y_1	Y_2

Chart 5

We'll begin by analyzing Chart 5, in which we assume that the first 6 sons are named after grandfathers and great-grandfathers. There are a total of N^8 possible name combinations in generation 3. For each of those, there are $2!$ orderings of sons A and B, and for each of those orderings there are $4!$ orderings of sons C, D, E and F. There are $\frac{8!}{6!}$ possibilities for the remaining two sons. All of this results in $2! * 4! * \frac{8!}{6!} * N^8$ possible combinations of 8 sons consistent with the ancestor assumptions. Next, following an argument similar to that for Example 1a, and using equation (2), we have a total of $2! * 4! * 2! * \frac{8!}{6!} * N^6$ combinations of ancestors and sons (chronologically ordered), consistent with what we know about the sons, and with what we assume about the ancestors. The selection factor is $\frac{1}{\binom{8}{8}} = 1$. We finally arrive at8

$$P(sons|chart\ 5) = \frac{2!*4!*2!*\frac{8!}{6!}*N^6}{2!*4!*\frac{8!}{6!}*N^8} = \frac{2}{N^2} \quad (9)$$

Interestingly, this is the same result as what we obtained in equation (3). In particular, there is no dependence on the number of males in generation 3, so we could considerably simplify the analysis by choosing, in this case, only a single male in generation 3.

Now generalize this to S sons, with $7 \leq S \leq 14$, where K sons are known (in other words, we have obtained, from records, the names of K sons). As before, the first 6 sons are named after the grandfathers and great-grandfathers. We have a total $T_{3a} = 2! * 4! * \frac{8!}{(14-S)!} * N^8$ combinations of S sons consistent with the ancestors. Similarly to previous derivations, we have a total $Q_{3a} = 2! * 4! * (S-6)! * \binom{14-K}{S-K} * \frac{8!}{(14-K)!} * N^{14-K}$ number of consistent combinations. The selection factor is $\frac{1}{\binom{S}{K}}$. This leads to

$$P(sons|example\ 3a) = \frac{Q_{3a}}{T_{3a}} * \frac{1}{\binom{S}{K}} = \frac{2!*4!*(S-6)!*\binom{14-K}{S-K}*K!*(S-K)!*\frac{8!}{(14-K)!}*N^{14-K}}{2!*4!*\left(\frac{8!}{(14-S)!}\right)*S!*N^8}$$

$$= \frac{(S-6)!}{(S-K)!*\binom{S}{K}} * N^{6-K} \quad (10)$$

Again, this result is independent of the number of males in generation 3.

Generalizing example 3a

Example 3 generalization

X_1 + wife	X_2 + wife	X_3 + wife	X_4 + wife	X_5 + wife	H + wife	I + wife	J + wife
C	wife of C	D	wife of D	E	wife of E	F	wife of F
	A		wife of A	B		wife of B	
			father	wife			

⬇

A	B	C	D	E	F	Y_1	Z_1

Chart 6

The son Y_1 is known, but does not have the same name as H, I or J,[11] and Z_1 is not known. Although in this chart, there are only 7 known sons (including the first 6) and one unknown son, and 3 names known in the third generation (H, I and J), we will assume a more general case. There are S sons ($7 \leq S \leq 14$), K of whose names we know ($7 \leq K \leq S$), including six of whom are named A, B, C, D, E and F. The other known sons are named Y_i and the unknown sons are named Z_i. We also examine the situation where there are G_3 known males from generation 3, where $0 \leq G_3 \leq 14 - S$. The other great-grandfathers are named X_i. The $Y_i{}'$s are not the names of any of the known great-great-grandfathers.

There are a total of N^{8-G_3} possible name combinations in generation 3. As in Example 3a, there are 2! orderings of sons A and B, and for each of those orderings there are 4! orderings of sons C, D, E and F. There are $\frac{8!}{(14-S)!}$ possibilities for the remaining sons. We therefore have a total $T_{Ex3} = T_{3a} = 2! * 4! * \frac{8!}{(14-S)!} * N^{8-G_3}$ combinations of S sons consistent with the ancestors.

Next, we more or less repeat that analysis done for the generalization of Example 2a. The first 6 sons are named after the 1st and 2nd generation, and there are $2! * 4!$ possibilities for those. There are $c_3 = \frac{(8-G_3)!}{(14-G_3-K)!} N^{14-G_3-K}$ possible configurations of males from the third generation, consistent with the names of the sons. None of the known sons are named after generation 3. For each of these, the number of possible (consistent with ancestors) combinations of sons are $2! * 4! * (S-6)! * \binom{14-K}{S-K}$. Putting all this together, we have the total number of consistent possibilities as $Q_{Ex3} = 2! * 4! * (S-6)! * \binom{14-K}{S-K} * \frac{(8-G_3)!}{(14-G_3-K)!} N^{14-G_3-K}$. The selection factor is $\frac{1}{\binom{S}{K}}$.

Finally, we have

$$P(sons|example\ 3) = \frac{Q_{Ex3}}{T_{Ex3}} * \frac{1}{\binom{S}{K}} = \frac{2!*4!*(S-6)!*\binom{14-K}{S-K}*K!*(S-K)!*\frac{(8-G_3)!}{(14-G-K)!}*N^{14-G_3-K}}{2!*4!*\left(\frac{8!}{(14-S)!}\right)*S!*N^{8-G_3}}$$

[11] The attentive reader might notice that I skipped over the letter G_2 in the list of ancestral names. That was done so as not to cause confusion with the later use of the letter G_2 to denote the <u>number</u> of known names in the ancestor generations 1 and beyond.

$$= \frac{(S-6)!*K!}{S!} * \frac{\binom{14-K}{G_3}}{\binom{8}{G_3}} * N^{6-K} \qquad (11)$$

We see that by introducing G_3 known names in the third generation, the result of equation (8) is modified by a factor $F = \frac{\binom{8-K'}{G_3}}{\binom{8}{G_3}}$ where $K' \equiv K - 6$. It is easy to see that when G_3 and K' are small, then $F \approx (\frac{8-K'}{8})^{G_3}$. In those cases, the results of equation (8) – which were independent of how many males there were in generation 3 – are not much altered by the knowledge of a few names from that generation.

Example 4 – Some of the first sons aren't known

In the examples considered thus far, the unknown sons were named exclusively after ancestors from a single generation (either generation 2 or 3). This has simplified the analysis but is actually quite rare in situations when researching Jewish civil records from the early 1800's.

Example 4a – Some unknown sons are named after generation 1

Example 4a

X_1 + wife	X_2 + wife	X_3 + wife	C + wife
A	wife of A	B	wife of B
	father	wife	

A	Z_1	Y_1	Z_2

Chart 7

Chart 7 illustrates the situation of interest. There are 2 known sons. One is A, and is named after generation 1. The other is Y_1, and since (by assumption) the name isn't B, we know that Y_1 is named after generation 2 (Also, we will assume here that Y_1 is not the same as C.) The names of the other two sons Z_1 and Z_2, are unknown. We'll begin by analyzing this situation, and then generalizing.

Just as we concluded for Chart 3, there are a total of $T_{4a-1} = 2! * \left(\frac{4!}{2!}\right) * N^3$ configurations of 4 sons consistent with the ancestors. Of these, which configurations are also consistent with the known information about the sons?

1.11. The son Z_1 is named after grandfather B. The sons A and Z_1 can be in either order, so that gives a factor of 2.
1.12. One of the great-grandfathers must have the name Y_1. There are $3 * N^2$ ways this can happen, consistent with what we know about generation 2.
1.13. For each such configuration of great-grandfathers, Z_2 is named after 1 of the 3 great-grandfathers who aren't named Y_1. There are 3 ways for this to happen.
1.14. The sons Z_2 and Y_1 can be in either order, so that gives another factor of 2.
1.15. Putting these all together, we get $Q_{4a-1} = 2 * 2 * 3 * 3 * N^2$.

Finally, including the selection factor, we get

$$P(sons|example\ 4a-1) = \frac{Q_{4a-1}}{T_{4a-1}} * \frac{1}{\binom{4}{2}} = \frac{2!*\ 2!*2*2*\ 3*3*N^2}{2!*\left(\frac{4!}{2!}\right)*4!*\ N^3} = \frac{1}{4} * \frac{1}{N} \qquad (12)$$

Notice in the analysis above, that up to 1.13, all the calculations proceed precisely as they would if we had known the names of 3 sons – where instead of the name Z_1, we would have had the name B. The only difference would have been in the selection factor where, instead of the factor $\frac{1}{\binom{4}{2}}$ that we used in equation (12), the selection factor would have been $\frac{1}{\binom{4}{3}}$. Thus another way of deriving equation (12) is to refer to equation (8) to obtain

$$P(sons|example\ 4a-1) = \ P(sons|example\ 2\ [S=4, K=3, G_2=1]) * \frac{\binom{4}{3}}{\binom{4}{2}}$$

$$= \frac{2!*3!}{4!} * \frac{\binom{3}{1}}{\binom{4}{1}} * \frac{1}{N} * \frac{2!*2!}{3!*1!} = \frac{1}{4} * \frac{1}{N} \qquad (13)$$

Now we can generalize. Let K_1 $(0 \leq K_1 \leq 2)$ be the number of known sons named for grandfathers (males of generation 1) and K_2 $(0 \leq 4)$ be the number of known sons named for males of generation 2. The total number of known sons is $K = K_1 + K_2$. The calculation of Q_{4a} (number of combinations of sons consistent with the selected known sons, and the ancestor assumptions) proceeds just as for 1.11 through 1.15. This also follows the derivation of equation (8), where we substitute in equation (8) the value $2 + K_2$ instead of K. The only difference in calculating the probability, is in the selection factor, just as explained in the derivation of equation (13). The result of all this is

$$P(sons|example\ 4a) = \frac{\binom{S}{K_2+2}}{\binom{S}{K}} * \frac{(S-2)!*(K_2+2)!}{S!} * \frac{\binom{4-K_2}{G_2}}{\binom{4}{G_2}} * N^{-K_2}$$

$$= \frac{K!*(S-K)!*(S-2)!}{(S-(K_2+2))!*S!} * \frac{\binom{4-K_2}{G_2}}{\binom{4}{G_2}} * N^{-K_2} \qquad (14)$$

Example 4b – Some unknown sons are named after generation 1 and generation 2

Example 4b

X_1 + wife	X_2 + wife	X_3 + wife	X_4 + wife	X_5 + wife	H + wife	I + wife	J + wife
C	wife of C	D	wife of D	E	wife of E	F	wife of F
	A		wife of A	B		wife of B	
			father	wife			

A	B	C	D	Z_1	Z_2	Y_1	Z_3

Chart 8

Chart 8 is an example of this situation, although we've chosen a case where there aren't any unknown sons named after generation 1. There are 5 known sons A, B, C, D and Y_1. A and B and are named after generation 1. C and D are named after generation 2. The last one Y_1, is named after generation 3 (i.e., it is not E or F) and is assumed to not be the same as H, I or J. The names of the other 3 sons Z_1, Z_2 and

Z_3 are unknown. The analysis is similar to that for Example 4a, so we will skip some of those steps. Begin by noticing that Z_1 and Z_2 are named after great-grandfathers E and F. The sons C, D, Z_1 and Z_2 can be in any order, so that gives a factor of 4!. The remainder of the analysis, other than for selection factors, is exactly the same as for equation (11) but where we treat Z_1 and Z_2 as though they were knowns. For Chart 8, the selection factor is $\frac{1}{\binom{8}{5}}$. On the other hand, for equation (11) – treating Z_1 and Z_2 as though they were knowns – the selection factor is $\frac{1}{\binom{8}{7}}$. So, substituting the values $K = 7, S = 8, G_3 = 3$ into equation (11), and then multiplying by the ratio of selection factors, we obtain

$$P(sons|\text{ Chart 8}) = \frac{\binom{8}{7}}{\binom{8}{5}} * \frac{2!*7!}{8!} * \frac{\binom{7}{3}}{\binom{8}{3}} = \frac{5}{224} * N^{-1} \qquad (15)$$

The general case is derived in a way very similar to equation (14). We use the same notation as in that equation, but we also introduce K_3 ($0 \le K_3 \le 8$), which is the number of known sons named for males of generation 3. The total number of sons is $K = K_1 + K_2 + K_3$.

$$P(sons|example\ 4b) = \frac{\binom{S}{K_3+6}}{\binom{S}{K}} * \frac{(S-6)!*(K_3+6)!}{S!} * \frac{\binom{8-K_3}{G_3}}{\binom{8}{G_3}} * N^{-K_3}$$

$$= \frac{K!*(S-K)!*(S-6)!}{(S-(K_3+6))!*S!} * \frac{\binom{8-K_3}{G_3}}{\binom{8}{G_3}} * N^{-K_3} \qquad (16)$$

Note that this result also holds if there are unknown sons named after generation 1. Only the selection process has any effect on the overall probability and the factors of equation (16) already account for that.

Example 5 – Some unknowns in each ancestral generation, but known sons aren't named after known ancestors

Often, we know the name of only one grandfather, and sometimes we don't know the names of any grandfathers. More generally, we may have unknown ancestors in more than one generation.

Example 5a – Some grandfathers and great-grandfathers aren't known

Example 5a

X_2 + wife	X_3 + wife	X_4 + wife	C + wife
A	wife of A	X_1	wife of X_1
	father	wife	

Z_1	Z_2	Y_1	Y_2

Chart 9

In Chart 9, one grandfather isn't known. We assume that there are exactly 4 sons, but we know the names of only 2 of those sons. None of the known sons are named after known ancestors. (Note that this assumption differs from assumptions made in previous examples, where some of the known sons had the same names as known grandfathers.) The row of sons shown in Chart 9 is not necessarily in chronological order.

As usual, there are a total of $T_{5a-1} = 2! * \left(\frac{4!}{2!}\right) * N^4$ configurations of 4 sons consistent with the ancestors. The following steps are taken to determine which of those configurations are also consistent with the known information about the sons.

1.16. Two of the X_i must have the names Y_1 and Y_2. There are 2 distinct types of ways in which this can happen.
   1.16.1. Both of those X_i (i.e. Y_1 and Y_2) are from generation 2. There are $2! * \binom{3}{2} * N^2$ ways for that to happen.
   1.16.2. One of the X_i, for example Y_1, is from generation 1, and the other is from generation 2. There are 2 choices (Y_1 or Y_2) for X_1, and for each of those, there are 3 possible X_i's from generation 2, which can be matched to the other Y_i. So altogether, there are $2 * \binom{3}{1} * N^2$ such combinations of grandfathers.

1.17. The analysis of sons is different for each of 1.16.1 and 1.16.2.
   1.17.1. (For 1.16.1) The first 2 sons (chronologically) are named after the grandfathers, and since the known sons are named after the great-grandfathers, then the two unknown sons must be the ones named after the grandfathers. This can happen in either order, hence a factor of 2!. For each order, the sons Y_1 and Y_2 can appear in either order hence another factor of 2! and therefore a total of $2! * 2!$ combinations of sons consistent with each configuration of 1.16.1.
   1.17.2. (For 1.16.2) Again, the first 2 sons (chronologically) are named after the grandfathers, so again there is a factor of 2! because the sons can be named in either order. However, one of those sons is known (for example Y_1)[12], so that leaves one unknown son to be named after generation 2. Since there is already a son (in this example, Y_2) named after one of the great-grandfathers then that leaves 3 possibilities for the name of the unknown son that is named after generation 2. For each of those possibilities, there are 2! ways of ordering the two sons named after generation 2. Putting all this together, there are a total of $3 * 2! * 2!$ combinations of sons consistent with each configuration of 1.16.2.

1.18. Now add up the results of 1.16 and 1.17. For the case considered in 1.16.1, the total number of consistent combinations are $2! * \binom{3}{2} * 2 * 2! * N^2$. For the case considered in 1.16.2, the total number of consistent combinations are $2 * \binom{3}{1} * 3 * 2! * 2! * N^2$. The total is $Q_{5a-1} = 2 * 2! * 2! * 12 * N^2$.

Finally, including the selection factor, we get

$$P(sons|example\ 5a-1) = \frac{Q_{5a-1}}{T_{5a-1}} * \frac{1}{\binom{4}{2}} = \frac{2!*2!*2!*\ 2!*2*12*N^2}{2!*\left(\frac{4!}{2!}\right)*\ 4!*N^4} = \frac{2}{3} * \frac{1}{N^2} \quad (17)$$

[12] This derivation might be clearer if we separately considered each of the two ways, in 1.16.2, of naming the unknown grandfather (either Y_1 or Y_2). Then, instead of representing those two choices in 1.16.2 by a factor of 2, consider each choice independently in 1.17.2 – which would effectively lead to an extra factor of 2 in 1.17.2. The net result is, of course, the same.

We now generalize. First we define G_1 to be the number of known grandfathers and as before, G_2 to be the number of known great-grandfathers. We define G to be the number of known ancestors (beyond generation 0) and in this example, we have $G = G_1 + G_2$. Note that the following inequalities must hold: $K_1 \leq 2 - G_1$ and $K_2 \leq 2 - G_2$. We will be able to enforce these inequalities in general equations, by setting terms to 0 if they violate those inequalities. It will turn out to be useful to introduce the notation

$$\overline{\binom{m}{n}} \equiv \begin{cases} \binom{m}{n}, & n \leq m \\ 0, & n > m \text{ or } n < 0 \end{cases} \qquad (18)$$

There are, as usual, a total of $T_{5a} = 2! * \left(\frac{4!}{(6-S)!}\right) * N^{6-G}$ configurations of S sons ($S \leq 6$) consistent with the ancestors. Next, we enumerate, as in 1.16, the different ways in which ancestors can be named consistent with the known sons, and for each of those ways we find how many combinations of sons are consistent with the known sons.

1.19. There are K known sons (and as before, we'll assume that none of them are named after known ancestors). We'll examine the situation where K_1 is the number of males in generation 1 who have the same names as known sons, and K_2 is the number of males in generation 2 to have the same names as known sons. $K = K_1 + K_2$. For each choice of K_1 names from the K known sons, there are $\overline{\binom{2-G_1}{K_1}} * K_1!$ ways of assigning those names to the males of generation 1. Also, there are $\overline{\binom{4-G_2}{K_2}} * K_2!$ ways of assigning the remaining $K_2 = K - K_1$ names to the males of generation 2. These two numbers are multiplied together. Then this number is multiplied by the number of ways $\binom{K}{K_1}$ in which K_1 names can be chosen from K names. Altogether, for K_1 names in generation 1, there are $\overline{\binom{2-G_1}{K_1}} * K_1! * \overline{\binom{4-G_2}{K_2}} * K_2! * \binom{K}{K_1}$ assignments of known names to males of generations 1 and 2. There are $N^{6-K-G_1-G_2}$ possible unknown (i.e. unassigned, nor previously identified) ancestors.

1.20. What combinations of sons are consistent with a specific configuration of (K_1, K_2)? Since there are K_1 known sons named after generation 1, there must be $2 - K_1$ unknown sons named after that generation. There are two orderings of the first 2 sons, so that is a factor of 2!. There are $S - K_2 - 2$ remaining unknown sons which are chosen from the $(4 - K_2)$ known ancestors of generation 2. The choice factor is $\overline{\binom{4-K_2}{S-K_2-2}}$. This must be multiplied by a permutation factor of $(S-2)!$. Altogether, the number of combinations of sons consistent with (K_1, K_2) are $q_{K_1,K_2} = 2! * \overline{\binom{4-K_2}{S-K_2-2}} * (S-2)!$.

Now we can sum up the results from 1.19 and 1.20 to obtain

$$Q_{5a} = \sum_{K_1=0}^{2-G_1} \overline{\binom{2-G_1}{K_1}} * K_1! * \overline{\binom{4-G_2}{K_2}} * K_2! * \binom{K}{K_1} * 2! * \overline{\binom{4-K_2}{S-K_2-2}} * (S-2)! * N^{6-K-G_1-G_2})$$

$$= 2! * (S-2)! * K! * N^{6-K-G_1-G_2} * \sum_{K_1=0}^{2-G_1}(\overline{\binom{2-G_1}{K_1}} * \overline{\binom{4-G_2}{K_2}} * \overline{\binom{4-K_2}{S-K_2-2}}) \qquad (19)$$

$$P(sons|example\ 5a) = \frac{Q_{5a}}{T_{5a}} * \frac{1}{\binom{S}{K}} = \frac{2!*(S-2)!*K!*N^{6-K-G_1-G_2}*\sum_{K_1=0}^{2-G_1}\left(\overline{\binom{2-G_1}{K_1}}*\overline{\binom{4-G_2}{K_2}}*\overline{\binom{4-K+K_1}{S-K_2-2}}\right)}{2!*\binom{S}{K}*\left(\frac{4!}{(6-S)!}\right)*N^{6-G}}$$

$$= \frac{K!*\sum_{K_1=0}^{2-G_1}\left(\overline{\binom{2-G_1}{K_1}}*\overline{\binom{4-G_2}{K-K_1}}*\overline{\binom{4-K+K_1}{6-S}}\right)}{\binom{S}{K}*\binom{4}{S-2}} * N^{-K} \qquad (20)$$

Example 5b – Some unknowns in each of generations 1, 2 and 3

Example 5b

X_4 + wife	X_5 + wife	X_6 + wife	X_7 + wife	X_8 + wife	H + wife	I + wife	J + wife
C	wife of C	D	wife of D	X_2	wife of X_2	X_3	wife of X_3
	A		wife of A	X_1		wife of X_1	
			father	wife			

Z_1	Z_2	Z_3	Z_4	Y_1	Y_2	Y_3	Y_4

Chart 10

In Chart 10, one grandfather isn't known, two great-grandfathers aren't known and five great-great-grandfathers aren't known. We assume that there are exactly 8 sons, but we know the names of only 4 of those sons. As in Example 5a, none of the known sons are named after known ancestors. The row of sons shown in Chart 10 is not necessarily in chronological order.

As deduced several times previously, there are a total of $T_{5b-1} = 2! * 4! * \left(\frac{8!}{(14-8)!}\right) * N^8$ configurations of 8 sons consistent with the ancestors. Now list those configuration that are consistent with the knowledge about the sons.

1.21. Four of the X_i must have the names Y_1, Y_2, Y_3 and Y_4. There are several distinct ways in which this can happen.
   1.21.1. All of the X_i (i.e., Y_1, Y_2, Y_3 and Y_4) are from generation 3. There are $4! * \binom{5}{4} * N^4$ ways for that to happen.
   1.21.2. Three X_i are from generation 3 and the remaining one is from generation 2. There are 4 ways of selecting the name of the remaining son from the original 4 Y_i. There are $3! * \binom{5}{3}$ ways of choosing (with order) X_i from generation 3, and for each of those there are $1! * \binom{2}{1}$ ways of choosing the remaining X_i from generation 2. Thus a total of $4 * 3! * \binom{5}{3} * 1! * \binom{2}{1} * N^4$.
   1.21.3. Three X_i are from generation 3 and the remaining one is from generation 1. There are 4 ways of selecting the name of the remaining son from the original 4 Y_i. There are $3! * \binom{5}{3}$ ways of choosing (with order) X_i from generation 3, and for each of those there are $1! *$

$\binom{1}{1}$ ways of choosing the remaining X_i from generation 1. Thus a total of $4 * 3! * \binom{5}{3} * 1! * \binom{1}{1} * N^4$.

1.21.4. Two X_i are from generation 3 and the remaining two are from generation 2. There are $\binom{4}{2}$ ways of selecting which 2 of the 4 X_i are in generation 3 (and the remaining ones are in generation 2). There are $2! * \binom{5}{2}$ ways of choosing (with order) X_i from generation 3, and for each of those there are $2! * \binom{2}{2}$ ways of choosing the remaining X_i from generation 2. Thus a total of $\binom{4}{2} * 2! * \binom{5}{2} * 2! * \binom{2}{2} * N^4$.

1.21.5. Two X_i are from generation 3, one of the remaining ones is from generation 2, and the last one is from generation 1. There are $\binom{4}{2}$ ways of selecting which of the 2 X_i are from generation 3. There are $2! * \binom{5}{2}$ ways of choosing (with order) X_i from generation 3. The third X_i could be in generation 2 or in generation 1, and the fourth one is in generation 1 or 2 respectively. So there are $\binom{2}{1}$ ways of choosing which of the remaining X_i are in generation 2. For the one in generation 2, there are $1! * \binom{2}{1}$ ways of choosing it, and for the one in generation 1, there are $1! * \binom{1}{1}$ way of choosing it. Thus a total of $\binom{4}{2} * 2! * \binom{5}{2} * \binom{2}{1} * 1! * \binom{2}{1} * 1! * (N^4)$.

1.21.6. One X_i is from generation 3, two of the remaining ones are from generation 2, and the last one is from generation 1. There are $\binom{4}{1}$ ways of selecting which X_i is from generation 3. There are $1! * \binom{5}{1}$ ways of choosing X_i from generation 3. There are $\binom{3}{2}$ ways of selecting the two remaining $X_i{'s}$ that will be assigned to generation 2 (and therefore of selecting the 4th X_i which is assigned to generation 1). For those assigned to generation 2, there are $2! * \binom{2}{2}$ choices, and for the one in generation 1, there is $1! * \binom{1}{1}$ choice. Thus a total of $\binom{4}{1} * 1! * \binom{5}{1} * \binom{3}{2} * 2! * \binom{2}{2} * 1! * \binom{1}{1} * N^4$.

1.22. The analysis of sons is different for each of 1.21.1 through 1.21.6.

1.22.1. Since there are 6 male ancestors amongst generations 1 and 2, and there are only 4 Z_i's, then at least 2 Y_i's must be named after ancestors in generations 1 and 2 (from assumption GA2). This therefore excludes options 1.21.1, 1.21.2 and 1.21.3.

1.22.2. (For 1.21.4) The first two sons (chronologically) are named after generation 1. Since the known sons are named after generations 2 and 3, then the first two sons are unknowns and they can be named in either order for the grandfathers – hence a factor of 2! Two known sons are named after the great-grandfathers Y_i and Y_j, and the remaining two unknown sons are named after the great-grandfathers C and D. There is a factor of 4! orderings. The remaining two known sons are named after the third generation males Y_l and Y_k. (Recall that the four names (Y_i, Y_j, Y_l, Y_k) are a permutation of (Y_1, Y_2, Y_3, Y_4).) This can happen in either order so a final factor of 2!. There are therefore a total of $2! * 4! * 2!$ possible combinations of sons, consistent with each configuration of 1.21.4.

1.22.3. (For 1.21.5) One of the known sons is named after generation 1, so one of the unknown sons must be named after the other grandfather in generation 1. This can happen in either order hence a factor of 2!. Another known son is named after generation 2, as are the remaining 3 unknown sons. There is a factor of 4! orderings. As in 1.22.2, the remaining two sons can appear in one of 2! ways. There are therefore a total of $2! * 4! * 2!$ possible combinations of sons, consistent with each configuration of 1.21.5.

1.22.4. (For 1.21.6) One of the known sons is named after generation 1, so one of the unknown sons must be named after the other grandfather in generation 1. This can happen in either order hence a factor of $2!$. Two known sons are named after generation 2, as are 2 of the remaining 3 unknown sons. There is a factor of $4!$ orderings. The last (4th) known son is named after generation 3, so the last (4th) unknown son is named after one of the remaining 7 males of generation 3. The 4th known and unknown sons can be named in either order, so there are $\binom{8-1}{2-1} * 2! = 14$ possibilities for those. There are altogether a total of $2! * 4! * 14$ possible combinations of sons, consistent with each configuration of 1.21.6.

1.23. Now add up the results of 1.21 and 1.22. For the cases considered in 1.21.1 through 1.21.3, the contributions add up to 0 as explained in 1.22.1. For the case considered in 1.21.4 and 1.22.2, the total number of consistent combinations are $\binom{4}{2} * 2! * \binom{5}{2} * 2! * \binom{2}{2} * 2! * 4! * 2! * N^4$. For the case considered in 1.21.5 and 1.22.3, the total number of consistent combinations are $\binom{4}{2} * 2! * \binom{5}{2} * \binom{2}{1} * 1! * \binom{2}{1} * 1! * 2! * 4! * 2! * \binom{1}{1} * N^4$. For the case considered in 1.21.6 and 1.22.4, the total number of consistent combinations are $\binom{4}{1} * 1! * \binom{5}{1} * \binom{3}{2} * 2! * \binom{2}{2} * 1! * \binom{1}{1} * 2! * 4! * 14 * N^4$. The total is $Q_{5b-1} = 4! * 4 * 10 * (4 + 8 + 21) * N^4$.

Finally, including the selection factor, we get

$$P(sons|example\ 5b - 1) = \frac{Q_{5b-1}}{T_{5b-1}} * \frac{1}{\binom{8}{4}} = \frac{4!*4!*4!*6!*4*10*(4+8+21)*N^4}{2!*4!*8!*8!*N^8} = \frac{39}{49} * \frac{1}{N^4} \qquad (21)$$

We now generalize the case where there are more than 6 sons ($6 \le S \le 14$), as we did for example 5a. We use notation developed in that example. The total number of known ancestors is $G = G_1 + G_2 + G_3$. There are a total of $T_{5b} = 2! * 4! * \frac{8!}{(14-S)!} * N^{14-G}$ configurations of S sons ($S \le 14$) consistent with the ancestors. Next, we enumerate, as in 1.21, the different ways in which ancestors can be named consistent with the known sons, and for each of those ways we find how many combinations of sons are consistent with the known sons.

1.24. There are K known sons (none named after known ancestors). We'll examine the situation where K_1 is the number of males in generation 1 who have the same names as known sons, K_2 is the number of males in generation 2 to have the same names as known sons, and K_3 is the number of males in generation 3 to have the same names as known sons $K = K_1 + K_2 + K_3$. For each choice of K_1 names from the K known sons, there are $\binom{2-G_1}{K_1} * K_1!$ ways of assigning those names to the males of generation 1. Also, there are $\binom{4-G_2}{K_2} * K_2!$ ways of assigning K_2 names to the males of generation 2. Finally, there are $\binom{8-G_3}{K_3} * K_3!$ ways of assigning the remaining $K_3 = K - K_1 - K_2$ names to the males of generation 3. These three numbers are multiplied together. Then this number is multiplied by the number of ways $\binom{K}{K_1}$ in which K_1 names can be chosen from K names, and is further multiplied by the number of ways $\binom{K-K_1}{K_2}$ in which K_2 names can be chosen from the $K - K_1$ names remaining after selecting K_1 names. (The last K_3 names are then completely determined.) Altogether, for K_1 names in generation 1, and K_2 names in generation 2, there are $\overline{\binom{2-G_1}{K_1}} * K_1! * \overline{\binom{4-G_2}{K_2}} * K_2! * \overline{\binom{8-G_3}{K_3}} * K_3! * \binom{K}{K_1} * \binom{K-K_1}{K_2} = \overline{\binom{2-G_1}{K_1}} * \overline{\binom{4-G_2}{K_2}} * \overline{\binom{8-G_3}{K-K_1-K_2}} * K!$ assignments of known names to males of generations 1, 2 and 3. There are N^{14-K-G} possible unknown ancestors.

1.25. What combinations of sons are consistent with a specific configuration of (K_1, K_2, K_3)? Since there are K_1 known sons named after generation 1, there must be $2 - K_1$ unknown sons named after that generation. There are two orderings of the first 2 sons, so that is a factor of 2!. Similarly, there are $4 - K_2$ unknown sons named after generation 2, with 4! orderings. There are $S - K_3 - 6$ remaining unknown sons which are chosen from the $(8 - K_3)$ known ancestors of generation 3. The choice factor is $\overline{\binom{8-K_3}{S-K_3-6}}$. This must be multiplied by a permutation factor of $(S-6)!$. Altogether, the number of combinations of sons consistent with (K_1, K_2, K_3) are $q_{K_1,K_2,K_3} = 2! * 4! * \overline{\binom{8-K_3}{S-K_3-6}} * (S-6)!$.

Now we can sum up the results from 1.24 and 1.25 to obtain

$$Q_{5b} = \sum_{K_1=0}^{2-G_1} \sum_{K_2=0}^{4-G_2} \left(\overline{\binom{2-G_1}{K_1}} * \overline{\binom{4-G_2}{K_2}} * \overline{\binom{8-G_3}{K-K_1-K_2}} * K! * 2! * 4! * \overline{\binom{8-K_3}{S-K_3-6}} * (S-6)! * N^{14-K-G} \right)$$

$$= 2! * 4! * (S-6)! * K! * N^{14-K-G} * \sum_{K_1=0}^{2-G_1} \sum_{K_2=0}^{4-G_2} \left(\overline{\binom{2-G_1}{K_1}} * \overline{\binom{4-G_2}{K_2}} * \overline{\binom{8-G_3}{K-K_1-K_2}} * \overline{\binom{8-K_3}{S-K_3-6}} \right)$$

(22)

$$P(sons|example\ 5b) = \frac{Q_{5b}}{T_{5b}} * \frac{1}{\binom{S}{K}} =$$

$$\frac{2! * 4! * (S-6)! * K! * N^{14-K-G} * \sum_{K_1=0}^{2-G_1} \sum_{K_2=0}^{4-G_2} \left(\overline{\binom{2-G_1}{K_1}} * \overline{\binom{4-G_2}{K_2}} * \overline{\binom{8-G_3}{K-K_1-K_2}} * \overline{\binom{8-K+K_1+K_2}{S-K_3-6}} \right)}{\binom{S}{K} * 2! * 4! * \frac{8!}{(14-S)!} * N^{14-G}}$$

$$= \frac{K! * \sum_{K_1=0}^{2-G_1} \sum_{K_2=0}^{4-G_2} \left(\overline{\binom{2-G_1}{K_1}} * \overline{\binom{4-G_2}{K_2}} * \overline{\binom{8-G_3}{K-K_1-K_2}} * \overline{\binom{8-K+K_1+K_2}{14-S}} \right)}{\binom{S}{K} * \binom{8}{S-6}} * N^{-K} \qquad (23)$$

Example 6 – Some unknowns in each ancestral generation, with some known sons named after known ancestors

We now consider the situation where we have found records for some of the sons, and their names might include names of ancestors that we know about. The analysis is very similar to that of the previous set of examples, but we will begin as usual with a concrete example and then proceed to the general case.

Example 6a

X_2 + wife	X_3 + wife	X_4 + wife	C + wife
A	wife of A	X_1	wife of X_1

	father	wife	

Z_1	Z_2	Y_1	Y_2	C

Chart 11

Chart 11 is like Chart 9 except there is one extra son whose name is C – the name of one of his great-grandfathers. We will proceed by noting the differences with the analysis of Chart 9. Since there are 3 sons named after generation 2, there are a total of $T_{6a-1} = 2! * \left(\frac{4!}{1!}\right) * N^4$ configurations of 5 sons consistent with the ancestors. Next, we enumerate the different ways in which ancestors can be named consistent with the known sons. This enumeration is identical to what we obtain in 1.16, since that analysis depends only on the number of known sons who are not named after any known ancestor. For each case in 1.16, we analyze the ways in which sons can be named.

1.26.1. (For ancestors named similarly to 1.16.1) The first 2 sons (chronologically) are named after the grandfathers, and since the known sons are named after the great-grandfathers, then the two unknown sons must be the ones named after the grandfathers. This can happen in either order, hence a factor of 2!. For each order, the sons Y_1, Y_2 and C can appear in either order hence another factor of 3! and therefore a total of $2! * 3!$ combinations of sons consistent with each configuration of the type in 1.16.1.

1.26.2. (For ancestors named similarly to 1.16.2) Again, the first 2 sons (chronologically) are named after the grandfathers, so again there is a factor of 2! because the sons can be named in either order. However, one of those sons is known (for example Y_1)[13], so that leaves one unknown son to be named after generation 2. Since there are already 2 sons (in this example, Y_2 and C) named after 2 of the great-grandfathers then that leaves 2 possibilities for the name of the unknown son that is named after generation 2. For each of those possibilities, there are 3! ways of ordering the 3 sons named after generation 2. Putting all this together, there are a total of $3! * 2 * 2!$ combinations of sons consistent with each configuration of the type in 1.16.2.

Now add up the results of 1.16, 1.26.1 and 1.26.2. The total is $Q_{6a-1} = 2! * \binom{3}{2} * 2! * 3! * N^2 + 2 * \binom{3}{1} * 3! * 2 * 2! * N^2$. Rather than continuing onwards to compute the probability, let's notice some patterns. First, define S_i to be the number of sons named after known ancestors in generation i. In computing the total T_{6a-1}, we used the general formula $T_2(K, S, G_1, G_2, S_1, S_2) = T_{5a} = 2! * \left(\frac{4!}{(6-S)!}\right) * N^{6-G}$. Next, in enumerating the various ways of naming ancestors, we used the same analysis as for 1.16, where the values of K_i and K, which refer to the known sons, are replaced respectively by $K_i - S_i$ and $K - S_1 - S_2$. Finally, in calculating the number of ways in which sons could be named after each ancestral configuration, we follow the analysis of 1.17 by where the values of K_i are those used for the ancestral configuration. Apply this to the results obtained in 1.19 and 1.20 as follows. From 1.19,

[13] This derivation might be clearer if we separately considered each of the two ways, in 1.16.2, of naming the unknown grandfather (either Y_1 or Y_2). Then, instead of representing those two choices in 1.16.2 by a factor of 2, consider each choice independently in 1.17.2 – which would effectively lead to an extra factor of 2 in 1.17.2. The net result is, of course, the same.

with (K_1, K_2) known names, there are $\overline{\binom{2-G_1}{K_1-S_1}} * (K_1 - S_1)! * \overline{\binom{4-G_2}{K_2-S_2}} * (K_2 - S_2)! * \binom{K-S_1-S_2}{K_1-S_1}$ assignments of known names to males of generations 1 and 2. There are $N^{6-K+S_1+S_2-G_1-G_2}$ possible unknown (i.e. unassigned, nor previously identified) ancestors. From 1.20, $q_{K_1,K_2} = 2! * \overline{\binom{4-K_2}{S-K_2-2}} * (S-2)!$ where we define q_{K_1,K_2} to be the number of combinations of sons consistent with the partitioning (K_1, K_2) used in the enumeration of ancestors. Now sum up those results to obtain

$$Q_2(K,S,G_1,G_2,S_1,S_2) = \sum_{K_1=S_1}^{2-G_1+S_1}\left(\overline{\binom{2-G_1}{K_1-S_1}} * (K_1-S_1)! * \overline{\binom{4-G_2}{K_2-S_2}} * (K_2-S_2)! * \binom{K-S_1-S_2}{K_1-S_1} * 2! * * (S-2)! * N^{6-K+S_1+S_2-G_1-G_2}\right)$$

$$= 2! * (S-2)! * (K-S_1-S_2)! * N^{6-K+S_1+S_2-G_1-G_2} * \sum_{K_1=S_1}^{2-G_1+S_1}\left(\overline{\binom{2-G_1}{K_1-S_1}} * \overline{\binom{4-G_2}{K_2-S_2}} * \overline{\binom{4-K_2}{S-K_2-2}}\right)$$

(24)

$$P(sons|example\ 6a) = \frac{Q_2(K,S,G_1,G_2,S_1,S_2)}{T(K,S,G_1,G_2,S_1,S_2)} * \frac{1}{\binom{S}{K}} = \frac{2!*(S-2)!*(K-S_1-S_2)!*N^{6-K+S_1+S_2-G_1-G_2}*\sum_{K_1=S_1}^{2-G_1+S_1}\left(\overline{\binom{2-G_1}{K_1-S_1}}*\overline{\binom{4-G_2}{K_2-S_2}}*\overline{\binom{4-K_2}{S-K_2-2}}\right)}{2!*\binom{S}{K}*\left(\frac{4!}{(6-S)!}\right)*N^{6-G}}$$

$$= \frac{(K-S_1-S_2)!*\sum_{K_1=S_1}^{2-G_1+S_1}\left(\overline{\binom{2-G_1}{K_1-S_1}}*\overline{\binom{4-G_2}{K-K_1-S_2}}*\overline{\binom{4-K+K_1}{6-S}}\right)}{\binom{S}{K}*\binom{4}{S-2}} * N^{-K+S_1+S_2} \quad (25)$$

We can proceed directly to the case, which we'll describe as example 6b, where there are more than 6 sons ($6 \leq S \leq 14$) and we consider 3 generations. In this analysis, S_3 will be the number of sons named after known ancestors in the 3rd generation. As before, we make substitutions but this time in example 5b. We have $T_3(K,S,G_1,G_2,G_3,S_1,S_2,S_3) = 2! * 4! * \frac{8!}{(14-S)!} * N^{14-G}$ and

$$Q_3(K,S,G_1,G_2,G_3,S_1,S_2,S_3) = 2! * 4! * (S-6)! * (K-S_1-S_2-S_3)! * N^{14-K+S_1+S_2+S_3-G} * \sum_{K_1=S_1}^{2-G_1+S_1} \sum_{K_2=S_2}^{4-G_2+S_2} \left(\overline{\binom{2-G_1}{K_1-S_1}} * \overline{\binom{4-G_2}{K_2-S_2}} * \overline{\binom{8-G_3}{K-K_1-K_2-S_3}} * \overline{\binom{8-K+K_1+K_2}{14-S}}\right) \quad (26)$$

$$P_{Ex6b}\big(sons\big|(K,S,G_1,G_2,G_3,S_1,S_2,S_3)\big) \equiv P(sons|example\ 6b) = \frac{Q_3(K,S,G_1,G_2,G_3,S_1,S_2,S_3)}{T(K,S,G_1,G_2,G_3,S_1,S_2,S_3)} * \frac{1}{\binom{S}{K}} = \frac{(K-S_1-S_2-S_3)!*\sum_{K_1=S_1}^{2-G_1+S_1}\sum_{K_2=S_2}^{4-G_2+S_2}\left(\overline{\binom{2-G_1}{K_1-S_1}}*\overline{\binom{4-G_2}{K_2-S_2}}*\overline{\binom{8-G_3}{K-K_1-K_2-S_3}}*\overline{\binom{8-K+K_1+K_2}{14-S}}\right)}{\binom{S}{K}*\binom{8}{S-6}} * N^{-K+S_1+S_2+S_3}$$

(27)

Example 7 – Generations where nothing is known

Sometimes nothing is previously known about the grandparents or great-grandparents. Yet, based on information about the descendants, we may attempt to construct some hypotheses about the ancestors, and then compare probabilities for various hypotheses. Even more frequently, there is some information about generations 1 and 2, but none about generation 3. Or there might be something known about generation 3 and not about earlier generations. We consider each of these situations. Even though generations 1, 2 and 3 are all covered by previous examples, there are some simplifications that can be applied and that are worth noting.

Example 7a – Nothing is known about generations 1 or 2

Example 7a

X_3 + wife	X_4 + wife	X_5 + wife	X_6 + wife
X_1	wife of X_1	X_2	wife of X_2
	father	wife	

Z_1	Z_2	Y_1	Y_2	Y_3

Chart 12

We will refer to as a special case but will derive the general case. The following identity will be useful in this example and later (Abramowitz & Stegun, 1964).[14]

$$\sum_{m=0}^{r+s} \overline{\binom{r}{m}} * \overline{\binom{s}{n-m}} = \binom{r+s}{n} \qquad (28)$$

We can use equation (24), together with equation (28) to obtain

$$Q_2(K,S,0,0,0,0) = 2! * (S-2)! * K! * N^{6-K} * \sum_{K_1=0}^{2}\left(\overline{\binom{2}{K_1}} * \overline{\binom{4}{K_2}} * \overline{\binom{4-K_2}{S-K_2-2}}\right)$$

$$= \frac{2!*4*K!}{(6-S)!} * \binom{S}{K} * N^{6-K}$$

Then

$$P(sons|example\ 7a) = \frac{Q_2(K,S,0,0,0,0)}{T_2(K,S,0,0,0,0)} * \frac{1}{\binom{S}{K}} = \frac{\frac{2!*4*K!}{(6-S)!}*\binom{S}{K}*N^{6-K}}{\binom{S}{K}*2!*\frac{4!}{(6-S)!}*N^{6}}$$

$$= K! * N^{-K} \qquad (29)$$

[14] This equation, with the tilde notation (i.e., $\overline{\binom{m}{n}} \equiv 0$ if $m < n$), differs slightly from what is given in the (9th printing of) the reference (Abramowitz & Stegun, 1964). The equation can be derived easily by considering the product $(x + y)^r * (x + y)^s$ where each term is expanded using binomial coefficients, and then comparing the product to the binomial expansion of $(x + y)^{r+s}$.

Example 7b – Nothing is known about generations 1, 2, 3

For this, we can use equation (26), together with equation (28) to obtain

$$Q_3(K, S, 0,0,0,0,0,0) = 2! * 4! * (S-6)! * K! * N^{14-K} * \sum_{K_1=0}^{2} \sum_{K_2=0}^{4} (\overline{\binom{2}{K_1}} * \overline{\binom{4}{K_2}} * \overline{\binom{8}{K-K_1-K_2}} * \overline{\binom{8-K+K_1+K_2}{14-S}})$$

$$= \frac{2!*4!*8!*K!}{(14-S)!} * \binom{S}{K} * N^{14-K}$$

$$P(sons|example\ 7b) = \frac{Q_3(K,S,0,0,0,0,0,0)}{T_3(K,S,0,0,0,0,0,0)} * \frac{1}{\binom{S}{K}} = \frac{\frac{2!*4!*8!*K!}{(14-S)!}*\binom{S}{K}*N^{14-K}}{2!*4!*\frac{8!}{(14-S)!}*\binom{S}{K}*N^{14-G}}$$

$$= K! * N^{-K} \qquad (30)$$

Since the results for example 7a and 7b are the same, it seems plausible that those results would extend to earlier generations.

Example 7c – Nothing is known about generation 2 or earlier

Again, we use equations (26) and (28).

$$Q_3(K, S, G_1, 0,0, S_1, 0,0) = 2! * 4! * (S-6)! * (K - S_1!) * N^{14-K+S_1-G} * \sum_{K_1=S_1}^{2-G_1+S_1} \sum_{K_2=0}^{4} (\overline{\binom{2-G_1}{K_1-S_1}} * \overline{\binom{4}{K_2}} * \overline{\binom{8}{K-K_1-K_2}} * \overline{\binom{8-K+K_1+K_2}{14-S}})$$

$$= \frac{2!*4!*8!*(K-S_1!)}{(14-S)!}*\binom{S-G_1}{K-S_1} * N^{14-K+S_1-G}$$

$$P(sons|example\ 7c) = \frac{Q_3(K,S,G_1,0,0,S_1,0,0)}{T_3(K,S,G_1,0,0,S_1,0,0)} * \frac{1}{\binom{S}{K}} = \frac{\frac{2!*4!*8!*(K-S_1)!}{(14-S)!}*\binom{S-G_1}{K-S_1}*N^{14-K+S_1-G}}{2!*4!*\frac{8!}{(14-S)!}*\binom{S}{K}*N^{14-G}}$$

$$= \frac{(K-S_1)!*\binom{S-G_1}{K-S_1}}{\binom{S}{K}} * N^{-K+S_1} \qquad (31)$$

Example 7d – Nothing is known about generations g or earlier

We now use the principle of induction to consider the effect of adding generations of ancestors whose names we don't know. We'll imagine starting with an ancestor-descendant analysis for ancestor-generations $1, \ldots, g$, and then adding a further ancestor generation $g+1$ where none of those ancestors are known (so we label them as X_j). That generation has $Gen(g+1) = 2^{g+1}$ males.[15] Also define $\beta(g) \equiv \sum_{i=1}^{g} Gen(i), \beta(0) = 0$ and $\widehat{K} \equiv \sum_{i=1}^{g} S_i$. Let $Q_g(K, S \ldots)$ be the number of combinations of sons and ancestors that are consistent with information known through generation g.

[15] Our results actually won't depend on the value of $Gen(g)$.

We continue to make the same assumptions as before, and assume that we may know some number of ancestors in each generation, and that there may be some known sons named after some of those ancestors. Setting the value G to be the total number of known ancestors, the inductive assumption is

$$Q_g(K, \beta(g) \dots) = (K - \widehat{K})! * \binom{\beta(g)-G}{K-\widehat{K}} * \ 2! * 4! * \dots * Gen(g)! * N^{\beta(g)-K+\widehat{K}-G} \qquad (32)$$

The first induction step is equation (31) for $Q_3(K, 6, G_1, 0,0, S_1, 0,0)$.

1.27. Let K_{g+1} be the number of known sons in generation $g+1$ ($K_{g+1} \leq K$). There are $\binom{K-\hat{S}}{K_{g+1}}$ ways of selecting those sons from amongst the K known sons. Then there are $\binom{Gen(g+1)}{K_{g+1}} *$ $K_{g+1}!$ ways of assigning those names to ancestors in generation $g+1$. For the other ancestors in generation $g+1$, there are $N^{Gen(g+1)-K_{g+1}}$ possibilities. So altogether, when there are K_{g+1} known sons in generation $g+1$, there are $\binom{K-\widehat{K}}{K_{g+1}} * \binom{Gen(g+1)}{K_{g+1}} * K_{g+1}! *$ $N^{Gen(g+1)-K_{g+1}}$ assignments of names to generation $g+1$.

1.28. For each situation above, and for each chronologically ordered set of sons, consider only the sons older than the son whose position in the sequence is $\beta(g)$. The number of those sons is $S - \beta(g)$, of which K_{g+1} are known, and $S - \beta(g) - K_{g+1}$ are unknown. The names of the unknown sons can be chosen arbitrarily from any of the names in generation $g+1$, other than those that have been assigned. There are $Gen(g+1) - K_{g+1}$ such names. Thus we have a factor of $\binom{Gen(g+1)- K_{g+1}}{S- \beta(g)- K_{g+1}}$. The aforementioned $S - \beta(g)$ sons can appear in any chronological order so there is another factor of $(S - \beta(g))!$.

1.29. Finally, for each situation above, we must multiply by the total number of ancestor-son combinations corresponding to generations 1 through g. Those have a total number of sons that are equal $\beta(g)$ and $K - K_{g+1}$ known sons. We use the inductive assumption (32), which gives a multiplicative factor of $\binom{\beta(g)-G}{K-\widehat{K}-K_{g+1}} * (K - \widehat{K} - K_{g+1})! * 2! * 4! * \dots * Gen(g)! *$ $N^{\beta(g)-K+\widehat{K}+K_{g+1}-G}$.

The result is

$$Q_{g+1}(K, S \dots) = \sum_{K_{g+1}=0}^{K} \binom{K-\hat{S}}{K_{g+1}} * \binom{Gen(g+1)}{K_{g+1}} * K_{g+1}! * N^{Gen(g+1)-K_{g+1}} * \binom{Gen(g+1)- K_{g+1}}{S- \beta(g)- K_{g+1}} *(S - \beta(g))! * \binom{\beta(g)-G}{K-\widehat{K}-K_{g+1}} * (K - \widehat{K} - K_{g+1})! * 2! * 4! * \dots * Gen(g)! * N^{\beta(g)-K+\widehat{K}+K_{g+1}-G}$$

$$= 2! * 4! * \dots * Gen(g)! * Gen(g+1)! * (K - \widehat{K})! * \binom{S-G}{K-\widehat{K}} * \frac{1}{(\beta(g+1)-S)!} * N^{\beta(g+1)-K+\widehat{K}-G} \qquad (33)$$

When we set (above) $S = \beta(g+1)$, we see that the inductive step is proven. In a similar fashion, it is easy to show that

$$T_{g+1}(K, S \dots) = \frac{2!*4!*\dots*Gen(g)!*Gen(g+1)!}{(\beta(g+1)-S)!} * N^{\beta(g+1)-G} \qquad (34)$$

Finally, putting all of this together, we find

$$P_{Ex7d}(sons|K,S\ \dots) = \frac{Q_{g+1}(K,S\dots)}{T_{g+1}(K,S\dots)} * \frac{1}{\binom{S}{K}} = \frac{2!*4!*\dots*Gen(g)!*Gen(g+1)!*(K-\widehat{K})!*\binom{S-G}{K-\widehat{K}}*\frac{1}{(\beta(g+1)-S)!}*N^{\beta(g+1)-K+\widehat{K}-G}}{\binom{S}{K}*\frac{2!*4!*\dots*Gen(g)!*Gen(g+1)!}{(\beta(g+1)-S)!}*N^{\beta(g+1)-G}}$$

$$= \frac{(K-\widehat{K})!*\binom{S-G}{K-\widehat{K}}}{\binom{S}{K}} * N^{-K+\widehat{K}} \qquad (35)$$

Combined formulas for preceding examples

All of the above examples can be covered by the following:

- The names aren't known of any male ancestors prior to the 3rd ancestor generation.
    - We know G_2 names from generation 2, G_2 names from generation 2 and G_3 names from generation 3.
- All names of male ancestors are distinct from one another (through the 3rd ancestor generation).
- We assume that there are S sons, of which we have discovered K names.
    - Of the known sons, K_1 are named after some of the G_1 known ancestral names from generation 1, K_2 are named after some of the G_2 known ancestral names from generation 2 and K_3 are named after some of the G_3 known ancestral names from generation 3.
- Selection rules are covered by **Error! Reference source not found.** through **Error! Reference source not found.**.

We use the notation $P_{ng}(K,S,G_1,G_2,G_3,S_1,S_2,S_3)$ to denote the probability of discovering the K names among S sons, given the known information about the first three generations of ancestors, and assuming ng generations of ancestors. Putting together equations (25), (27) and **Error! Reference source not found.**, together with an easily-derived equation for situations with 2 sons or less

$$P_{ng}(K,S,G_1,G_2,G_3,S_1,S_2,S_3)$$

$$= \frac{(K-S_1)!*\overline{\binom{2-G_1}{K-S_1}}*\overline{\binom{2-K}{2-S}}}{\binom{S}{K}*\binom{2}{S}} * N^{-K+S_1} \quad [\text{if } 0 < S \le 2]$$

$$= \frac{(K-S_1-S_2)!*\sum_{K_1=S_1}^{2-G_1+S_1}\left(\binom{2-G_1}{K_1-S_1}*\overline{\binom{4-G_2}{K-K_1-S_2}}*\overline{\binom{4-K+K_1}{6-S}}\right)}{\binom{S}{K}*\binom{4}{S-2}} * N^{-K+S_1+S_2} \quad [\text{if } 2 < S \le 6]$$

$$= \frac{(K-S_1-S_2-S_3)!*\sum_{K_1=S_1}^{2-G_1+S_1}\ \sum_{K_2=S_2}^{4-G_2+S_2}\left(\overline{\binom{2-G_1}{K_1-S_1}}*\overline{\binom{4-G_2}{K_2-S_2}}*\overline{\binom{8-G_3}{K-K_1-K_2-S_3}}*\overline{\binom{8-K+K_1+K_2}{14-S}}\right)}{\binom{S}{K}*\binom{8}{S-6}} * N^{-K+S_1+S_2+S_3} \quad [\text{if } 6 < S \le 14]$$

$$= \frac{(K-S_1-S_2-S_3)!*\binom{S-G_1-G_2-G_3}{K-S_1-S_2-S_3}}{\binom{S}{K}} * N^{-K+S_1+S_2+S_3} \quad [\text{if } 14 < S] \qquad (36)$$

Equation (36) is valid provided the following inequalities hold and we set the value of P_3 to 0 when the inequalities are violated.

- $0 \le S_i \le G_i$
- $0 \le G_1 \le 2, 0 \le G_2 \le 4, 0 \le G_3 \le 8$

- $\sum_{i=1}^{3} S_i \leq K$
- $0 < K \leq S$
- If $S_2 \neq 0$ then $S > 2$
- If $S_3 \neq 0$ then $S > 6$

We also introduce some notation that will prove to be useful later. We define functions P_{Tree} and P_{Bayes} as

$$P_{Tree}(mean, K, G_1, G_2, G_3, S_1, S_2, S_3) \equiv \sum_{S=1}^{\infty} P_{ng}(K, S, G_1, G_2, G_3, S_1, S_2, S_3) * f_{Poisson}(S, mean) \quad (37)$$

$$P_{Bayes}(p, p_A, p_B) \equiv \frac{p * p_A}{(p * p_A + (1-p) * p_B)} \quad (38)$$

There is frequently a situation where names appear twice (or even more often) amongst the known ancestors (including ancestor generation 0). In those instances, a son would be named after the later ancestor with that name, and no-one would be named after the earlier ancestor. This has the same effect as if there were less members of that earlier generation – namely, $Gen(g) < 2^g$. It isn't difficult to modify equation (36) to account for more general values of $Gen(g)$.

$$P_{ng}^{Gen}(K, S, G_1, G_2, G_3, S_1, S_2, S_3, Gen(1), \ldots, Gen(ng))$$

$$= \frac{(K-S_1)! * \overline{\binom{Gen(1)-G_1}{K-S_1}} * \overline{\binom{Gen(1)-K}{\beta(1)-S}}}{\binom{S}{K} * \binom{Gen(1)}{S}} * N^{-K+S_1} \text{ [if } 0 < S \leq \beta(1)]$$

$$= \frac{(K-S_1-S_2)! * \sum_{K_1=S_1}^{Gen(1)-G_1+S_1} \left(\overline{\binom{Gen(1)-G_1}{K_1-S_1}} * \overline{\binom{Gen(2)-G_2}{K-K_1-S_2}} * \overline{\binom{Gen(2)-K+K_1}{\beta(2)-S}} \right)}{\binom{S}{K} * \binom{Gen(2)}{S-\beta(1)}} * N^{-K+S_1+S_2} \text{ [if } \beta(1) < S \leq \beta(2)]$$

$$= \frac{(K-S_1-S_2-S_3)! * \sum_{K_1=S_1}^{Gen(1)-G_1+S_1} \sum_{K_2=S_2}^{Gen(2)-G_2+S_2} \left(\overline{\binom{Gen(1)-G_1}{K_1-S_1}} * \overline{\binom{Gen(2)-G_2}{K_2-S_2}} * \overline{\binom{Gen(3)-G_3}{K-K_1-K_2-S_3}} * \overline{\binom{Gen(3)-K+K_1+K_2}{\beta(3)-S}} \right)}{\binom{S}{K} * \binom{Gen(3)}{S-\beta(2)}} *$$

$$N^{-K+S_1+S_2+S_3} \text{ [if } \beta(2) < S \leq \beta(3)]$$

$$= \frac{(K-S_1-S_2-S_3)! * \binom{S-G_1-G_2-G_3}{K-S_1-S_2-S_3}}{\binom{S}{K}} * N^{-K+S_1+S_2+S_3} \text{ [if } \beta(3) < S] \quad (39)$$

There will be situations where it will turn out to be useful to also have equations for Q_{ng}^{Gen} and T_{ng}^{Gen}, where Q_{ng}^{Gen} is the number of ways that ancestors and sons are consistent with the information we have, and T_{ng}^{Gen} is the number of ways that ancestors and sons are consistent with the information we have about ancestors only. Unlike the equations for P_{ng}^{Gen}, there are factors of N that depend on how many generations are included of ancestors about whom we know nothing. We will denote the factors for generations 4 and higher as N^{τ}. Since they are the same in both Q_{ng}^{Gen} and T_{ng}^{Gen}, they will ultimately cancel out when calculating probabilities.

$$Q_{ng}^{Gen}(K, S, G_1, G_2, G_3, S_1, S_2, S_3, Gen(1), \ldots, Gen(ng))$$

$$= S!\,(K-S_1)! * \overline{\binom{Gen(1)-G_1}{K-S_1}} * \overline{\binom{Gen(1)-K}{\beta(1)-S}} * N^{Gen(1)+Gen(2)+Gen(3)+\tau-K+S_1-G_1} \text{ [if } 0 < S \leq \beta(1)]$$

$$= Gen(1)! * (S - \beta(1))! * (K - S_1 - S_2)! * \sum_{K_1=S_1}^{Gen(1)-G_1+S_1} (\overline{\binom{Gen(1)-G_1}{K_1-S_1}} * \overline{\binom{Gen(2)-G_2}{K-K_1-S_2}} * \overline{\binom{Gen(2)-K+K_1}{\beta(2)-S}}) * N^{Gen(1)+Gen(2)+Gen(3)+\tau-K+S_1+S_2-G_1-G_2} \text{ [if } \beta(1) < S \leq \beta(2)]$$

$$= Gen(1)! * Gen(2)! * (S - \beta(2))! * (K - S_1 - S_2 - S_3)! * \sum_{K_1=S_1}^{Gen(1)-G_1+S_1} \sum_{K_2=S_2}^{Gen(2)-G_2+S_2} \left(\overline{\binom{Gen(1)-G_1}{K_1-S_1}} * \overline{\binom{Gen(2)-G_2}{K_2-S_2}} * \overline{\binom{Gen(3)-G_3}{K-K_1-K_2-S_3}} * \overline{\binom{Gen(3)-K+K_1+K_2}{\beta(3)-S}} \right) * N^{Gen(1)+Gen(2)+Gen(3)+\tau-K+S_1+S_2+S_3-G_1-G_2-G_3}$$

$$\text{[if } \beta(2) < S \leq \beta(3)]$$

$$= Gen(1)! * Gen(2)! * \ldots * Gen(g+1)! \, (K - S_1 - S_2 - S_3)! * \frac{\binom{S-G_1-G_2-G_3}{K-S_1-S_2-S_3}}{(\beta(g+1)-S)!} * N^{Gen(1)+Gen(2)+Gen(3)+\tau-K+S_1+S_2+S_3-G_1-G_2-G_3} \, [\beta(g) < S \leq \beta(g+1)] \text{ and } 3 < g \quad (40)$$

$$T_{ng}^{Gen}(K, S, G_1, G_2, G_3, S_1, S_2, S_3, Gen(1), \ldots, Gen(ng))$$

$$= \frac{Gen(1)! * Gen(2)! * \ldots * Gen(g+1)!}{(\beta(g+1)-S)!} * N^{Gen(1)+Gen(2)+Gen(3)+\tau-G_1-G_2-G_3} \text{ [if } \beta(g) < S \leq \beta(g+1)]$$

(41)

Equations (39), (40) and (41) are valid provided the following inequalities hold and we set the values of P_3^{Gen}, Q_3^{Gen} and T_3^{Gen} to 0 when the inequalities are violated.

- $0 \leq S_i \leq G_i$
- $0 \leq G_1 \leq Gen(1), 0 \leq G_2 \leq Gen(2), 0 \leq G_3 \leq Gen(3)$
- $\sum_{i=1}^{3} S_i \leq K$
- $0 < K \leq S$
- If $S_2 \neq 0$ then $S > \beta(1)$
- If $S_3 \neq 0$ then $S > \beta(2)$

Similar to equation (37) we define P_{Tree}^{Gen} as

$$P_{Tree}^{Gen}(mean, K, G_1, G_2, G_3, S_1, S_2, S_3, Gen(1), Gen(2), Gen(3)) \equiv \lim_{ng \to \infty} \left(\sum_{S=1}^{\beta(ng)} P_{ng}^{Gen}(K, S, G_1, G_2, G_3, S_1, S_2, S_3, Gen(1), \ldots, Gen(ng)) * f_{Poisson}(S, mean) \right)$$

(42)

A case study – Jonatan of Drobin

Introduction of the null hypothesis

The methods of statistical and combinatorial analysis will be illustrated in what follows, with the examination of a hypothesis regarding Jonas (JD), a.k.a. Rabbi Jonatan Eybeschuetz of Drobin, and his wife Dwojra. Dwojra's father was (according to certain Drobin vital records) named Jacob. Jonas lived from approximately 1720 to approximately 1768. As it happens, there are several family histories (typically found in the prefaces to 19th century rabbinical works) from which one would deduce that JD's father was named Aron, and that Aron's father-in-law was named Meyer.[16] However, such histories and deductions are occasionally inaccurate. For the purposes of the analyses that follow, I will assume that there is some possibility of error in these documents regarding the claim that Aron was Jonas's father, and I will somewhat arbitrarily propose a 20% probability that the claim was invalid (by which I mean this: whether or not it turns out that Aron really was the father of Jonas, that conclusion cannot be drawn from the aforementioned documents, since their conclusions are based on erroneous information). In addition to the anecdotal information we have about Aron, there are also many vital records from Drobin and elsewhere, dating from approximately 1808, of JD's descendants. Then, based on the information in these vital records, I will apply Bayes' (equation (1)) theorem to derive the probability that Aron was JD's father. In applying Bayes' theorem, I will assume the above proposed 20% probability that, based on anecdotal evidence, Aron was JD's father.[17]

There are reasons to believe that there was in fact an Aron who, if not the father of JD, was his uncle – and that Aron's father-in-law was Meyer. So the assumption here will be that there is 100% probability that Meyer is JD's father's father-in-law provided that Aron is JD's father. Separately, based on anecdotal information including some gravestone inscriptions, also assume that JD has (with 100% probability) a grandfather named Nuta. Finally, also based on anecdotal information, assume that JD has two great-grandfathers named Eyzyk and Mosze. (The names of the great-grandfathers will be ignored in most analyses below, and will only be considered in certain cases where they are explicitly mentioned.) All of this information is summarized in the following chart.

Documented Ancestors of Jonas and Dwojra

Eyzyk + wife	Mosze + wife	X_3 + wife	X_4 + wife	X_5 + wife	X_6 + wife	X_7 + wife	X_8 + wife
Nuta	wife of Nuta	Meyer	wife of Meyer	X_1	wife of X_1	X_2	wife of X_2
	Aron		wife of Aron	Jacob		wife of Jacob	
			Jonas	Dwojra			

Chart 13

The question to be answered is whether the records of descendants of JD support or reject the following null hypothesis:

$$H_{\emptyset}^{JD}: \quad \text{"Aron is the father of JD" with } P\left(H_{\emptyset}^{JD}\right) = 0.80 \qquad (43)$$

with an alternative hypothesis:

[16] Most information about JD's family, has been provided to me by Dr. Heshel Teitelbaum. An example (amongst many) of such anecdotal information can be found in (Blokherovitsh, 1939) page 16

ט.) חתן הרה״נ מוח״ר יונתן אייבעשיץ אבר״ק דרובנין

י) בהרה״ג 'חמופלג מוה״ר אהרן פירר אייבעשיץ אבד״ק שטעדבורז

[17] It's important to take account, as we have done, of the anecdotal information about Aron. Otherwise, we would be examining a very different scenario where we would be making probabilistic analyses based on the assumed frequency of the first name Aron in the general Jewish population. That number can by hypothesized to be about 2%, based on a study of Ellis Island records as published by Yannay Spitzer in 2012 at https://yannayspitzer.net/2012/07/24/most-common-jewish-names/ with similar results at http://www.jewishgen.org/databases/USA/1890nyNames.htm.

$$H_A^{JD}: \quad \text{“}\textit{Aron} \text{ is not the father of JD”} \text{ with } P\left(H_A^{JD}\right) = 0.20 \qquad (44)$$

The null hypothesis will be rejected if, based on statistical inference, the probability of $H_\emptyset^{JD}$ given the evidence from records of JD's descendants, is less than the *level of significance.* For this paper, the *level of significance* will be set to 5%.

Children and grandsons of Jonas of Drobin and Dwojra

The analysis which follows will be based on records of the children and grandsons of JD and his wife Dwojra. It's important to note that those records are somewhat fluid. As time passes, it's reasonably likely that new records will be found which will either reveal new children or grandchildren. Also, there are a few sub-branches for which the names or connections to JD are moderately speculative. There may or may not eventually be further evidence one way or the other about these sub-branches. Here is the current tree (limited to children and grandsons) to be used in this paper.

2. Jonas (~1720 – ~1768) + Dwojra
    - 2.1. Jacob (1743 –?) + wife
        - 2.1.1. Nachman-Wulf (1778 –?) (*double names will be hyphenated*)
        - 2.1.2. Meyer-Nuta (1791 – before 1866)
    - 2.2. Daughter + Jacob-Lipman (1743 –?) (son of Haim)
        - 2.2.1. Israel (? – ?)
        - 2.2.2. Jonas (1774 –?)
    - 2.3. Abraham (? – before 1810) + Laja (? – ~1770)
        - 2.3.1. Leyzor (1749 – 1830)
        - 2.3.2. Jonas (~1769 – 1821)

            Abraham (? – before 1810) + Szajndel (? – ?)

        - 2.3.3. Mosze-Haim (1770 – 1833)
        - 2.3.4. Chaskel (1780 – 1831)
    - 2.4. Fajga (1743 – 1823) + Lewek (1743 – 1819) (son of Jacob)
        - 2.4.1. Jacob (1771 –?)
        - 2.4.2. Jonas (1799 – 1863)
    - 2.5. Eyzyk (1768 – 1827) + Hudes (1780 – 1826) (daughter of Haim Meyer (son of Jacob))
        - 2.5.1. Jonas (1809 – 1810)
        - 2.5.2. Hersz (1811 –?)
        - 2.5.3. Jacob (1817 –?)
        - 2.5.4. Mathias (1819 –?)
        - 2.5.5. Eliasz-Lewek (1823 –?)
    - 2.6. Tanchum (? – ?) + wife (*Assume equal probability that JD's child is either Tanchum or Tanchum's wife)*
        - 2.6.1. Abraham (1778 – 1847)
    - 2.7. Mosze (? – ?) + wife (*Assume equal probability that JD's child is either Mosze or Mosze's wife)*

2.8. Hersz (aka Naftali Hersz) (? – ?) + wife
   2.8.1. Szmuel (? – ?)

The default set of sons is

Jacob	Abraham	Eyzyk	Tanchum	Mosze	Hersz

Chart 14

The stage is now set for analysis.

Sons of Rabbi Jonas of Drobin

To the list of General Assumptions above, and to the known facts mentioned above about the ancestors of Jonas and his wife, add the following assumptions that are specific to JD.

JD1. The fathers of Jonas and his wife died before the birth of their first-born son, the grandfathers of Jonas and his wife died before the birth of the third-born son and the great-grandfathers of Jonas and his wife died before the birth of the seventh-born son. (This assumption permits an analysis in which the names of all male ancestors, are available for the naming of sons.)

JD2. Jonas and Dwojra enjoyed, as a couple, 25 child-bearing years.

JD3. Nothing is known about any male ancestors earlier than the great-grandfathers (that's actually not true, but the assumption will be made here for simplicity of analysis).

From Chart 13, and from list 2 above, we can set the parameters of equation (36) to be $K = 6, G_1 = 2, G_2 = 2, G_3 = 2, S_1 = 1, S_2 = 0, S_3 = 2$. The probability, $q(S)$, that Jonas had S sons is, from assumption GA3, $f_{Poisson}(S, 5)$. Therefore, using the notation of equation (37)

$$P\left("JD\ sons"\middle|H_{\emptyset}^{JD}\right) = P_{Tree}(5,6,2,2,2,1,0,2) = 0.0015 * N^{-5} \quad (45)$$

where the term "JD sons" means "known sons of Jonas" and N has the same meaning as earlier

Next, analyze $P\left("JD\ sons"\middle|H_{A}^{JD}\right)$, i.e., the probability of observing the known sons of Jonas, given the alternate hypothesis that Aron was not the father of Jonas. In this case, the parameters of equation (36) must be adjusted to remove both Aron and his father-in-law Meyer, and we have

$$P\left("JD\ sons"\middle|H_{A}^{JD}\right) = P_{Tree}(5,6,1,1,2,1,0,2) = 0.0094 * N^{-5} \quad (46)$$

As above, we can now apply the Bayes equation (equation (38)).

$$P\left(H_{\emptyset}^{JD}\middle|"JD\ sons"\right) = P_{Bayes}\left(P\left(H_{\emptyset}^{JD}\right), P\left("JD\ sons"\middle|H_{\emptyset}^{JD}\right), P\left("JD\ sons"\middle|H_{A}^{JD}\right)\right)$$

Substituting the values in equations from (43), (44), (45) and (46)

$$P\left(H_{\emptyset}^{JD}\middle|"JD\ sons"\right) = 0.39 \quad (47)$$

In words, our analysis shows a 39% probability, of the validity of the null hypothesis (Aron is the father of JD), given the names of the 6 sons of Jonas found in the records. This is certainly much higher than the level of significance (5%) that has been chosen in this paper, and therefore we cannot reject the null hypothesis on the basis of the sons of Jonas.

Now we can see how $P\left(H_{\emptyset}^{JD} \middle| "JD\ sons"\right)$ varies depending on variations of the family tree. Some variations of interest, including the one above, are given in the following table.

Table 1 – Sons of Jonas

Differences from Chart 13 and Chart 14	$\emptyset: (\boldsymbol{M}, \boldsymbol{K}, \boldsymbol{G_1}, \boldsymbol{G_2}, \boldsymbol{G_3}, \boldsymbol{S_1}, \boldsymbol{S_2}, \boldsymbol{S_3})$ $\boldsymbol{p_\emptyset = P_{Tree}(\emptyset)}$	$\boldsymbol{A}: (\boldsymbol{M}, \boldsymbol{K}, \boldsymbol{G_1}, \boldsymbol{G_2}, \boldsymbol{G_3}, \boldsymbol{S_1}, \boldsymbol{S_2}, \boldsymbol{S_3})$ $\boldsymbol{p_A = P_{Tree}(A)}$	$\boldsymbol{P_{Bayes}(0.8, p_\emptyset, p_A)}$
No difference	(5,6,2,2,2,1,0,2) $p_\emptyset = 0.0015 * N^{-3}$	(5,6,1,1,2,1,0,2) $p_A = 0.0094 * N^{-3}$	0.39
Null hypothesis doesn't include Meyer	(5,6,2,1,2,1,0,2) $p_\emptyset = 0.0042 * N^{-3}$	(5,6,1,1,2,1,0,2) $p_A = 0.0094 * N^{-3}$	0.64
Chart 14 without Tanchum	(5,5,2,2,2,1,0,2) $p_\emptyset = 0.00070 * N^{-2}$	(5,5,1,1,2,1,0,2) $p_\emptyset = 0.0021 * N^{-2}$	0.58
Chart 14 without Mosze	(5,5,2,2,2,1,0,1) $p_\emptyset = 0.00070 * N^{-3}$	(5,5,1,1,2,1,0,2) $p_\emptyset = 0.00070 * N^{-3}$	0.34
Chart 14 without Mosze or Tanchum	(5,4,2,2,2,1,0,1) $p_\emptyset = 0.0045 * N^{-2}$	(5,4,2,2,2,1,0,1) $p_\emptyset = 0.015 * N^{-2}$	0.55

Although the analysis of children of Jonas have included the 3 ancestor-generations of Jonas and his wife, we will, in the analysis of children of Jonas, ignore the information we have about Jonas' great-great-grandparents (these are in the 4th ancestor-generation of the children of Jonas. This will simplify the calculations and will not greatly alter the results.

Sons of Jacob

There are two significant ways in which the analysis of Jacob's family tree differs from the cases studied previously. First of all, the known sons have double names. Secondly, Jacob has the same name as one of his grandfathers, and therefore no son is named after that grandfather. In order to proceed, we must therefore add to the list of general assumptions above, and to the known information mentioned above about the ancestors of Jonas and his wife. The following assumptions are specific to Jacob, the son of JD.

J1. When sons are (mostly) given name-pairs[18]:
  a. Name-pairs correspond to ancestors from the same generation when possible.
  b. When it's not possible for name-pairs to be from a single ancestor-generation, then only single names are used.
  c. Assume also that names are chosen in order so that near-generations are exhausted before the next ancestor-generation is used – but that within a generation, names are chosen at random.

[18] These assumptions might not be borne out through a detailed survey of naming patterns. We will discuss this later on in the paper.

J2. The fathers of Jacob and his wife died before the birth of their first-born son, the grandfathers of Jacob and his wife died before the birth of the third-born son, and the great-grandfathers of Jacob and his wife died before the birth of the seventh-born son. (This assumption permits an analysis in which the names of all male ancestors, are available for the naming of sons.)

J3. Jacob and his wife enjoyed, as a couple, 25 child-bearing years.

J4. Nothing is known about any male ancestors earlier than the great-grandfathers (that's actually not true, but the assumption will be made here for simplicity of analysis and could be checked later, to presumably remove it as an assumption).

Jacob's family tree, with and without the null hypothesis is shown in Chart 15 and Chart 16.

Null hypothesis $H_{\emptyset}^{JD}$: Ancestors of Jacob and wife

Nuta + wife	Meyer + wife	X_4+wife	X_5+wife	X_6+wife	X_7+wife	X_8+wife	X_9+wife
Aron	wife	Jacob	wife	X_2	wife	X_3	wife
Jonas		wife		X_1		wife	
Jacob				wife			

Chart 15

As usual, the names X_i aren't known. It will be convenient below to let $Z_1 =$ Meyer, $Z_2 =$ Nuta, and $Z_i = X_{i+1}$ for $3 \leq i \leq 8$. The steps of the analysis follow.

JN1. We will see shortly that Jacob must have had at least 4 sons. For $4 \leq S \leq 7$, we compute the total number of possible combinations of sons.

a. Jacob's first son would be named after Jacob's father and father-in-law, with a factor of N possible names for X_1.
   i. Jonas-X_1

b. His next two sons would be named in one of the following combinations (noting that Jacob would not give the name Jacob to one of his sons) and in either order (resulting in a factor of 2), with a factor of N^2 possible names for X_2 and X_3.
   i. Aron-X_2, X_3
   ii. Aron-X_3, X_2
   iii. Aron, X_2-X_3

c. The 4th son would be named after a pair (Z_i, Z_j) for $1 \leq i < j \leq 8$. There are $\binom{8}{2}$ such possibilities. The 5th son would be named after a pair selected from the remaining 6 names and so on. The names can be in any order therefore resulting in an extra factor of $(S-3)!$. Altogether, for the 4th through 7th sons, the number of combinations is $(S-3)! * \frac{\prod_{i=4}^{S}\binom{16-2*i}{2}}{(S-3)!}$ with a factor of N^6 possible names for X_4 through X_9.[19]

JN2. There is a factor of N for the unknown ancestor of generation 1. The son Nachman-Wulf is either named after the 2nd generation or the 3rd generation of ancestors.

---

[19] We will frequently encounter the expression $\prod_{i=m}^{n}\binom{k-2*i}{2}$. This can be simplified although in general I have preferred to leave the expression in its unsimplified form. The simplified form is $\prod_{i=m}^{n}\binom{k-2*i}{2} = \frac{(k-2*m)!}{(k-2*(n+1))!*2^{n+1-m}} = \binom{k-2*m}{2*(n+1-m)} * \frac{(2*(n+1-m))!}{2^{n+1-m}}$.

a. If he was named after the 2nd generation, then the names of X_2 and X_3 are Nachman and Wulf in either order (so there are $2 * N^7$ options).

b. If he was named after the 3rd generation, then there are $\frac{6!}{4!}$ possible ways for 2 of the 6 unknown great-grandfathers to be named (so there are $\frac{6!}{4!} * N^7$ possible options)

JN3. Now analyze, for each of JN2.a and JN2.b, the number of combinations of sons consistent with the information we have about the sons of Jacob.

a. (For each combination of ancestors named as in JN2.a) The first son is Jonas-X_1. The next two sons are Aron and Nachman-Wulf, in either order (so a factor of 2). Amongst the next sons, one is named Meyer-Nuta, and if there are other sons (up to 7 sons), their names are pairs chosen from the remaining 6 great-grandfathers. So for 5 or more sons, (using an analysis similar to JN1.c) there are $(S-3)! * \frac{\prod_{i=5}^{S}\binom{16-2*i}{2}}{(S-4)!}$ options.

b. (For each combination of ancestors as named as in JN2.b). The first son is Jonas-X_1. The next two sons are in one of three combinations – (Aron-X_2, X_3), (Aron-X_3, X_2), (Aron, X_2-X_3) – in either order (so a factor of $2 * 3$). The remaining $S-3$ sons can be named in any order, so there is a factor of $(S-3)!$. Two of those sons are Meyer-Nuta and Nachman-Wulf. For the remaining sons, there are (similar to before) $\frac{\prod_{i=6}^{S}\binom{16-2*i}{2}}{(S-5)!}$ options.

JN4. Putting together the above, the results for 4 through 7 sons are therefore:

a. $P_{Jacob}(4\ sons|\emptyset) = \frac{2}{3*\binom{8}{2}*\binom{4}{2}} * N^{-2}$

b. $P_{Jacob}(5\ sons|\emptyset) = \frac{2}{3*\binom{8}{2}*\binom{5}{2}} * 16 * N^{-2}$

c. $P_{Jacob}(6\ sons|\emptyset) = \frac{2}{3*\binom{8}{2}*\binom{6}{2}} * 21 * N^{-2}$

d. $P_{Jacob}(7\ sons|\emptyset) = \frac{2}{3*\binom{8}{2}*\binom{7}{2}} * 40 * N^{-2}$

JN5. For greater than 7 sons, we will proceed by induction. Also, for simplicity, ignore terms of the form N^{τ} of the kind used in equation (40). These cancel out in the computation of probabilities.

a. The derivation uses the following notation:

i. $Q_J(S|\emptyset) \equiv$ the number of combinations of S sons, consistent with Chart 15 and with further ancestor generations (as required for naming the sons) consisting of unknown ancestors.

ii. $Q_J(S|\emptyset') \equiv$ the number of combinations of S sons, consistent with Chart 15 and with further ancestor generations (as required for naming the sons) consisting of unknown ancestors, except assuming Jacob's only known son is Meyer-Nuta.

iii. $T_J(S|\emptyset) \equiv$ the number of combinations of S sons, consistent with the known information about Jacob's ancestors (only) in Chart 15 and with further ancestor generations (as required for naming the sons) consisting of unknown ancestors.

iv. $Gen_J(i) \equiv$ the number of males in ancestor generation i.

v. $\beta_J(g) \equiv \sum_{i=1}^{g}\lceil Gen_J(i)/2 \rceil$ where the notation ⌈⌉ denotes the rounding-up of fractions.[20]

b. The inductive assumptions for $S \geq 7$ and $\beta_J(g) < S \leq \beta_J(g+1)$, are:

   i. $Q_J(S|\emptyset) = 16 * (3 * S - 11) * \prod_{i=5}^{7}\binom{16-2*i}{2} *$
   $\dots * \prod_{i=\beta_J(g)+1}^{S}\binom{2*(\beta_J(g+1)+1-i)}{2} * N^{2*\beta_J(g+1)-7}$

   ii. $Q_J(S|\emptyset') = 24 * \prod_{i=5}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g)+1}^{S}\binom{2*(\beta_J(g+1)+1-i)}{2} * N^{2*\beta_J(g+1)-5}$

   iii. $T_J(S|\emptyset) = 6 * \prod_{i=4}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g)+1}^{S}\binom{2*(\beta_J(g+1)+1-i)}{2} * N^{2*\beta_J(g+1)-5}$

c. The initial step of the inductive proof will be $S = 7$. Compare the values above, for $S = 7$, with the values from JN1 through JN3. The inductive proofs follow.

d. Consider $\beta_J(g+1) < S \leq \beta_J(g+2)$. Partition the cases into those where Nachman-Wulf is named after one of the first $g+1$ generations, and those where Nachman-Wulf is named after generation $g+2$.

   i. If he was named after one of the first $g+1$ generations, then the first $\beta_J(g+1)$ sons are named as in the calculation of $Q_J(S|\emptyset)$ for $\beta_J(g) < S \leq \beta_J(g+1)$ and the later sons are named after generation $g+2$. The number of ways those later sons can be named are $\frac{(S- \beta_J(g+1))! * \prod_{i= \beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2}}{(S- \beta_J(g+1))!} *$
   $N^{2*(\beta_J(g+2)- \beta_J(g+1))}$. The total contribution from this partition is therefore
   $Q_J(\beta_J(g+1)|\emptyset) * \frac{(S- \beta_J(g+1))! * \prod_{i= \beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2}}{(S- \beta_J(g+1))!} *$
   $N^{2*(\beta_J(g+2)- \beta_J(g+1))} = 16 * (3 * \beta_J(g+1) - 11) * \prod_{i=5}^{7}\binom{16-2*i}{2} *$
   $\dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-7}$.

   ii. If Nachman-Wulf was named after an ancestor in generation $g+2$ then the first $\beta_J(g+1)$ sons are named as in the calculation of $Q_J(S|\emptyset')$ for $\beta_J(g) < S \leq \beta_J(g+1)$. In generation $g+2$, two ancestors are named Nachman and Wulf, and this can happen in $2 * \binom{Gen_J(g+2)}{2} * N^{2*(\beta_J(g+2)-\beta_J(g+1)-1)}$ ways. The sons beyond number $\beta_J(g+1)$ (in other words, sons who are younger than the oldest ≥sons) are named after generation $g+2$, but one of them is named after Nachman and Wulf. The number of such combinations is $(S - \beta_J(g+1))! *$
   $\frac{\prod_{i=\beta_J(g+1)+2}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2}}{(S-\beta_J(g+1)-1)!}$ where I will adopt the convention that this term equals 1 if $S = \beta_J(g+1)+1$. So the total contribution from this partition is
   $Q_J(\beta_J(g+1)|\emptyset') * 2 * \binom{Gen_J(g+2)}{2} * (S - \beta_J(g+1)) *$
   $\prod_{i=\beta_J(g+1)+2}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*(\beta_J(g+2)-\beta_J(g+1)-1)} = 24 * 2 * (S -$
   $\beta_J(g+1)) * \prod_{i=5}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-7}$.

[20] As an example, $\left\lceil\frac{3}{2}\right\rceil = 2$. We use this here because in generation 2, there are only 3 ancestors whose names can be used for naming the sons of Jacob, and therefore one son has a name-pair, and the other has a single name.

e. For $\beta_J(g+1) < S \le \beta_J(g+2)$, the calculation of $Q_J(S|\emptyset')$ is straightforward. The total of combinations is $Q_J\big(\beta_J(g+1)\big|\emptyset'\big) * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\left(\beta_J(g+2)-\beta_J(g+1)\right)} = 24 * \prod_{i=5}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-5}$.

f. Again, consider $\beta_J(g+1) < S \le \beta_J(g+2)$. The total of possible combinations consistent with ancestors, is $T_J\big(\beta_J(g+1)\big|\emptyset\big) * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\left(\beta_J(g+2)-\beta_J(g+1)\right)} = 6 * \prod_{i=4}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-5}$.

g. Finally, putting this all together, we obtain for $\beta_J(g+1) < S \le \beta_J(g+2)$, the proof of the induction step:

$$Q_J(S|\emptyset) = \left(16 * \left(3 * \beta_J(g+1) - 11\right) + 24 * 2 * (S - \beta_J(g+1))\right) \prod_{i=5}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-7}$$

$$= 16 * (3 * S - 11) * \prod_{i=5}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-7}$$

$$Q_J(S|\emptyset') = 24 * \prod_{i=5}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-5}$$

$$T_J(S|\emptyset) = 6 * \prod_{i=4}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-5}$$

h. We now can use the inductive assumptions above (and remembering the selection factor $\binom{S}{2}$), to derive the probability equation for $S > 7$.

$$P_{Jacob}(S\ sons|\emptyset) = \frac{Q_J(S|\emptyset)}{T_J(S|\emptyset)} = \frac{16*(3*S-11)}{6*\binom{8}{2}*\binom{S}{2}} * N^{-2}$$

JN6. The final result for the null hypothesis is then

$$P(\text{"}Jacob\ sons\text{"}|H_\emptyset^{JD}) = \sum_{S=4}^{\infty} P_{Jacob}(S\ sons|\emptyset) * f_{Poisson}(N,5) = 0.023 * N^{-2} \qquad (48)$$

Next, analyze $P(\text{"}J\ sons\text{"}|H_A^{JD})$, i.e., the probability of observing the known sons of Jacob, given the alternate hypothesis that Aron was not the father of Jonas.

Alternate hypothesis H_A^{JD}: Ancestors of Jacob and wife

Nuta + wife	X_5 + wife	X_6+wife	X_7+wife	X_8+wife	X_9+wife	X_{10}+wife	X_{11}+wife
X_2	wife	Jacob	wife	X_3	wife	X_4	wife
Jonas		wife		X_1		wife	
Jacob				wife			

The names X_i aren't known. It will be convenient below to let $Z_1 =$ Nuta, and $Z_i = X_{i+3}$ for $1 \leq i \leq 8$. The steps of the analysis follow.

JA1. We will see shortly that Jacob must have had at least 4 sons. For $4 \leq S \leq 7$, we compute the total number of possible combinations of sons.

- b. Jacob's first son would be named after his father and father-in-law, with a factor of N possible names for X_1.
  - i. Jonas-X_1
- b. His next two sons would be named in one of the following combinations (noting that Jacob would not give the name Jacob to one of his sons) and in either order (resulting in a factor of 2), with a factor of N^3 possible names for $X_{1,}X_2$ and X_3.
  - i. X_i-X_j, X_k where $1 \leq i < j \leq 3$, $1 \leq k \leq 3$, and $i \neq k \neq j$
- c. The 4th son would be named after a pair (Z_i, Z_j) for $1 \leq i < j \leq 8$. There are $\binom{8}{2}$ such possibilities. The 5th son would be named after a pair selected from the remaining 6 names and so on. The names can be in any order therefore resulting in an extra factor of $(S-3)!$. Altogether, for the 4th through 7th sons, the number of combinations is $(S-3)! * \frac{\prod_{i=4}^{S}\binom{16-2*i}{2}}{(S-3)!}$ with a factor of N^7 possible names for X_5 through X_{11}.

JA2. There is a factor of N for the unknown ancestor of generation 1. The son Nachman-Wulf is either named after the 2nd generation or the 3rd generation of ancestors.

- a. If he was named after the 2nd generation, then the possible name assignments of Nachman and Wulf are (X_1, X_2), (X_1, X_3) and (X_2, X_3) in either order (so there are $6 * N$ options). Furthermore, the name Meyer would be assigned to one of the X_i's in the 3rd generation ($7 * N^6$ options).
- b. If he was named after the 3rd generation, then there are N^3 possible options for the names in the 2nd generation. In the 3rd generation, one of the X_i's is assigned to Meyer, and of the remaining six X_i's, two are assigned – in either order – to Nachman and Wulf, hence a total of $7 * \binom{6}{2} * 2 * N^4$ possible options.

JA3. Now analyze, for each of JA2.a and JA2.b, the number of combinations of sons consistent with the information we have about the sons of Jacob.

- a. (For each combination of ancestors named as in JA2.a) The first son is Jonas-X_1. The next two sons are X_i and Nachman-Wulf, in either order (so a factor of 2). Amongst the next sons, one is named Nuta-X_j, and if there are other sons (up to 7 sons), their names are pairs chosen from the remaining 6 great-grandfathers. So for 5 or more sons, (using an analysis similar to JA1.c) there are $(S-3)! * \frac{\prod_{i=5}^{S}\binom{16-2*i}{2}}{(S-4)!}$ options.
- b. (For each combination of ancestors as named as in JA2.b). The first son is Jonas-X_1. The next two sons are in one of three combinations – $[(X_1 - X_2), X_3]$, $[(X_1 - X_3), X_2]$, $[(X_2 - X_3), X_1]$ – in either order (so a factor of $2 * 3$). The remaining $S-3$ sons can be named in any order, so there is a factor of $(S-3)!$. Two of those sons are Nuta-X_j and Nachman-Wulf. For the remaining sons, there are (similar to before) $\frac{\prod_{i=6}^{S}\binom{16-2*i}{2}}{(S-5)!}$ options.

JA4. Putting together the above, the results for 4 through 7 sons are therefore:

a. $P_{Jacob}(4\ sons|A) = \frac{14}{\binom{8}{2}*\binom{4}{2}} * N^{-3}$

b. $P_{Jacob}(5\ sons|A) = \frac{42}{\binom{8}{2}*\binom{5}{2}} * N^{-3}$

c. $P_{Jacob}(6\ sons|A) = \frac{42}{\binom{8}{2}*\binom{6}{2}} * 3 * N^{-3}$

d. $P_{Jacob}(7\ sons|A) = \frac{56}{\binom{8}{2}*\binom{7}{2}} * 4 * N^{-3}$

JA5. For greater than 7 sons, we will proceed by induction similarly to JN5.

a. The derivation uses the following notation:

i. $Q_J(S|A) \equiv$ the number of combinations of S sons, consistent with Chart 16 and with further ancestor generations (as required for naming the sons) consisting of unknown ancestors.

ii. $Q_J(S|A') \equiv$ the number of combinations of S sons, consistent with Chart 16 and with further ancestor generations (as required for naming the sons) consisting of unknown ancestors, except assuming Jacob's only known son is Meyer-Nuta.

iii. $T_J(S|A) \equiv$ the number of combinations of S sons, consistent with the known information about Jacob's ancestors (only) in Chart 16 and with further ancestor generations (as required for naming the sons) consisting of unknown ancestors.

iv. $Gen_J(i) \equiv$ the number of males in ancestor generation i.

v. $\beta_J(g) \equiv \sum_{i=1}^{g}[Gen_J(i)/2]$ where the notation [] denotes the rounding-up of fractions.[21]

b. The inductive assumptions for $S \geq 7$ and $\beta_J(g) < S \leq \beta_J(g+1)$, are:

i. $Q_J(S|A) = 336 * (S-3) * \prod_{i=5}^{7}\binom{16-2*i}{2} * \ldots * \prod_{i=\beta_J(g)+1}^{S}\binom{2*(\beta_J(g+1)+1-i)}{2} * N^{2*\beta_J(g+1)-6}$

ii. $Q_J(S|A') = 168 * \prod_{i=5}^{7}\binom{16-2*i}{2} * \ldots * \prod_{i=\beta_J(g)+1}^{S}\binom{2*(\beta_J(g+1)+1-i)}{2} * N^{2*\beta_J(g+1)-3}$

iii. $T_J(S|A) = 6 * \prod_{i=4}^{7}\binom{16-2*i}{2} * \ldots * \prod_{i=\beta_J(g)+1}^{S}\binom{2*(\beta_J(g+1)+1-i)}{2} * N^{2*\beta_J(g+1)-3}$

c. The initial step of the inductive proof will be $S = 7$. Compare the values above, for $S = 7$, with the values from JA1 through JA3. The inductive proofs follow.

d. Consider $\beta_J(g+1) < S \leq \beta_J(g+2)$. Partition the cases into those where Nachman-Wulf is named after one of the first $g+1$ generations, and those where Nachman-Wulf is named after generation $g+2$.

i. If he was named after one of the first $g+1$ generations, then the first $\beta_J(g+1)$ sons are named as in the calculation of $Q_J(S|A)$ for $\beta_J(g) < S \leq \beta_J(g+1)$ and the later sons are named after generation $g+2$. The number of ways those later sons can be named are $\frac{\left(S-\beta_J(g+1)\right)! * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2}}{(S-\beta_J(g+1))!} * N^{2*(\beta_J(g+2)-\beta_J(g+1))}$. The total contribution from this partition is therefore

[21] As an example, $\left\lceil\frac{3}{2}\right\rceil = 2$. We use this here because in generation 2, there are only 3 ancestors whose names can be used for naming the sons of Jacob, and therefore one son has a name-pair, and the other has a single name.

$Q_J(\beta_J(g+1)|A) * \frac{(S-\beta_J(g+1))! * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2}}{(S-\beta_J(g+1))!} *$
$N^{2*(\beta_J(g+2)-\beta_J(g+1))} = 336 * (\beta_J(g+1)-3) * \prod_{i=5}^{7}\binom{16-2*i}{2} *$
$\dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-6}$.

ii. If Nachman-Wulf was named after an ancestor in generation $g+2$ then the first $\beta_J(g+1)$ sons are named as in the calculation of $Q_J(S|A')$ for $\beta_J(g) < S \leq \beta_J(g+1)$. In generation $g+2$, two ancestors are named Nachman and Wulf, and this can happen in $2 * \binom{Gen_J(g+2)}{2} * N^{2*(\beta_J(g+2)-\beta_J(g+1)-1)}$ ways. The sons beyond number $\beta_J(g+1)$ (in other words, sons who are younger than the oldest ≥sons) are named after generation $g+2$, but one of them is named after Nachman and Wulf. The number of such combinations is $(S-\beta_J(g+1))! * \frac{\prod_{i=\beta_J(g+1)+2}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2}}{(S-\beta_J(g+1)-1)!}$ where I will continue to follow the convention that this term equals 1 if $S = \beta_J(g+1)+1$. So the total contribution from this partition is $Q_J(\beta_J(g+1)|A') * 2 * \binom{Gen_J(g+2)}{2} * (S-\beta_J(g+1)) * \prod_{i=\beta_J(g+1)+2}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*(\beta_J(g+2)-\beta_J(g+1)-1)} = 168 * 2 * (S - \beta_J(g+1)) * \prod_{i=5}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-6}$.

e. For $\beta_J(g+1) < S \leq \beta_J(g+2)$, the calculation of $Q_J(S|A')$ is straightforward. The total of combinations is $Q_J(\beta_J(g+1)|A') * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*(\beta_J(g+2)-\beta_J(g+1))} = 168 * \prod_{i=5}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-3}$.

f. Again, consider $\beta_J(g+1) < S \leq \beta_J(g+2)$. The total of possible combinations consistent with ancestors, is $T_J(\beta_J(g+1)|A) * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*(\beta_J(g+2)-\beta_J(g+1))} = 6 * \prod_{i=4}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-3}$.

g. Finally, putting this all together, we obtain for $\beta_J(g+1) < S \leq \beta_J(g+2)$, the proof of the induction step:

$$Q_J(S|A) = (336 * (\beta_J(g+1)-3) + 168 * 2 * (S-\beta_J(g+1))) \prod_{i=5}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-6}$$

$$= 336 * (S-3) * \prod_{i=5}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-6}$$

$$Q_J(S|A') = 168 * \prod_{i=5}^{7}\binom{16-2*i}{2} * \dots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-3}$$

$$T_J(S|A) = 6 * \prod_{i=4}^{7}\binom{16-2*i}{2} * \ldots * \prod_{i=\beta_J(g+1)+1}^{S}\binom{2*(\beta_J(g+2)+1-i)}{2} * N^{2*\beta_J(g+2)-3}$$

h. We now can use the inductive assumptions above (and remembering the selection factor $\binom{S}{2}$), to derive the probability equation for $S > 7$.

$$P_{Jacob}(S\ sons|A) = \frac{Q_J(S|A)}{T_J(S|A)} = \frac{336 * (S-3)}{6 * \binom{8}{2} * \binom{S}{2}} * N^{-3}$$

JA6. The final result for the alternative hypothesis is then

$$P(\text{"}Jacob\ sons\text{"}|H_A^{JD}) = \sum_{S=4}^{\infty} P_{Jacob}(S\ sons|A) * f_{Poisson}(N,5) = 0.17 * N^{-3} \qquad (49)$$

Again, we can apply the Bayes equation (equation (1)).

$$P(H_\emptyset^{JD}|\text{"}Jacob\ sons\text{"}) = \frac{P(H_\emptyset^{JD}) * P(\text{"}Jacob\ sons\text{"}|H_\emptyset^{JD})}{P(H_\emptyset^{JD}) * P(\text{"}Jacob\ sons\text{"}|H_\emptyset^{JD}) + P(H_A^{JD}) * P(\text{"}Jacob\ sons\text{"}|H_A^{JD})}$$

Substitute the values in equations from (43), (44), (48) and (49)

$$P(H_\emptyset^{JD}|\text{"}Jacob\ sons\text{"}) = \frac{0.018}{0.018 + \frac{0.034}{N}} \approx 1.0 \qquad (50)$$

The interpretation of this result is straightforward: Since one of Jacob's sons was named Nuta-Meyer, and since the null hypothesis has one of the great-grandfathers named Meyer, then the above result essentially tells us that, in the alternate case where we don't know of any ancestors named Meyer, it would have been 'too much of a coincidence' for Jacob to have given the name Meyer to one of his two known sons. How much of a coincidence? That depends on the value of N. From examining the references footnote 17, it would be reasonable to assume that N is on the order of about 100 and therefore that it would be extremely unlikely for an unknown ancestor (through generation 3) to randomly be named Meyer.

Now consider the situation where Meyer isn't part of the null hypothesis – in other words, assume as in Chart 15 that Aron is the father of Jonas, but do not assume that Meyer is the father-in-law of Aron. We can follow the kind of analysis done above to obtain the following:

$$P_{Jacob}(4\ sons|\emptyset - Meyer) = \frac{14}{3 * \binom{8}{2} * \binom{4}{2}} * N^{-3}$$

$$P_{Jacob}(5\ sons|\emptyset - Meyer) = \frac{14}{3 * \binom{8}{2} * \binom{5}{2}} * 8 * N^{-3}$$

$$P_{Jacob}(6\ sons|\emptyset - Meyer) = \frac{14}{3 * \binom{8}{2} * \binom{6}{2}} * 21 * N^{-3}$$

$$P_{Jacob}(7\ sons|\emptyset - Meyer) = \frac{14}{3 * \binom{8}{2} * \binom{7}{2}} * 40 * N^{-3}$$

$$P_{Jacob}(S\ > 7\ sons|\emptyset - Meyer) = \frac{14}{3*\binom{8}{2}*\binom{S}{2}} * 4 * (3 * S - 11) * N^{-3}$$

From these equations, we obtain

$$P\left(\text{"}Jacob\ sons\text{"}\middle|H^{JD}_{\emptyset-Meyer}\right) = \sum_{S=4}^{\infty} P_{Jacob}(S\ sons|\emptyset - Meyer) * f_{Poisson}(N, 5) = 0.14 * N^{-3} \quad (51)$$

Then applying Bayes theorem, and substituting from (43), (44), (49) and (51)

$$P\left(H^{JD}_{\emptyset-Meyer}\middle|\text{"}Jacob\ sons\text{"}\right) = \frac{0.11}{0.11+0.03} = 0.76 \qquad (52)$$

The results above are summarized in the following table:

Table 2 – Sons of Jacob

Differences from Chart 15 and Chart 16	$p_\emptyset$	p_A	$P_{Bayes}(0.8, p_\emptyset, p_A)$
No difference	$p_\emptyset = 0.023 * N^{-2}$	$p_A = 0.17 * N^{-3}$	0.96 if N=50
Null hypothesis doesn't include Meyer	$p_\emptyset = 0.14 * N^{-3}$	$p_A = 0.17 * N^{-3}$	0.76

Alternative assumptions about double-names

In the analysis of Jacob's sons, we relied heavily on assumption J1. This assumption is rather weak in several respects. To begin with, there are many Jewish records showing double names where the first name in the pair corresponds to an ancestor from a different generation than the second name in the pair. In addition, there are many records – including in JD's family tree – where some sons have double names and some sons appear to have single names. Moreover, even where sons appear to have single names, subsequent discoveries may reveal that the sons actually had double names but were most often known by only one of the two names. Needless to say, the general situation appears to be unmanageably complicated. Still, if a son has the name Nachman-Wulf, it seems reasonable to infer that he had ancestors with the names Nachman and Wulf and therefore we need to try and assign some probability to this. The above analysis was the result of such an attempt.

Here is an approach based on different assumptions which hopefully are reasonably realistic,[22] but which are computationally more tractable than the assumptions used above. If none of the sons are known to have double names, then we'll assume that all sons have single names. However, if any of the known sons have double names, we'll assume they all do (however, in some of the situations examined later, there will be times when we explicitly treat a double-name as a singular event where we assume for example, that the son was named after an ancestor with the same double name). This is the situation analyzed below.

D1. Assumptions: Even though some of the known sons might only be known by one of their two names, we will assume that they have second names which are unknown.
   1. If there are S sons, then there are $2 * S$ ancestors after whom those sons are named. We will assume that the naming rules for each of those $2 * S$ names, to be the same rules (GA1 through GA4GA6) we would use if there were $2 * S$ sons each with a single name.

[22] What ultimately needs to be examined, is how sensitive to the assumptions the conclusions are.

2. Recall that the above naming rules will result in chronologically ordered sets of names for the sons. Since these are double-names, we will adopt the following chronological rule: a correct chronologically-ordered set of double names is one in which the first names of the pairs are in the same chronological order as they were amongst the $2 * S$ names.
3. All correct chronologically-ordered pairs are equally likely to occur.[23]
4. The random selection rule GA5 can be applied.

D2. Other than for points below, the analysis of probabilities follows that which leads to Equation (39). Recall that in computing the probabilities of Equation (39), we arrived at an intermediate step where we enumerated the ways in which S sons could be named in chronological order. For the case of double-names, that enumeration counts the ways in which $2 * S$ names can be assigned "in chronological order" to the S sons. However, following D1.2 and D1.3, only some of those pairings will be regarded as "correct chronological order".

1. Start by considering the total, T, of configurations of S sons consistent with the known ancestors. There are $2 * S$ names $(V_1, V_2, V_3, \dots, V_{2*S})$. V_1 can be paired with any of the names that follow, so there are $2 * S - 1$ such pairings. Let V_a be the first (in numerical order) name not already assigned to the first pair. Then V_a can be paired with any of the $2 * S - 3$ remaining names that follow. This method of pairing then continues until there are no names left. The total number of pairs is[24]

$$T_D(S) = (2 * S - 1) * (2 * S - 3) * \dots * 1 = \frac{(2*S)!}{S! * 2^S} \quad (53)$$

2. Next, consider the total, Q, of configurations of S sons consistent with the known ancestors <u>and</u> the known sons. Again, designate the sons by the set $(V_1, V_2, V_3, \dots, V_{2*S})$. Let K_s be the number of known sons where we know only one of their two names (recall that in this analysis we are assuming all sons have double names, even if they are known by only one of their names). Then $K_D = K - K_s$ is the number of known sons where we know both of their two names. Let $Q_D(S, K_D, K_S)$ be the number of correct pairs that can be constructed out of the $2 * S$ names. (Note that the total number, T, of pairs

---

[23] This assumption seems flawed, but appears to greatly simplify the calculations. As an example of the flaw, consider a family with 5 sons, each with double-names $V_i - V_j$, where the V_i are in chronological order. Here are two possible examples.

- $(V_1 - V_2, V_3 - V_4, V_5 - V_6, V_7 - V_8, V_9 - V_{10})$
- $(V_1 - V_6, V_2 - V_7, V_3 - V_8, V_4 - V_9, V_5 - V_{10})$

The first example seems reasonable. The first son is born and since the family doesn't know if other sons will be born, they give a double name that covers the 'most important' ancestors (generally the most recent ancestors). This situation is similar to the one considered earlier when analyzing the sons of Jacob. In the second example, the first son is named after an 'important' ancestor and is also named after a 'less important' (i.e., earlier) ancestor. If the family were confident that there would be other sons, then that might be all right, but how could they be that confident?

[24] The equality can be found in (Abramowitz & Stegun, 1964) and can be deduced by noticing the following: $(2 * S) * (2 * S - 2) * (2 * S - 4) * \dots * 2 = 2 * S * 2 * (S - 1) * 2 * (S - 2) \dots = 2^S * S!$. If we divide $(2 * S)!$ by $(2 * S) * (2 * S - 2) * \dots$, we are left with $(2 * S - 1) * (2 * S - 3) * \dots$

computed above in part (1) is $Q_D(S,0,0)$.) When counting the number of correctly ordered pairs $V_i - V_j$, we exclude the pairs corresponding to the K_D known double-names. That would leave us with the total number of 'potentially' correct pairs as $\overline{Q_D}(S,K_D) = \frac{(2*(S-K_D))!}{(S-K_D)!*2^{(S-K_D)}}$. However, that needs to be adjusted by excluding all pairs from amongst the K_S names, since we know that those names cannot be paired with one another. Suppose that $K_S = 2$. We would need to adjust $\overline{Q_D}$ by subtracting all correct pairs which include the pair that consists of the two single names, so $Q_D(S,K_D,2) = \overline{Q_D}(S,K_D) - \overline{Q_D}(S-2,K_D)$. What about if $K_S = 3$? There are $\binom{3}{2}$ ways of selecting a pair from the three single names. In each case, one of the names is left over so no further pairs can be created from those names. The result is therefore $Q_D(S,K_D,3) = \overline{Q_D}(S,K_D) - \overline{Q_D}(S-2,K_D) * \binom{3}{2}$. For $K_S = 4$, there are $\binom{4}{2}$ ways of selecting the 'first' pair that can be constructed from the 4 single names. It is instructive to look at an example. Suppose the 4 names are (V_1,V_2,V_3,V_4) and that for each of the $\binom{4}{2}$ ways of selecting the first pair, we subtract all collections of pairs which include that first pair. One of the ways of choosing the first pair is $V_1 - V_2$ so we would subtract all collections of pairs which include $V_1 - V_2$. Another way of choosing the first pair is $V_3 - V_4$ so we would also subtract all collections of pairs which include $V_3 - V_4$. But this would result in over-subtraction, since that collection would also include – amongst others – the pair $V_1 - V_2$ which was already subtracted. Therefore, we need to add back all collections which have two pairs from (V_1,V_2,V_3,V_4). The number of ways of picking two pairs without repetition, is $\frac{\binom{4}{2}*\binom{2}{2}}{2!}$. So altogether the result for $K_S = 4$ is $Q_D(S,K_D,4) = \overline{Q_D}(S,K_D) - \overline{Q_D}(S-2,K_D) * \binom{4}{2} + \overline{Q_D}(S-4,K_D) * \frac{\binom{4}{2}*\binom{2}{2}}{2!}$. This can be generalized.

$$Q_D(S,K_D,2*M) = \overline{Q_D}(S,K_D) + \sum_{k=1}^{M}(-1)^k \overline{Q_D}(S-2*k,K_D) * \frac{(2*M)!}{2^k*k!*(2*M-2*k)!} \tag{54}$$

$$Q_D(S,K_D,2*M+1) = \overline{Q_D}(S,K_D) + \sum_{k=1}^{M}(-1)^k \overline{Q_D}(S-2*k,K_D) * \frac{(2*M+1)!}{2^k*k!*(2*M+1-2*k)!} \tag{55}$$

3. We are now ready to calculate the probability of selecting known double-named sons consistent with known information about ancestors, where K_S of the known sons are known by one name, and K_D of the known sons are known by two names. We refer to equations (39), (53)
4. for definitions, and we note that the selection factor should be $\binom{S}{2}$ rather than $\binom{2*S}{2}$.

$$P_{3-ng}^{Gen}(K_D + K_S, S, G_1, G_2, G_3, S_1, S_2, S_3, Gen(1), \ldots, Gen(ng), K_S.K_D) \equiv$$

$$P_{ng}^{Gen}(2*K_D + K_S, 2*S, G_1, G_2, G_3, S_1, S_2, S_3, Gen(1), \ldots, Gen(ng)) * \frac{Q_D(S,K_D,K_S)}{T_D(S)} * \frac{\binom{2*S}{2}}{\binom{S}{2}} \tag{56}$$

As for equations (37) and (42) define P_{3Tree}^{Gen-D} as

$$P_{Tree}^{Gen-D}(mean, K_D + K_S, G_1, G_2, G_3, S_1, S_2, S_3, Gen(1), Gen(2), Gen(3), K_S.K_D) \equiv$$

$$\lim_{ng\to\infty}\left(\sum_{S=1}^{\infty} P_{ng-D}^{Gen}(K_D+K_S,S,G_1,G_2,G_3,S_1,S_2,S_3,Gen(1),\ldots,Gen(ng),K_S.K_D) * f_{Poisson}\,(S,mean)\right) \tag{57}$$

We can apply this equation to the analysis of Jacob's sons, and compare those results to Table 2.

Table 2a – Sons of Jacob, analyzed using assumptions D1

Differences from Chart 15 and Chart 16	$p_\emptyset$	p_A	$P_{Bayes}(0.8, p_\emptyset, p_A)$
No difference	$p_\emptyset = 0.0021 * N^{-2}$	$p_A = 0.043 * N^{-3}$	0.91 if N=50
Null hypothesis doesn't include Meyer	$p_\emptyset = 0.026 * N^{-3}$	$p_A = 0.043 * N^{-3}$	0.71

Table 2 and Table 2a are in reasonable agreement for the Bayes probabilities, which ultimately is what we care most about. However, the values of $p_\emptyset$ and p_A in Table 2a are both an order of magnitude smaller than in Table 2. That would indicate the possibility of significant differences between predictions based on assumptions J1 and D1.

The other single-spouse children of JD

From the family tree of JD, we note that following children of JD are assumed – based on existing records – to have had only one spouse:

- Daughter and Jacob-Lipman
- Fajga and Lewek
- Eyzyk and Hudes
- Tanchum and wife
- Mosze and wife
- Naftali-Hersz and wife

For each of those couples, we make the following assumptions, to be added to general assumptions and other information used previously about Jonas and his wife.

JS1. The fathers died before the birth of their first-born son, the grandfathers died before the birth of the third-born son, and the great-grandfathers died before the birth of the seventh-born son.

JS2. The couple had 25 child-bearing years.

Wife of Jacob-Lipman

Null hypothesis $H_\emptyset^{JD}$: Ancestors of Jacob-Lipman and wife

X_3 + wife	X_4 + wife	X_5 + wife	X_6 + wife	Nuta + wife	Meyer + wife	X_7 + wife	X_8 + wife
X_1	wife of X_1	X_2	wife of X_2	Aron	wife of Aron	Jacob	wife of Jacob
	Haim		wife of Haim	Jonas		Dwojra	
			Jacob-Lipman	wife			

Chart 17

Sons

Jonas	Israel

Alternative hypothesis H_A^{JD}: Ancestors of Jacob-Lipman and wife

X_4 + wife	X_5 + wife	X_6 + wife	X_7 + wife	Nuta + wife	X_8 + wife	X_9 + wife	X_{10} + wife
X_1	wife of X_1	X_2	wife of X_2	X_3	wife of X_3	Jacob	wife of Jacob
	Haim		wife of Haim	Jonas		Dwojra	
			Jacob-Lipman	wife			

Chart 18

Although this situation is very similar to the one assumed in the derivation of Equation (36), it differs in one important respect. Since the father's name is Jacob-Lipman, then (according to our assumptions) none of his sons would be named Jacob – which is also the name of Jonas' father-in-law. This is the situation considered in equation (39) with $Gen(1) = 2, Gen(2) = 3, Gen(3) = 8$. Define

$$P'_3(K, S, G_1, G_2, G_3, S_1, S_2, S_3) \equiv P_3^{Gen}(K, S, G_1, G_2, G_3, S_1, S_2, S_3, 2,3,8)$$

Also note that the G_2 does not include the name Jacob. Also define (see equation (42))

$$P'_{Tree}(mean, K, G_1, G_2, G_3, S_1, S_2, S_3) \equiv P_{Tree}^{Gen}(mean, K, G_1, G_2, G_3, S_1, S_2, S_3, 2,3,8) \quad (58)$$

Now, using Equations (38) and (58) we can compute the various probabilities for Chart 17 and Chart 18 as well as the variant where Meyer is not included in the null hypothesis.

Table 3 – Sons of Jacob-Lipman

Differences from Chart 17 and Chart 18 Chart 14	$\emptyset: (M, K, G_1, G_2, G_3, S_1, S_2, S_3)$ $p_\emptyset = P'_{Tree}(\emptyset)$	$A: (M, K, G_1, G_2, G_3, S_1, S_2, S_3)$ $p_A = P'_{Tree}(A)$	$P_{Bayes}(0.8, p_\emptyset, p_A)$
No difference	(5,2,2,1,2,1,0,0) $p_\emptyset = 0.17 * N^{-1}$	(5,2,2,0,1,1,0,0) $p_A = 0.24 * N^{-1}$	0.73
Null hypothesis doesn't include Meyer	(5,2,2,1,1,1,0,0) $p_\emptyset = 0.17 * N^{-1}$	(5,2,2,0,1,1,0,0) $p_A = 0.24 * N^{-1}$	0.74

Fajga

Null hypothesis $H_\emptyset^{JD}$: Ancestors of Lewek and Fajga

X_3 + wife	X_4 + wife	X_5 + wife	X_6 + wife	Nuta + wife	Meyer + wife	X_7 + wife	X_8 + wife
X_1	wife of X_1	X_2	wife of X_2	Aron	wife of Aron	Jacob	wife of Jacob
	Jacob		wife of Jacob	Jonas		Dwojra	
			Lewek	Fajga			

Chart 19

Sons

Jonas	Jacob

Alternative hypothesis H_A^{JD}: Ancestors of Lewek and Fajga

X_4 + wife	X_5 + wife	X_6 + wife	X_7 + wife	Nuta + wife	X_8 + wife	X_9 + wife	X_{10} + wife
X_1	wife of X_1	X_2	wife of X_2	X_3	wife of X_3	Jacob	wife of Jacob
	Jacob		wife of Jacob	Jonas		Dwojra	
			Lewek	Fajga			

Chart 20

This situation is very similar to that of the family of Lipman-Jacob. The first two sons would be named Jacob and Jonas, therefore none of the later sons would be named after Dwojra's father Jacob. We can use Equations (38) and (58) to compute the various probabilities for Chart 19 and Chart 20 as well as the variant where Meyer is not included in the null hypothesis.

Table 4 – Sons of Lewek

Differences from Chart 19 and Chart 20	$\boldsymbol{\emptyset}: (\boldsymbol{M}, \boldsymbol{K}, \boldsymbol{G_1}, \boldsymbol{G_2}, \boldsymbol{G_3}, \boldsymbol{S_1}, \boldsymbol{S_2}, \boldsymbol{S_3})$ $\boldsymbol{p_\emptyset = P'_{Tree}(\emptyset)}$	$\boldsymbol{A}: (\boldsymbol{M}, \boldsymbol{K}, \boldsymbol{G_1}, \boldsymbol{G_2}, \boldsymbol{G_3}, \boldsymbol{S_1}, \boldsymbol{S_2}, \boldsymbol{S_3})$ $\boldsymbol{p_A = P'_{Tree}(A)}$	$\boldsymbol{P_{Bayes}(0.8, p_\emptyset, p_A)}$
No difference	(5,2,2,1,2,2,0,0) $p_\emptyset = 0.20$	(5,2,2,0,1,2,0,0) $p_A = 0.20$	0.80
Null hypothesis doesn't include Meyer	(5,2,2,1,1,2,0,0) $p_\emptyset = 0.20$	(5,2,2,0,1,2,0,0) $p_A = 0.20$	0.80

Eyzyk

Null hypothesis $H_\emptyset^{JD}$: Ancestors of Eyzyk and Hudes

X_3 + wife	X_4 + wife	X_5 + wife	X_6 + wife	Nuta + wife	Meyer + wife	X_7 + wife	X_8 + wife
Aron	wife of Aron	Jacob	wife of Jacob	X_1	wife of X_1	X_2	wife of X_2
	Jonas		Dwojra	Haim-Meyer		Wife of Haim-Meyer	
			Eyzyk	Hudes			

Chart 21

Sons

Jonas	Hersz	Jacob	Mathias	Eliasz-Lewek

Alternative hypothesis H_A^{JD}: Ancestors of Eyzyk and Hudes

X_4 + wife	X_5 + wife	X_6 + wife	X_7 + wife	Nuta + wife	X_8 + wife	X_9 + wife	X_{10} + wife
X_1	wife of X_1	Jacob	wife of Jacob	X_2	wife of X_2	X_3	wife of X_3
	Jonas		Dwojra	Haim-Meyer		Wife of Haim-Meyer	
			Eyzyk	Hudes			

Chart 22

Eyzyk's family has a few features that differ from previous families whose probabilities have been calculated.

- One of the ancestors, Haim-Meyer, has a double name. I will assume that there wouldn't be two sons Haim and Meyer named after this one ancestor. If a son is named after him, then that son will have the name Haim-Meyer, or Meyer or Haim. Therefore, the ancestor Haim-Meyer can be treated in just the same way as any ancestor with a single name but the three possible sons' names will need to be regarded as three independent possibilities.
- If a son is named either Haim-Meyer or Meyer, after the ancestor in generation 1, then there won't be another son named Meyer after the known ancestor in generation 3 (null hypothesis only). In that case, we will treat the third generation as having only 7 males.
- We will also consider separately the possibility that a son is named Haim after the ancestor in generation 1, in which case a later son might be named Meyer after the known ancestor in generation 3 (null hypothesis only).
- The son Eliasz-Lewek will be either assumed to have been named after two ancestors Eliasz and Lewek, or will be assumed to have been named after a single ancestor Eliasz-Lewek. In the first case, we will analyze the situation using assumptions of D1.

Based on the above, we will compute the probabilities for the following cases.

E1. A son named Haim – assign this event as having a probability of 1/3. In this case, we will treat all generations g as having the standard number of male ancestors – 2^g.

E2. A son named Meyer or a son named Haim-Meyer – assign each of these events as having a probability of 1/3. The third generation will be treated as having 7 male ancestors. (Note that this applies only to the null hypothesis.) In the case of Haim-Meyer, we will treat this exactly as though it were a single, rather than double, name. We can do this because we know that the son would be named after an ancestor with the name Haim-Meyer.

E3. For each of the cases above, Eliasz-Lewek will be treated as a single name, named after an ancestor whose name was Eliasz-Lewek.

E4. For each of the first two cases above, Eliasz-Lewek will be treated as a double name following assumptions of D1.

Define $p_\emptyset^E(x,y)$ to be the probability of the null hypothesis assuming assumptions x and y selected from assumptions E1 through E4. Similarly define $p_A^E(x,y)$ as the probability of the alternate hypothesis. Here are some examples using equation(39), where the null hypothesis has both Aron and Meyer included amongst ancestors as in Chart 21.

$$p_\emptyset^E(E1,E3) = P_{Tree}^{Gen}(5,5,2,2,2,1,1,0,2,4,8) = 0.058 * N^{-3} \quad (59)$$

$$p_\emptyset^E(E2,E3) = P_{Tree}^{Gen}(5,5,2,2,2,1,1,0,2,4,7) = 0.054 * N^{-3} \quad (60)$$

$$p_\emptyset^E(E1,E4) = P_{Tree}^{Gen-D}(5,5,2,2,2,1,1,0,2,4,8,4,1) = 0.27 * N^{-3} \quad (61)$$

Cases E1 and E2 are combined, leading to (for instance)

$$p_\emptyset^E(E3) = \frac{1}{3} * p_\emptyset^E(E1,E3) + \frac{2}{3} * p_\emptyset^E(E2,E3) = .055 * N^{-3}$$

Since $p_\emptyset^E(E1,E3)$ has almost the same value as $p_\emptyset^E(E2,E3)$, it doesn't make much difference what probabilities we assign to cases E1 and E2, so long as they add up to 1.

We can now compute the various probabilities for Chart 21 and Chart 20Chart 22 as well as the variant where Meyer is not included in the null hypothesis.

Table 5a – Sons of Eyzyk – case E3 (Eliasz-Lewek treated as single name)

$\boldsymbol{Arg}: (\boldsymbol{M}, \boldsymbol{K}, \boldsymbol{G_1}, \boldsymbol{G_2}, \boldsymbol{G_3}, \boldsymbol{S_1}, \boldsymbol{S_2}, \boldsymbol{S_3}, \boldsymbol{Gen1}, \boldsymbol{Gen2}, \boldsymbol{Gen3})$

Differences from Chart 21 and Chart 22	$\boldsymbol{p_\emptyset(E1) \equiv P^{Gen}_{Tree}(Arg(\emptyset, E1))}$ $\boldsymbol{p_\emptyset(E2) \equiv P^{Gen}_{Tree}(Arg(\emptyset, E2))}$ $\boldsymbol{p_\emptyset = \frac{1}{3} * p_\emptyset(E1) + \frac{2}{3} * p_\emptyset(E2)}$	$\boldsymbol{p_A(E1) \equiv P^{Gen}_{Tree}(Arg(A, E1))}$ $\boldsymbol{p_A(E2) \equiv P^{Gen}_{Tree}(Arg(A, E2))}$ $\boldsymbol{p_A = \frac{1}{3} * p_A(E1) + \frac{2}{3} * p_A(E2)}$	$\boldsymbol{P_{Bayes}(0.8, p_\emptyset, p_A)}$
No difference	$Arg(\emptyset, E1)$: (5,5,2,2,2,1,1,0,2,4,8,) $p_\emptyset(E1) = 0.058 * N^{-3}$ $Arg(\emptyset, E2)$: (5,5,2,2,2,1,1,0,2,4,7) $p_\emptyset(E2) = 0.054 * N^{-3}$ $p_\emptyset = 0.055 * N^{-3}$	$Arg(A, E1)$: (5,5,2,1,1,1,1,0,2,4,8) $p_A(E1) = 0.21 * N^{-3}$ $Arg(\emptyset, E2)$: (5,5,2,1,1,1,1,0,2,4,8) $p_A(E2) = 0.21 * N^{-3}$ $p_\emptyset = 0.21 * N^{-3}$	0.51
Null hypothesis doesn't include Meyer	$Arg(\emptyset, E1)$: (5,5,2,2,1,1,1,0,2,4,8) $p_\emptyset(E1) = 0.073 * N^{-3}$ $Arg(\emptyset, E2)$: (5,5,2,2,1,1,1,0,2,4,8) $p_\emptyset(E2) = 0.073 * N^{-3}$ $p_\emptyset = 0.073 * N^{-3}$	$Arg(A, E1)$: (5,5,2,1,1,1,1,0,2,4,8) $p_A(E1) = 0.21 * N^{-3}$ $Arg(\emptyset, E2)$: (5,5,2,1,1,1,1,0,2,4,8) $p_A(E2) = 0.21 * N^{-3}$ $p_\emptyset = 0.21 * N^{-3}$	0.58

Table 5b – Sons of Eyzyk – case E3 (Eliasz-Lewek treated as double name)

$\boldsymbol{Arg}: (\boldsymbol{M}, \boldsymbol{K}, \boldsymbol{G_1}, \boldsymbol{G_2}, \boldsymbol{G_3}, \boldsymbol{S_1}, \boldsymbol{S_2}, \boldsymbol{S_3}, \boldsymbol{Gen1}, \boldsymbol{Gen2}, \boldsymbol{Gen3}, \boldsymbol{K_S}, \boldsymbol{K_D})$

Differences from Chart 21 and Chart 22	$\boldsymbol{p_\emptyset(E1) \equiv P^{Gen-D}_{Tree}(Arg(\emptyset, E1))}$ $\boldsymbol{p_\emptyset(E2) \equiv P^{Gen-D}_{Tree}(Arg(\emptyset, E2))}$ $\boldsymbol{p_\emptyset = \frac{1}{3} * p_\emptyset(E1) + \frac{2}{3} * p_\emptyset(E2)}$	$\boldsymbol{p_A(E1) \equiv P^{Gen-D}_{Tree}(Arg(A, E1))}$ $\boldsymbol{p_A(E2) \equiv P^{Gen-D}_{Tree}(Arg(A, E2))}$ $\boldsymbol{p_A = \frac{1}{3} * p_A(E1) + \frac{2}{3} * p_A(E2)}$	$\boldsymbol{P_{Bayes}(0.8, p_\emptyset, p_A)}$
No difference	$Arg(\emptyset, E1)$: (5,5,2,2,2,1,1,0,2,4,8,4,1) $p_\emptyset(E1) = 0.27 * N^{-3}$ $Arg(\emptyset, E2)$: (5,5,2,2,2,1,1,0,2,4,7,4,1) $p_\emptyset(E2) = 0.24 * N^{-3}$ $p_\emptyset = 0.25 * N^{-3}$	$Arg(A, E1)$: (5,5,2,1,1,1,1,0,2,4,8,4,1) $p_A(E1) = 0.91 * N^{-3}$ $Arg(A, E2)$: (5,5,2,2,2,1,1,0,2,4,8,4,1) $p_A(E2) = 0.91 * N^{-3}$ $p_A = 0.91 * N^{-3}$	0.54
Null hypothesis doesn't include Meyer	$Arg(\emptyset, E1)$: (5,5,2,2,1,1,1,0,2,4,8,4,1) $p_\emptyset(E1) = 0.44 * N^{-3}$ $Arg(\emptyset, E2)$: (5,5,2,2,1,1,1,0,2,4,8,4,1) $p_\emptyset(E2) = 0.44 * N^{-3}$ $p_\emptyset = 0.44 * N^{-3}$	$Arg(A, E1)$: (5,5,2,1,1,1,1,0,2,4,8,4,1) $p_A(E1) = 0.91 * N^{-3}$ $Arg(A, E2)$: (5,5,2,2,2,1,1,0,2,4,8,4,1) $p_A(E2) = 0.91 * N^{-3}$ $p_A = 0.91 * N^{-3}$	0.66

Tanchum (or his wife)

Null hypothesis $H_\emptyset^{JD}$: Ancestors of Tanchum and wife

Nuta + wife	Meyer + wife	X_4 + wife	X_5 + wife	X_6 + wife	X_7 + wife	X_8 + wife	X_9 + wife
Aron	wife of Aron	Jacob	wife of Jacob	X_2	wife of X_2	X_3	wife of X_3
	Jonas		Dwojra	X_1		wife of X_1	
			Tanchum	wife of Tanchum			

Chart 23

Sons

Abraham

Alternative hypothesis H_A^{JD}: Ancestors of Tanchum and wife

Nuta + wife	X_5+ wife	X_6 + wife	X_7 + wife	X_8 + wife	X_9 + wife	X_{10} + wife	X_{11} + wife
X_2	wife of X_2	Jacob	wife of Jacob	X_3	wife of X_3	X_4	wife of X_4
	Jonas		Dwojra	X_1		wife of X_1	
			Tanchum	wife of Tanchum			

Chart 24

Note that the analysis of probabilities is independent of whether Tanchum is the son of JD, or whether his wife is the daughter of JD. We can compute probabilities as before.

Table 6 – Sons of Tanchum

Differences from Chart 23 and Chart 24	$\emptyset: (M,K,G_1,G_2,G_3,S_1,S_2,S_3)$ $p_\emptyset = P_{Tree}(\emptyset)$	$A: (M,K,G_1,G_2,G_3,S_1,S_2,S_3)$ $p_A = P_{Tree}(A)$	$P_{Bayes}(0.8, p_\emptyset, p_A)$
No difference	(5,1,1,2,2,0,0,0) $p_\emptyset = 0.13 * N^{-1}$	(5,1,1,1,1,0,0,0) $p_A = 0.17 * N^{-1}$	0.76
Null hypothesis doesn't include Meyer	(5,1,1,2,1,0,0,0) $p_\emptyset = 0.14 * N^{-1}$	(5,1,1,1,1,0,0,0) $p_A = 0.17 * N^{-1}$	0.77

Hersz (or his wife)

Null hypothesis $H_\emptyset^{JD}$: Ancestors of Hersz and wife

Nuta + wife	Meyer + wife	X_4 + wife	X_5 + wife	X_6 + wife	X_7 + wife	X_8 + wife	X_9 + wife
Aron	wife of Aron	Jacob	wife of Jacob	X_2	wife of X_2	X_3	wife of X_3
	Jonas		Dwojra	X_1		wife of X_1	
			Hersz	wife of Hersz			

Chart 25

Sons

Szmuel

Alternative hypothesis H_A^{JD}: Ancestors of Hersz and wife

Nuta + wife	X_5+ wife	X_6 + wife	X_7 + wife	X_8 + wife	X_9 + wife	X_{10} + wife	X_{11} + wife
X_2	wife of X_2	Jacob	wife of Jacob	X_3	wife of X_3	X_4	wife of X_4
	Jonas		Dwojra	X_1		wife of X_1	
			Hersz	wife of Hersz			

Chart 26

Note that the analysis of probabilities is independent of whether Hersz is the son of JD, or whether his wife is the daughter of JD. We can compute probabilities as before. The results are identical to those for Tanchum.

Table 7 – Sons of Hersz

Differences from Chart 25 and Chart 26	$\emptyset: (M,K,G_1,G_2,G_3,S_1,S_2,S_3)$ $p_\emptyset = P_{Tree}(\emptyset)$	$A: (M,K,G_1,G_2,G_3,S_1,S_2,S_3)$ $p_A = P_{Tree}(A)$	$P_{Bayes}(0.8, p_\emptyset, p_A)$
No difference	(5,1,1,2,2,0,0,0) $p_\emptyset = 0.13 * N^{-1}$	(5,1,1,1,1,0,0,0) $p_A = 0.17 * N^{-1}$	0.76
Null hypothesis doesn't include Meyer	(5,1,1,2,1,0,0,0) $p_\emptyset = 0.14 * N^{-1}$	(5,1,1,1,1,0,0,0) $p_A = 0.17 * N^{-1}$	0.77

Mosze (or his wife)

Mosze did not have any known sons. There is therefore no information whatsoever that can be derived from information about his sons. We can formalize this by setting, for Mosze, $p_\emptyset = p_A = 1$ and therefore $P_{Bayes}(0.8, p_\emptyset, p_A) = 0.8$.

Sons of Abraham and his two wives

From the family tree of JD, we see that Abraham had sons by two wives, Laja and Szajndel. Furthermore, his son Jonas was born after his son Leyzor since JD was still alive at the time of the birth of Leyzor. These facts change the analysis considerably from what has been done previously in this paper. The following assumptions will be used.

A1. Assumptions:

1. Abraham and Laja had 20 child-bearing years together. Abraham and Szajndel had 15 child-bearing years together.
2. Jonas was close to being Laja's last son (note that JD died a year later), but all other sons are named in the *standard* order (the assumption being that all other grandparents, etc. all predeceased the birthdates of Laja's other sons). This assumption will be implemented by dividing Laja's child-bearing years into period $P(1)$ of 18 years prior to the death of JD and period $P(2)$ of 2 years subsequent to the death of JD. Jonas was born in that second period.
3. Szajndel's children were named in the *standard* order.
4. Mosze-Chaim was the only one of Abraham's known children with a double name (separately from information presented in this paper, there is a considerable amount of

information available about Chaskel with no evidence that he had a second name). We'll assume that the name Chaim was either added on after Mosze recovered from an illness, or that Mosze-Chaim was named after an ancestor with the same name (i.e., we'll treat Mosze-Chaim as a single name).

Null hypothesis $H_{\emptyset}^{JD}$: Ancestors of Abraham and wives

$X4$	$X5$	$X6$	$X7$	Nuta	Meyer	$XY1$	$XY2$	$X11$	$X12$	$X13$	$X14$
$X2$	w-$X2$	$X3$	w-$X3$	Aron	w-Aron	Jacob	w-Jacob	$X9$	w-$X9$	$X10$	w-$X10$
	$X1$		w-$X1$		Jonas		Dwojra		$X8$		w-$X8$
		Laja				Abraham				Szajndel	

Chart 27

Sons of Abraham and Laja

Leyzor	Jonas

Sons of Abraham and Szajndel

Mosze-Chaim	Chaskel

Alternate hypothesis H_{A}^{JD}: Ancestors of Abraham and wives

$X4$	$X5$	$X6$	$X7$	Nuta	$XY2$	$XY3$	$XY4$	$X11$	$X12$	$X13$	$X14$
$X2$	w-$X2$	$X3$	w-$X3$	$XY1$	w-$XY1$	Jacob	w-Jacob	$X9$	w-$X9$	$X10$	w-$X10$
	$X1$		w-$X1$		Jonas		Dwojra		$X8$		w-$X8$
		Laja				Abraham				Szajndel	

Chart 28

A2. Compute $T_{ng}^{A}(S^L(1), S^L(2), S^S)$, the total number of configurations, consistent with Chart 27, of $S^L(i)$ sons of Laja during period $P(i)$ (see A1.2) and of S^S sons of Szajndel. The subscript ng denotes the number of ancestor-generations that are included in the analysis.

1. For $S^L(1) = 1$, $S^L(2) = 0$ Laja's son must be named X_1. For $S^L(1) = 0$, $S^L(2) = 1$ his name could be Jonas or X_1. So there is one possible configuration of Laja's sons in the first period and two possible configurations in the second period. We will use the notation introduced earlier of $Gen1, Gen2, Gen3$ to denote the number of *relevant* males in each ancestor-generation of Abraham and Szajndel. By a *relevant* male, I mean a male ancestor for which a son may be named. If Laja had exactly one son whose name was X_1 then Szajndel could name one of her sons Jonas – so $(Gen1, Gen2, Gen3) = (2,4,8)$. On the other hand, if Laja had exactly one son named Jonas, then Szajndel couldn't name one of her sons Jonas – so $(Gen1, Gen2, Gen3) = (1,4,8)$.
   Let $T_S(2, S^S, Gen1, Gen2, Gen3)$ be the number of configurations of Szajndel's S^S sons consistent with a possible configuration of ancestors. Then we get (noting as usual that there are N possibilities for each unknown ancestor)

$$T_3^A(1,0,S^S) = T_S(S^S, 2, 4, 8) * N^{16} \qquad (62)$$

$$T_3^A(0,1,S^S) = 2 * T_S(S^S, 1, 4, 8) * N^{16} \qquad (63)$$

2. For $S^L(1) = 2$, $S^L(2) = 0$ one of Laja's sons must be named X_1 and the other must be named after one of the 4 grandfathers, and the naming can be in either order – so a total of $1 * 4 * 2$. However, we need to distinguish between grandfathers of Abraham and grandfathers of Laja. So more precisely, there are 4 configurations that include a grandfather of Abraham, and 4 configurations that include a grandfather of Laja. In the first case, we need to multiply by $T_S(S^S, 2, 3, 8)$ and in the second case we multiply by $T_S(S^S, 2, 4, 8)$.

$$T_3^A(2,0,S^S) = (4 * T_S(S^S, 2, 3, 8) + 4 * T_S(S^S, 2, 4, 8)) * N^{16} \qquad (64)$$

3. For $S^L(1) = 1$, $S^L(2) = 1$ one of Laja's sons is named X_1 and the other is named Jonas. Exactly the same conclusion occurs when $S^L(1) = 0$, $S^L(2) = 2$, except that X_1 and Jonas can be in either order. We have

$$T_3^A(1,1,S^S) = T_S(S^S, 1, 4, 8) * N^{16} \qquad (65)$$

$$T_3^A(0,2,S^S) = 2 * T_S(S^S, 1, 4, 8) * N^{16} \qquad (66)$$

4. For the more general cases, we introduce yet more notation[25]. We set $t_{ng}\left(k, l, G_1, G_2, G_3, S(1), \dots, S(3), GenY(1), GenX(1), \dots, GenY(ng), GenX(ng)\right)$ to be the number of possible configurations consisting of k sons who are named after a possible configuration of ancestors of Abraham and l sons who are named after a possible configuration of ancestors of Laja, given $GenY(g)$ male ancestors of Abraham in generation g and $GenX(g)$ male ancestors of Laja in generation g – and furthermore consisting of $S(g)$ sons with names of known generation-g ancestors of Abraham, G_1 known 1st-generation ancestors of Abraham, G_2 known 2nd-generation ancestors of Abraham and G_3 known 3rd-generation ancestors of Abraham. Here, when we refer to ancestors, we mean *relevant* male ancestors (ones for whom sons can be named). We also define $\beta Y(g) \equiv \sum_{i=1}^{g} GenY(i)$ and $\beta X(g) \equiv \sum_{i=1}^{g} GenX(i)$. We also define $\beta Y(0) = \beta X(0)$. As usual, sons aren't named after generation g until previously born sons have been named after all males in generations prior to g (regardless of whether those sons were named after Laja's or Abraham's ancestors). Therefore

$$t_3\left(k, l, G_1, G_2, G_3, S(1), S(2), S(3), GenY(1), GenX(1), GenY(2), GenX(2), GenY(3), GenX(3)\right) = \prod_{g=1}^{3} \overline{\binom{GenY(g)-G(g)}{\alpha Y(k,g)-S(g)}} * \overline{\binom{GenX(g)}{\alpha X(l,g)}} * \overline{\binom{G(g)}{S(g)}} * \left(\alpha Y(k,g) + \alpha X(l,g)\right)! \text{ where}$$

$$\left(\alpha Y(k,1), \alpha Y(k,2), \alpha Y(k,3), \alpha X(l,1), \alpha X(l,2), \alpha X(l,3)\right) \equiv$$
$(k, 0,0, l, 0,0)$ if $0 \le k \le \beta Y(1)$ and $0 \le l \le \beta X(1)$,
$(GenY(1), k - \beta Y(1), 0, GenX(1), l - \beta X(1), 0)$ if $\beta Y(1) \le k \le \beta Y(2)$ and $\beta X(1) \le l \le \beta X(2)$,
$\left(GenY(1), GenY(2), k - \beta Y(2), GenX(1), GenX(2), l - \beta X(2)\right)$ if $\beta Y(2) \le k \le \beta Y(3)$ and $\beta X(2) \le l \le \beta X(3)$ a

[25] The notation chosen in this section on the sons of Abraham, assumes that for the entire analysis, we don't change our assumptions about what is known of the sons of Abraham, nor do we change what we know about the first-generation ancestors of Abraham.

and $G(i) \equiv 0$ for $i \neq 1, i \neq 2, i \neq 3$, $G(1) \equiv G_1, G(2) \equiv G_2, G(3) \equiv G_3$ (67)

Equation (67) can be simplified and easily generalized to t_g as follows. The above definitions of αY and αX result, for each of the three sets of inequalities, in one of the following for each of two values of g: either $(\alpha Y(k_{1,}g), \alpha X(l_1, g)) = (0,0)$ or $(\alpha Y(k_{1,}g), \alpha X(l_1, g)) = (GenY(g), GenX(g))$. In either case, $\overline{\binom{GenY(g)-G(g)}{\alpha Y(k,g)-S(g)}} * \overline{\binom{GenX(g)}{\alpha X(l,g)}} = 1$ if $(\alpha Y(k_{1,}g), S(g)) = (GenY(g), G(g))$ or if $(\alpha Y(k_{1,}g), S(g) = (0,0))$ otherwise $\overline{\binom{GenY(g)-G(g)}{\alpha Y(k,g)-S(g)}} * \overline{\binom{GenX(g)}{\alpha X(l,g)}} = 0$. For the remaining value of g, we have $\alpha Y(k_{1,}g) = k_1 - \beta Y(g-1)$ and $\alpha X(l_{1,}g) = l_1 - \beta X(g-1)$. This generalizes to

$$t_g\big(k, l, G_1, G_2, G_3, S(1), \dots S(3), GenY(1), GenX(1), \dots, GenY(ng), GenX(ng)\big) = \overline{\binom{GenY(gg)-G(gg)}{\alpha Y(k,gg)-S(gg)}} * \overline{\binom{GenX(gg)}{\alpha X(l,gg)}} * \prod_{g=1}^{gg-1} \big(GenY(g) + GenX(g)\big)! * * \overline{\binom{G(gg)}{S(gg)}} * \big(\alpha Y(k, gg) + \alpha X(l, gg)\big)!$$

gg is implicitly defined by $\beta Y(gg-1) \leq k \leq \beta Y(gg)$ and $\beta X(gg-1) \leq l \leq \beta X(gg)$. $\alpha Y(k_{,}gg) = k - \beta Y(gg-1), \alpha X(l, gg) = l - \beta X(gg-1)$ and if $(gg > 3,$ $i \leq gg)$ we have $\alpha Y(k_{,}i) = GenY(i), \alpha X(l, i) = GenX(i), S(i) = G(i)$. $S(i) = 0$ if $i >$ 3. (68)

For values of k and l which don't satisfy the set of lower and upper bounds above, the value of the function $t()$ will be set to 0. Also, note that when, for example, $k = \beta Y(i)$, then that value of k might appear on two of the lines above. However, it is easy to verify that there is no inconsistency between the two lines.

5. In computing $T_{ng}^{A}(S^L(1), S^L(2), S^S)$, the sequence of multiplicands is as follows:
   i. For period 1, compute $t_{ng}\big(k_1, l_1, 0, G_2, G_3, 0, S(2), S(3), 0, GenX(1), GenY(2), GenX(2), \dots, GenY(ng), GenX(3ng)\big)$ for each value of k_1 and l_1 such that $k_1 + l_1 = S^L(1)$, and for each value of $S(2)$ and $S(3)$ such that $0 \leq S(2) \leq G_2, 0 \leq S(3) \leq G_3$ and $0 \leq S(2) \leq \alpha Y(k_1, 2), 0 \leq S(3) \leq \alpha Y(k_1, 3)$.
   ii. For period 2, compute $t_{ng}\big(k_2, l_2, 1, G_2 - S(2), G_3 - S(3), 1, S'(2), S'(3), 1 - \alpha Y(k_1, 1), GenX(1) - \alpha X(l_1, 1), GenY(2) - \alpha Y(k_1, 2), GenX(2) - \alpha X(l_1, 2), \dots, GenY(ng) - \alpha Y(k_1, ng), GenX(ng) - \alpha X(l_1, ng)\big)$ for each value of k_2 and l_2 such that $k_2 + l_2 = S^L(2)$ and for each value of $S'(2)$ and $S'(3)$ such that $0 \leq S'(2) \leq G_2 - S(2), 0 \leq S'(3) \leq G_3 - S(3)$ and $0 \leq S'(2) \leq \overline{\alpha Y(k_2, 2)}, 0 \leq S'(3) \leq \overline{\alpha Y(k_2, 3)}$. Note that definitions of αY and αX etc. are the definitions applicable to the arguments of t_3 for period 1.
   iii. Then, for the sons of Szajndel, the factor is $T_{ng}^{Gen}\Big(2, S^S, 0, G_2 - S(2) - S'(2), G_3 - S(3) - S'(3), 0,0,0, Gen(1) - \alpha Y(k_1, 1) - \overline{\alpha Y(k_2, 1)}, Gen(2) - \alpha Y(k_1, 2) - \overline{\alpha Y(k_2, 2)}, \dots, Gen(ng) - \alpha Y(k_1, ng) - \overline{\alpha Y(k_2, ng)}\Big) *$

$N^{1+G_2+G_3+k_1+k_2-\tau-14}$, where T_{ng}^{Gen} is given by Equation (41) and $\overline{\alpha Y}$ is defined by

$$\overline{\alpha Y(k_2,i)} = k_2 - \overline{\beta Y(i-1)} \quad \text{where } \overline{\beta Y(i)} \equiv \sum_{j=1}^{i}(GenY(j) - \alpha Y(k_1,j))$$
$$\text{and } \overline{\beta Y(i-1)} \leq k_2 \leq \overline{\beta Y(i)} \qquad \text{and for } j < i,$$
$$\overline{\alpha Y(k_2,j)} = GenY(j) - \alpha Y(k_1,j)$$

The factor of N is present in order to cancel the factor in equation (41), where the term τ is defined.

iv. Finally, sum over all $k_1, l_1, k_2, l_2, S(2), S'(2), S(3), S'(3)$ such that $k_1 + l_1 = S^L(1)$, $k_2 + l_2 = S^L(2)$, $0 \leq S(2) + S'(2) \leq G_2$, $0 \leq S(3) + S'(3) \leq G_3$. Then multiply by a factor of $N^{20+\tau-G_2-G_3}$.

A3. We next analyze the computation of $Q_{ng}^A(S^L(1), S^L(2), S^S)$, the number of combinations consistent with known ancestors, as well as known sons, given the previous definitions of the arguments of $S^L(1), S^L(2)$ and S^S. Start by noting that Leyzor was born in period 1, and Jonas was born in period 2. Therefore, we need only consider situations where $S^L(1) \geq 1$ and $S^L(2) \geq 1$. We know that the son Jonas was named after Abraham's father Jonas. Leyzor could have been named after any of the unknown ancestors. Define $q_{ng}(k, l, G_1, G_2, G_3, S(1), S(2), S(3), GenY(1), GenX(1), \dots, GenY(ng), GenX(ng))$ to be the number of possible configurations in period 1 – consistent with the one known son – consisting of k sons who are named after possible[26] ancestors of Abraham and l sons who are named after possible ancestors of Laja, given $GenY(g)$ male ancestors of Abraham in generation g and $GenX(g)$ male ancestors of Laja in generation g – and furthermore consisting of $S(g)$ sons with names of known generation-g ancestors of Abraham, G_2 known 2nd-generation ancestors of Abraham and G_3 known 3rd-generation ancestors of Abraham We enumerate the possibilities by considering, in turn, the generation of the ancestor after whom Leyzor is named.

1. Leyzor is named after generation 1: In that case, Leyzor must be the name of Laja's father. It is easy to see that for any configuration of sons,

$$q_{ng}^{1,X}(k, l, G_1, G_2, G_3, S(1), S(2), S(3), GenY(1), GenX(1), \dots, GenY(ng), GenX(ng)) = t_{ng}(k, l, G_1, G_2, G_3, S(1), S\,(2), S\,(3), GenY(1), GenX(1), \dots, GenY(ng), GenX(ng))$$

$$q_{ng}^{1,Y}(\dots) = 0$$

where the superscript on $q_{ng}^{i,X}$ denotes the restriction of q_{ng} to the case where Leyzor is named after generation i of Laja's ancestors and $q_{ng}^{i,Y}$ denotes the restriction of q_{ng} to the case where Leyzor is named after generation i of Abraham's ancestors.

---

[26] Here, the term 'possible' refers to a configuration consistent with what is known. In particular, since the son Leyzor is named after one of his ancestors, then the possible ancestor-configurations need to include Leyzor. If we are limiting ourselves to sons named after g generations, then Leyzor would need to be an ancestor amongst those generations.

2. Leyzor is named after one of Abraham's ancestors in generation 2. The calculation of sons is the same as for the calculation of t_g in equation (68) except for three things: (a) multiply by the number of ways of assigning Leyzor to unknown ancestors in generation 2 – in this case $GenY(2) - G_2$ (b) There is one less *relevant* 2nd-generation ancestor of Abraham's from which to choose the sons that aren't named Leyzor (c) there is one less son to consider, from amongst the sons named after Abraham's ancestors but since Leyzor can appear in any order amongst sons named after generation 2, the term $\big(kY(k,g) + lX(l,g)\big)!$ in equation (68) needs to include Leyzor. The term which changes in equation (68) is $\overline{\binom{GenY(2)-G(2)}{\alpha Y(k,2)-S(2)}}$. That term becomes $\overline{\binom{GenY(2)-G(2)-1}{\alpha Y(k,2)-S(2)-1}} = \overline{\binom{GenY(2)-G(2)}{\alpha Y(k,2)-S(2)}} * \overline{\frac{\alpha Y(k,2)-S(2)}{GenY(2)-G(2)}}$.[27] We therefore get

$$q_{ng}^{2,Y}\big(k,l,G_1,G_2,G_3,S(1),S(3),GenY(1),GenX(1),\ldots\ GenY(ng),GenX(ng)\big) =$$
$$t_{ng}\big(k,l,G_1,G_2,G_3,S(1),S\ (2),S(3),GenY(1),GenX(1),\ldots,GenY(ng),GenX(ng)\big) * (GenY(2) - G_2) * \overline{\frac{\alpha Y(k,2)-S(2)}{GenY(2)-G(2)}} =$$
$$t_{ng}\big(k,l,G_1,G_2,G_3,S(1),S\ (2),S(3),GenY(1),GenX(1),\ldots,GenY(ng),GenX(ng)\big) * \ (\alpha Y(k,2) - S(2))$$

3. Leyzor is named after one of Laja's ancestors in generation 2. This can happen in one of 2 ways ($GenX(2) = 2$), provided that at least one of Laja's sons was born previously (therefore in period 1) and named after generation 1. Following the same logic as above,

$$q_{ng}^{2,X}\big(k,l,G_1,G_2,G_3,S(1),S(2),S(3),GenY(1),GenX(1),\ldots,GenY(ng),GenX(ng)\big) =$$
$$t_{ng}\big(k,l,G_1,G_2,G_3,S(1),S\ (2),S(3),GenY(1),GenX(1),\ldots,GenY(ng),GenX(ng)\big) * \alpha X(l,2)$$

4. Leyzor is named after one of Abraham's ancestors in generation 3. We can use the logic above.

$$q_{ng}^{3,Y}\big(k,l,G_1,G_2,G_3,S(1),S(2),S(3),GenY(1),GenX(1),\ldots,GenY(ng),GenX(ng)\big) =$$
$$t_{ng}\big(k,l,G_1,G_2,G_3,S(1),S(2),S\ (3),GenY(1),GenX(1),\ldots,GenY(ng),GenX(ng)\big) * \big(\alpha Y(k,3) - S(3)\big)$$

5. Leyzor is named after one of Laja's ancestors in generation 3. As before

$$q_{ng}^{3,X}\big(k,l,G_1,G_2,G_3,S(1),S(2),S(3),GenY(1),GenX(1),\ldots,GenY(ng),GenX(ng)\big) =$$
$$t_{ng}\big(k,l,G_1,G_2,G_3,S(1),S\ (2),S(3),GenY(1),GenX(1),\ldots,GenY(ng),GenX(ng)\big)$$

[27] I will introduce the notation $\overline{\frac{a}{b}}$ to denote $\frac{a}{b}$ when $b \neq 0$ but when $b = 0$, then $\overline{\frac{a}{b}} = 0$. In the context being used here, it is easy to see that this notation is required for the equality to hold.

$$* \alpha X(l, 3))$$

6. Since $S(g) = 0$ for $g > 3$ we can generalize the above to get

$$q_{ng}^{g,Y}\big(k, l, G_1, G_2, G_3, S(1), S(2), S(3), GenY(1), GenX(1), \dots, GenY(ng), GenX(ng)\big) = t_{ng}\big(k, l, G_1, G_2, G_3, S(1), S\ (2), S(3), GenY(1), GenX(1), \dots, GenY(ng), GenX(ng)\big) * \alpha Y(k, g)$$

and

$$q_{ng}^{g,X}\big(k, l, G_1, G_2, G_3, S(1), S(2), S(3), GenY(1), GenX(1), \dots, GenY(ng), GenX(ng)\big) = t_{ng}\big(k, l, G_1, G_2, G_3, S(1), S\ (2), S(3), GenY(1), GenX(1), \dots, GenY(ng), GenX(ng)\big) * \alpha X(l, g)$$

7. We now sum all of this to obtain q_3.

$$q_{ng}\big(k, l, G_1, G_2, G_3, S(1), S(2), S(3), GenY(1), GenX(1), \dots, GenY(ng), GenX(ng)\big) = t_{ng}\big(k, l, G_1, G_2, G_3, S(1), S\ (2), S(3), GenY(1), GenX(1), \dots, GenY(ng), GenX(ng)\big) * (k + l - S(1) - S(2) - S(3)\ ) \quad (69)$$

8. In computing $Q_{ng}^{A}(S^L(1), S^L(2), S^S)$, the sequence of multiplicands is as follows:
   i. For period 1, compute $q_{ng}\big(\ k_1, l_1, 0, G_2, G_3, 0, S(2), S(3), GenY(1), GenX(1), \dots, GenY(ng), GenX(ng)\big)$ for each value of k_1 and l_1 such that $k_1 + l_1 = S^L(1)$ and $l_1 > 0$ and for each value of $S(2)$ and $S(3)$ such that $0 \le S(2) \le G_2, 0 \le S(3) \le G_3$ and $0 \le S(2) \le \alpha Y(k_1, 2), 0 \le S(3) \le \alpha Y(k_1, 3)$. (Note that we set $GenY(1) = 0$ in period 1 because Jonas is Abraham's ancestor in the first generation, and since a son was named, in period 2, after Jonas, then the ancestor Jonas isn't *relevant* in period 1.)
   ii. For period 2, the number of combinations consistent with k_2 sons named after ancestors of Abraham and l_2 sons named after Laja, is $t_{ng}\big(\ k_2, l_2, 1, G_2 - S(2), G_3 - S(3), 1, S'(2), S'(3), 1 - \alpha Y(k_1, 1), GenX(1) - \alpha X(\ l_1, 1), GenY(2) - \alpha Y(k_1\ , 2), GenX(2) - \alpha X(\ l_1, 2), \dots, GenY(ng) - \alpha Y(k_1, ng), GenX(ng) - \alpha X(\ l_1, ng)\big)$ for each value of k_2 and l_2 such that $k_2 + l_2 = S^L(2)$ and for each value of $S'(2)$ and $S'(3)$ such that $0 \le S'^{(2)} \le G_2 - S(2), 0 \le S'(3) \le G_3 - S(3)$ and $0 \le S'(2) \le \overline{\alpha Y(k_2, 2)}, 0 \le S'(3) \le \overline{\alpha Y(k_2, 3)}$. (Note that we set to 1, the number of relevant Generation-1 ancestors of Abraham, since we know that Jonas was the relevant ancestor.) Note that definitions of αY and αX etc. are the definitions applicable to the arguments of q_{ng} for period 1.
   iii. Then, for the sons of Szajndel, the factor is $Q_{ng}^{Gen}\big(2, S^S, 0, G_2 - S(2) - S'(2), G_3 - S(3) - S'(3), 0,0,0, Gen(1) - \alpha Y(k_1, 1) - \overline{\alpha Y(k_2, 1)}, Gen(2) -$

$\alpha Y(k_1, 2) - \overline{\alpha Y(k_2, 2)}, \dots,\ Gen(ng) - \alpha Y(k_1, ng) - \overline{\alpha Y(k_2, ng)}\Big) *$
$N^{3+G_2+G_3+k_1+k_2-\tau-14}$ where Q_{ng}^{Gen} is given by equation(40) and $\overline{\alpha Y}$ is defined as in A2.5.iii. The factor of N is present in order to cancel the factor in equation(40) where the term τ is defined.

iv. Finally, sum over all $k_1, l_1, k_2, l_2, S(2), S'(2), S(3), S'(3)$ such that $k_1 + l_1 = S^L(1)$, $k_2 + l_2 = S^L(2)$, $0 \le S(2) + S'(2) \le G_2$, $0 \le S(3) + S'(3) \le G_3$. Then multiply by a factor of $N^{17+\tau-G_2-G_3}$.

A4. Next, we compute partial probabilities $P_{ng}^A(S^L(1), S^L(2), S^S)$, the probability – for each of $S^L(1), S^L(2)$ and S^S – of selecting the sons Jonas, Leyzor, Chaskel and Mosze-Chaim given the known ancestors. We use the previous definitions of the arguments of $S^L(1), S^L(2)$ and S^S, as well as Q_{ng}^A and T_{ng}^A.

$$P_{ng}^A(S^L(1), S^L(2), S^S) = \frac{Q_{ng}^A(S^L(1), S^L(2), S^S)}{T_{ng}^A(S^L(1), S^L(2), S^S) * \binom{S^L(1)}{1} * \binom{S^L(2)}{1} * \binom{S^S}{2}} \qquad (70)$$

A5. The final step is to compute the probability of selecting the sons Jonas, Leyzor, Chaskel and Mosze-Chaim given the known ancestors.

$$P_{Tree}^{Ab}(Sons, M1, M2, M3, G_2, G_3) = \sum_{s=1}^{\infty} \sum_{t=1}^{\infty} \sum_{u=1}^{\infty} P_{ng}^A(s, t, u) * f_{Poisson}(s, M1) * f_{Poisson}(t, M2) * f_{Poisson}(u, M3)$$

where $P_{ng}^A(s, t, u)$ is given by (70) (many of the arguments are suppressed) where $M1, M2$ and $M3$ are the Poisson expected values based on the time frames given in A1 as well as on assumption GA3.

$$M1 = 3.6, \quad M2 = 0.4, \quad M3 = 3.0 \qquad (71)$$

These results were computed (as were others previously) within Microsoft Excel, using the Visual Basic feature. The computation has considerably more complexity than the previous ones used to obtain probabilities for other sons of JD. Unlike the previous cases, I have not been able to show that results are independent of assumptions made about the number of ancestors in generations earlier than the 3rd generation. Therefore, the calculations have been explicitly done through 5 ancestral generations with the assumption that there are 16 males in generation 4 and 32 males in generation 5. The Poisson factors in equation (71) sharply suppress contributions from the earlier generations, so those contributions are negligible. In fact, by comparing results when changing the number of 4th-generation ancestors, it turns out that those differences are also negligible (but don't appear to be exactly 0). As another practical consideration, sums aren't taken out beyond 20 sons for each of $S^L(1), S^L(2)$ and S^S. This corresponds to 8000 separate calculations of the Poisson function. Within two significant figures, those results aren't different than when sums are taken up to 10 sons (for each of the 3 variables above). A more rigorous analysis could be done by examining asymptotic behavior of the various equations we've used.

Table 8 – Sons of Abraham

Differences from Chart 27 and Chart 28	$\emptyset$: $(\boldsymbol{Sons, M1, M2, M3, G2, G3})$ $\boldsymbol{p_\emptyset = P^{Ab}_{Tree}(\emptyset)}$	A: $(\boldsymbol{Sons, M1, M2, M3, G2, G3})$ $\boldsymbol{p_A = P^{Ab}_{Tree}(A)}$	$\boldsymbol{P_{Bayes}(0.8, p_\emptyset, p_A)}$
No difference	(Sons,3.6,0.4,3.0,2,2) $p_\emptyset = 0.048 * N^{-3}$	(Sons,3.6,0.4,3.0,1,1) $p_A = 0.079 * N^{-3}$	0.71
Null hypothesis doesn't include Meyer	(Sons,3.6,0.4,3.0,2,1) $p_\emptyset = 0.045 * N^{-3}$	(Sons,3.6,0.4,3.0,1,1) $p_A = 0.079 * N^{-3}$	0.69

What is a bit disconcerting, is that the Bayes probability is slightly higher for the case where Meyer is included as a known ancestor, than in the case when Meyer isn't included. Usually the Bayes probabilities are higher when less information is known. This situation appears to be an artifact of the Poisson weightings (without them, the unweighted probabilities are indeed a bit higher for the case where Meyer isn't included).

Joint Probabilities

We have determined, for each of the children of Jonas, the probability that Aron (and separately Aron and Meyer) was a specific ancestor of theirs, and we have also obtained that probability based on the information about the sons of Jonas. Since these probabilities are all independent, we can compute the joint probability by first calculating $P_\emptyset = \prod_i p_\emptyset(i)$ where $p_\emptyset(i)$ is the value of $p_\emptyset$ previously calculated for either Jonas, or one of Jonas's children – and then calculating $P_A = \prod_i p_A(i)$ where $p_A(i)$ is the value of p_A previously calculated for either Jonas, or one of Jonas's children. Then we can apply the Bayes equation to get

$$P(H_\emptyset|"JD\ descendants") = P_{Bayes}(0.8, P_\emptyset, P_A) \qquad (72)$$

Table 9 below summarizes all of the values of $p_\emptyset(i)$ and $p_A(i)$ that we've obtained in Tables 1 through 8 for the key variants – null hypothesis (Aron and Meyer), null hypothesis without Meyer, alternate hypothesis. We'll also discuss later some of the other variants that were examined.

Table 9 – Summary

JD and children	$p_\emptyset$	$p_\emptyset$ -- no Meyer	p_A
Jonas	$0.0015 * N^{-3}$	$0.0042 * N^{-3}$	$0.0094 * N^{-3}$
Jacob (Table 2)	$0.023 * N^{-2}$	$0.14 * N^{-3}$	$0.17 * N^{-3}$
Wife of Jacob-Lipman	$0.17 * N^{-1}$	$0.17 * N^{-1}$	$0.24 * N^{-1}$
Abraham	$0.048 * N^{-3}$	$0.045 * N^{-3}$	$0.079 * N^{-3}$
Fajga	0.20	0.20	0.20
Eyzyk	$0.055 * N^{-3}$	$0.073 * N^{-3}$	$0.21 * N^{-3}$
Tanchum	$0.13 * N^{-1}$	$0.14 * N^{-1}$	$0.17 * N^{-1}$
Mosze	1.0	1.0	1.0
Naftali Hersz	$0.13 * N^{-1}$	$0.14 * N^{-1}$	$0.17 * N^{-1}$

From this summary table, we can then multiply down the columns and then derive from equation (72):

$$P(H_\emptyset|"JD\ descendants") = 0.22 \text{ assuming } N = 50 \text{ (the value is larger if } N \text{ is larger)} \quad (73)$$

$$P(H_\emptyset - Meyer|"JD\ descendants") = 0.12 \quad (74)$$

Earlier in this paper (and earlier in the course of this investigation – long before I derived the above probabilities), it was suggested that the null hypothesis should be rejected if the probability of $H_\emptyset^{JD}$, given evidence from records of JD's descendants, is less than a level of significance of 0.05. Quite clearly, the probability of $H_\emptyset^{JD}$ (with or without the inclusion of Meyer) is much greater than the level of significance. However, if it turned out that the documentary evidence for Meyer was considerably weakened, then we might conclude from the value in equation **(**74) **(**74), that some reasonable doubt could be cast on the claim that Aron was the father of Jonatan of Drobin.

For some of the individuals, such as Jonas and Eyzyk and Jacob, we explored a few alternative scenarios. For example, for Jacob we examined different ways of accounting for double-names, and for Jonas we considered the possibility that some of his putative sons, were actually sons-in-law. None of these alternative scenarios significantly change the above conclusions.**(**74)

References

Abramowitz, M., & Stegun, I. A. (1964). *Handbook of Mathematical Functions.* New York: Dover Publications, Inc.

Blokherovitsh, Y. A. (1939). *Zikne Mahaneh Yehudah.* Peitrikow: see hebrewbooks.org.

Derrida, B., Manrubia, S. C., & Zanette, &. D. (1999). Statistical Properties of Genealogical Trees. *Physical Review Letters 82, 9*, 1987-90.

Freund, J. E. (1988). *Modern Elementary Statistics, Seventh Edition.* Englewood Cliffs, New Jersey 07632: Prentice-Hall.