# The Shared cM Project Version 4.0 (March 2020) 

Blaine T. Bettinger, Ph.D., J.D.

The Shared cM Project is a collaborative data collection and analysis project created to understand the ranges of shared centimorgans associated with various known relationships. For this update, total shared cM data for nearly 60,000 known relationships were provided.

For more information:

- The Shared cM Project - http://thegeneticgenealogist.com/2015/05/29/the-shared-cm-project/
- Bettinger, Blaine T., "The Shared cM Project: A Demonstration of the Power of Citizen Science." Journal of Genetic Genealogy (2016): 38-42. http://jogg.info/pages/vol8/editorial/bettinger/bettingersharedcMProject.html

To provide your data for subsequent updates:

- The Shared cM Project: https://forms.gle/sg45AMGynMsvxfzU6
- Also consider submitting data to the "Pedigree Collapse, Double/Multiple Cousin, and ROH Shared cM Project": https://forms.gle/zDR49xjgB1nKLpq99

Possible issues with user-provided data:

- Data entry errors - some of the information entered by participants is affected by data entry errors (for example, a longest segment greater than the total shared $\mathrm{cM})$. When these entries could be definitively determined, they were removed.
- Incorrect relationships (known or unknown) - some relationships were almost certainly entered incorrectly, which might be due to misunderstandings of 'removed' relationships in genealogy. Other relationship errors were clearly due to misattributed parentage events resulting in the believed relationship being incorrect.
- Endogamy and Pedigree Collapse - Some relationships will be affected by endogamy and/or pedigree collapse, which will increase the amount of DNA shared by test-takers having a certain genealogical relationship. Although the collection form requests information about known endogamy and/or pedigree collapse, many contributors will not be aware of the endogamy and pedigree collapse in their tree. Additionally, some participants may have selected only one relationship although there were several known relationships.
- Company Thresholds - Each of the DNA testing companies applies a different matching threshold to maximize the identification of genetic cousins while minimizing false positives. These thresholds may impact the total amount of DNA shared by two test-takers, especially at more distant relationships.


## Differences Between Version 3.0 And Version 4.0

There are numerous changes in this version of the Shared cM Project compared to previous versions. These changes were made to improve legibility and presentation of data, to add additional types of data, and to eliminate unnecessary data. Following is a list of some of the major changes to the data analysis and presentation in Version 4.0 of the Shared cM Project:

- Added 32,999 data points - thanks to submissions from thousands of generous genealogists, this update represents an increase of $147 \%$;
- Added a methods section - a methods section was added to provide information about how the submitted data was processed;
- Changed "Clusters" to "Groupings" - due to the growing popularity of shared match clustering, I've changed the name of the meiosis clusters in the former Cluster Chart to Meiosis Groupings. A description of meiosis groupings was added;
- Added meiosis grouping histograms and line graphs - these provide additional useful information about ranges and relationship overlaps;
- Added standard deviation - standard deviations were added to provide additional information about ranges and variation within a relationship range;
- Standardized histogram bins from random ranges to fixed ranges - the bins for all histograms were previously generated by Excel, but in this version are generated by me in an attempt to have consistent ranges for similar relationships;
- Added several histograms containing multiple relationships - these special relationship groupings are common problems for genealogists, and showing the overlap between these relationships provides additional information;
- Removed company and endogamy breakdowns - these breakdowns were only rarely utilized and did not have a significant impact on relationship predictions; and
- Provided a short description and chart of differences between version 3.0 and version 4.0 regarding the minimum, average, and maximum values - this summary highlights some of the major differences in this new version.


## Methods

## Data Collection

Data was collected from participants using Google Forms, which collected the submissions into a spreadsheet. The Google Form contained data entry fields for required information ("Known Relationship," "Total Shared cM," "Number of Shared Segments," "Endogamy or Known Cousin Marriage" (YES/NO) and "Source" (23andMe, AncestryDNA, Family Tree DNA, MyHeritage, GEDmatch, or Other)), and optional data entry fields ("Longest Block," "Notes," and "Email Address").

A total of 59,714 submissions were made to the Shared cM Project as of 8 July 2019 (beginning March 4, 2015). For analysis, the submissions were downloaded as an Excel spreadsheet on 8 July 2019.

## Initial Data Curation

Because "Known Relationship" was a text entry field, submissions varied considerably regarding the naming of various relationships. In this initial data curation stage, all decipherable relationships were converted to a uniform format (where "C" equals cousin and "R" equals removed). Submissions with indecipherable relationships were eliminated. Submissions with obvious data entry errors were also eliminated, such as those where the longest segment was longer than the total shared cM , or where there was text in the cM field instead of a number.

This initial data curation eliminated a total of 1,739 submissions (2.9\%), bringing the total to 57,975 data points used for statistical analysis (although there were submissions included in this total for relationships not analyzed by the project).

A total of 48 different relationships ranging from Parent/Child to 8C were analyzed individually. The total number of submissions for each relationship varied, with a low of 33 for $5 \mathrm{C}_{3} \mathrm{R}$, and a high of 5,281 for 2 C 1 R .

## Outlier Removal

Each relationship was analyzed individually, and obvious errors were removed (for example, 7 cM for a parent/child relationship). Then, a total of $1 \%$ of the submissions for each relationship was removed, removing $0.5 \%$ of the submissions at each end of the range. For example, if there were 200 submissions, 2 submissions were removed (the highest submission and the lowest submission).

## Data Analysis

The dataset contained 55,418 submissions for the 48 different relationships analyzed by the project. Following outlier removal, the minimum, average, and maximum values of the remaining data points were identified for each relationship using standard methodology. Standard deviation was calculated using Excel.

For relationships where the minimum value was o cM shared, the averages were calculated only for cM amounts greater than o cM . Accordingly, these averages represent the average only for cousins sharing a detectable amount of DNA.

A histogram was created for each relationship. The histograms were created in Excel using the data for each relationship after outliers were removed.

## Previous Versions of the Shared cM Project

|  | Launch Date | Total Submissions |
| :--- | :---: | :---: |
| Version 1.0 | May 2015 | $>6,000$ |
| Version 2.0 | June 2016 | $>10,000$ |
| Version 3.0 | August 2017 | $>25,000$ |
| Version 4.0 | March 2020 | $>59,000$ |

## Thank You

Thank you to EVERYONE that has submitted data to the Shared cM Project, whether one submission or many. YOU make this project possible!

Thank you to members of the write-up review team that provided valuable data consistency checks, documentation and formatting review, and wonderful suggestions for improving this document (B.J. Jamieson, Bob Danovich, Darrin Chambers, Deanna Eckman Korte, Elizabeth Heise, Eva Dahlberg, Fiona Brooker, Graham Hart, Jarrett Ross, Jim Owston, John Collins, Mary Kathryn Crews Kozy, Mia Bennett, Michelle Patient, Paula Williams, R S Vivs Laliberte, Randy W Whited, Rob Warthen). A very special thank you to Anne Bettinger for her generous donation of many hours to the project! And a thank you to Jonny Perl for so generously creating and hosting the interactive version of the Shared cM Project at DNA Painter (www.dnapainter.com)!

## Using the Shared cM Project

Step 1: How much DNA do two people share?

- Determine how much DNA you share with a genetic match (in this AncestryDNA example, I share 95 cM with this match)

3rd-4th Cousin
Shared DNA: 95 cM across 4 segments

Step 2: Which Meiosis Grouping(s) does the total shared cM fit into?

- Review the Meiosis Grouping Table to see into which Meiosis Grouping(s) the total shared cM fits (see next page for more information about "Meiosis Groupings")
- In this example, 95 cM fits into each of Meiosis Groupings \#5, 6, 7, 8,9 , and 10

Step 3: Which Meiosis Grouping(s) does the total shared cM best fit into?

- Based on the average, which Meiosis Grouping(s) does the total shared cM most closely match?
- In this example, 95 cM best fits into Meiosis Groupings \#6 and 7 (i.e., 95 cM is closest to the average for these Meiosis Grouping(s)

Step 4: Review the histograms for the relationships in the best fit Meiosis Grouping(s)

- Using the relationships listed in the Meiosis Grouping(s), go to the histograms and see for which relationship(s) the total shared cM is nearest the peak of the histogram. Note that this only provides clues as to the most likely relationships; your actual relationship may vary.
- For this example, for which relationships in Meiosis Groupings \#6 and 7 is 95 cM closest to the peak of the histogram?

This method helps you narrow in on most likely relationships, but cannot definitely identify a specific relationship.

## Meiosis Groupings

In the table on the following page, some of the closest relationships are gathered into Groupings based on the number of separating meioses (which could also be called recombination events or reproductive events) for that relationship. For example, there are two meioses that separate full siblings, and they form their own Meiosis Grouping (Grouping 1). There are three meioses that separate half siblings, Aunts/Uncles/Nieces/Nephews, and Grandparents/Grandchildren, so they form a second Meiosis Grouping (Grouping 2). Since the number of separating meioses are the same for each relationship in the chart, there is expected to be commonality of the minimum, average, and maximum for that group.

In the table, the average, minimum, maximum, and standard deviation values were calculated using the combined submissions for each relationship within that Meiosis Grouping. Minimums were automatically set to ocM for Meiosis Grouping 6-10.

Thirty-six of the 48 relationships analyzed by the Shared cM Project were utilized to generate the 10 Meiosis Groupings.

The Expected values in the table are based on perfect $50 \%$ inheritance of shared DNA at each generation. For example, the Expected value for half-siblings is $25 \%$ (or $1,700 \mathrm{cM}$ ). However, the Expected value does not consider the natural variation that occurs in human genetics. A table of expected values with more information can be found at the ISOGG Wiki (https://isogg.org/wiki/Autosomal DNA statistics).

## Meiosis Groupings

| The Shared cM Project - Version 4.0 March 2020 |  |  |  |  | Blaine T. Bettinger www.TheGeneticGenealogist.com CC 4.0 Attribution License |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Grouping | Relationships | Total \# | Average | Minimum | Maximum | Standard Deviation | Expected |
| Grouping \#1 | Siblings | 2465 | 2613 | 1613 | 3488 | 204 | 2550 |
| Grouping \#2 | Half Sibling, Aunt/Uncle/Niece/Nephew, and Grandparent/Grandchild | 5067 | 1749 | 984 | 2461 | 201 | 1700 |
| Grouping \#3 | 1C, Half Aunt/Uncle/Niece/Nephew, Great-Grandparent/Great-Grandchild, and Great-Aunt/Uncle/Niece/Nephew | 5134 | 865 | 330 | 1486 | 165 | 850 |
| Grouping \#4 | 1 C 1 R , Half 1 C , <br> Half Great-Aunt/Uncle/Niece/Nephew, and Great-Great Aunt/Uncle/Niece/Nephew | 4766 | 435 | 102 | 980 | 124 | 425 |
| Grouping \#5 | 1 C 2 R , Half $1 \mathrm{C} 1 \mathrm{R}, 2 \mathrm{C}$, and <br> Half Great-Great-Aunt/Uncle/Niece/Nephew | 5998 | 227 | 33 | 592 | 85 | 213 |
| Grouping \#6 | $1_{1} \mathrm{C}_{3}$, Half 1 C 2 R , Half 2 C , and $2 \mathrm{C}_{1} \mathrm{R}$ | 6160 | 122 | o | 353 | 59 | 106 |
| Grouping \#7 | Half $1_{1} \mathrm{C}_{2}$, Half $2 \mathrm{C} 1 \mathrm{R},{ }_{2} \mathrm{C} 2 \mathrm{R}$, and ${ }_{3} \mathrm{C}$ | 6686 | 72 | O | 244 | 43 | 53 |
| Grouping \#8 | Half $2 \mathrm{C} 2 \mathrm{R},{ }_{2} \mathrm{C}_{3} \mathrm{R}$, Half 3 C , and ${ }_{3} \mathrm{C}_{1} \mathrm{R}$ | 5077 | 48 | O | 192 | 32 | 27 |
| Grouping \#9 | Half $3 \mathrm{C}_{12}$, $3^{\text {C }} 2 \mathrm{R}$, and 4 C | 3966 | 35 | O | 166 | 24 | 13 |
| Grouping \#10 | Half ${ }_{3} \mathrm{C} 2 \mathrm{R},{ }_{3} \mathrm{C}_{3} \mathrm{R}$, Half 4 C , and 4 C 1 R | 2633 | 28 | o | 126 | 19 | 7 |

## Histograms

Histogram = a graphical representation of the distribution of numerical data
The following steps are followed to create a histogram:

1. Divide entire range of total shared cM (from the smallest amount to the largest amount) into a series of intervals of equal size (called "bins"); and
2. Count how many data points fall within each of the intervals.

How to read these histograms:

> The number at the top of each bar is the total number of data entries for each interval (see "bin," below). For example, 653 data entries are in the interval of 701 to 800 cM .

These are "bins," ranges of total shared cM. For example, the bin of 301 to 400 cM contains 2 data entries for this relationship (1C). That indicates that this amount is very, very rare for 1 C .


## Meiosis Groupings Histograms

On the following pages are histograms for each of the ten Meiosis Groupings from the Meiosis Grouping Table on page 6 .




The Shared cM Project - Version 4.0 (March 2020)








## Meiosis Grouping Line Graphs

These following line graphs are another way to visualize the Meiosis Grouping histogram distributions from the previous pages. In the first graph, all ten Meiosis Groupings are shown. In the second and third graphs, the Meiosis Groupings are broken down in the first five and the last five.



The Shared cM Project - Version 4.0 (March 2020)


## The Relationship Chart

## The Shared cM Project - Version 4.0 (March 2020)

| Blaine T. Bettinger www.TheGeneticGenealogist.com CC 4.0 Attribution License |  |  | How to read this chart: |  |  |  |  | Great-G Gran | t-Greatarent | GGGGAunt/Uncle | Other Relationships |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Great-Gre | andparent | $\begin{gathered} \text { GGG- } \\ \text { Aunt/Uncle } \end{gathered}$ |  |  |
| Half GG- Aunt/Uncle 208 $103-284$ | $\begin{gathered} \text { Great-Grandparent } \\ 887 \\ 485-1486 \\ \hline \end{gathered}$ |  |  |  |  |  |  | $\begin{gathered} \hline \text { Great-Great } \\ \text { Aunt/Uncle } \\ 420 \\ 186-713 \\ \hline \end{gathered}$ | $\begin{gathered} \mathbf{1 C 3 R} \\ 117 \\ 25-238 \end{gathered}$ | $\begin{gathered} \mathbf{2 c 3} \mathbf{R} \\ 51 \\ \mathbf{o}-154 \end{gathered}$ |  |
| $\begin{gathered} \text { Half 1C2R } \\ 125 \\ 16-269 \end{gathered}$ | $\begin{gathered} \hline \text { Half Great- } \\ \text { Aunt/Uncle } \\ 431 \\ 184-668 \\ \hline \end{gathered}$ | $\begin{gathered} \text { Grandparent } \\ 1754 \\ 984-2462 \end{gathered}$ |  |  |  |  | Great Aunt/Uncle 850 $330-1467$ |  | $\begin{gathered} \text { 2c2R } \\ 71 \\ \text { o- } 244 \end{gathered}$ | $\begin{gathered} \mathbf{3 C 2 R} \\ 36 \\ o-166 \end{gathered}$ | $\begin{gathered} \mathbf{6 C} \\ 18 \\ 0-71 \end{gathered}$ |
| $\begin{gathered} \text { Half 2c1R } \\ 66 \\ 0-190 \end{gathered}$ | $\begin{gathered} \text { Half 1C1R } \\ 224 \\ 62-469 \end{gathered}$ | Half Aunt/Uncle 871 $492-1315$ |  |  |  |  | $\begin{gathered} \text { Parent } \\ 3485 \\ 2376-3720 \end{gathered}$ |  |  | $\begin{gathered} \text { Aunt/Uncle } \\ 1741 \\ 1201-2282 \end{gathered}$ | $\begin{gathered} \mathbf{1 C 1 R} \\ 433 \\ 102-980 \\ \hline \end{gathered}$ | $\begin{gathered} \mathbf{2 c 1 R} \\ 122 \\ 14-353 \end{gathered}$ | $\begin{gathered} \mathbf{3 C 1 R} \\ 48 \\ \mathrm{o}-192 \end{gathered}$ | $\begin{gathered} \mathbf{4 C 1 R} \\ 28 \\ o-126 \end{gathered}$ | $\begin{gathered} \mathbf{6 C 1 R} \\ 15 \\ o-56 \end{gathered}$ |
| Half $3 \mathbf{c}$ 48 o-168 | Half 2c <br> 120 <br> $10-325$ | $\begin{gathered} \text { Half 1C } \\ 449 \\ 156-979 \end{gathered}$ | $\begin{gathered} \text { Half-Sibling } \\ 1759 \\ 1160-2436 \end{gathered}$ | $\begin{gathered} \text { Sibling } \\ 2613 \\ 1613-3488 \end{gathered}$ | SELF | $\begin{gathered} \text { 1C } \\ 866 \\ 396-1397 \end{gathered}$ | $\begin{gathered} \mathbf{2 c} \\ 229 \\ 41-592 \end{gathered}$ | $\begin{gathered} \mathbf{3 c} \\ 73 \\ \mathrm{o}-234 \end{gathered}$ | $\begin{gathered} \mathbf{4 c} \\ 35 \\ 0-139 \end{gathered}$ | $\begin{gathered} \mathbf{5 c} \\ 25 \\ \mathrm{o}-117 \end{gathered}$ | $\begin{gathered} \text { 6C2R } \\ 13 \\ o-45 \end{gathered}$ |
| $\begin{gathered} \text { Half 3c1R } \\ 37 \\ \mathbf{o - 1 3 9} \end{gathered}$ | $\begin{gathered} \text { Half 2c1R } \\ 66 \\ 0-190 \\ \hline \hline \end{gathered}$ | $\begin{gathered} \text { Half 1C1R } \\ 224 \\ 62-469 \\ \hline \hline \end{gathered}$ | Half <br> Niece/Nephew <br> 871 <br> $492-1315$ | $\begin{gathered} \text { Niece/Nephew } \\ 1740 \\ 1201-2282 \end{gathered}$ | $\begin{gathered} \text { Child } \\ 3487 \\ 2376-3720 \end{gathered}$ | $\begin{gathered} \mathbf{1 C 1 R} \\ 433 \\ 102-980 \end{gathered}$ |  | $\begin{gathered} \mathbf{3 C 1 R} \\ 48 \\ \mathrm{o}-192 \\ \hline \end{gathered}$ | $\begin{gathered} \mathbf{4 C 1 R} \\ 28 \\ o-126 \end{gathered}$ | $\begin{gathered} \text { 5C1R } \\ 21 \\ 0-80 \end{gathered}$ | $\begin{gathered} 7 \mathrm{C} \\ 14 \\ \mathrm{o}-57 \end{gathered}$ |
| $\begin{gathered} \text { Half 3c2R } \\ 27 \\ 0-78 \\ \hline \end{gathered}$ | $\begin{gathered} \text { Half } 2 \mathbf{c} 2 \mathbf{R} \\ 48 \\ 0-144 \\ \hline \end{gathered}$ | $\begin{gathered} \text { Half 1C2R } \\ 125 \\ 16-269 \\ \hline \end{gathered}$ | Half Great <br> Niece/Nephew <br> 431 <br> $184-668$ | Great- <br> Niece/Nephew <br> 850 <br> $330-1467$ | $\begin{gathered} \text { Grandchild } \\ 1754 \\ 984-2462 \\ \hline \end{gathered}$ |  | $\begin{gathered} \text { 2c2R } \\ 71 \\ \text { o- } 244 \end{gathered}$ | $\begin{gathered} \mathbf{3 C 2 R} \\ 36 \\ o-166 \end{gathered}$ | $\begin{gathered} 4 \mathrm{C} 2 \mathrm{R} \\ 22 \\ 0-93 \\ \hline \end{gathered}$ | $\begin{gathered} \text { 5C2R } \\ 18 \\ o-65 \end{gathered}$ | $\begin{gathered} 7 \mathbf{C 1 R} \\ 12 \\ 0-50 \\ \hline \end{gathered}$ |
| Half 3c3R | Half 2c3R | $\begin{gathered} \text { Half } 1 C_{3} R \\ 60 \\ 0-120 \end{gathered}$ | Half GG <br> Niece/Nephew <br> 208 <br> $103-284$ | Great-GreatNiece/Nephew $\begin{gathered} 420 \\ 186-713 \\ \hline \end{gathered}$ | Great- Grandchild 887 $485-1486$ | $\begin{gathered} \mathbf{1 C 3 R} \\ 117 \\ 25-238 \end{gathered}$ | $\begin{gathered} \mathbf{2 c 3 R} \\ 51 \\ o-154 \end{gathered}$ | $\begin{gathered} \mathbf{3 C 3 R} \\ 27 \\ o-98 \end{gathered}$ | $\begin{gathered} 4 \mathbf{C}_{3} \mathbf{R} \\ 19 \\ 0-60 \end{gathered}$ | $\begin{gathered} \mathbf{5 C 3 R} \\ 13 \\ \mathrm{o}-30 \end{gathered}$ | $\begin{gathered} \mathbf{8 C} \\ 11 \\ \mathrm{o}-42 \end{gathered}$ |

Minimum was automatically set to 0 cM for relationships more distant than Half 2C, and averages were determined only for submissions in which DNA was shared
NOTE: for this and other charts or diagrams in this document, the minimum was automatically set to "o cM" for relationships more distant than Half 2C, and averages were determined only for relationships in which DNA was shared.

## Individual Relationship Histograms

| Relationship | \# | Min | Average | Max | SD | Histogram |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Parent/Child | 2412 | 2376 | 3485 | 3720 | 115 |  |

Page 23 of 56

| Relationship | \# | Min | Average | Max | SD | Histogram |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sibling (Grouping \#1) | 2465 | 1613 | 2613 | 3488 | 203 |  |
| Aunt/Uncle/Niece/Nephew (Grouping \#2) | 2695 | 1201 | 1741 | 2282 | 180 |  |

Page 24 of 56

| Relationship | \# | Min | Average | Max | SD | Histogram |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Grandparent/Grandchild (Grouping \#2) | 1106 | 984 | 1754 | 2462 | 236 |  |
| Half Sibling (Grouping \#2) | 1266 | 1160 | 1759 | 2436 | 207 |  |

Page $\mathbf{2 5}$ of 56

| Relationship | \# | Min | Average | Max | SD | Histogram |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\begin{gathered} 1 \mathrm{C} \\ \text { (Grouping \#3) } \end{gathered}$ | 3337 | 396 | 866 | 1397 | 161 |  |
| Great- <br> Aunt/Uncle/Niece/Nephew <br> (Grouping \#3) | 833 | 330 | 850 | 1467 | 118 |  |


| Relationship | \# | Min | Average | Max | SD | Histogram |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Great-Grandparent/Great Grandchild (Grouping \#3) | 149 | 485 | 887 | 1486 | 181 |  |
| ```Half Aunt/Uncle/Niece/Nephew (Grouping #3)``` | 815 | 492 | 871 | 1315 | 165 |  |



Page $\mathbf{2 8}$ of 56


Page 29 of 56


Page $\mathbf{3 0}$ of 56


Page 31 of 56

| Relationship | \# | Min | Average | Max | SD | Histogram |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\begin{gathered} 1 \mathrm{C} 3 \mathrm{R} \\ \text { (Grouping \#6) } \end{gathered}$ | 202 | 25 | 117 | 238 | 48 |  |
| $\begin{gathered} 2 \mathrm{C} 1 \mathrm{R} \\ \text { (Grouping \#6) } \end{gathered}$ | 5281 | 14 | 122 | 353 | 59 |  |

Page 32 of 56


Page 33 of 56


Page 34 of 56


Page 35 of 56


Page 36 of 56


Page 37 of 56

| Relationship | \# | Min | Average | Max | SD | Histogram |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\begin{gathered} 3 \mathrm{C} 2 \mathrm{R} \\ \text { (Grouping \#9) } \end{gathered}$ | 965 | 0 | 36 | 166 | 27 |  |
| $\begin{gathered} 4 \mathrm{C} \\ \text { (Grouping \#9) } \end{gathered}$ | 2752 | 0 | 35 | 139 | 23 |  |

Page 38 of 56


Page 39 of 56

| Relationship | \# | Min | Average | Max | SD | Histogram |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\begin{gathered} 4 \mathrm{C} 1 \mathrm{R} \\ \text { (Grouping \#10) } \end{gathered}$ | 2383 | 0 | 28 | 126 | 20 |  |
| Half 3C2R <br> (Grouping \#10) | 55 | 0 | 27 | 78 | 15 |  |

Page 40 of 56


Page 41 of 56


Page 42 of 56


Page 43 of 56

\begin{tabular}{|c|c|c|c|c|c|c|c|c|c|c|c|c|}
\hline \multirow[t]{2}{*}{Relationship} \& \# \& Min \& Average \& Max \& SD \& \multicolumn{7}{|c|}{Histogram} \\
\hline \& 441 \& o \& 18 \& 71 \& 13 \&  \&  \& \begin{tabular}{l}
15
\(\square\) \\
50
\end{tabular} \& \[
\begin{gathered}
5 \\
60
\end{gathered}
\] \& \begin{tabular}{|}
5 \\
\hline 70
\end{tabular} \& \(\frac{2}{80}\) \& \[
\begin{gathered}
0 \\
90
\end{gathered}
\] \\
\hline 5 C 3 R \& 34 \& o \& 13 \& 30 \& 7 \&  \& \begin{tabular}{l}
15 \\
20
\end{tabular} \& \& 5

30 \& \& \& <br>
\hline
\end{tabular}

Page 44 of 56


Page 45 of 56


Page 46 of 56

| Relationship | $\#$ | Min | Average | Max | SD | Histogram |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |

## Histograms with Multiple Relationships (Special Cases)



[^0]

[^1]


## Relationship Chart

| Grouping \# | Relationship | \# | Min | Average | Max | SD | Expected |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Parent/Child | 2412 | 2376 | 3485 | 3720 | 115 | Varies |
| 1 | Sibling | 2465 | 1613 | 2613 | 3488 | 203 | 2550 |
| 2 | Aunt/Uncle/Niece/Nephew | 2695 | 1201 | 1741 | 2282 | 180 | 1700 |
| 2 | Grandparent/Grandchild | 1106 | 984 | 1754 | 2462 | 236 | 1700 |
| 2 | Half Sibling | 1266 | 1160 | 1759 | 2436 | 207 | 1700 |
| 3 | 1C | 3337 | 396 | 866 | 1397 | 161 | 850 |
| 3 | Great- Aunt/Uncle/Niece/Nephew | 833 | 330 | 850 | 1467 | 118 | 850 |
| 3 | Great-Grandparent/Great- Grandchild | 149 | 485 | 887 | 1486 | 181 | 850 |
| 3 | Half Aunt/Uncle/Niece/Nephew | 815 | 492 | 871 | 1315 | 165 | 850 |
| 4 | 1C1R | 3700 | 102 | 433 | 980 | 123 | 425 |
| 4 | Great-Great- <br> Aunt/Uncle/Niece/Nephew | 137 | 186 | 420 | 713 | 121 | 425 |
| 4 | Half 1C | 691 | 156 | 449 | 979 | 134 | 425 |
| 4 | Half Great- Aunt/Uncle/Niece/Nephew | 238 | 184 | 431 | 668 | 105 | 425 |
| 5 | 1-2R | 1187 | 33 | 221 | 471 | 83 | 213 |
| 5 | 2 C | 4172 | 41 | 229 | 592 | 86 | 213 |
| 5 | Half 1C1R | 605 | 62 | 224 | 469 | 80 | 213 |
| 5 | Half Great-GreatAunt/Uncle/Niece/Nephew | 34 | 103 | 208 | 284 | 43 | 213 |
| 6 | $1 \mathrm{C}_{3} \mathrm{R}$ | 202 | 25 | 117 | 238 | 48 | 106 |
| 6 | 2 C 1 R | 5281 | 14 | 122 | 353 | 59 | 106 |
| 6 | Half 1C2R | 165 | 16 | 125 | 269 | 54 | 106 |
| 6 | Half 2C | 512 | 10 | 120 | 325 | 62 | 106 |
| 7 | 2 C 2 R | 1404 | o | 71 | 244 | 42 | 53 |
| 7 | 3 C | 4775 | o | 73 | 234 | 43 | 53 |
| 7 | Half 1C3R | 33 | 0 | 60 | 120 | 32 | 53 |
| 7 | Half 2C1R | 474 | o | 66 | 190 | 39 | 53 |
| 8 | ${ }_{2} \mathrm{C}_{3} \mathrm{R}$ | 163 | o | 51 | 154 | 32 | 27 |
| 8 | $3 \mathrm{C1R}$ | 4514 | 0 | 48 | 192 | 32 | 27 |
| 8 | Half 2C2R | 98 | o | 48 | 144 | 32 | 27 |
| 8 | Half 3 C | 302 | o | 48 | 168 | 33 | 27 |
| 9 | 3 C 2 R | 965 | o | 36 | 166 | 27 | 13 |
| 9 | 4 C | 2752 | o | 35 | 139 | 23 | 13 |
| 9 | Half 3C1R | 249 | 0 | 37 | 139 | 24 | 13 |
| 10 | $3 \mathrm{C}_{3} \mathrm{R}$ | 106 | o | 27 | 98 | 19 | 7 |
| 10 | 4 C 1 R | 2383 | 0 | 28 | 126 | 20 | 7 |


| Grouping \# | Relationship | $\#$ | Min | Average | Max | SD | Expected |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 10 | Half 3C2R | 55 | 0 | 27 | 78 | 15 | 7 |
| 10 | Half 4C | 89 | 0 | 30 | 74 | 18 | 7 |
| $\mathrm{n} / \mathrm{a}$ | 4 C 2 R | 603 | 0 | 22 | 93 | 16 | 3.3 |
| $\mathrm{n} / \mathrm{a}$ | 5 C | 1130 | 0 | 25 | 117 | 18 | 3.3 |
| $\mathrm{n} / \mathrm{a}$ | 4 C 3 R | 89 | 0 | 19 | 60 | 12 | 1.7 |
| $\mathrm{n} / \mathrm{a}$ | 5 C 1 R | 975 | 0 | 21 | 80 | 15 | 1.7 |
| $\mathrm{n} / \mathrm{a}$ | 5 C 2 R | 242 | 0 | 18 | 65 | 12 | 0.8 |
| $\mathrm{n} / \mathrm{a}$ | 6 C | 441 | 0 | 18 | 71 | 13 | 0.8 |
| $\mathrm{n} / \mathrm{a}$ | 5 C 3 R | 34 | 0 | 13 | 30 | 7 | 0.4 |
| $\mathrm{n} / \mathrm{a}$ | 6 C 1 R | 441 | 0 | 15 | 56 | 9 | 0.4 |
| $\mathrm{n} / \mathrm{a}$ | 6 C 2 R | 161 | 0 | 13 | 45 | 8 | 0.2 |
| $\mathrm{n} / \mathrm{a}$ | 7 C | 258 | 0 | 14 | 57 | 9 | 0.2 |
| $\mathrm{n} / \mathrm{a}$ | 7 C 1 R | 330 | 0 | 12 | 50 | 7 | 0.1 |
| $\mathrm{n} / \mathrm{a}$ | 8C | 239 | 0 | 11 | 42 | 5 | 0.05 |

## Differences in Minimum, Average, and Maximum Values Between Version 3.0 Vs. Version 4.0

There are many changes to the minimum, average, and maximum values for relationships in Version 4.0 of the Shared cM Project relative to the prior Version 3.0. As the number of submissions for a relationship grows, the distribution of cM values for that relationship is more clearly defined. This allows for improved definition and elimination of outliers for each relationship. In some cases, the very large increase in submissions moved the minimum and/or maximum values further outward for a broader distribution in this version, and in other cases it moved the minimum and/or maximum values inward for a tighter distribution in this version.

For example, the large increase in the number of submissions for GreatAunt/Uncle/Niece/Nephew allowed for significantly better definition of the distribution for that relationship, and thus better definition of outliers. The maximum cM value for that relationship moved from $2,108 \mathrm{cM}$ in Version 3.0 to $1,467 \mathrm{cM}$ in Version 4.0.

Interestingly, the average cM values were very consistent from Version 3.0 to Version 4.0. For the 48 analyzed relationships, only nine changed by more than 10 cM , and only three changed by more than 20 cM .

| $\underset{\#}{\text { Grouping }}$ | Relationship | $\begin{aligned} & \text { Version } 3.0 \\ & \text { \# of } \\ & \text { Submissions } \end{aligned}$ | $\begin{aligned} & \text { Version } 4.0 \\ & \text { \# of } \\ & \text { Submissions } \end{aligned}$ | Version 3.0 Min | Version 4.0 Min | Change | Version <br> 3.0 <br> Average | Version <br> 4.0 <br> Average | Change | Version 3.0 Max | Version 4.0 Max | Change |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Parent/Child | 1378 | 2412 | 3330 | 2376 | -29\% | 3487 | 3485 | 0\% | 3720 | 3720 | 0\% |
| 1 | Sibling | 1345 | 2465 | 2209 | 1613 | -27\% | 2629 | 2613 | -1\% | 3384 | 3488 | 3\% |
| 2 | Aunt/Uncle/Niece/Nephew | 1411 | 2695 | 1349 | 1201 | -11\% | 1750 | 1741 | -1\% | 2175 | 2282 | 5\% |
| 2 | Grandparent/Grandchild | 611 | 1106 | 1156 | 984 | -15\% | 1766 | 1754 | -1\% | 2311 | 2462 | 7\% |
| 2 | Half Sibling | 451 | 1266 | 1317 | 1160 | -12\% | 1783 | 1759 | -1\% | 2312 | 2436 | 5\% |
| 3 | 1C | 1512 | 3337 | 553 | 396 | -28\% | 874 | 866 | -1\% | 1225 | 1397 | 14\% |
| 3 | GreatAunt/Uncle/Niece/Nephew | 413 | 833 | 251 | 330 | 31\% | 914 | 850 | -7\% | 2108 | 1467 | -30\% |
| 3 | Great-Grandparent/GreatGrandchild | 61 | 149 | 464 | 485 | 5\% | 881 | 887 | 1\% | 1486 | 1486 | 0\% |
| 3 | Half <br> Aunt/Uncle/Niece/Nephew | 275 | 815 | 500 | 492 | -2\% | 891 | 871 | -2\% | 1446 | 1315 | -9\% |
| 4 | 1C1R | 1594 | 3700 | 141 | 102 | -28\% | 439 | 433 | -1\% | 851 | 980 | 15\% |
| 4 | Great-Great- <br> Aunt/Uncle/Niece/Nephew | 25 | 137 | 191 | 186 | -3\% | 427 | 420 | -2\% | 885 | 713 | -19\% |
| 4 | Half 1C | 177 | 691 | 137 | 156 | 14\% | 457 | 449 | -2\% | 856 | 979 | 14\% |
| 4 | Half GreatAunt/Uncle/Niece/Nephew | 44 | 238 | 125 | 184 | 47\% | 432 | 431 | 0\% | 765 | 668 | $-13 \%$ |
| 5 | 1C2R | 481 | 1187 | 43 | 33 | -23\% | 229 | 221 | -3\% | 531 | 471 | -11\% |
| 5 | 2 C | 1590 | 4172 | 46 | 41 | -11\% | 233 | 229 | -2\% | 515 | 592 | 15\% |
| 5 | Half 1C1R | 138 | 605 | 57 | 62 | 9\% | 226 | 224 | -1\% | 530 | 469 | -12\% |
| 5 | Half Great-GreatAunt/Uncle/Niece/Nephew | 15 | 34 | 12 | 103 | 758\% | 187 | 208 | 11\% | 383 | 284 | -26\% |
| 6 | 1C3R | 38 | 202 | 0 | 25 | n/a | 123 | 117 | -5\% | 283 | 238 | -16\% |
| 6 | 2C1R | 2064 | 5281 | 0 | 14 | n/a | 123 | 122 | -1\% | 316 | 353 | 12\% |
| 6 | Half 1C2R | 39 | 165 | 37 | 16 | -57\% | 145 | 125 | -14\% | 360 | 269 | -25\% |
| 6 | Half 2C | 143 | 512 | 9 | 10 | 11\% | 117 | 120 | 3\% | 397 | 325 | -18\% |
| 7 | 2 C 2 R | 564 | 1404 | 0 | 0 | n/a | 74 | 71 | -4\% | 261 | 244 | -7\% |

Page 55 of 56

| 7 | 3 C | 1791 | 4775 | O | O | $\mathrm{n} / \mathrm{a}$ | 74 | 73 | -1\% | 217 | 234 | 8\% |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 7 | Half 1C3R | 19 | 33 | O | O | n/a | 87 | 60 | -31\% | 191 | 120 | -37\% |
| 7 | Half 2C1R | 116 | 474 | O | O | n/a | 73 | 66 | -10\% | 341 | 190 | -44\% |
| 8 | 2 C 3 R | 50 | 163 | 0 | O | n/a | 57 | 51 | -11\% | 139 | 154 | 11\% |
| 8 | 3 C 1 R | 1736 | 4514 | 0 | O | n/a | 48 | 48 | 0\% | 173 | 192 | 11\% |
| 8 | Half 2C2R | 24 | 98 | O | O | n/a | 61 | 48 | -21\% | 353 | 144 | -59\% |
| 8 | Half 3C | 55 | 302 | 0 | O | $\mathrm{n} / \mathrm{a}$ | 61 | 48 | -21\% | 178 | 168 | -6\% |
| 9 | 3C2R | 459 | 965 | 0 | O | n/a | 35 | 36 | 3\% | 116 | 166 | 43\% |
| 9 | 4C | 998 | 2752 | 0 | O | n/a | 35 | 35 | 0\% | 127 | 139 | 9\% |
| 9 | Half 3C1R | 71 | 249 | 0 | O | n/a | 42 | 37 | -12\% | 165 | 139 | -16\% |
| 10 | 3C3R | 47 | 106 | 0 | O | n/a | 22 | 27 | 23\% | 48 | 98 | 104\% |
| 10 | 4C1R | 934 | 2383 | 0 | O | n/a | 28 | 28 | 0\% | 117 | 126 | 8\% |
| 10 | Half 3C2R | 26 | 55 | 0 | O | $\mathrm{n} / \mathrm{a}$ | 34 | 27 | -21\% | 96 | 78 | -19\% |
| 10 | Half 4C | 32 | 89 | 0 | O | n/a | 36 | 30 | -17\% | 120 | 74 | -38\% |
| $\mathrm{n} / \mathrm{a}$ | 4C2R | 245 | 603 | 0 | O | n/a | 22 | 22 | 0\% | 109 | 93 | -15\% |
| $\mathrm{n} / \mathrm{a}$ | 5 C | 422 | 1130 | 0 | O | $\mathrm{n} / \mathrm{a}$ | 25 | 25 | 0\% | 94 | 117 | 24\% |
| $\mathrm{n} / \mathrm{a}$ | 4C3R | 25 | 89 | O | O | n/a | 29 | 19 | -34\% | 86 | 60 | -30\% |
| $\mathrm{n} / \mathrm{a}$ | 5C1R | 354 | 975 | 0 | O | n/a | 21 | 21 | 0\% | 79 | 80 | 1\% |
| $\mathrm{n} / \mathrm{a}$ | 5C2R | 92 | 242 | 0 | O | n/a | 17 | 18 | 6\% | 43 | 65 | 51\% |
| n/a | 6C | 122 | 441 | 0 | O | n/a | 18 | 18 | 0\% | 86 | 71 | -17\% |
| $\mathrm{n} / \mathrm{a}$ | 5 C 3 R | 12 | 34 | 0 | O | n/a | 11 | 13 | 18\% | 44 | 30 | -32\% |
| $\mathrm{n} / \mathrm{a}$ | 6 C 1 R | 138 | 441 | 0 | O | n/a | 16 | 15 | -6\% | 72 | 56 | -22\% |
| n/a | 6C2R | 59 | 161 | O | O | $\mathrm{n} / \mathrm{a}$ | 17 | 13 | -24\% | 75 | 45 | -40\% |
| n/a | 7 C | 54 | 258 | 0 | O | n/a | 13 | 14 | 8\% | 57 | 57 | 0\% |
| $\mathrm{n} / \mathrm{a}$ | 7C1R | 89 | 330 | 0 | O | n/a | 13 | 12 | -8\% | 53 | 50 | -6\% |
| $\mathrm{n} / \mathrm{a}$ | 8C | 80 | 239 | O | O | n/a | 12 | 11 | -8\% | 50 | 42 | -16\% |
|  |  | 22419 | 55418 |  |  |  |  |  |  |  |  |  |


[^0]:    * This chart suggests that full sibling relationships in the range of 1600 to 2000 or so are likely to be unknown half-sibling relationships (and potentially some full sibling relationships that are believed to be half-sibling relationships).

[^1]:    * This chart suggests that total shared cM cannot be utilized to differentiate between half-sibling and aunt/uncle/niece/nephew relationships.

