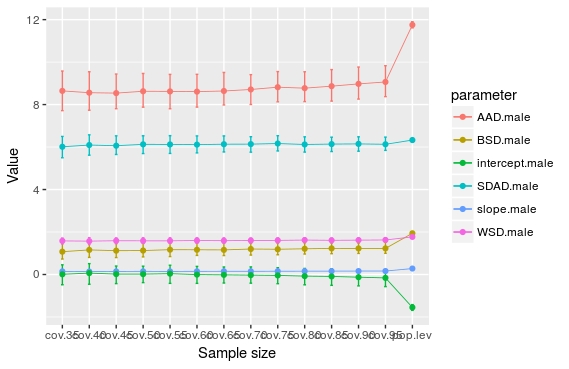
**(i) Age mixing characteristic in partnerships**

When considering sampled individuals (who are infected between 36 and 37 time interval) at different sequence coverage scenarios we have, the age mixing characteristics



We can see that the average age difference from infected individuals in their partnership is not far from the mean of preferred age differences distribution for men/women which was 10; the difference between the two is slightly greater than 1. Same fact is observed from standard deviation of age difference in partnership of infected individuals and standard deviation of preferred age differences distribution for men, which was 5 ; the difference between the two is slightly greater than 1.



***Figure:*** *Partnership age mixing statistics in infected and general populations*

**(ii) Proportion of men/women in pairings**

The true pairings at 100% coverage don’t change as in MCAR scenario.



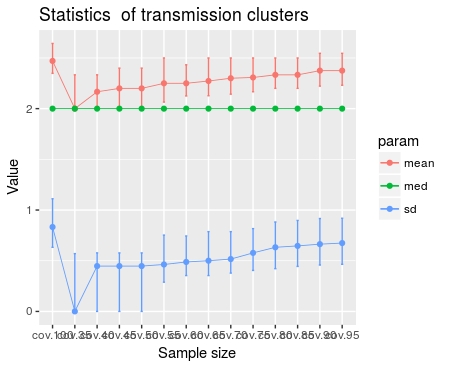
True aggregated pairings across men and women in different age groups were



In transmission clusters pairings across men and women in different age groups were



We computed transmission clusters’ statistics, mainly mean, median, and standard deviation of sizes of transmission clusters.



Proportions of pairing of men/women across different age groups from transmission clusters:

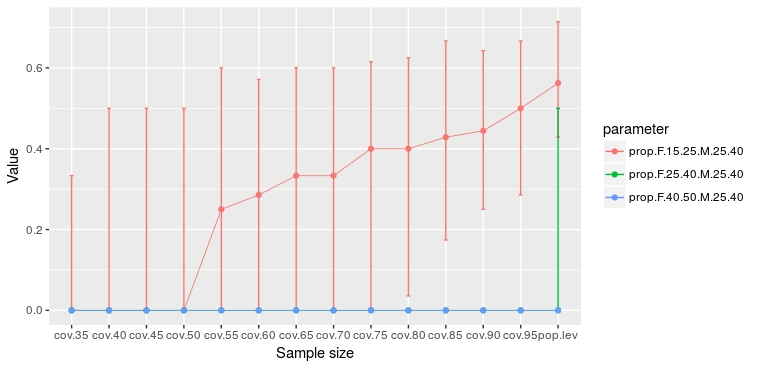


True proportions of pairing of men/women across different age groups from transmission clusters:



One of the highlight from the two tables above is that values of proportions of men and women in any given age group who are pairing with women and men in another age group obtained from transmission network built from transmission clusters are close to true values obtained from same individuals in the true transmission network, and this is true across all sequence coverage scenarios.

Another important fact is that as the sequence coverage increase the estimate proportion approaches the true value at 100% coverage, and also the defined confidence intervals[[1]](#footnote-2) are reduced as it can be seen in the figure below.



***Figure****: Example for proportion of men of 25-40 paired with women in different age groups inferred from transmission clusters*

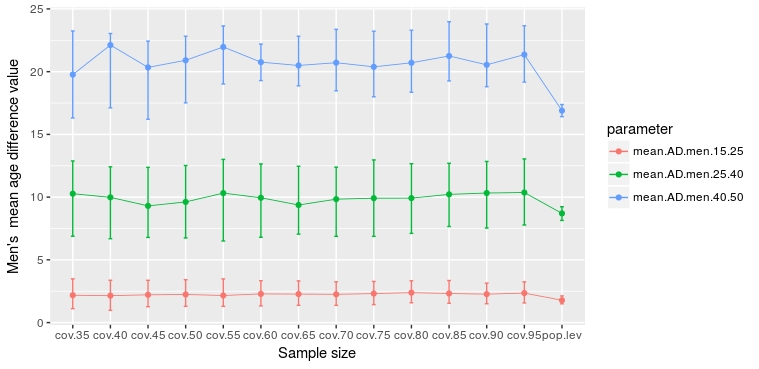
Another highlight is that the truth value at 100% coverage can never be greater than the maximum of the estimates values; hence estimates of proportions at any sequence coverage scenario can overestimate or underestimate the truth.

**(iii) Mean, median, and standard deviation of age gap in pairings**

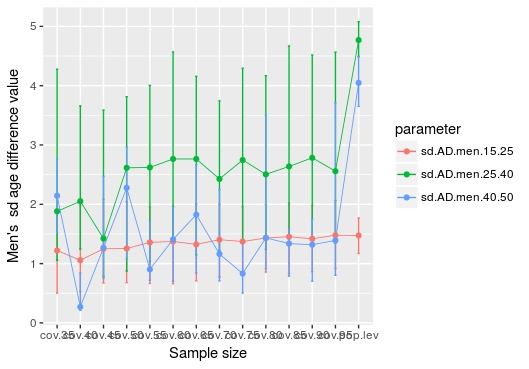
From the transmission network constructed from transmission clusters, we compute mean, median, and standard deviation of age gap of females/males in a given age group paired with males/females regardless of age group.



The minimum age gap value for men in pairings is at 100% coverage (being the true value as well) as it can be seen in the figure below.



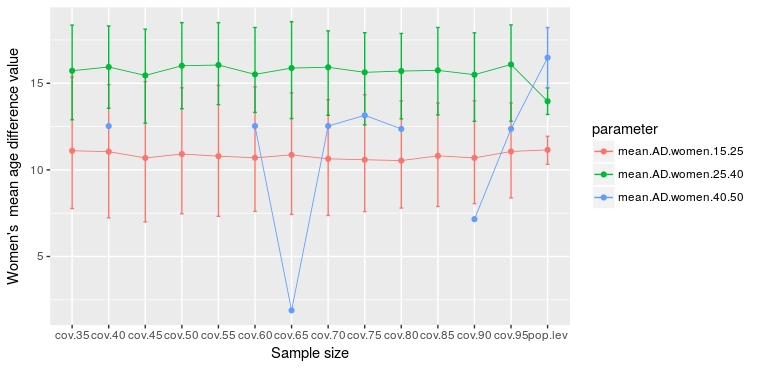
***Figure****: Men’s average age gap in pairings*



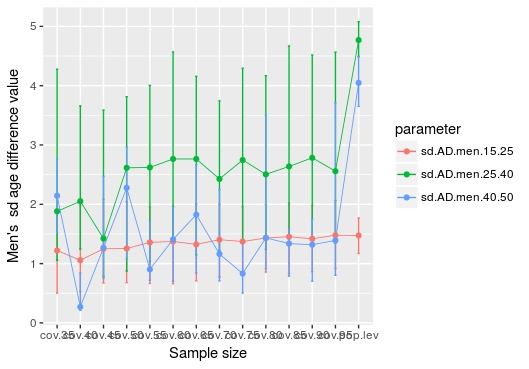
***Figure****: Men’s standard deviation age gap in pairings*

Men of 40 and 50 years old, 25 and 40 years, and 15 and 25 years had an average age gap of almost 20, 10, and 2 respectively. This means old men paired with younger women.

For women, we can see that women of 40 and 50 years old had much variable age gap with their pairs, spanning between 2 and 8. This means that sometimes they were paired with younger men and men of same age. For other women in 25 and 40 years, and 15 and 25 years, they had an average age gap of almost 10, and 13 respectively.



***Figure****: Women’s average age gap in pairings*

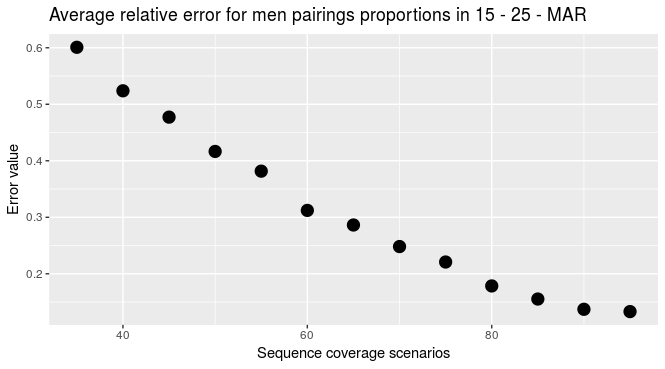


***Figure****: Women’s standard deviation age gap in pairings*

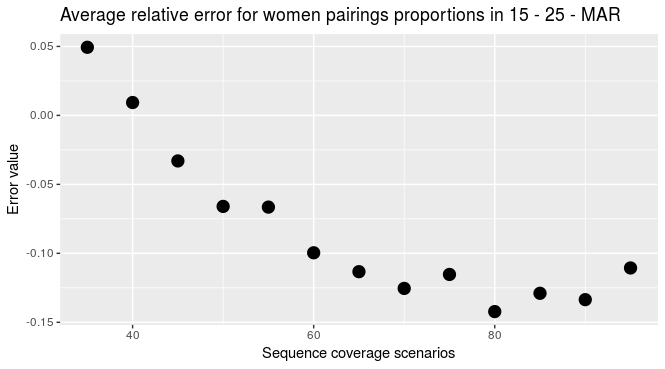
**(iv) Error: Goodness of fit**

The aim was to measure how far/close the inferences from transmission clusters are from the truth for all age mixing measurements (proportions of pairing and statistics of age gap) mentioned above. These values are computed by taking the true values from 100% of coverage and values from inference in transmission clusters. In some examples below, we considered mean relative error (MRE), which was computed by averaging the vector of difference between true value at 100% coverage and inferred values at different sequence coverage.

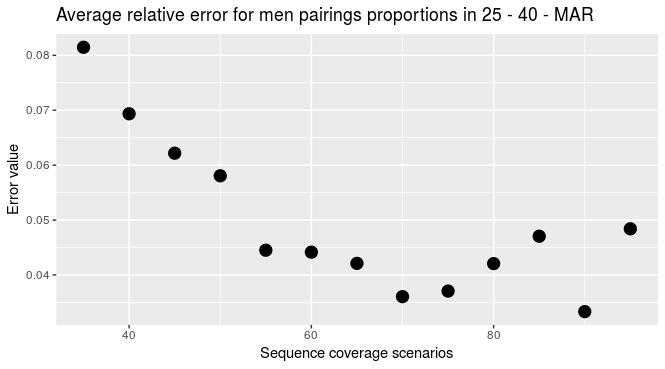
**For proportions: some decrease systematically other are sporadic**



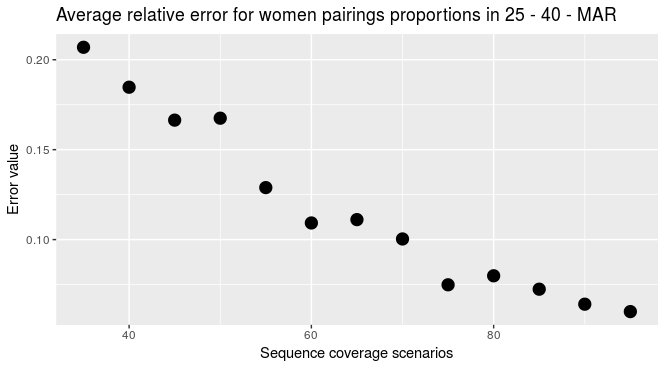
For example, for proportion of men aged between 15 – 25, we had a systematic decrease of the error as the sequence coverage increase as it can be seen in the figure above, whereas the mean relative error decrease sporadically for women in 15 and 25 years.



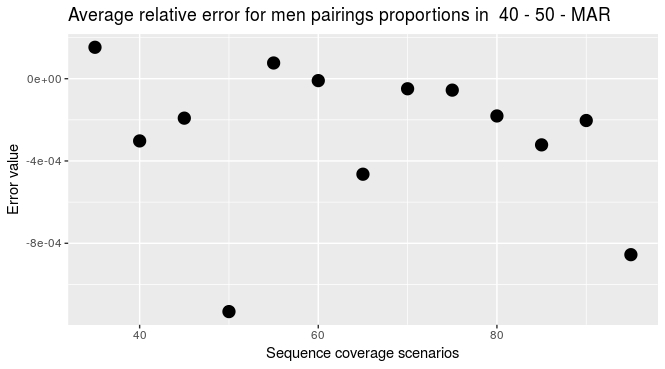
Same sporadic decrease of error is observed in pairings of men in 25 and 40 years in the figure below:

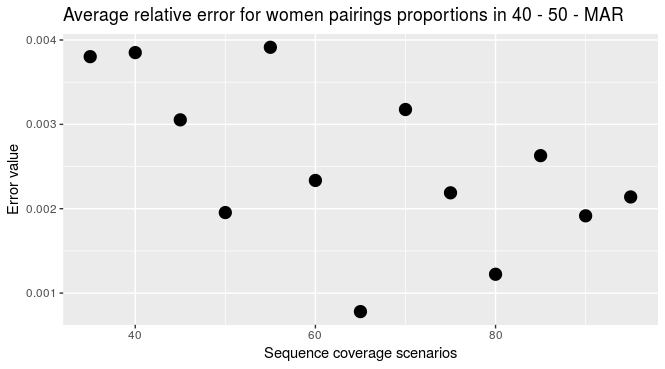


Surprisingly the error decrease systematically for women in that same age group as it can be seen in the figure below



For men and women in 40 and 50 years, the error does not seems to decrease or increase, it varies across age groups as it can be seen in the following two figures:





One explanation of a weird behavior of MRE for proportion of men/women pairings in 40 and 50 years, it may be the lack of clients in these age groups.

1. Which is the 1st and 3rd quantiles [↑](#footnote-ref-2)