

# Sex Partner Choice and Age Mixing Bias in Mathematical Models of Human Papillomavirus

Caleb Easterly<sup>1</sup>, Fernando Alarid-Escudero<sup>2</sup>, MS, Eva Enns<sup>2</sup>, PhD., & Shalini Kulasingam<sup>2</sup>, PhD.

<sup>1</sup>Macalester College, <sup>2</sup> University of Minnesota

## Introduction

Mathematical models of HPV disease dynamics are popular methods of evaluating vaccination strategies, and were used to inform the vaccination policy recommendations of Gardasil [?], for example. A key feature of many models is their description of sexual behavior; that is, how a person with certain characteristics chooses their partners.

## The Assortative-Proportionate Paradigm

In a model that does not stratify by age or sexual activity, the total population of potential partners is just the proportion of the opposite sex. However, many models stratify the population by age and sexual activity, to allow for the input of data specific to each subgroup [?, ?]. Once the group-specific data are defined, many models make the allowance for a mix of assortative and proportionate sexual mixing [?, ?].

**Assortative Mixing:** persons in a specific subgroup of the population (for example, a specific age and sexual activity) only mix with opposite sex members of that subgroup.

**Proportionate Mixing:** chooser has no preference for any subgroup and chooses a partner solely based on the proportions of the population in each subgroup and the number of partnerships they offer.

## A-P Mixture Model

The following is a common model for representing sexual mixing for one demographic characteristic, where  $p$  is the group of the partner and  $c$  is the group of the chooser:

$$P_{G_p|G_c}(p|c) = \begin{cases} \epsilon + (1 - \epsilon) \frac{c_p N_p}{\sum_{\sigma} c_{\sigma} N_{\sigma}}, & \text{if } p = c \\ (1 - \epsilon) \frac{c_p N_p}{\sum_{\sigma} c_{\sigma} N_{\sigma}}, & \text{otherwise} \end{cases} \quad (1)$$

The parameter  $\epsilon$  is the proportion of people who are completely assortative, while  $c_i$  is the number of sexual partners of group  $i$  and  $N_i$  is the proportion of the population in group  $i$ .

For sexual activity and age groups, the sexual mixing matrix used in models such as [?] is the following:

$$\begin{aligned} \hat{\rho}_{klmsj} = & \epsilon_A \epsilon_S \delta_{ls} \delta_{mj} \\ & + \epsilon_A (1 - \epsilon_S) \frac{c_{k'sj} N_{k'sj}}{\sum_{\sigma=1}^{n_S} c_{\sigma j} N_{k'\sigma j}} \delta_{mj} \\ & + (1 - \epsilon_A) \epsilon_S \frac{c_{k'sj} N_{k'sj}}{\sum_{\alpha=1}^{n_A} c_{k'\sigma\alpha} N_{k'\sigma\alpha}} \delta_{ls} \\ & + (1 - \epsilon_A) (1 - \epsilon_S) \frac{c_{k'sj} N_{k'sj}}{\sum_{\alpha=1}^{n_A} \sum_{\sigma=1}^{n_S} c_{k'\sigma\alpha} N_{k'\sigma\alpha}} \end{aligned} \quad (2)$$

Where  $\epsilon_A$  and  $\epsilon_S$  denote the proportion of the population who is assortative by age and sexual activity, respectively.

## Criticisms of The A-P Paradigm

- Difficult to interpret.
- Can not be made to fit empirical data easily.
- Computationally expensive.

## References

## NATSAL-3 Data Analysis

The British population probability survey National Survey on Sexual Attitudes and Lifestyles has a wide variety of detailed sexual behavior data [?]. For this analysis, we use two types of response data:

- 1 Age preference data, based on respondents' most recent sexual partners (up to 3).
- 2 Sex partner acquisition rate (SPAR), based on the respondents' number of new sexual partners in the past year.

The analysis focuses exclusively on heterosexual respondents, but the next planned phase of the project is to analyze the patterns of all respondents.

## Age Preference Data

We model the age preference data as a collection of Gamma distributions - one for each age group - using the method in [?] to analyze heteroscedasticity. The process is as below:

- 1 Perform a linear regression on chooser's age versus age of partner.
- 2 Model the squared residuals as a log-linear function of chooser's age, and predict variance with this model.
- 3 Take the mean age of each age group and use that to find the average mean and average variance for that age group.
- 4 Use method of moments (MoM) on the mean and variance of each age group. Assume the probability of partner's age,  $P_p$  Gamma( $\alpha_i, \beta_i$ ), where  $\alpha_i$  and  $\beta_i$  are the shape and rate parameters of the Gamma distribution for age group  $i$ , which has estimated mean partner age  $\hat{\mu}_i$  and variance  $\hat{\sigma}_i^2$ . Then

$$\hat{\beta}_i = \frac{\hat{\mu}_i}{\hat{\sigma}_i^2}$$

$$\hat{\alpha}_i = \frac{\hat{\mu}_i^2}{\hat{\sigma}_i^2}$$

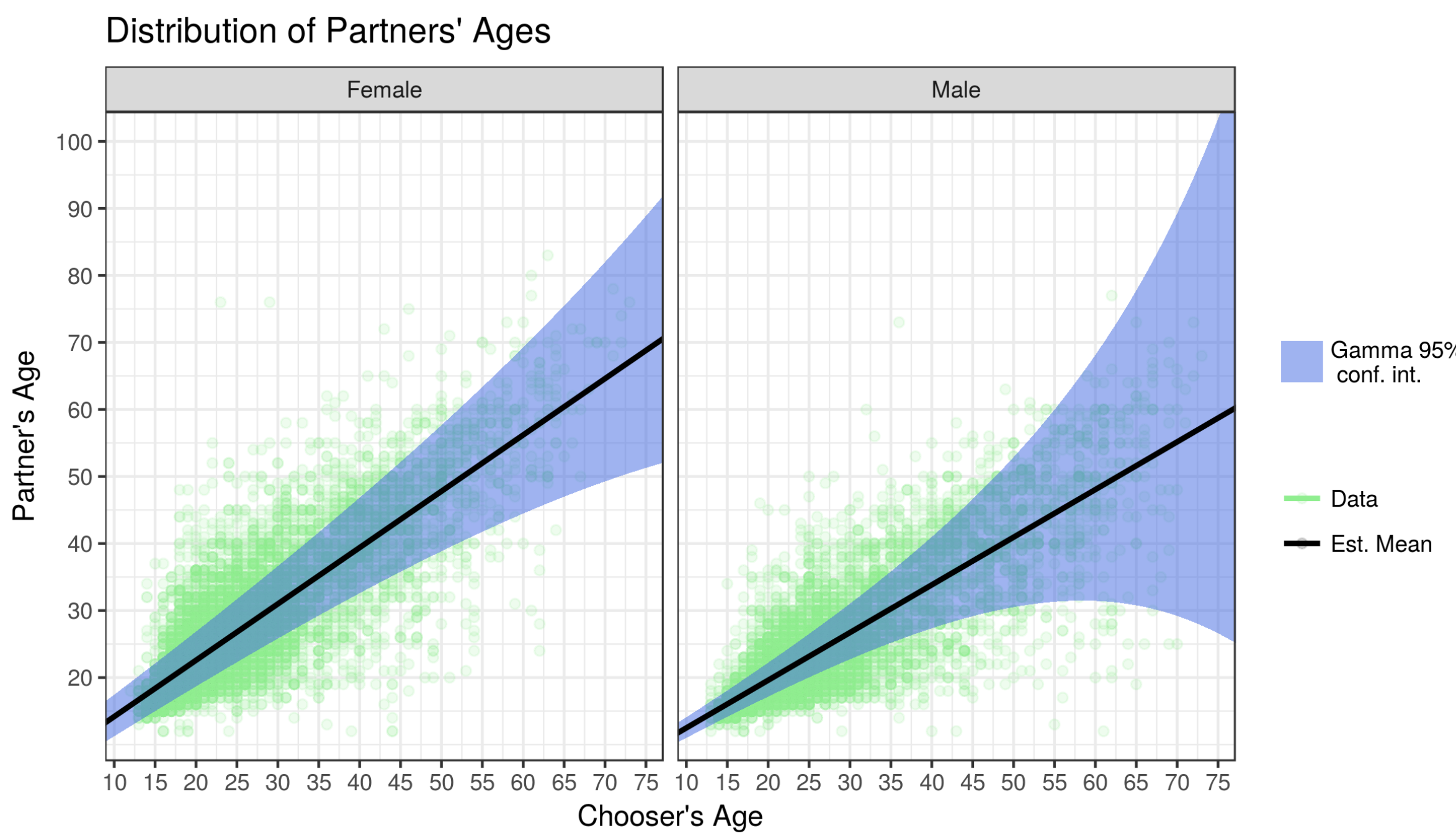


Figure 1: Regression of Mean and Variance by Age. 95% Confidence Intervals are for a Gamma distribution

## SPAR Analysis

- We model the SPAR data as a Poisson regression by age.
- In addition, we follow [?] in assuming that the proportion of the population in the high, medium, and low sexual activity groups are 0.03, 0.15, and 0.82, respectively.

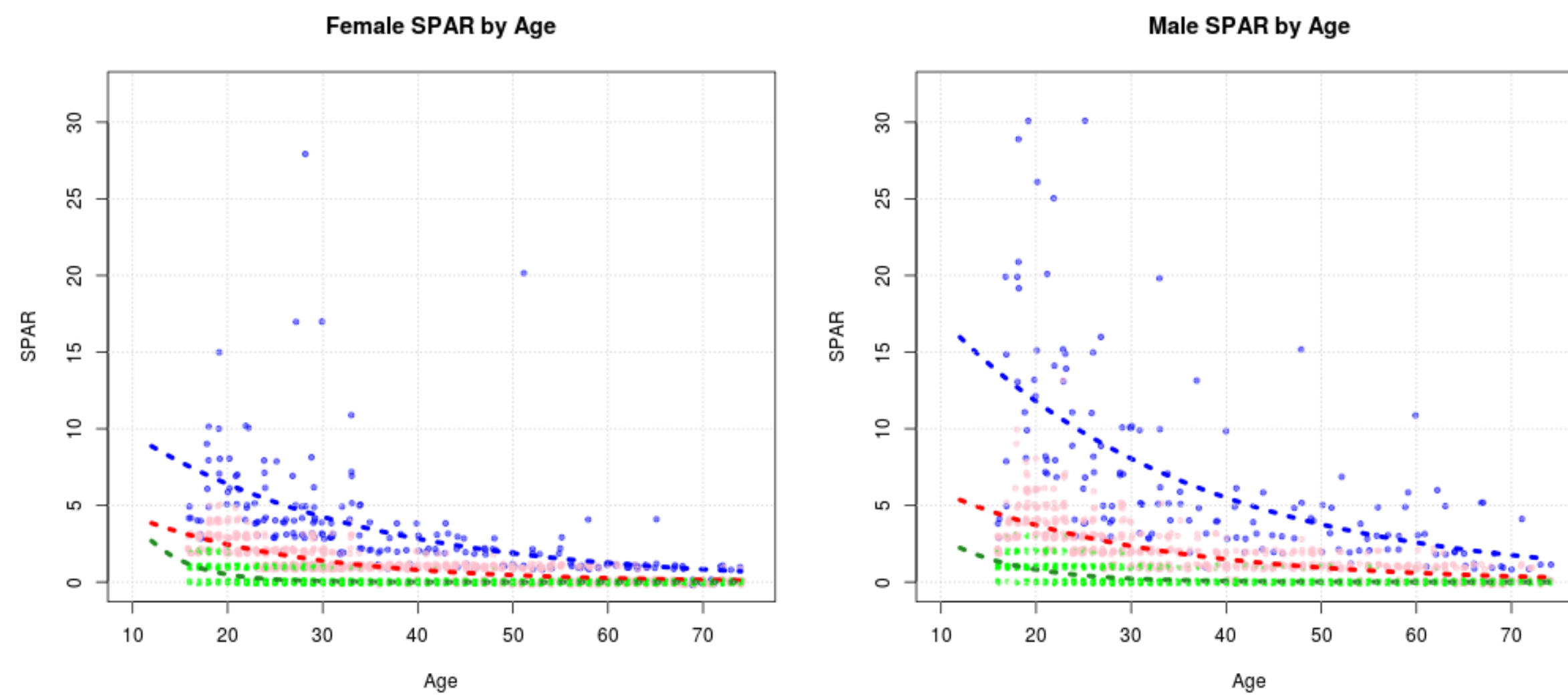


Figure 2: Male SPAR's are generally much higher than those of females. The plots display high variability.

## Preferential Mixing

We define  $\Omega$  to be the age preference matrix derived from the Gamma distributions, and  $S$  to be the sexual activity preference matrix. Then, the partnerships offered from group  $sj$  to  $lm$  is  $\Omega_{k'jm} S_{k'sl} c_{k'sj} N_{k'sj}$ . The new mixing matrix, the probability of group  $lm$  mixing with group  $sj$ , is

$$\begin{aligned} P_{\{S_p, A_p|S_c, A_c\}}(s, j | l, m) &= \rho_{klmsj} \\ &= \frac{S_{k'sl} \Omega_{k'jm} c_{k'sj} N_{k'sj}}{\sum_{\alpha=1}^{n_A} \sum_{\sigma=1}^{n_S} S_{k'\sigma l} \Omega_{k'\sigma m} c_{k'\sigma\alpha} N_{k'\sigma\alpha}}, \end{aligned} \quad (3)$$

The differences in the matrix are fairly large, as shown in the below figures.

## Differences in Mixing

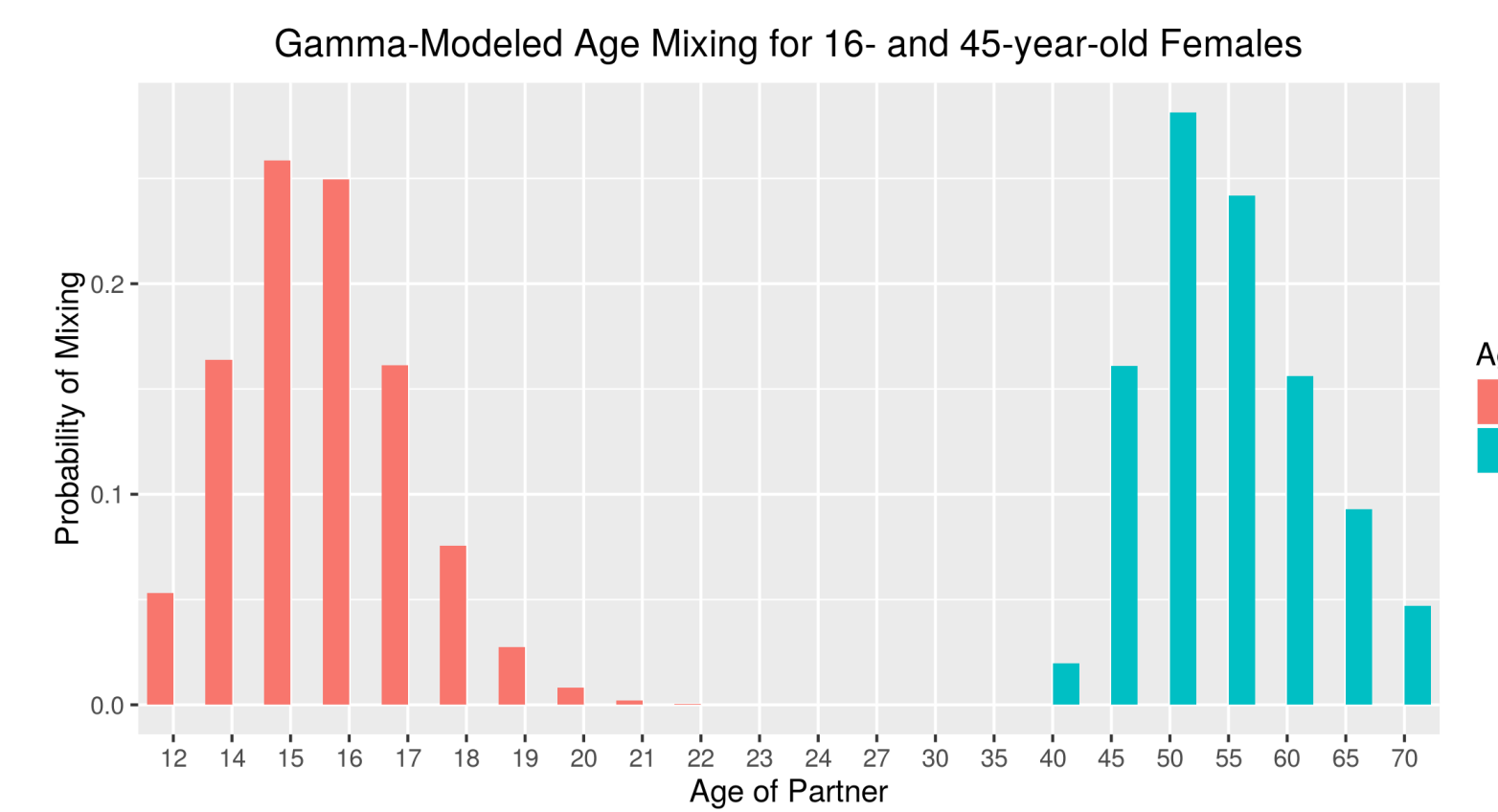


Figure 3: Preferential age mixing based on the empirical survey data in Natsal-3 [?]. The probability of mixing falls off quickly as the potential partners' ages get farther from the chooser.

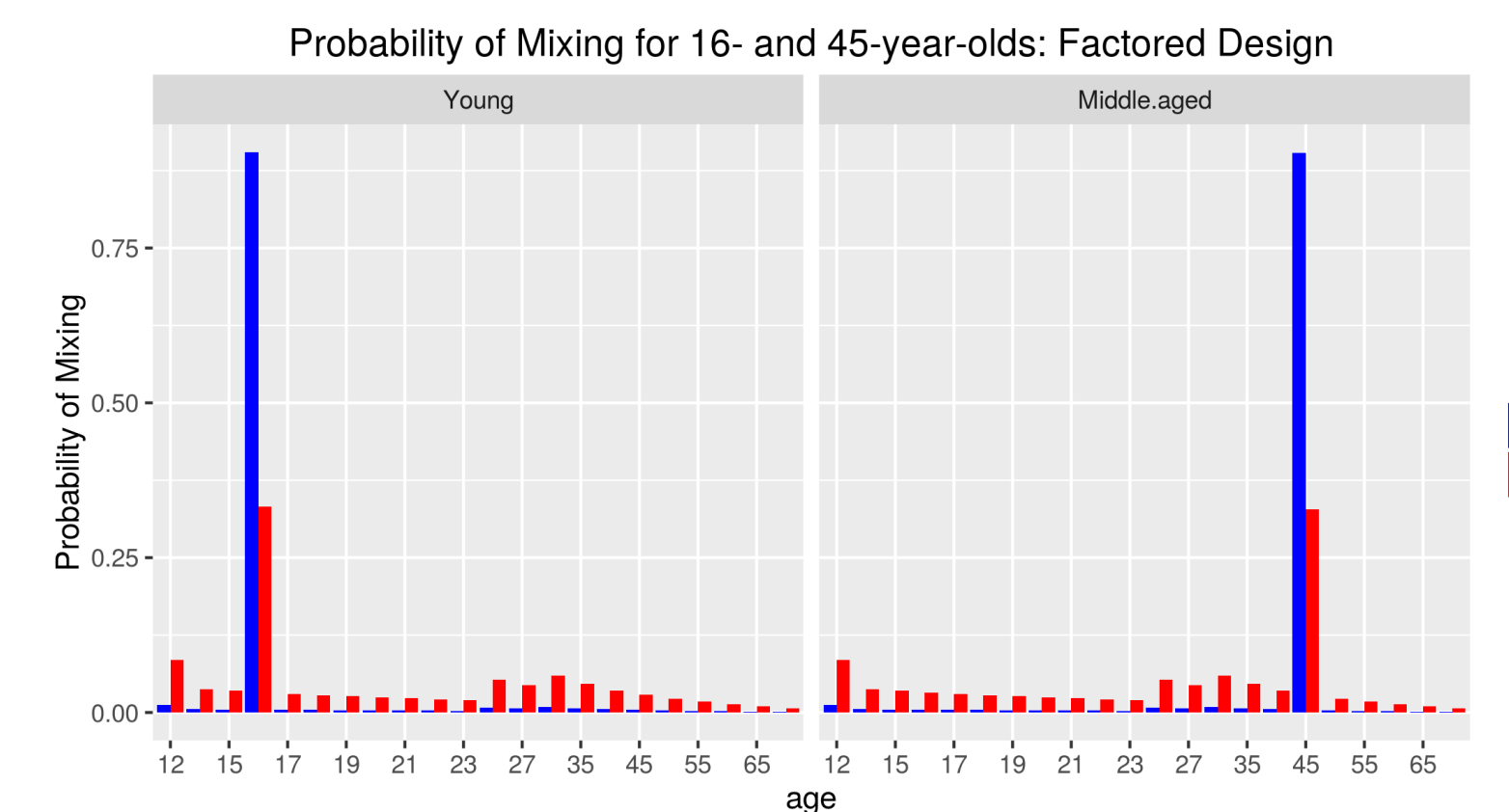


Figure 4: Age mixing based on a factored matrix (equation (2)). For most values of  $\epsilon$  there is a relatively high probability of mixing at even distant ages, while very high values of  $\epsilon$  preclude mixing at close ages.

## Pre-Vaccination Prevalence of Oncogenic HPV Types

The goal of any model is to reproduce observed data. To accomplish that, we calibrate the model by minimizing the squared difference between the endemic steady-state of the model and the observed prevalence curves:

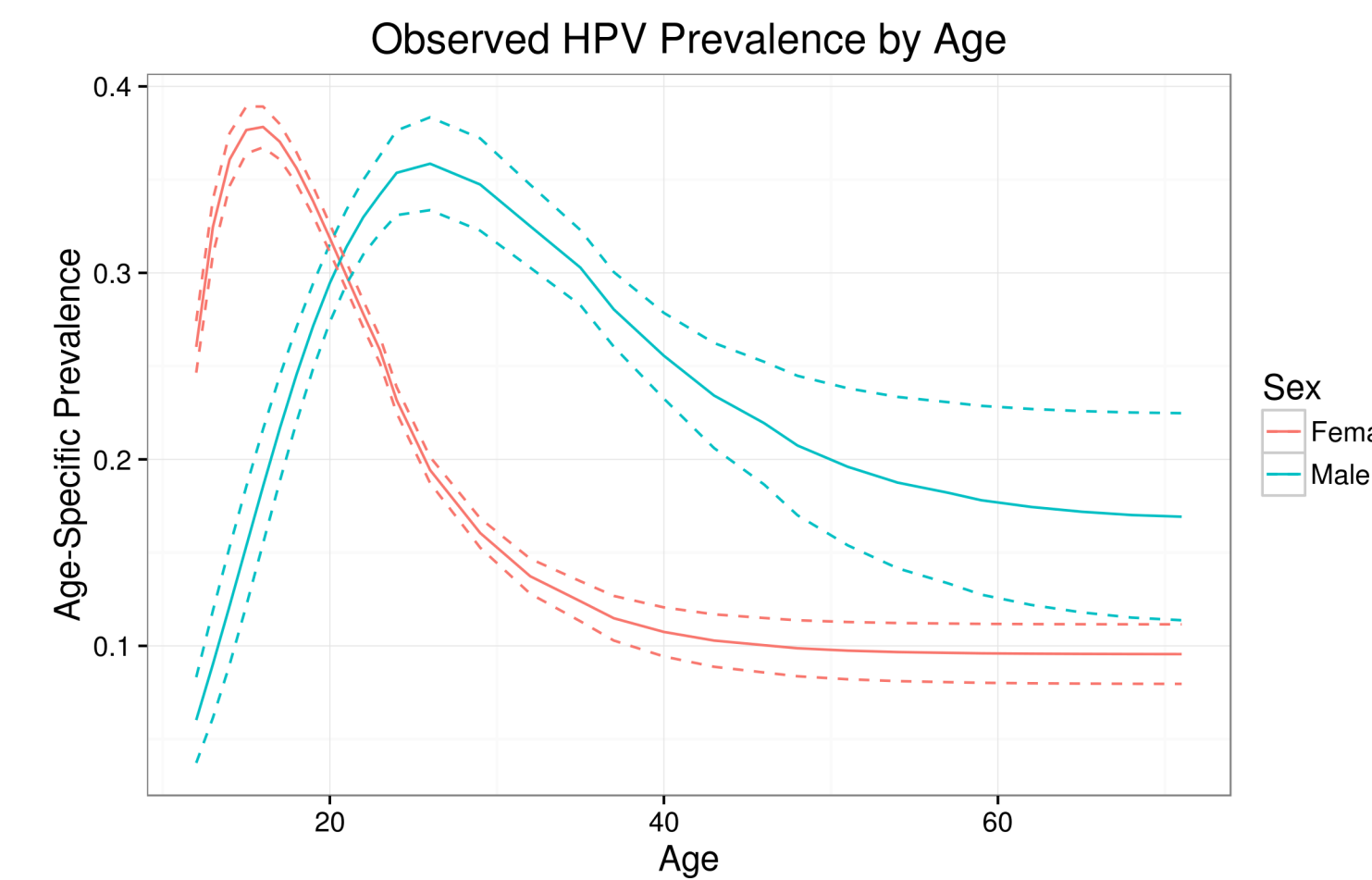


Figure 5: Calculated with exponentially-damped polynomials with data from [?] for women and [?] for men. Error bounds calculated with Markov Chain Monte Carlo error propagation. This is used as target data in the HPV calibration

## Acknowledgements

This work was supported by the Howard Hughes Medical Institute Off-Campus Data Scientist Fellowship. Many thanks to HHMI and to Shalini, Fernando, and Eva.