Reference sample mean age at transition is strongly influenced by sample size

Valerie Sgheiza and Helen M. Liversidge

Here I will be discussing the impact of sample size, relative to other variables such as sex and ethnic group, on reference sample mean age at transition.

Background

- Thevissen (2009): When a population-specific reference sample doesn't exist, the largest available reference sample should be used

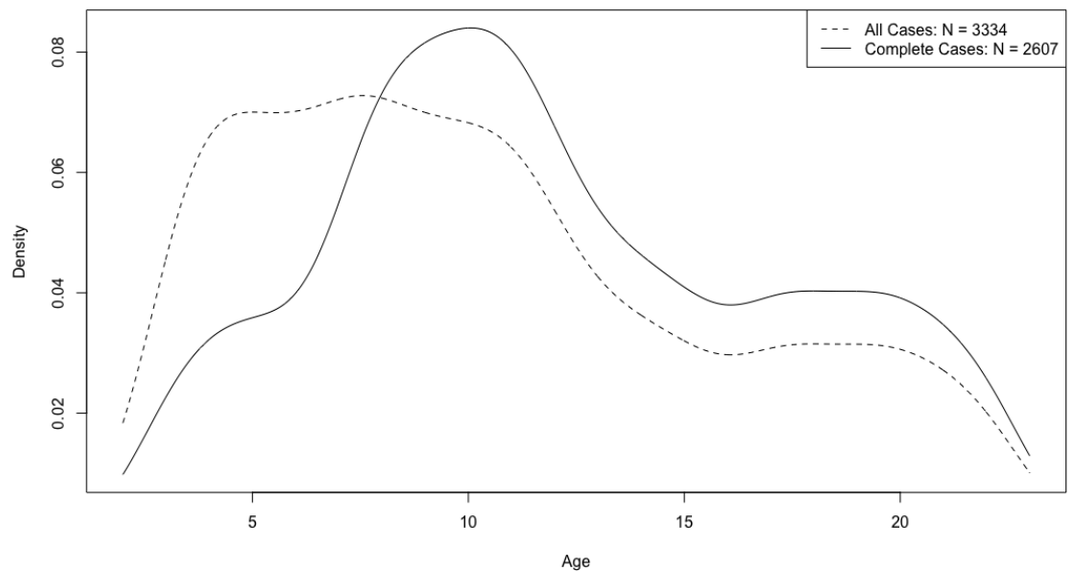
- Liversidge (2011): The majority of stages were not significantly different between ethnic groups. Sex differences were observed in the canine, premolars, and third molar.

  - Standard error of mean age at transition may be driven by within-stage sample size rather than overall sample size

    - Potential argument for collapsing stages when reference sample size is small

Prior research has suggested that sample size effects should be ruled out before making a determination of population specific differences in dental development.
Thevissen (2009) found that when a country-specific reference sample for third molar age estimation was unavailable, the best option was to use a large Belgian reference sample.
Liversidge (2011) showed that for the permanent mandibular teeth, mean age within stage was not significantly different between two ethnic groups for most stages, however there were larger sex-based differences.
One possible explanation is that apparent stage differences are being driven by small within-stage sample sizes. This is potentially the case even when the overall reference sample is in the thousands, due to large numbers of stages and a non-uniform age distribution. An immediate remedy would be to collapse stages or use a statin system with a smaller number of stages.
Our objective is therefore to test the effect of both total and within stage sample size on reference sample mean age at transition and compare this to the effects of sex and ethnic group.

Dataset



We use the same dataset presented in Liversidge (2011) These data consist of Moorrees et al. (1963) scores of the left permanent mandibular dentition from panoramic radiographs of 3334 London dental patients between 2 and 23 years of age. Scoring was performed by HL. The dataset was constructed such that there were approximately equal numbers of boys and girls at each year of age. Only the individuals with complete scores were used, for a total of 2607 individuals. Many of the incomplete scores were due to blurred images, primarily in younger patients whose small size causes challenges in positioning within the imaging device.

Stage Conversion

- Collapse stages to increase sample size within stage without increasing total sample size
  - Use established staging system ⟶ Demirjian et al. (1973)

Moorrees et al.	Demirjian et al.
Crypt Absent	O
Crypt	
Ci	A
Cco	B
Coc	
C1/2	C
C3/4	
Cc	D
Ri	
Rcl	E
R1/4	
R1/2	F
R3/4	
Rc	G
A1/2	
Ac	H

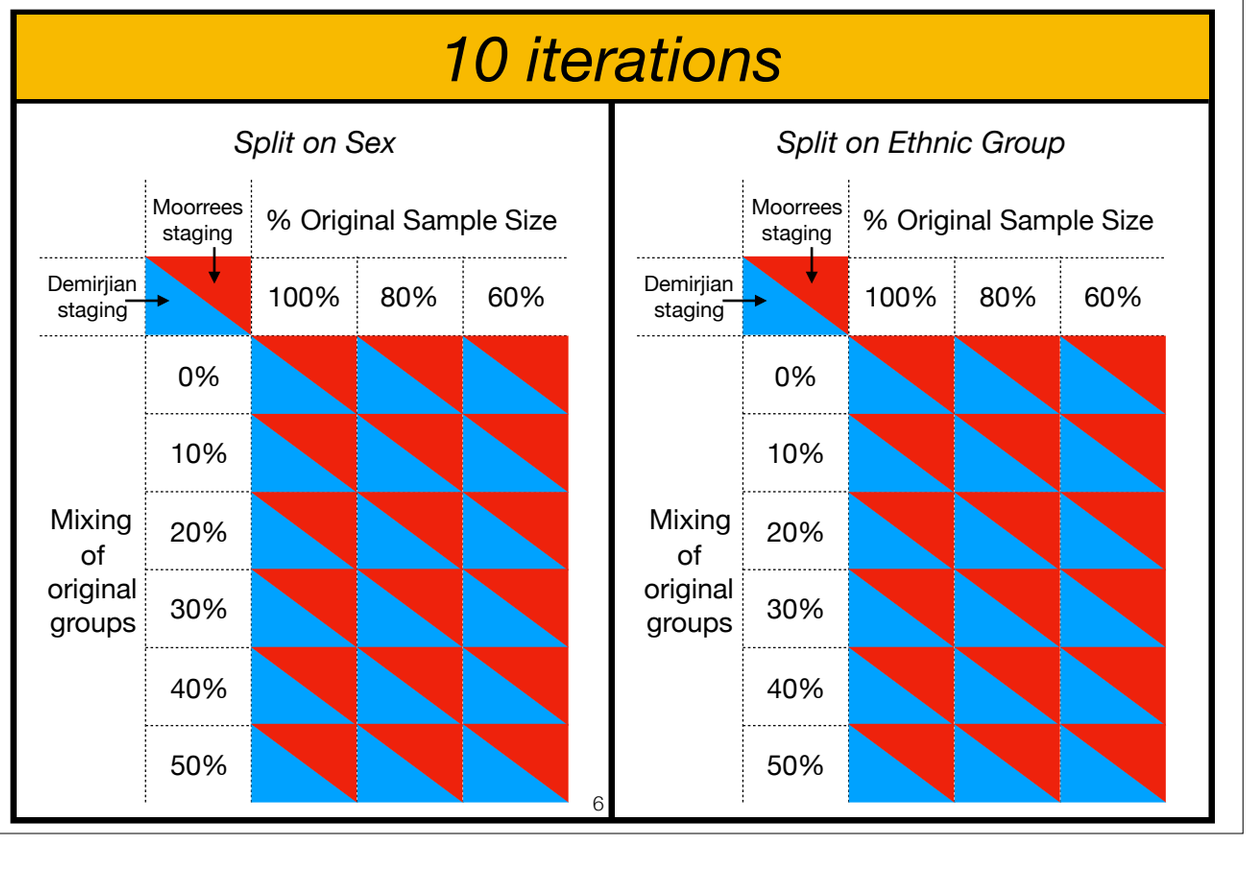
In order to study total sample size and within-stage sample size independently, we collapsed Moorrees et al. stages into Demirjian et al. stages using the conversion shown here. This provided a consistent and non-arbitrary secondary staging system.

Experimental Design

1. Separate individuals into two groups by sex or ethnic group
2. Randomly trim larger group to size of smaller group
3. Fit a cumulative probit model to each group for both staging systems
4. Compute standard error of the difference in mean age at transition between groups
5. For each tooth, report mean SE of all stage transitions
6. Randomly trim each group to 80% and then 60% of original size, repeat steps 3-5
7. Randomly swap 10% of each original group, repeat steps 3-6
8. Iterate 10 times, report mean and SD of step 5 SE
9. Repeat entire process separating by other variable (sex or ethnic group)

The experimental design was used to manipulate group composition, total sample size, and within stage sample size with as much mutual independence as possible. Our test metric was standard error of the difference in mean age at transition between groups. This metric should reflect uncertainty in model parameters due to the reference sample composition between both groups.

Design for 1 tooth



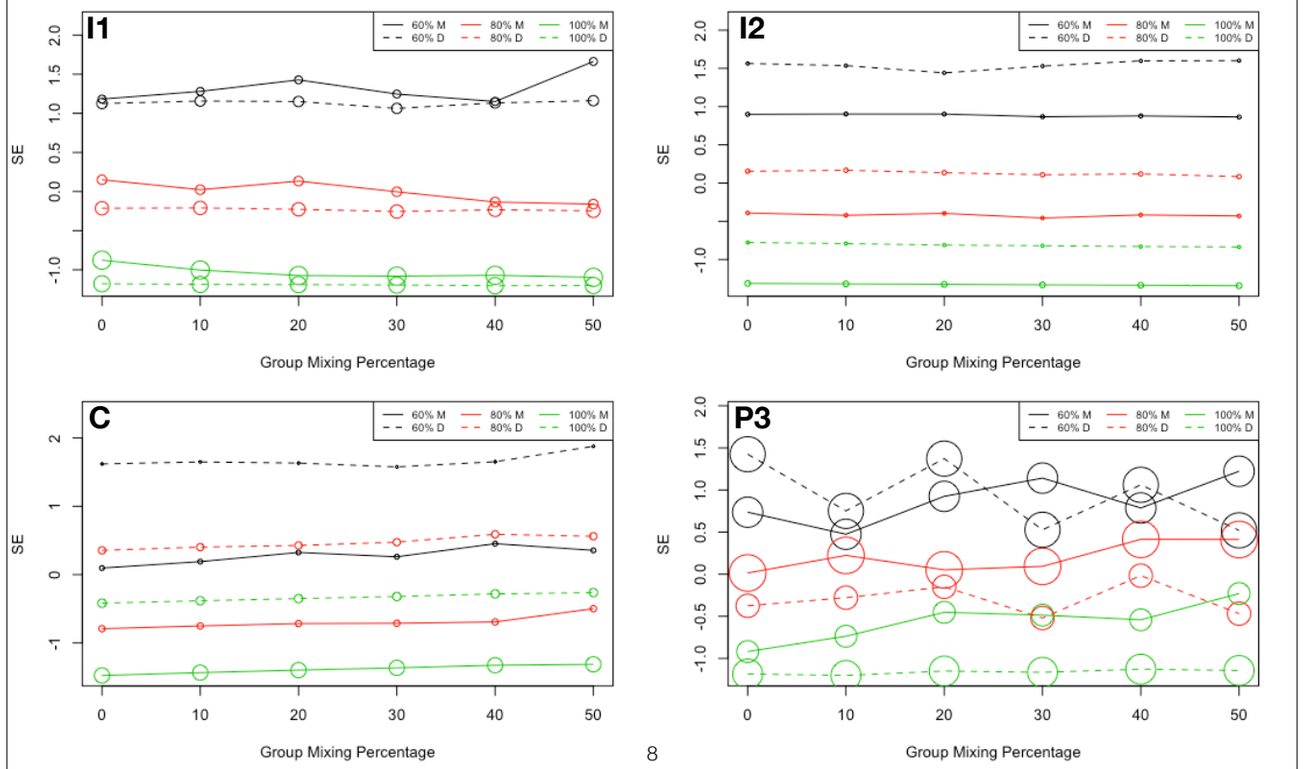
For each iteration, SE was calculated in a complete design for three sample sizes, two staging systems, and six mixing percentages, as shown here. Since groups were trimmed randomly, the process was iterated ten times for each tooth.

Predictions

AIC Result	Prediction
Sex is significant for all teeth	SE will increase with increased mixing by sex
Ethnic group significant for 1st premolar and 3rd molar	SE will increase with mixing by ethnic group for these teeth only
NA	Demirjian stages will produce lower SE
NA	Reduced sample sizes will increase SE

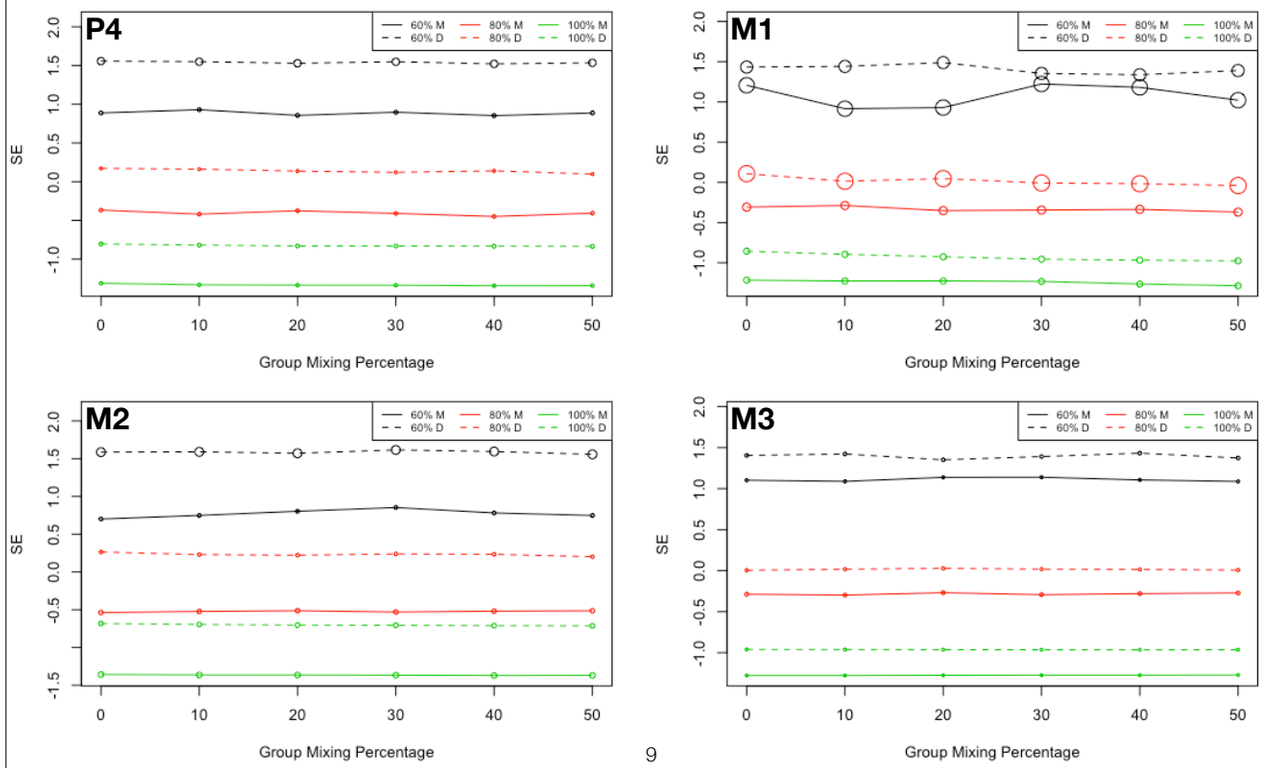
We began the analysis by fitting cumulative probit models with sex, ethnic group, and age as explanatory variables for tooth stage. Stepwise AIC of the of the models found that sex was a significant explanatory variable for all teeth. Based on this result, we would expect that standard error will increase with increased mixing by sex. Ethnic group was significant for the first premolar and third molar only. This corresponds to Liversidge (2011) in which the first premolar and third molar were the only teeth with stage differences at $p<0.01$. The expectation here is that these teeth will show increased standard error with mixing by ethnic group.
We predict that Demirjian stages will produce lower SE than Moorrees et al. stages because collapsing stages will increase within-stage sample size. We also predict that lower overall sample size will increase standard error through small sample size effects.

Results by sex

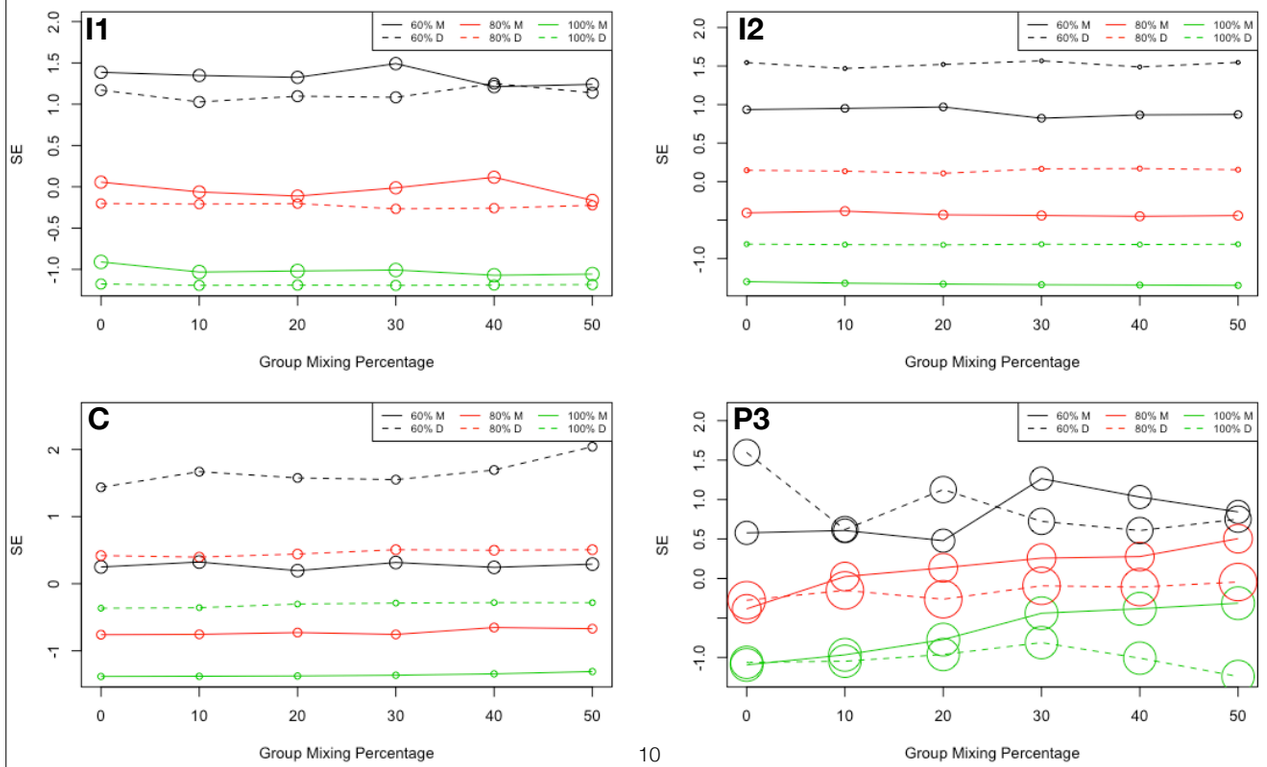


Plots are colored by percentage of original sample size. Circles are proportional to the standard deviation of the standard error measurement across ten iterations. Standard error is transformed to the standard normal distribution to facilitate comparison between teeth.

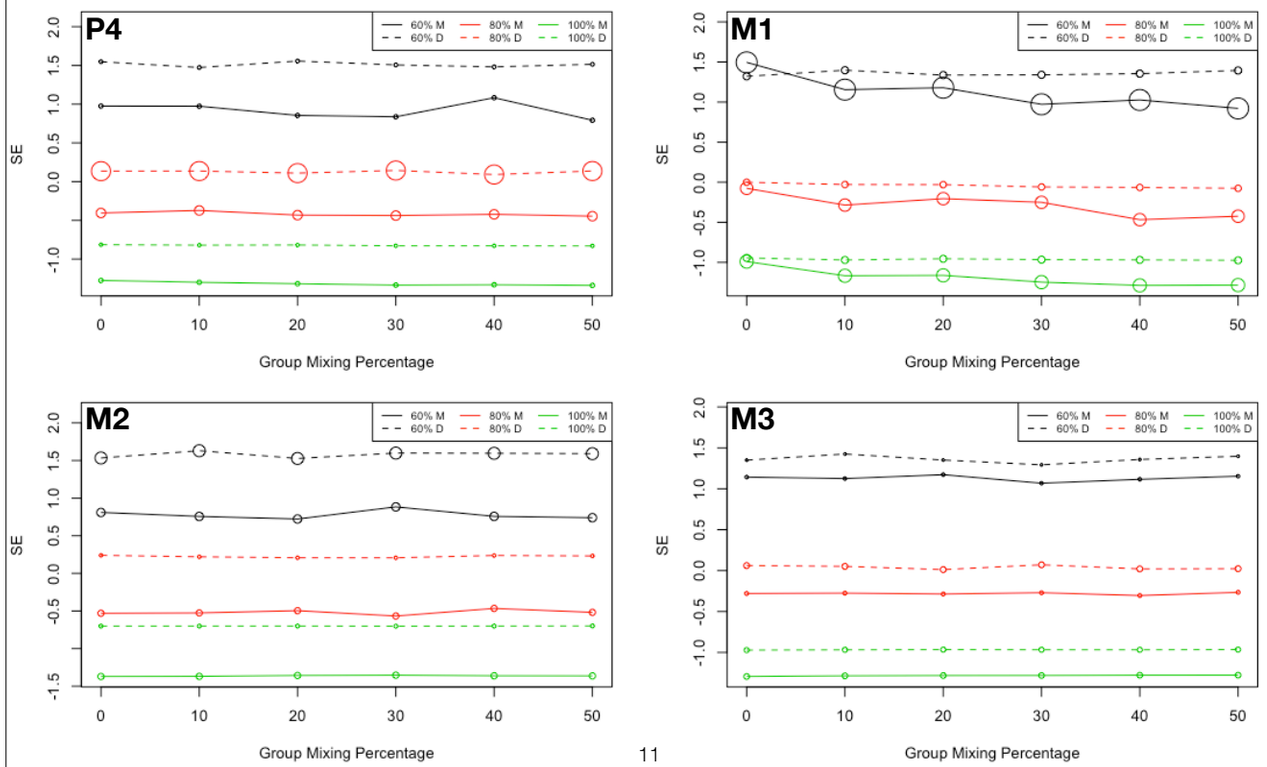
Results by sex



Results by ethnic group



Results by ethnic group



Predictions Revisited

Prediction	Reality
SE will increase with increased mixing by sex	Mixing by sex has no consistent effect
SE will increase with mixing by ethnic group for P3 and M3	Mixing by ethnic group has no consistent effect
Demirjian stages will produce lower SE	Demirjian stages produce higher SE for all but I1 and P3
Reduced sample sizes will increase SE	TRUE!

Results show that standard error was far more affected by tooth than by what variable was being mixed. Tooth by tooth patterns are consistent between the two mixing variables. There was no consistent upward trend in SE by mixing percentage, SE remained generally flat. Demirjian et al. stages generally produced higher rather than lower SE, however SE increased with decreasing overall sample size.

What does it mean?

- Small, demographically homogenous reference samples produce less consistent models than larger, heterogenous samples
  - Suggests that many population differences could be the result of small sample sizes
  - Support for pooling reference samples
- Scoring systems with fewer stages are not necessarily better when sample size is low
  - Evaluate model fit, then collapse stages

The results suggest that larger reference samples produce more stable models even if they are pooled across demographic variables. Interestingly, increasing within-stage sample sizes by reducing number of stages did not improve model consistency for most teeth. This argues against a one-size-fits-all approach to stage collapsing. Instead, goodness-of-fit tests should be used for each tooth to evaluate the best staging system.

Next Steps

- SE of difference in mean age at transition is indirect measure of model performance
  - Using reference sample design presented here:
    1. Test model fits
    2. Performance test age estimation models

Standard error is a reflection of model consistency, but does not directly test model performance. The next step is therefore to evaluate model fits and then perform direct testing of the performance of different reference sample configurations in estimating ages.