Author's response to reviews

Title: Insights into age- and sickle-cell-disease- interaction using principal components analysis

Authors:

Mamta Sharma (mamtassharma@rediffmail.com)
Manju R Mamtani (manjumamtani@rediffmail.com)
Manik Amin (saiamin@yahoo.com)
Tushar P Thakre (tthakre@hsc.unt.edu)
Smita Sharma (smitasssharma@yahoo.co.uk)
Amit Amin (amit_p_amin@yahoo.com)
Hemant Kulkarni (hemant_kulkarnius@yahoo.com)

Version: 2 **Date:** 16 August 2006

Author's response to reviews: see over

To
The Editor
BMC Blood Disorders

Re:
MS: 5446203421090990
Insights into age- and sickle-cell-disease- interaction using principal components analysis
Mamta Sharma, Manju R Mamtani, Manik Amin, Tushar P Thakre, Smita Sharma, Amit Amin and Hemant Kulkarni

Dear Sir

First, we are extremely thankful to you for a quick and critical review of our abovementioned manuscript. We are also very thankful to both the Reviewers for extremely encouraging words. We will like to place on record our special thanks to Dr. Hasan for the kind words. We are including in this letter our response to the two questions that Dr. van der Merwe has asked. Since neither reviewer recommended any change in the manuscript, we are persisting with the version of the manuscript that we had originally submitted.

We hereby resubmit our manuscript for publication in your esteemed Journal.

Sincerely

Hemant Kulkarni, MD
7458 Louis Pasteur Dr
Apt #1205
San Antonio, TX 78229

Response to questions of Reviewer #2

Reviewer's question 1
"Firstly, why did they derive the principal components for the hemoglobin AA subjects only instead of for the whole sample?"

Our Response
Our main research question was to assess if the factor structure obtained from hematological parameters was influenced by the sickle cell genotype. Clearly, an answer to this question needed to be arrived at in two steps. First, we needed to understand the principal components derived from normal subjects and second we needed to see if this derived factor structure is altered in subjects with the sickle cell gene – either in homozygous or heterozygous state. Hemoglobin AA genotype represents the normal non-mutated adult hemoglobin. Therefore this group of subjects served as the reference group for deriving the principal components. Moreover, to our knowledge, there is currently no study that has derived principal components based on the parameters we studied. Thus the normal factor structure based on hematological parameters is currently unknown. Further, if we had derived the factor structure based on all the subjects, we anticipated that one of the retained factors would have been the sickle cell genotype itself. In that case, the interpretation from the results would have been that sickle cell genotype influences the peripheral blood hematological parameters. This inference would have been neither compatible with our research question nor novel by disposition. Thus we would have missed the more interesting research question that was the focus of our investigation.

Reviewer's question 2
"Secondly, why did they categorise the ages into age groups instead of treating age as a numerical confounder? It would give them much more power."

Our Response
This is a very pertinent question and it is well established that the process of discretization will invariably lead to a loss of information contained within a continuously distributed variable if an assumption of linear association is valid. However, our decision to categorize age was based on the following reasons considered in unison:

a. We wished to tease out if particular clinically relevant age groups influenced the factor structure rather than whether age is a predictor of the factor scores.
b. If age would have been used as a continuous variable then it would have constrained our analysis around an implicit assumption of a linear association between factor scores and age. However, as our age categorization indicates, we had no *a priori* basis for assuming a linear relationship.
c. We conducted analysis of variance using factor scores as the dependent variable and the sickle cell genotype and age as the predictors. We first used age categories and then used age as a continuous variable (in which case it amounted to an analysis of covariance). The results of these analyses indicated (especially for the second factor score) that the model fit was better when age categories were

used as the predictor. These results suggested that the assumption of linearity for our dataset may be questionable. The table below shows the adjusted R^2 values based on the ANOVA/ANCOVA models.

Factor score	*Age categorized*	*Age used as a continuous variable*
First factor	0.3458	0.3345
Second factor	0.0226	0.0073
Third factor	0.2169	0.1746

We believe that these reasons together justify our use of age categories.