Research Article **Open Access**

Statistical Approaches to Assess the Effects of Disease on Neurocognitive Function Over Time

Tracy L Bergemann[1,6]*, Paul Bangirana[2,3], Michael J Boivin[4,5], John E Connett[1], Bruno J Giordani[5] and Chandy C John[7]

[1]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA

[2]Department of Psychiatry, Makerere University School of Medicine, Kampala, Uganda

[3]Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden

[4]International Neurologic and Psychiatric Epidemiology Program, College of Osteopathic Medicine, Michigan State University, East Lansing, Michigan, USA

[5]Neuropsychology Section, Department of Psychiatry, University of Michigan, Ann Arbor, Michigan

[6]Cardiac Rhythm Disease Management, Medtronic, Mounds View, Minnesota, USA

[7]Division of Global Pediatrics, Department of Pediatrics, University of Minnesota Medical School, Minneapolis, Minnesota, USA

Abstract

Introduction: Assessment of the effects of disease on neurocognitive outcomes in children over time presents several challenges. These challenges are particularly pronounced when conducting studies in low-income countries, where standardization and validation is required for tests developed originally in high-income countries. We present a statistical methodology to assess multiple neurocognitive outcomes over time. We address the standardization and adjustment for age in neurocognitive testing, present a statistical methodology for development of a global neurocognitive score, and assess changes in individual and global neurocognitive scores over time in a cohort of children with cerebral malaria.

Methods: Ugandan children with cerebral malaria (CM, N = 44), uncomplicated malaria (UM, N = 54) and community controls (N = 89) were assessed by cognitive tests of working memory, executive attention and tactile learning at 0, 3, 6 and 24 months after recruitment. Tests were previously developed and validated for the local area. Test scores were adjusted for age, and a global score was developed based on the controls that combined the assessments of impairment in each neurocognitive domain. Global normalized Z-scores were computed for each of the three study groups. Model-based tests compare the Z-scores between groups.

Results: We found that continuous Z-scores gave more powerful conclusions than previous analyses of the dataset. For example, at all four time points, children with CM had significantly lower global Z-scores than controls and children with UM. Our methods also provide more detailed descriptions of longitudinal trends. For example, the Z-scores of children with CM improved from initial testing to 3 months, but remained at approximately the same level below those of controls or children with UM from 3 to 24 months. Our methods for combining scores are more powerful than tests of individual cognitive domains, as testing of the individual domains revealed differences at only some but not all time points.

Introduction

Cognitive impairment has been reported in children affected by a number of infectious and non-infectious diseases in sub-Saharan Africa [1-3]. For example, our studies have shown that ~25% of children surviving an episode of cerebral malaria will develop long-term impairment in one or more neuropsychological domains [4,5]. These data suggest that almost 200,000 African children a year develop cognitive impairment after cerebral malaria. Other studies demonstrate that iron deficiency, which is very common in children in low-income countries, is frequently associated with cognitive impairment [6,7]. It is likely that millions of children are at risk for cognitive deficits from infectious and non-infectious diseases, but the long-term cognitive effects of disease on children have been studied in only a few contexts. Longitudinal studies to assess the effects of disease on childhood neurocognitive development, and the pathogenesis of cognitive impairment in children with specific diseases, are urgently needed, so that the magnitude of the problem can be defined and so that new interventions to decrease cognitive deficits can be planned.

The longitudinal assessment of the effects of disease on cognition in children also presents a number of challenges, particularly for children in low-income countries, where validated and standardized instruments have not been developed. A limited number of cognitive tests have been validated in such settings [8,9], meaning that assessing cognitive deficits may require more careful study design and statistical modeling, especially when this assessment happens over time. An assessment of the general effects of disease on cognition needs to account for the different ways in which disease can affect the brain and the psychometric properties of the different tests being used. This is further complicated when instruments to measure cognitive function variables report results on different scales, with different ranges and different distributions [10]. When assessing the effects of disease on cognition, a global score that incorporates the scores from each domain may be useful for quantification of the effects of disease on general neurocognition. Such an approach has been developed and practiced

in several other contexts such as neurofibromatosis and fetal alcohol spectrum disorders [11-13]. A common metric, such as a Z-score, may also be helpful to compare test results across different groups, studies, and test types.

The utility of combining these tests is problematic when the relationships between variables are non-linear, and tests require adjustment according to local norms of test performance as well as adjustment for age. Additional potential problems arise when a small number of age-matched controls are available to provide a norm within each age category. Traditional approaches like ANCOVA analysis that incorporate confounding variables are inappropriate if the raw score performance distribution deviates from the normal and functional transformation of raw scores to a linear mapping is insufficient.

We have encountered each of these problems in our assessments of the relationship between cerebral malaria and long-term neurocognitive deficit in children [5,10]. In studies of development and cognition in low-income countries, a number of different strategies have been used to assess cognition over time [8,14-16], but many of the issues outlined above are not addressed or are only partially addressed in these studies. In the present paper, we propose a step-by-step analytic strategy to assess neurocognition that addresses each of the challenges we have outlined, using data first presented in Boivin et al., and John et al., [5,10]. The global neurocognitive score featured in the present study is derived from (1) a working memory score using the Kaufman Assessment Battery for Children [K-ABC], (2) executive attention using the visual form of the computerized Tests of Variables of Attention [TOVA], and (3) tactile-based learning using the Tactual Performance Test [TPT]. It is reasonable to combine these three domains into a global neurocognitive score because these domains have been implicated as especially vulnerable to persisting neurocognitive disability in the aftermath of cerebral malaria [10,17,18]. As such, the global neurocognitive score can be interpreted as an index of a neuropsychological deficit profile that might be expected specifically with school-age children surviving cerebral malaria. We address the standardization and adjustment for age in neurocognitive testing, present a statistical methodology for development of a global neurocognitive score, and describe assessment of changes in individual and global neurocognitive scores over time. The methods used are presented in detail, as they may provide a roadmap for other studies to conduct comparative assessments of neurocognitive function over time between groups of children.

Materials and Methods

Data description

The variables in our analysis were provided by children who had participated in earlier studies examining the cognitive and neurological outcomes of cerebral malaria, with testing at 0, 3, 6 and 24 months of follow-up [5,10]. Children 4 to 12 years of age were recruited as part of two studies assessing the complications of cerebral malaria. Longitudinal assessment of test performance was conducted only for children 5 to 12 years of age, because the ability of children 4 years of age to perform the different tests was highly variable. Children with cerebral malaria (coma, Plasmodium falciparum on blood smears, and no other cause of coma) or uncomplicated malaria (fever, P. falciparum infection on blood smear, and no World Health Organization criteria for severe malaria or evidence of other acute illness) were recruited for the study. A third group of community controls (CC) was recruited from the extended family or neighbourhood of children with cerebral malaria (CM) or uncomplicated malaria (UM). Controls and children

with UM were recruited to be in the same age range as children with CM. The three study groups were comparable with respect to potential confounders (Additional file 1: Table S1, 10). Complete details of enrolment criteria and study groups have been previously published [5,10].

One hundred and eighty-seven children (44 children with CM, 54 children with UM, and 89 CC) were enrolled in the study of neurocognition and were willing and able to undergo neurocognitive assessment (Table 1). Neurocognitive testing was performed at discharge for children with CM, 3 days after treatment for children with UM, and at the time of enrolment for CC children. Testing was also performed at 3 and 6-month and 24-month follow-up [5,10]. Follow-up in each study group was excellent (Table 1). Missing data at follow-up visits is assumed to be missing at random. There is no evidence that the absence of a data point is correlated with any other study outcomes. For example, the ages, group membership and previous neurocognitive test scores were similar in children without 24 month follow-up visits and the full sample (data not shown). Written informed consent was obtained from the parents or guardians of study participants. The institutional review boards for human studies at Makerere University School of Medicine, University of Minnesota, Indiana Wesleyan University and University Hospitals of Cleveland and Case Western Reserve University granted ethical approval for the study.

Cognitive assessments focused on (1) working memory using the Kaufman Assessment Battery for Children [K-ABC], (2) executive attention using the visual form of the computerized Tests of Variables of Attention [TOVA], and (3) tactile-based learning using the Tactual Performance Test [TPT]. Primary outcome measures used to define neurocognitive deficits were summary variables that assessed working memory (sequential processing of the K-ABC), executive attention (D prime test of the TOVA), and tactile-based learning (total time per block of the TPT). Complete details of the test batteries used are described in Boivin et al., [10]. These tests have been previously shown to be sensitive to the underlying neuropsychological constructs, robust across cultural contexts, and consistent in the manner in which they assess those underlying constructs across the age span for our school-age samples [17].

Data transformation

Some neurocognitive tests show improving performance with age, on the average. In order to compare children with cerebral malaria to other groups, test scores were adjusted for age first and then combined into a global score. Since the age and grouping variables were uncorrelated, each variable was modelled separately. The ultimate goal of our analysis is to convert multiple age-adjusted neurocognitive scores to a normalized global Z-score for each time point and then fit an appropriate longitudinal model.

The first step in our analysis was to adjust test scores for age. Thus,

Time of study	Study group			Total
	CM	UM	CC	
Baseline	44	54	89	187
3 months	42 (95%)	53 (98%)	87 (98%)	182
6 months	42 (95%)	52 (96%)	87 (98%)	181
2 years	38 (86%)	48 (89%)	84 (94%)	170

CM= children with cerebral malaria; UM = children with uncomplicated malaria; CC= community children

Table 1: Children enrolled in a study of neurodevelopmental impairment in cerebral malaria. At follow-up visits, the number of children with data available and the percentage of the original sample are provided.

we needed to determine if the relationship between age and each test score was linear or non-linear. A two-dimensional plot of age in years versus test score at a particular time point can assess this relationship. In addition to graphical assessment, a Box-Cox analysis can determine if a linear model is preferable to a polynomial transformation [19]. Figure 1 indicates a linear relationship between age and the working memory score. Similar figures for other time points and neurocognitive scores are provided in Additional File 1 (Figures S1, S3, and S5). From these figures, we concluded that the association between age and working memory was linear, age and executive attention was linear, and age and tactile learning was non-linear. We used linear regression for the two variables with linear relationships, and locally weighted scatter plot smoother (loess) regression for the variable with a non-linear relationship, to adjust for age at each time point [19]. All available samples in the control group at a particular time point are used in the age adjustment. Any missing data from neurocognitive tests are not a concern in the regressions because sufficient representation exists across the continuum of ages and test scores. The resulting residuals from these models act as Z-score values that were then combined into a global neurocognitive score and used to compare performance in groups over time.

For the two variables (working memory and executive attention) for which age and test score were linearly related, a linear regression was fit between age in years and the test score and the resulting residuals were used. Comparative tests based on residuals usually lead to bias unless the variables in question are orthogonal (uncorrelated) [20]. The Spearman and Pearson correlation estimates of age and group were both 0.03 in our dataset, and these values, along with the scatter plots in Figures S1, S3, and S5, show that age and group were unrelated. An ANOVA model of age by group also showed no significant differences. Thus, we were able to use age-adjusted residuals for comparative tests. The regression was fit for the control group only and then applied to subjects in both malaria groups to obtain the residuals. Using model residuals as Z-scores allowed us to combine several neurocognitive test scores into a single variable that describes global neurocognitive function.
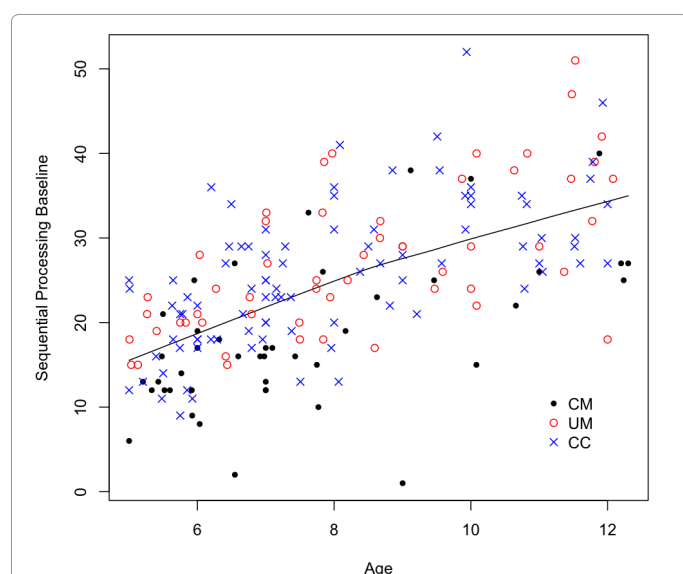
Studentized residuals from a regression are often preferred to ordinary residuals. The studentized residuals are centered and scaled and therefore follow a standard normal distribution when linear regression model assumptions are approximately met [19]. Therefore we have made use of the studentized residuals as Z-scores. With a Z-score, a value above 2 or below -2 falls outside of the range of 95% of the data. The normality of the Z-score was assessed with a histogram (Additional File 1: Figures S2 and S4).

For the tactile learning score, which had a non-linear relationship with age, a loess regression between age and score was fit [19]. A linear regression would not adequately estimate the relationship between age and the tactile learning score, and there was not sufficient evidence to suggest an obvious function of the relationship between covariate and response. A loess regression was used to adjust for age because it allowed for irregular curvature in the association. The residuals from the loess curve were converted to Z-scores. This was achieved by fitting a loess curve to the control group only and then applying this fit to the malaria subjects also to obtain residuals and then Z-scores. The Z-scores followed an approximate standard normal that was, again, assessed with a histogram (Additional File 1: Figure S6).

Statistical testing and estimation of confidence intervals for longitudinal models of continuous variables generally require that the model residuals be normally distributed. The combination of Z-scores into a global score also requires that the individual test scores follow the same distribution and have the same scale. So the next step was to determine if residuals from the linear and loess regression models roughly follow a normal distribution. Figure 2 shows a histogram for loess residuals of the raw tactile learning score. These residuals are clearly non-normal with a strong positive skew. While the majority of values are close to zero, the range of the data extends up to 800. We observed the same non-normality after either linear adjustment or non-linear regression adjustment. For distributions of this shape, a log transform of the outcome will often result in residuals that more closely resemble a normal distribution, and indeed Figure 2 indicates that after the log transform, the distribution of the residuals is closer to normal.

After conversion of each of the three testing scores to standard normal distributions, the scores fell on the same scale. Thus, the three scores could be summed together and scaled by an estimate of the square root of the variance to compute a global neurocognitive score at each time point. This scaling was necessary to obtain another standard normal distribution. The variance estimate uses the pair wise correlations between each of the three tests, denoted ρ_{ij}.

Figure 1: The Working Memory (Sequential Processing) Score collected at baseline versus age. The line indicates the loess fit using a span of 1. The disease groups are CM = cerebral malaria, UM = uncomplicated malaria, and CC = community controls.
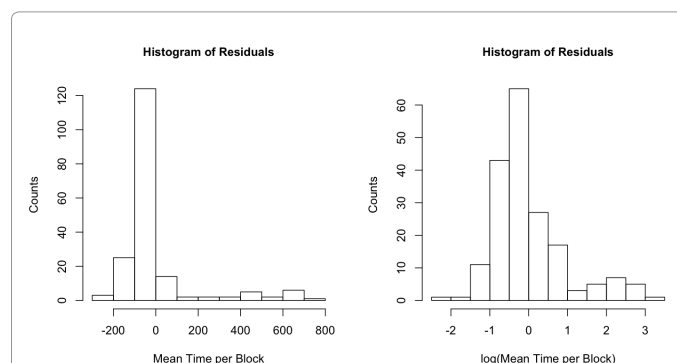

Figure 2: Histograms of the residuals derived from the age adjustment of the Tactile-based Learning Score, before and after log function transformation.

A linear combination of three standard normal distributions also yields a normal distribution with an expectation of zero. The variance of this normal distribution is

$$\begin{aligned}
\text{Var}(Z_1 + Z_2 - Z_3) &= \text{Var}(Z_1) + \text{Var}(Z_2) + \text{Var}(Z_3) \\
&\quad + 2\text{Cov}(Z_1, Z_2) - 2\text{Cov}(Z_1, Z_3) - 2\text{Cov}(Z_2, Z_3) \\
&= 1 + 1 + 1 + 2\rho_{12} - 2\rho_{13} - 2\rho_{23} \\
&= 3 + 2(\rho_{12} - \rho_{13} - \rho_{23})
\end{aligned}$$

In order to obtain another standard normal, we simply divide the linear combination by the square root of the variance. Replacing the variance with its estimates will yield an approximate but not exact standard normal distribution. The pair wise correlations can be estimated in a straightforward manner using a Pearson correlation coefficient or similar.

The scores were summed so that a higher score indicates higher neurocognitive function. The tactile learning score takes a negative value in the sum because a higher score reflects worse outcome. Each of the three neurocognitive tests contributed equally to the final score:

Cognitive Z =

$$\frac{(\text{Working Memory Z} + \text{Executive Attention Z} - \text{Tactile Learning Z})}{\sqrt{3 + 2(\hat{\rho}_{12} - \hat{\rho}_{13} - \hat{\rho}_{23})}}$$

Histograms indicated that these scores had approximate standard normal distributions at each time point (Additional File 1: Figure S7). Note that the above Z-score can be updated when additional samples are made available in a study.

While the global Z-score above assumed that each test contributed equally to neurocognitive function, it is possible to estimate a linear combination using the data in order to give preference to one neurocognitive score over another. Using eigenvalue decomposition, principal components analysis (PCA) will estimate linear combinations of the data such that the first component of the set explains the largest proportion of variance. Thus, in addition to the estimation of a neurocognitive Z-score, we also examined a score estimated using the first component in a PCA. The variance explained and the component loadings for our dataset are given in the Results section. The principal component score is scaled so as to have an approximate standard normal distribution, requiring input test scores to again have a normal distribution. We refer to this estimate as the neurocognitive PCA-score. Additional File 1: Figure S9 shows that the global neurocognitive Z-scores and PCA-scores were very similar.

Longitudinal models

A linear mixed effects (LME) model was then used to assess changes in scores over time between groups of interest. The global neurocognitive Z-score or the PCA-score, computed above, was the primary endpoint in the model. The model assumes measurements at the four time points were correlated within individual. The model also assumes that missing data is missing at random. In this study, the assumption is thought to be reasonable, especially given the low dropout rate over time. The LME is equivalent to a composite multilevel model for change, as described by Singer and Willett, where the structural component fits a group by time interaction and the stochastic component fits an individual intercept [21]. The LME model allowed time trends to vary by group using a time by group interaction term in addition to their main effects. An F-test for the group main effect compares the three groups of interest: cerebral malaria, uncomplicated malaria, and community controls. Model-based t-tests were used to examine pairwise comparisons or contrasts.

The resulting p-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction [22]. The model was fit in R using the nlme package [23].

The approach outlined above, assessing a cumulative or global neurocognitive score over time, was compared to a set of three independent linear mixed effects models. In these three independent models, each one of the three neurocognitive tests was fit separately. The model's fit used the same equations as those used for the global Z-scores but the model outcome was the raw test score and so there was an additional covariate for age. When the global Z-score was constructed, the age adjustment differs for each individual test and so the age adjustment and longitudinal models are performed separately. When the models are fit separately for each neurocognitive test, however, the age adjustment, longitudinal modelling, and group comparisons can be performed simultaneously [24].

The results of the linear mixed effects models were also compared with results reported previously that analyzed the same dataset with different methods [5,10]. In our earlier assessments, to account for age, each raw outcome was converted into an age-specific standardized Z-score based on the scores of community controls for each year of age. In each area of cognitive testing, a child was considered to have a cognitive deficit if the child's Z-score was < -2 (for working memory and attention, where a lower score was a worse outcome) or > 2 (for tactile learning, where a higher score reflected a worse outcome). The primary outcome (neurocognitive deficit) was defined as the presence of a deficit in one or more of the areas of neurocognition tested (working memory, executive attention, tactile learning). Outcomes were recorded independently at each time point. A Fisher's Exact or χ^2 test, depending on sample sizes, compared the frequency of cognitive deficits between groups.
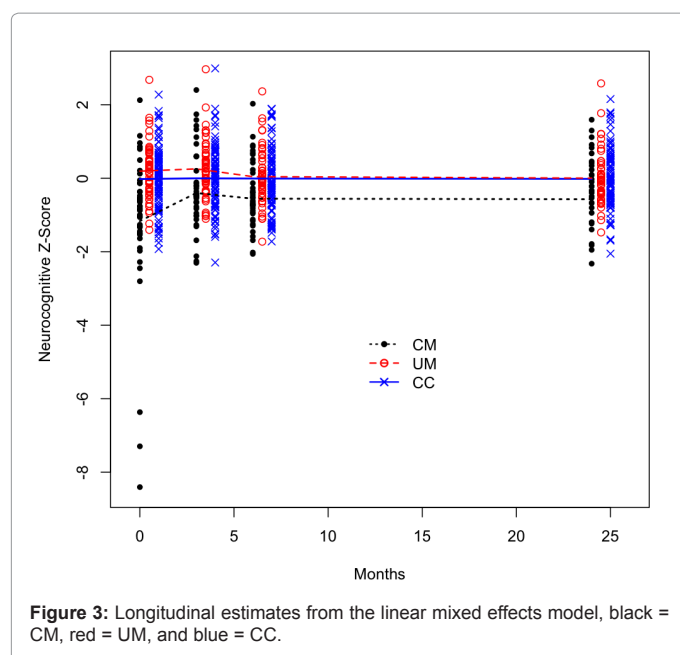
Results

Linear mixed effect model of global neurocognitive Z-score

The linear mixed effects model of global neurocognitive Z-scores estimates the change in cognitive function over time for each group of interest. The model estimates the average global neurocognitive Z-score at each time point within each group. The value of the Z-score for any study subject reflects the number of standard deviations away from the mean of the control group. The estimated average Z-score in the control group is zero at each time point because they are the reference group for comparison. An F-test of the group main effect showed that there was a significant difference between the three groups (p < 0.001). Figure 3 shows the data included in the model and the fitted trend lines for each group. Note that this figure does not show the raw development progression of the three groups but rather, how the CM group global Z-scores compares to the global Z-scores of the other groups (CC as the reference).

This figure demonstrates that on average the cerebral malaria group has lower cognitive scores at all time points, with the worst relative performance at baseline. The fitted model estimate in Figure 3 indicates that the CM group had lower scores than the other two groups at baseline but made a sharp recovery by the month 3 assessment. Nonetheless, despite this recovery, a constant deficit persisted after 3 months compared to the other two study groups.

The actual estimates from the linear mixed effects model of Z-scores are provided in Table 2. The slope estimates demonstrate the change in the global cognitive Z-score over time and between groups. These changes correspond to the amount of the standard deviation from the

Figure 3: Longitudinal estimates from the linear mixed effects model, black = CM, red = UM, and blue = CC.

Variable	Difference in the Z-score	95% Confidence Interval	Unadjusted p-value
CM – CC at each time point			
CM – CC: baseline	-1.14	(-1.51, -0.77)	< 0.0001
CM – CC: 3 months	-0.40	(-0.77, -0.02)	0.04
CM – CC: 6 months	-0.55	(-0.92, -0.17)	0.005
CM – CC: 2 years	-0.55	(-0.94, -0.17)	0.005
UM – CC at each time point			
UM – CC: baseline	0.21	(-0.13, 0.56)	0.22
UM – CC: 3 months	0.26	(-0.09, 0.60)	0.15
UM – CC: 6 months	0.05	(-0.30, 0.40)	0.77
UM – CC: 2 years	0.02	(-0.33, 0.37)	0.91
CM trend over time			
CM: 3 months – baseline	0.76	(0.54, 0.98)	< 0.0001
CM: 6 months - 3 months	-0.15	(-0.38, 0.07)	0.17
CM: 2 years - 6 months	-0.02	(-0.24, 0.21)	0.89

Table 2: Estimates of the difference in neurocognitive Z-scores over time in children with cerebral malaria (CM), uncomplicated malaria (UM) and community children (CC) from the linear mixed effects model.

mean of the standard normal distribution. The comparison between community controls and cerebral malaria at baseline, for example, gave a difference in the Z-score of 1.14, meaning that community controls had cognitive scores on average that are 1.14 standard deviations higher than the cerebral malaria patients. Other estimates in Table 2 are interpreted similarly.

The t-tests in Table 2 assessed the significance of the pairwise comparisons listed. At each time point, the comparison between community controls and cerebral malaria is statistically significant, even after adjusting for multiple testing. There is no detected difference between children with uncomplicated malaria and community controls at any time point. We also tested for a difference between time

points within the CM group. Within the cerebral malaria group, we found a significant difference in cognitive Z-scores between baseline and three months, but not thereafter. This shows some neurocognitive recovery after baseline in the CM group. From three months to two years, the slope estimates within the CM group are quite close to zero. Figure 3 reiterates this phenomenon, showing a persistent and constant deficit from 3 months out to 24 months between children with cerebral malaria compared to both community controls and children with uncomplicated malaria.

Given that slopes are nearly constant after three months for all study groups, the model results suggest a potential reduction in the number of parameters that need to be fit to describe the longitudinal behaviour. Thus, based on these results, a post-hoc analysis was also conducted with a model that estimates the slopes for each study group before and after a change-point set at 3 months. The models were fit for both the global neurocognitive Z-scores and for the PCA-scores. Results of the model fit are provided in Supplemental Tables S6 and S7. The results of the change-point model confirm those provided in Table 2, Table 3 and Figure 3. Namely, the CM group differs from the CC group at baseline, followed by a statistically significant increase in scores between baseline and three months and then no detectable change in scores thereafter. No difference between the UM and CC groups are detected and no change over time is detected for the UM or CC group.

Figure 3 indicates that three patients in the CM group had neurocognitive Z-scores below -3 at baseline. We had no statistical or informational reason to discard these potential outliers from our dataset and therefore retain them in our analysis and results. Indeed, the poor performance of these children at baseline suggests that the initial neuropsychological assessment was sensitive to immediate post-illness malaise and subsequent assessment of these children shows a marked recovery in neurocognitive performance. A supplemental analysis was also performed that excluded these three patients and the results are given in Additional File 1: Table S2. This table reports slightly different conclusions for specific comparisons than those in the

Variable	Difference in the Z-score	95% Confidence Interval	Unadjusted p-value
CM – CC at each time point			
CM – CC: baseline	-0.71	(-1.08, -0.33)	0.0003
CM – CC: 3 months	-0.35	(-0.73, 0.02)	0.07
CM – CC: 6 months	-0.54	(-0.91, -0.16)	0.006
CM – CC: 2 years	-0.58	(-0.97, -0.20)	0.003
UM – CC at each time point			
UM – CC: baseline	0.24	(-0.10, 0.58)	0.17
UM – CC: 3 months	0.30	(-0.04, 0.65)	0.09
UM – CC: 6 months	0.10	(-0.25, 0.44)	0.58
UM – CC: 2 years	0.04	(-0.31, 0.39)	0.83
CM trend over time			
CM: 3 months – baseline	0.24	(0.01, 0.47)	0.04
CM: 6 months - 3 months	-0.08	(-0.32, 0.15)	0.48
CM: 2 years - 6 months	-0.05	(-0.29, 0.20)	0.71

Table 3: Estimates of the difference in PCA-scores of cognitive tests in children with cerebral malaria (CM), uncomplicated malaria (UM) and community children (CC) from the linear mixed effects model.

full dataset, yet the overall message from the results of the modified analysis are the same.

Linear mixed effects model of principal components analysis (PCA) scores

The linear mixed effects model of PCA-scores is shown in Table 3. This table shows that the results of the model based on PCA-scores were very similar to the results for the Z-scores shown in Table 2. The first component in the PCA at baseline explained 71% of the variance in the three baseline test results. At all four time points, this percentage ranged from 58% to 71%. The first component at baseline weighted each of the three test scores roughly equal in the linear combination: 0.4 for working memory, 0.9 for executive attention, and -0.2 for tactile learning. At 3, 6, and 24 months, the weights of the first component were: 0.7 for working memory, 0.7 for executive attention, and -0.1 for tactile learning. Thus in the PCA, the tactile-based learning score is somewhat down-weighted compared to the other two tests at 3, 6, and 24 months.

Comparison to previous dichotomous analysis and to linear mixed effects models of individual domain scores

In Boivin et al., we found a significant differences in the frequency of cognitive deficits at 6 months between cerebral malaria patients (21.4%) and community controls (5.7%) with a corresponding p-value of 0.01 from the χ^2 test [10]. As shown in Table 2, p-values at this time point and all other time points for difference between children with CM and CC are smaller for the present method of analysis than for the prior dichotomized analysis, indicating that the new approach has more power to detect differences in cognitive outcome between groups.

The results of the LME model of global cognitive Z-scores were also compared to three independent LME models for working memory, executive attention, and tactile-based learning. The results of the individual models of cognitive tests are shown in Additional File 1: Tables S3, S4, and S5. While for the most part the results of the individual models coincide with the cumulative model shown in Table 2, there are some differences. Notably, in the cumulative model, all comparisons of community controls versus cerebral malaria are statistically significant at all time points. This does not hold up in the individual models. A global score such as the one devised may detect the overall or cumulative effects of disease on neurocognition better than the results for testing in a single area of neurocognition.

Discussion

This manuscript illustrates methods to summarize multiple cognitive measurements on children into a meaningful metric that can be used to test for association over time between clinical variables of interest and cognitive outcome. The methods adjust for age of the child and allow for the adjustment to differ with each cognitive measurement. The model of a global Z-score compares children with cerebral malaria who survive the illness after appropriate management and that are able to undergo neurocognitive assessment to uncomplicated malaria and community controls. The results suggest that the children with cerebral malaria recover some neurocognitive function within 3 months but then retain a persistent deficit thereafter compared to other groups. This opens up the intriguing speculation that more early and aggressive intervention in children with cerebral malaria within the first 3 months may help them regain better neurocognitive function that will persist over time. Though the model used is promising for this reason, it needs to be validated in other populations in future studies.

In the statistical methods we propose, the summary of cognitive tests into a single continuous score enables the modelling of longitudinal studies of cognition in a succinct interpretable way. The continuous score has advantages over previously used dichotomous scores that divide participants into two groups, for example cognitive deficit and normal function. Though dichotomous scores have the advantage of clear interpretability and obvious use for clinical decision-making, the disadvantage of these scores is that they lose information about the size of the cognitive deficit and therefore power to detect subtle differences. We provide two continuous alternatives: a global score and a PCA-based score.

The global scores equally weight the tests and the result is a continuous measure. The advantage of this approach is that we use all of our information in a summary statistic. If a deficit exists in only one test score, but not another, we will see this reduction on the continuous scale. They do not cancel each other out, just as they do not cancel out in a dichotomous setting. The reductions are more obvious when they occur in multiple tests, but they are still discernible when they occur in only one test. The disadvantage of the global score is that each test is equally weighted in the combination and it may be the case that one test is more clinically relevant than another. The PCA-based score has all of the power advantages of the global scores because it too is a linear combination of continuous measurements. It carries the further advantage over the global scores of using the data to determine the best weights on the test scores. This advantage can, however, also be a limitation because a data-driven weight estimate will change with each dataset. Therefore, we lose the consistency of our application.

We found that global and PCA-based scores were more powerful than dichotomous scores. We also ran analyses of each individual test score to ensure that there was not a wash-out effect over multiple tests. When we looked at PCA-based scores, we found that the data-driven weight estimates were roughly equal and the scores were very similar to the global Z-scores. So, the data suggest that our three test scores can be considered equally. This suggests that future analysis can use the global score only, because the weights will be consistent across future datasets in the same research context. If new batteries of tests are to be considered and different populations are to be studied, then another PCA-based analysis can be performed to determine weights on the tests in a new context.

In this research, we opted to adjust cognitive scores for age using either a linear or non-linear regression and develop models for the subsequent residuals. The mechanism by which age acted on score differed notably for each cognitive test. Our approach has the advantage of using all available samples in the control group to perform an age adjustment via a linear or non-linear function. This also allows the developmental pattern of a cognitive test to vary with increasing age. Since age had a linear relationship with two tests and a non-linear relationship with the third test, it was not possible to use statistical models like MANOVA or multivariate multi-level models of longitudinal data to analyze the three cognitive tests. When a neurocognitive score like tactile-based learning is non-linear and does not follow a polynomial function, an ANCOVA model would also be inappropriate. Accounting for this non-linearity with our approach will permit simultaneous modelling with other variables under study. ANCOVA is also not appropriate for studies that use raw score distributions with non-normality and different ranges or when standardized norms for the study group do not exist [25]. Our approach, by contrast, will facilitate the comparison of test results

between groups, across studies, and across cognitive performance measures by mapping raw scores to a common age-adjusted metric.

An important caveat to our approach exists, however, when there are interactions between the grouping variable and the age variable. Additional File 1: Figure S8 illustrates one hypothetical scenario where, because of this interaction, the studentized residuals from a linear regression model would be misleading. King also provides a useful review of when it is inappropriate to construct a regression on residuals [20]. A plot of age versus score within the grouping variables is helpful in this assessment as well as estimates of correlation or differences in the age distribution by group. If the pattern of age by score is consistent over groups, then the studentized residuals will not be biased. If the pattern is inconsistent, then a separate slopes ANCOVA model for each cognitive test, or a non-linear model with an interaction term, may be more appropriate.

Earlier studies of cognitive and neurodevelopmental impairment in African children have used a number of methods to analyze data on cognition and development, including [8,10-16,26-30]. Each of these methods has its strengths, but each also has limitations that may have less power or more bias for assessment of cognitive data, particularly longitudinal cognitive data. In the case of [30], data is log-transformed, but no record is provided about whether the transformed data meets assumptions of normality, or what was done if it did not meet these assumptions. In the current study, assumptions are tested and an alternative method of assessment provided, such as variable transformation or non-linear regression, if the assumptions are not met. Other previous studies controlled for covariates with linear regression analysis, but the cognitive tests were dichotomized into impairment groups [14,15].

The methods in this paper fit a linear mixed effects model for the assessment of longitudinal cognitive data. Other studies of longitudinal data may require variants of this model, depending on the variables involved. In other contexts, for example, it may be important to account for gender, SES, or other confounders. For clarity, we did not consider other confounders in this manuscript, but they could easily be added to the linear mixed effects model we have described. In general, more complicated longitudinal studies will require a greater degree of sophistication in a multilevel model of change.

The K-ABC Sequential Processing (memory) and TOVA (visual attention) measures significantly differed between exposure and control groups in a retrospective study of the effects of cerebral malaria (CM) in Senegalese children [18]. For this reason, we used these tests in our prospective study of cerebral malaria in Ugandan children [5,10,31]. The Tactual Performance Test (TPT) is part of the Halstead-Reitan Neuropsychological test battery for adults and for children. It measures tactile form recognition, spatial thinking, learning, and incidental memory [32]. Although TPT performance did not differ significantly between CM and controls in Senegal, overall TPT performance (time-per-block) was significantly related to anthropometric measures of nutritional wellbeing in children from DR Congo. Improvement in performance between the preferred and non-preferred hand trials of the TPT is a measure of global brain inter-hemispheric development [33] and was significantly correlated with K-ABC Sequential Processing in Congolese children [34]. Since TPT performance is one of the most robust measures of overall neuropsychological performance in a factor analysis of the Halstead-Reitan tests [35], we included TPT overall time-per-block performance as our third global indicator for neurocognitive function (along with K-ABC Sequential Processing and TOVA D prime).

In our prospective study of CM, our initial assessment was just before the children with CM were released from hospital. Since they were still recovering from a serious illness, we did not do a neuropsychological assessment such as the Halstead-Reitan that could easily take 3 to 4 hours to administer. We chose the K-ABC, TOVA, and TPT because together they provided a reasonable profile of the core domains of neurocognitive performance, and included measures that in the past had proven sensitive to the effects of CM. Further, based on converging lines of evidence, these three ability domains reflect a foundational brain/behaviour omnibus [17]. The present paper evaluates novel statistical methods for evaluating the differences among exposure groups for several core neurocognitive domains: K-ABC Memory, TOVA attention, and TPT tactile learning. It is important to do separate comparisons for these three measures so as to determine if a particular neurocognitive domain is particularly affected by CM, as we saw was the case with attention. However, we also combined these three measures into a global neurocognitive performance score for between-group comparison and assessment over time.

Scores from various neuropsychological tests are often combined in determining whether a patient has a global deficit or brain injury or neurodisability. The most widely used neuropsychological assessment battery is the Halstead-Reitan [36]. Performance measures from the Category Test (executive function), TPT (tactile learning and memory), Seashore Rhythm Test (nonverbal auditory discrimination), Speech Sounds Perception Test (verbal auditory discrimination), and Finger Oscillation Test (lateralized fine motor control) are combined to form the Halstead Impairment Index. This index represents the proportion of the 7 measures from these five tests for which the patient has performed below the normal cut-off for that test. An index score of 1.00 means that the patient is below the range of normal performance for all 7 measures.

The value of the Halstead-Reitan impairment index decreases as the number of test scores (or normative data for establishing a cut-off) declines. The index is of little value when the number of scores available is less than 5 [36]. However, even when all 7 measures are available for determining the impairment index, a patient with a score of 1.0 may only be mildly impaired on most of the measures, having missed the cut-off for all seven scores by only a small margin. Compare this patient to one with an impairment index of only 0.14, but with a profound deficit due to brain injury, on tactile-based learning as measured by the TPT. The composite score for our three tests used in the present analysis circumvents these limitations by providing for a quantitative composite score of disparate neurocognitive measures on a normalized scale, for which cut-offs were based on normative data for that cultural context (derived from tests of otherwise healthy community children). The composite index score provided us with sufficient statistical power to detect important effects of cerebral malaria disease on overall neurocognitive function.

In summary, the proposed approach to data assessment and the linear mixed effects model outlined for longitudinal assessment of neurocognition provide a more robust and accurate measure of neurocognitive changes over time than those we have used previously [5,10-15]. The methods of comparing groups with local age-appropriate control data, or comparing between groups, are particularly useful in low-income countries. Although there are important caveats to use of this approach, particularly in regard to types of psychometric testing used and the method of combining these tests, we believe that correctly implemented, it has the potential to be a useful tool in the analysis of cognitive outcomes in children not only with cerebral malaria but with

other disorders affecting the central nervous system, such as seizures [11,37], neurofibromatosis [12], and fetal alcohol spectrum disorders [13], among others. In conclusion, the statistical approach we propose has the potential to contribute significantly to improved analysis and interpretation of longitudinal cognitive assessment in children.

Acknowledgements

References

1. Carter JA, Murira GM, Ross AJ, Mung'ala-Odera V, Newton CR (2003) Speech and language sequelae of severe malaria in Kenyan children. Brain Inj 17: 217-224.

2. Olness K (2003) Effects on brain development leading to cognitive impairment: a worldwide epidemic. J Dev Behav Pediatr 24: 120-130.

3. Walker SP, Wachs TD, Gardner JM, Lozoff B, Wasserman GA, et al. (2007) Child development: risk factors for adverse outcomes in developing countries. Lancet 369: 145-157.

4. Carter JA, Ross AJ, Neville BG, Obiero E, Katana K, et al. (2005) Developmental impairments following severe falciparum malaria in children. Trop Med Int Health 10: 3-10.

5. John CC, Bangirana P, Byarugaba J, Opoka RO, Idro R, et al. (2008) Cerebral malaria in children is associated with long-term cognitive impairment. Pediatrics 122: e92-e99.

6. Beard JL, Connor JR (2003) Iron status and neural functioning. Annu Rev Nutr 23: 41-58.

7. Lozoff B (2007) Iron deficiency and child development. Food Nutr Bull 28: 560S-571S.

8. Holding PA, Taylor HG, Kazungu SD, Mkala T, Gona J, et al. (2004) Assessing cognitive outcomes in a rural African population: development of a neuropsychological battery in Kilifi District, Kenya. J Int Neuropsychol Soc 10: 246-260.

9. Gladstone M, Lancaster GA, Umar E, Nyirenda M, Kayira E, et al. (2010) The Malawi Developmental Assessment Tool (MDAT): the creation, validation, and reliability of a tool to assess child development in rural African settings. PLoS Med 7: e1000273.

10. Boivin MJ, Bangirana P, Byarugaba J, Opoka RO, Idro R, et al. (2007) Cognitive impairment after cerebral malaria in children: a prospective study. Pediatrics 119: e360-e366.

11. Fastenau PS, Johnson CS, Perkins SM, Byars AW, deGrauw TJ, et al. (2009) Neuropsychological status at seizure onset in children: risk factors for early cognitive deficits. Neurology 73: 526-534.

12. Krab LC, Aarsen FK, de Goede-Bolder A, Catsman-Berrevoets CE, Arts WF, et al. (2008) Impact of neurofibromatosis type 1 on school performance. J Child Neurol 23: 1002-1010.

13. Rasmussen C, Bisanz J (2009) Executive functioning in children with Fetal Alcohol Spectrum Disorders: profiles and age-related differences. Child Neuropsychol 15: 201-215.

14. Holding PA, Stevenson J, Peshu N, Marsh K (1999) Cognitive sequelae of severe malaria with impaired consciousness. Trans R Soc Trop Med Hyg 93: 529-534.

15. Holding PA, Kitsao-Wekulo PK (2004) Describing the burden of malaria on child development: what should we be measuring and how should we be measuring it? Am J Trop Med Hyg 71: 71-79.

16. Kihara M, Carter JA, Newton CR (2006) The effect of Plasmodium falciparum on cognition: a systematic review. Trop Med Int Health 11: 386-397.

17. Boivin MJ, Giordani B (2009) Neuropsychological assessment of African children: evidence for a universalbrain/behavior omnibus within a coconstructivist paradigm. Progress in Brain Research 178: 113-135.

18. Boivin MJ (2002) Effects of early cerebral malaria on cognitive ability in Senegalese children. J Dev Behav Pediatr 23: 353-364.

19. Kutner M, Nachtsheim C, Neter J, Li W (2004) Applied Linear Statistical Models (5thedn). McGraw-Hill/Irwin, New York.

20. King G (1986) How not to lie with statistics: avoiding common mistakes in quantitative political science. Am J Pol Sci 30: 666-687.

21. Singer JD, Willett JB (2003) Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press, New York.

22. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. JR Statist Soc B 57: 289-300.

23. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core team (2009) Linear and Nonlinear Mixed Effects Models. R package version 3: 1-93.

24. Colugnati FA, Louzada-Neto F, Taddei JA (2005) An application of bootstrap resampling method to obtain confidence interval for percentile fatness cutoff points in childhood and adolescence overweight diagnoses. Int J Obes (Lond) 29: 340-347.

25. Boivin MJ, Green SD, Davies AG, Giordani B, Mokili JK, et al. (1995) A preliminary evaluation of the cognitive and motor effects of pediatric HIV infection in Zairian children. Health Psychol 14: 13-21.

26. Van Rie A, Mupuala A, Dow A (2008) Impact of the HIV/AIDS epidemic on the neuro-development of preschool-aged children in Kinshasa, Democratic Republic of the Congo. Pediatrics 122: e123-e128.

27. Gladstone MJ, Lancaster GA, Jones AP, Maleta K, Mtitimila E, et al. (2008) Can Western developmental screening tools be modified for use in a rural Malawian setting? Arch Dis Child 93: 23-29.

28. Ruffieux N, Njamnshi AK, Mayer E, Sztajzel R, Eta SC, et al. (2010) Neuropsychology in Cameroon: first normative data for cognitive tests among school-aged children. Child Neuropsychol 16: 1-19.

29. Bangirana P, John CC, Idro R, Opoka RO, Byarugaba J, et al. (2009) Socioeconomic predictors of cognition in Ugandan children: implications for community interventions. PLoS One 4: e7898.

30. John CC, Panoskaltsis-Mortari A, Opoka RO, Park GS, Orchard PJ, et al. (2008) Cerebrospinal fluid cytokine levels and cognitive impairment in cerebral malaria. Am J Trop Med Hyg 78: 198-205.

31. Spreen O, Strauss E (1998) A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. Oxford University Press, Oxford, New York.

32. Boivin MJ, Ryan KA, Aldridge KA, Giordani B (2011) Transfer-of-Learning effect with the Tactual Performance Test using familiar and unfamiliar shapes with American, Lao, and Senegalese children. Dev Neuropsychol 36: 552-565.

33. Boivin MJ, Giordani B, Bornefeld B (1995) Use of the Tactual Performance Test for cognitive ability testing with African children. Neuropsychology 9: 409-417.

34. Lezak MD (2004) Neuropsychological Assessment. 4th edition, Oxford University Press, New York.

35. Jarvis PE, Barth JT (1994) The Halstead-Reitan neuropsychological battery: A guide to interpretation and clinical applications. Psychological Assessment Resources, Odessa, FL.

36. Aldenkamp AP, Weber B, Overweg-Plandsoen WCG, Reijs R, van Mil S (2005) Educational underachievement in children with epilepsy: a model to predict the effects of epilepsy on educational achievement. J Child Neurol 20: 175-180.

37. Hyman SL, Shores A, North KN (2005) The nature and frequency of cognitive deficits in children with neurofibromatosis type 1. Neurology 65: 1037-1044.