Data management and statistical analysis

(a) Nature of data and storage plans

Data collected in hard copy files will be entered at site and transmitted to a database manager at KEMRI-Wellcome Programme who will store it in the secure programme servers. The hard copy will also be transported to Kilifi for storage in cabinets under lock and key fireproof cabinets.

This research will generate personal, laboratory, clinical and neuropsychological data. The personal data will be collected by fieldworkers in the community as part of routine census for the demographic surveillance systems. The laboratory data will mostly comprise malaria parasitemia and immunological markers collected from integrated studies of cohort studies of malaria, the clinical data will be obtained as medical history and examinations by trained clinicians, but additional data will be obtained from clinical admissions to hospitals linked to demographic surveillance systems. Neuropsychological data will be obtained by trained assessors using locally validated questionnaire and tools. The data will be cleaned and inspected for completeness by data entry clerk and will then be double entered into a computer database (MySQL (Oracle Corp. USA)). All the entered data will be linked to the Demographic Surveillance Systems of Kilifi, Kombewa and Siaya. The data management database will be set-up in such a way that it can be centrally managed from Kilifi, Kenya where my host organization is based and where the data manager will be based. The database manager will travel to Kombewa and Siaya to set up the database.

The storage will employ a system that we have used in other studies of malaria, neurological and mental disorders at the KEMRI-Wellcome Trust Research Programme. Our database managers have extensive experience in database management in Kilifi having previously managed data from multisite studies. All data will be backed up in portable and secure hard disks or DVDs which will be stored locally. Data storage will be password protected and extremely confidential personal details of the patients will exclusively be in the custody of the core data manager. We will store hard copy files and questionnaires in local archives and in offsite archiving security firms such as G4S for files that will be over seven years or will not have been referred to during this period.

(b) Statistical analysis

Data will be analyzed using R and Stata statistical software. Risk ratios of the association of vaccination with RTS,S/AS01 or protection by ITN with neurodevelopment measures and school participation will be computed using either generalized estimating equations with appropriate link functions (for analysis of panel data from the primary study in which active surveillance is done at multiple time points) or generalized linear models with appropriate link functions (for analysis of cross-sectional data from the secondary studies in Siaya and Kombewa where most variables will be inferred retrospectively). These association models will be accounted for age, sex, and potential confounders measured at time of vaccination or during surveillance such as HIV, provoked seizures, meningitis, and malnutrition. Likelihood ratio tests will be used to test if vaccination with RTS,S/AS01 interacts with dosage, malaria episodes, bed-net use or clinical admissions with malaria to determine the neurodevelopment or school participation.

Then, it will be examined if anti-circumsporozoite antibodies measured during vaccination or surveillance or antibodies responses to antigens such as merozoite surface proteins and apical membrane protein measured for children in the ITN trial mediates the association between vaccination with RTS, S/AS01 and neurodevelopment or school participation. Mediation can be

said to occur when: (i) the explanatory variable (vaccination status) affects the mediator (titers to anti-circumsporozoite antibodies); (ii) the explanatory variable (vaccination status) significantly affects the response variable (neurobehavioral outcomes and school participation) in the absence of the mediator variable (titers to anti-circumsporozoite antibodies) in the model; (iii) the mediator variable (titers to anti-circumsporozoite antibodies) has a significant unique effect on the response variable, and (iv) the effect of the explanatory variable (vaccination status) on the response variable (neurobehavioral outcomes and school participation) shrinks upon the addition of the mediator variable (titers to anti-circumsporozoite antibodies) to the model.

The long-term efficacy of RTS,S/AS01 or ITN in improving neurodevelopment and school participation will be measured by subtracting risk ratios from 1. We will also test the hypothesis that the efficacy of phase III RTS,S/AS01 will be greatest with the rabies vaccine controls than with meningococcal serogroup C conjugate vaccine (may protect some controls against sequelae from meningitis) using ratio of risk ratios. These analyses will be accounted for potential confounders specifically any hospital visits, exposure to malaria, reason for admission and nutritional status during the trials; this information will be obtained from the databases or collected through medical history. Descriptive statistics will be provided as means, medians and proportions and presented in graphs and contingency tables. Comparison of continuous measures between two groups will be done using student t-test or Mann-Whitnney test where applicable. Discrete variables will be compared using Pearson's Chi-squared test or Fisher's exact where applicable.