Introduction

The Biostatistics and Epidemiology group within the Department of Public Health and Epidemiology at the STI conducts long-term research projects in epidemiology, in collaboration with biomedical scientists. It also carries out methodological research and develops and applies new methods in biostatistics and epidemiology. The main areas of focus are simulation modelling of malaria epidemiology, and statistical modelling of the geographical distribution of parasitic infections. The group offers teaching, consulting and data services to the other groups in the Swiss Tropical Institute, including training in statistics for MSc and PhD students, consultation on study design, statistical analysis and bioinformatics support.

Modelling malaria epidemiology and control

Mathematical models are important tools for decision making in the control of infectious diseases. Malaria was one of the first infections for which such modelling was applied, but the existing models need updating in the context of renewed worldwide efforts to address the malaria burden in different ecological settings, using different mixes of interventions.

The STI's current malaria modelling work began in 2003 with a project to simulate malaria vaccines. We are now developing models of a full range of interventions and integrated control programmes in settings characterised by a wider range of transmission patterns. This involves (i) developing general models of the natural history and epidemiology of *Plasmodium falciparum* malaria; (ii) detailed modelling of health systems, including different delivery systems for interventions; (iii) comparing predictions of different models where there is substantial uncertainty about aspects of the epidemiology (e.g. decay of immunity); and (iv) carrying out probabilistic sensitivity analyses of the results. All this will feed into economic analyses and comprehensive assessments of uncertainty presented in a policy-relevant form. The work is carried out in close collaboration with the Interventions and Health Systems unit (section 8) and the Health Systems and Economics group of the Swiss Centre for International Health (section 12, page 70).

To model the natural history and epidemiology of *Plasmodium falciparum* malaria, we are using individual-based stochastic simulation models, with 1-day or 5-day time steps. At each step we model the parasite density of each malaria infection, basing these on a description of the time courses of infection in neurosyphilis patients treated therapeutically with malaria. The simulated infections are introduced into populations of simulated hosts at rates based on those measured in the field. Each host is characterised by a vector of time-dependent variables, including age, health and immune status. Further stochastic elements of the model determine whether the simulated host becomes ill or dies.

Because we need to make quantitative predictions of effectiveness, we fit our models to an extensive library of field datasets that measure different malariological parameters. To fit such models is extremely computer-intensive, and for this we use computing power provided over the Internet by volunteers with spare computing capacity (see www.malariacontrol.net). We are now engaged in fitting many different models via iterative processes, investigating their predictions and carrying out sensitivity analyses.

Although there is currently no licensed malaria vaccine, we can use our models to investigate the likely effects of malaria vaccination programmes. This should help elucidate what the minimal vaccine profile should be before registration is worthwhile; how to allocate resources between different candidates with different profiles; which candidates to consider combining; and what deployment strategies might be best. We consider a range of endemic malaria settings with deployment of vaccines via the Expanded Programme on Immunisation (EPI), with and without additional booster doses, and also via mass campaigns for a range of coverages. The simulation scenarios account for the dynamic effects of natural and vaccine-induced immunity, for treatment of clinical episodes, and for births, ageing and deaths in the cohort. Simulated pre-erythrocytic vaccines have greatest benefits in low endemic settings [annual Entomological Inoculation Rate (EIR) <10.5] where between 12 and 14% of all deaths are averted when initial efficacy is 50%. In some high-transmission scenarios (EIR >84) pre-erythrocytic vaccines (PEV) may lead to increased incidence of severe disease in the long term, if efficacy is moderate to low (<70%). Blood-stage vaccines (BSVs) are most useful in high-transmission settings, and are comparable to PEV for low-transmission settings. Combinations of PEV and BSV generally perform little better than the best of the contributing components. A minimum half-life of protection of 2–3 years appears to be a precondition for substantial epidemiological effects. Herd immunity effects can be achieved with even moderately effective (>20%) malaria vaccines (either PEV or BSV) when deployed through mass campaigns targeting all age groups as well as EPI, and especially if combined with highly efficacious transmission-blocking components. These results raise several issues for vaccine clinical development, in particular appropriateness of vaccine types for different transmission settings; the need to assess transmission to the vector and duration of protection; and the importance of deployment additional to the EPI, which again may make the issue of number of doses required more critical.

In addition to vaccines, a broad selection of novel malaria control tools are entering the field. We have also modelled the likely effects of intermittent preventive treatment in infants (IPTi), exploring hypotheses of how IPTi might work, and used the best-fitting model to make predictions of the impact of IPTi in different epidemiological settings. Our original epidemiological model was able to reproduce the pattern of trial results reasonably well, while models based on alternative hypotheses did not improve fit to IPTi trial data, suggesting that variation in trial results from known differences between trial sites. These models are consistent with IPTi having short-term impacts on immunity but suggest that there will be no long-term complicated effects of IPTi implementation, and that community incidence of illness will remain constant over many years, in contrast to effects of vaccines or vector control, which have longer-term dynamics. It is thus reasonable to predict the long-term impact of IPTi from static models. In addition we recently started to compare the impact and cost-effectiveness of intermittent preventive treatment in children (IPTc) and IPTi in different settings. Initial simulations suggest that in a setting such as Niakhar, Senegal, with a short transmission season and an annual EIR of approximately 10 per annum, IPTc would avert more clinical episodes and mortality than IPTi, but at a higher cost per episode averted.

In the same way that the field of climate modelling is moving towards making predictions from ensembles of different models, we will be most confident about the predictions if we see that they remain substantially unchanged across a range of different sets of assumptions. To this end, we hope to use the malariacontrol.net platform to implement models proposed by other research groups, taking full advantage of the massive computational power that is available.

Scientists:	N. Chitnis, L. Conteh, C. Lengeler, N. Maire, M. Penny, A. Ross, A. Schapira, T. Smith, A. Studer, B. Genton, D. de Savigny, M. Tanner, F. Tediosi, P. Vounatsou
Students:	S. Bhardawaj, M. Bretscher, V. Crowell, A. Lutambi
Collaboration:	Liverpool School of Tropical Medicine (I. Hastings); European Organisation for Nuclear Research (CERN, F. Grey); Ifakara Health Research and Development Centre (G. Killeen, S. Abdulla); Papua New Guinea Institute of Medical Research (I. Müller); Bocconi University, Milan (F. Tediosi)
Funding:	Bill & Melinda Gates Foundation (BMGF); Swiss National Science Foundation (SNSF); PATH-MACEPA

Spatio-temporal statistical modelling of infectious diseases

Spatio-temporal modelling is a main research focus of the STI. It involves developing statistical methods that can handle the complexities of infectious disease data as well as map the epidemiology of malaria, schistosomiasis and other helminthic infections to further our understanding of the causes of infectious disease-related death in a spatially explicit manner.

Our work focuses on analysing geostatistical data collected over a very large number of fixed locations (e.g. more than 18,000) at different spatial scales such as households and villages. These data comprise continuous measures of disease transmission (entomological inoculation rates – EIRs for malaria), prevalence data (field surveys on the presence or absence of an infection and sporozoite rates), counts (mosquito densities and helminth infection intensities), multiple categories (multinomial data with categories presenting multiple infections or the frequency distribution of mosquito species) or time to death (survival data). We employ so-called Bayesian variogram models fitted by standard or Reversible Jump Markov chain Monte Carlo (MCMC) simula-

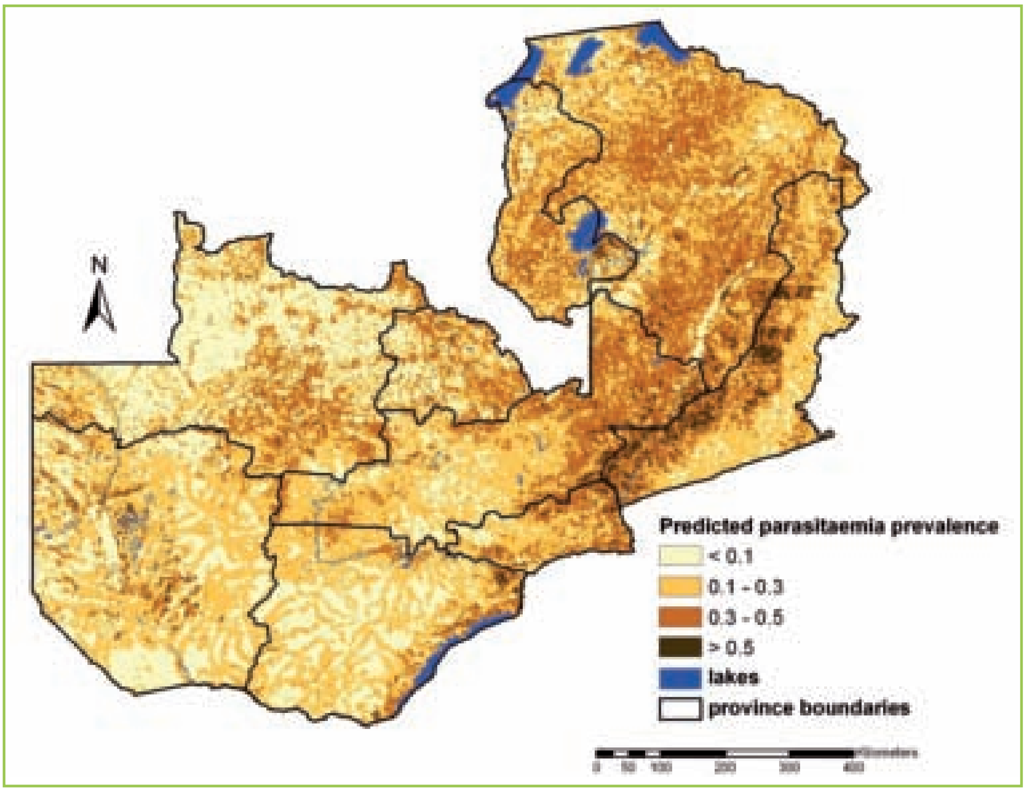

Figure 1: Predicted parasitaemia risk in children under 5 years old in Zambia based on a Bayesian geostatistical model with environmental predictors. The map is based on predictions over 100,000 pixels.

tion. Our research in geostatistical methodology involves modelling (i) very large geostatistical datasets, (ii) multivariate spatial processes, (iii) prevalence data with diagnostic error, (iv) temporal count data with seasonal variation and (v) data with spatial correlation depending on the distance between locations as well as the location itself (non-stationary). We also develop tools to validate models. We describe bellow the main areas of application of our work.

Mapping malaria transmission

We analyse malaria survey data to estimate malaria transmission at high spatial resolution in Africa. The data are obtained either from recent nationwide malaria indicator surveys (MISs) or compiled from historical surveys extracted from published and unpublished sources.

MIS data are collected from household surveys and include individual parasitological data on children under 5 years of age and pregnant women, fever history, demographic and socio-economic information, and malaria interventions. We develop geostatistical models to assess the effects of interventions taking into account climatic and environmental factors extracted from satellites. Using a technique known as Bayesian kriging, we predict the malaria risk and number of people at risk at various spatial scales. The models consider elapsing time between the effects of environmental predictors and malaria risk (lag time) as well non-linearity in the malaria-climate relation. In collaboration with MACEPA (Malaria Control and Evaluation Partnership in Africa) we have completed the analysis of the Zambia MIS for 2006 and also analysed the MEASURE Demographic and Health Survey (DHS) MIS of Angola for 2007.

Historical data include counts of screened and parasite-positive individuals over different age groups from surveys carried out during different seasons in a variety of locations. We adjust for age and seasonality by combining geostatistical and mathematical malaria transmission models to convert prevalence data to yearly EIRs and then predict EIRs at high spatial resolution. Using the mathematical model, we convert the EIR estimates to age- and season-adjusted estimates of parasitaemia risk which we further model via non-stationary geostatistical models to produce smooth estimates of malaria transmission in West, Central and East Africa. Non-stationarity is an important feature of spatial malaria outcomes because malaria interventions, health system efficiency and other unmeasured factors may influence spatial correlation differently at various parts of the study surface. Therefore the standard geostatistical assumption that spatial correlation is independent of location (stationarity) does not hold. We consider non-stationarity by partitioning the space into fixed or random Voronoi tiles, assume within-tile stationarity and model spatial correlation as a mixture of the locally stationary spatial processes.

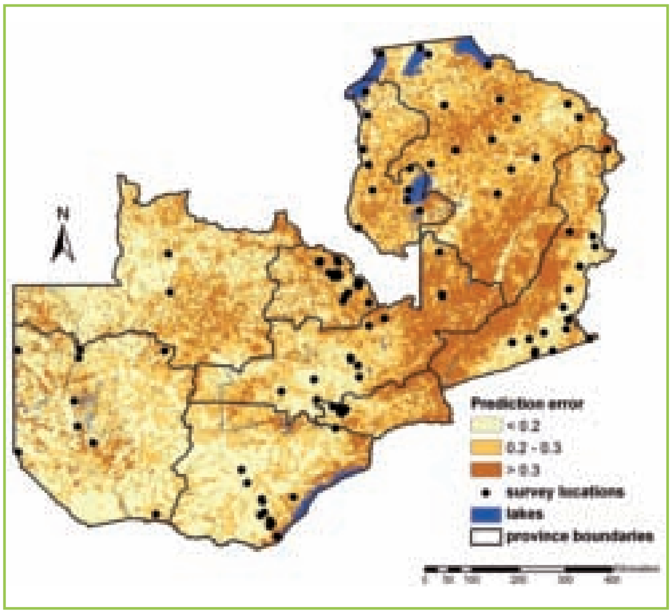

Figure 2: Prediction error of the parasitaemia risk estimates given in Figure 1.

In collaboration with the Medical Research Council (MRC) in Durban, South Africa, and with funds from the Bill and Melinda Gates Foundation (BMGF) we are updating the Mapping Malaria Risk in Africa (MARA) database. MARA is the most comprehensive malariometric database in Africa to date. It includes historical data from 1900 onwards at over 13,000 locations.

Mapping malaria seasonality

Mapping malaria transmission depends on accurate maps of malaria occurrence according to the time of year (seasonality) because (i) environmental predictors vary over season and year, (ii) use of malaria transmission models in malaria risk mapping requires knowledge of the transmission months at each location and (iii) effective malaria control depends on timely targeting during the transmission season. In collaboration with MRC Durban, we derived malaria seasonality indices and estimated the months and duration of transmission in Africa by extracting the seasonal yearly patterns within incidence and EIR data via Fourier transformations. Additional work involved modelling malaria seasonality cycles via cosine terms within Bayesian spatio-temporal models.

Mapping malaria vectors

Mapping the geographical distribution of mosquito species and chromosomal forms is helpful for effective malaria control interventions since insecticide resistance is related to specific mosquito subspecies. In collaboration with the Malaria Research and Training Center (MRTC) in Bamako we have analysed unique datasets on the distribution and ecology of malaria vectors in Mali, both at the national level, and within the Office du Niger irrigation project in Niono district. In particular, we developed multivariate geostatistical models (i) to identify environmental factors that explain spatial patterns of the major species in Mali *(An. gambiae* s.s and *An. arabiensis)* and subspecies of *An. gambiae* s.s and (ii) to predict their distribution in the country.

Forecasting malaria outbreaks

Analyses of temporal data on seasonal malaria incidence lack rigorous statistical modelling methods. In collaboration with O. Briet from the International Water Management Institute (IWMI) in Sri Lanka we have developed Bayesian negative binomial Seasonal Autoregressive Integrated Moving Average (SARIMA) models for forecasting malaria incidence in Sri Lanka on the basis of climate. These models can supplement malaria early warning systems.

Applications in schistosomiasis

Our research on schistosomiasis is carried out in collaboration with J. Utzinger and his group from the STI's Ecosystems Health Sciences unit.

We are currently compiling schistosomiasis survey data all over Africa from published and unpublished sources with the aim of producing schistosomiasis infection risk maps in Africa at different spatial scales using Bayesian geostatistical models. This work is part of the European Union (EU)-funded CONTRAST project under the leadership of T. Kristensen from the Danish Billharziasis Laboratory (DBL) Institute of Veterinary Parasitology, Faculty of Life Science, at the University of Copenhagen, Denmark, and carried out in collaboration with C. Simonga from the University of Zambia.

One example of the application of statistical methods to modelling multivariate geostatistical data is modelling co-infection between schistosomiasis, malaria and filariasis in Uganda. This project is carried out with A. Stensgaard from the University of Copenhagen. Another example is geostatistical modelling of infection risk data observed with diagnostic error and of overdispersed schistosomiasis intensity data from Côte d'Ivoire. The transmission intensity of schistosomiasis is a function of the parasitic worm load within a group of individuals, which can indirectly be quantified by the number of eggs that are excreted. Only a few individuals harbour large numbers of worms, whilst the majority of individuals are uninfected or only carry a low worm burden. In addition, widely used diagnostic approaches for schistosomiasis (e.g. the Kato-Katz) fail to detect some infected individuals, particularly for light infections. Hence, a large proportion of individuals are considered as "zero egg excretors". We introduced zero-inflated negative binomial models to analyse overdispersed intensity data with large numbers of zeros, spatially structured.

In a separate line of research we are developing Bayesian formulations of the immigration-death model to estimate the age-specific prevalence of *Schistosomiasis Mansoni.*

External collaborators other than those mentioned above include G. Raso from the Queensland Institute of Medical Research in Brisbane, Australia, and X. Zhou and his group from the National Institute of Parasitic Diseases, Chinese Centre for Disease Control and Prevention, in Shanghai, China.

Applications in mortality

Reliable statistics on mortality, its causes and trends are in great demand for assessing health and for devising appropriate interventions. In most developing countries, monitoring mortality is difficult due to the lack or scarcity of vital statistics (e.g. registries of births and deaths). However, mortality can be estimated from national census data, DHS data and from other information that is regularly collected via demographic surveillance systems (DSSs). A typical DSS site monitors longitudinally a population size of 30,000–100,000 people, collecting demographic and health-related data on pregnancy, births, mortality (including causes of death) and migration. Verbal autopsies determine cause-specific mortality. We are developing models for very large geostatistical datasets using sparse Gaussian process approximations and applying them (i) to assess risk factors and spatio-temporal patterns of child mortality within the Agincourt DSS in South Africa, and (ii) to estimate the relationship between child mortality and malaria transmission at seven DSS sites in Africa which collected entomological data as part of the MTIMBA project. This is a collaborative effort with the INDEPTH Network and K. Kahn from Witwatersrand University. In addition, we are estimating geographical patterns of child mortality in Tanzania using census and DHS data and assessing the relation between malaria and mortality by linking the DHS and Mapping Malaria Risk in Africa (MARA) data.

Scientists: K. Boutsika, D. de Savigny, L. Gosoniu, T. Jäggi, C. Lengeler, P. Odermatt, T. Smith, J. Utzinger, P. Vounatsou
Students: M. Craig, N. Chitnis, N. Sogoba, D. Gosoniu, S. Kasasa, N. Riedel, S. Rumisha, A. Ombek, N. Onyiri, S. Nakashima, A. Msengwa
Trainee: N. Köhler
Collaboration: Ifakara Health Institute, Tanzania (S. Abdulla); INDEPTH Network; IWMI, Sri Lanka (O. Briët); MRTC, Mali (N. Sogoba); SAMRC, South Africa (M. Mabaso); MACEPA (J. Miller, R. Steketee); Witwatersrand University, South Africa (K. Kahn); DBL, University of Copen-hagen (T. Kristensen, A. Stensgaard); Queensland Institute of Medical Research (G. Raso); National Institute of Parasitic Diseases, China CDC, Shanghai, China (X. Zhou); University of Zambia (C. Simonga)
Funding: BMGF; EU; SNSF

Publications

Abdullah S, Adazu K, Masanja H, Diallo D, Hodgson A, Ilboudo-Sanogo E, Nhacolo A, Owusu-Agyei S, Thompson R, **Smith T** & Binka FN (2007) Patterns of age-specific mortality in children in endemic areas of sub-Saharan Africa. *Am J Trop Med Hyg* 77, 99–105.

Aponte JJ, Menendez C, Schellenberg D, Kahigwa E, Mshinda H, **Vounatsou P, Tanner M** & Alonso PL (2007) Age interactions in the development of naturally acquired immunity to *Plasmodium falciparum* and its clinical presentation. *PLoS Med* 4, e242.

Briët OJT, Vounatsou P & Amerasinghe PH (2008) Malaria seasonality and rainfall seasonality are correlated in space. *Geospat Health* 2, 183–190.

Briët OJT, Vounatsou P, Gunawardena DM, Galappaththy GN & Amerasinghe PH (2008) Temporal correlation between malaria and rainfall in Sri Lanka. *Malar J* 7, 77.

Briët OJT, Vounatsou P, Gunawardena DM, Galappaththy GN & Amerasinghe PH (2008) Models for short term malaria prediction in Sri Lanka. *Malar* J 7, 76.

Chitnis N, Smith T & Steketee RW (2008) A mathematical model for the dynamics of malaria in mosquitoes feeding on a heterogeneous host population. *J Biol Dynamics* 2, 259–285.

Chitnis N, Hyman JM, & Cushing JM (2008) Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model. *Bull Math Biol* 70, 1272–1296.

Gosoniu L, Vounatsou P, Sogoba N & **Smith T** (2006) Bayesian modelling of geostatistical malaria risk data. *Geospat Health* 1, 126–139.

Gosoniu L, Vounatsou P, Sogoba N, Maire N & **Smith T** (2008) Mapping malaria risk in West Africa using a Bayesian nonparametric non-stationary model. *Comput Stat Data Anal* special issue on spatial statistics (in press).

Gosoniu L, Vounatsou P, Tami A, Nathan R, Grundmann H & **Lengeler C** (2008). Spatial effects of mosquito bednets on child mortality. *BMC Public Health* (in press).

Haghdoost AA, Alexander N & **Smith T** (2007) Maternal malaria during pregnancy and infant mortality rate: critical literature review and a new analytical approach. *J VectBorne Dis* 44, 98–104.

Hastings IM, **Smith TA** (2008) MalHaploFreq: a computer program for estimating malaria haplotype frequencies from blood samples. *Malar J* 7, 130.

Killeen GF & **Smith TA** (2007) Exploring the contributions of bed nets, cattle, insecticides and excitorepellency to malaria control: a deterministic model of mosquito host-seeking behaviour and mortality. *Trans R Soc Trop Med Hyg* 101, 867–880.

Lines J, **Schapira A** & **Smith T** (2008) Malaria elimination and eradication: risks and opportunities in endemic countries. *BMJ* (in press).

Mabaso ML, Craig M, Ross A & **Smith T** (2007) Environmental predictors of the seasonality of malaria transmission in Africa: the challenge. *Am J Trop Med Hyg* 76, 33–38.

Mabaso ML, Kleinschmidt I, Sharp B & **Smith T** (2007) El Nino Southern Oscillation (ENSO) and annual malaria incidence in Southern Africa. *Trans R Soc Trop Med Hyg* 101, 326–330.

Penny M, Maire N, Studer A & **Smith T** (2008) What should vaccine developers ask? Simulation of the effectiveness of malaria vaccines. *PloS One* (in press).

Raso G, **Vounatsou P,** McManus DP, N'Goran EK & **Utzinger J** (2007) A Bayesian approach to estimate the age-specific prevalence of *Schistosoma mansoni* and implications for schistosomiasis control. *Int J Parasitol* 37, 1491–1500.

Ross A, Penny M, Maire N, Studer A, Carneiro I, Schellenberg D, Greenwood B, **Tanner M** & **Smith T** (2008) Modelling the impact of intermittent preventive treatment in infants. *PloS One* 3, e2661.

Smith TA (2007) Measures of clinical malaria in field trials of interventions against *Plasmodium falciparum. Malar J* 6, 53.

Smith TA (2008) Estimation of heterogeneity in malaria transmission by stochastic modelling of apparent deviations from mass action kinetics. *Malar J* 7, 12.

Smith T (2007) Health care Swiss style. *Significance* 4, 45.

Smith T, Maire N, Ross A, Penny M, Chitnis N, Schapira A, Studer A, Genton B, Lengeler C, Tediosi F, de Savigny D & **Tanner M** (2008) Towards a comprehensive simulation model of malaria epidemiology and control. *Parasitology* (in press).

Sogoba N, Doumbia S, **Vounatsou P,** Baber I, Keita M, Maiga M, Traore SF, Toure A, Dolo G, **Smith T** & Ribeiro JM (2007) Monitoring of larval habitats and mosquito densities in the Sudan savanna of Mali: implications for malaria vector control. *Am J Trop Med Hyg* 77, 82–88.

Sogoba N, Doumbia S, **Vounatsou P,** Bagayoko MM, Dolo G, Traore SF, Maiga HM, Toure YT & **Smith T** (2007) Malaria transmission dynamics in Niono, Mali: the effect of the irrigation systems. *Acta Trop* 101, 232–240.

Sogoba N, Vounatsou P, Bagayoko MM, Doumbia S, Dolo G, **Gosoniu L,** Traore SF, Toure YT & **Smith T** (2007) The spatial distribution of *Anopheles gambiae sensu stricto* and *An. arabiensis* (Diptera: Culicidae) in Mali. *Geospat Health* 1, 213–222.

Sogoba N, Vounatsou P, Doumbia S, Bagayoko M, Toure MB, Sissoko IM, Traore SF, Toure YT & **Smith T** (2007) Spatial analysis of malaria transmission parameters in the rice cultivation area of Office du Niger, Mali. *Am J Trop Med Hyg* 76, 1009–1015.

Sogoba N, Vounatsou P, Bagayoko M, Doumbia S, Dolo G, **Gosoniu L,** Traore SF, **Smith T,** Toure YT (2008). Spatial distribution of the chromosomal forms of *Anopheles gambiae senso stricto* (Diptera: Cullicidae) in Mali. *Malaria J* (in press).

Vounatsou P, Raso G, Tanner M, N'Goran EK, Utzinger J (2008). Bayesian geostatistical modelling for mapping schistosomiasis transmission. *Parasitology, Supplement* (in press).

Wang XH, Zhou XN, **Vounatsou P,** Chen Z, **Utzinger J,** Yang K, **Steinmann P** & Wu XH (2008) Bayesian spatio-temporal modeling of *Schistosoma japonicum* prevalence data in the absence of a diagnostic "gold" standard. *PLoS Negl Trop Dis* 2, e250.

Yang GJ, Utzinger J, Sun LP, Hong QB, **Vounatsou P, Tanner M** & Zhou XN (2007). Effect of temperature on the development of *Schistosoma japonicum* within *Oncomelania hupensis,* and hibernation of *O. hupensis. Parasitol Res* 100, 695–700.