MRC LID Studentships: 2023-24 Research Project

TITLE OF PROJECT		
Data science platforms for enhanced analytics and monitoring of female genital schistosomiasis and related health impacts		
SUPERVISORY TEAM		
Supervisor	Dr Chrissy Roberts Email: Chrissy.Roberts@lshtm.ac.uk	LSHTM
Co-Supervisor	Professor Amaya Bustindy Email: Amaya.Bustinduy@lshtm.ac.uk	LSHTM
PROJECT SUMMARY		
Project summary	Female Genital Schistosomiasis is a disabling neglected gynaecological disease. This exciting and innovative PhD will (1) explore how convolutional neural networks (CNN) can be trained to support diagnosis, surveillance and decision-making for communities affected by FGS and (2) develop electronic diary devices for use in longitudinal research in low resource settings.	
Project key words	Machine learning Technology Female Genital Schistosomiasis Mixed methods Programming Sexual and reproductive health	
MRC LID themes	Health Data Science Infectious Disease Global Health	
MRC Core Skills developed through this project	Quantitative skills Interdisciplinary skills	
Skills we expect a student to develop/acquire whilst pursuing this project	• Quantitative Epidemiology & Statistics • Machine Learning, prognostics, diagnostics • Programming & Software design for public good • Data analysis, longitudinal analysis • Research methods development • Research study design • Ethics, Data stewardship & Good research practice	
Is this project available for students applying for the	Route	1+4 = Yes +4 = Yes

1+4 route? And possible Master's options identified by supervisory team	Suitable Master's programmes	LSHTM – MSc Health Data Science
Particular prior educational requirements for a student undertaking this project	This project requires a high level of computer literacy and will suit candidates with existing skills in computer sciences • Some experience with Python is required. • Some experience working with either R or Pandas is required • Experience with Keras, & Tensorflow or similar end-to-end ML platforms is desirable.	
PROJECT IN MORE DETAIL		
Scientific description of this research project	Female genital schistosomiasis (FGS) is a disabling gynaecological condition caused by *Schistosoma haematobium.* A neglected disease of poverty, FGS affects the lives of ~40,000,000 people, mostly living in Africa. FGS has impacts on sexual and reproductive health (SRH); increases risks of HIV transmission and associates with increased incidence/severity/sequelae of cervical dysplasia, human papillomavirus (HPV) and Trichomonas Vaginalis (*Tv*). Accurate diagnoses of FGS, HPV and *Tv* are crucial for effective treatment, surveillance, control and prevention, but few health workers are appropriately trained. Symptomatic and syndromic overlap with co-endemic cervico-vaginal health conditions complicates diagnosis. State of the art diagnostics use rapid tests, digital colposcopic photography, PCR and other methods. Whilst efforts to characterise the epidemiology/endemicity of FGS have been a recent success, less is known about the relationship between symptoms, severity, progression and other disease features across time. Temporal factors of interest may be biological (menstrual cycle), socio-behavioural (sexual, cultural) and environmental (seasonal, transmission). **Project Objectives** In two work-packages, we will leverage emerging technologies to answer questions about FGS from the perspective of (a) diagnostics and (b) temporal dynamics. **Workpackage 1 – Machine Learning approaches to diagnosis, decision support and prevalence estimation** The availability of high-quality digital photographs of cervixes in health and disease presents an opportunity to use open tools for machine learning based computer vision in the diagnostic process. This project will explore how convolutional neural networks (CNN) can be trained on colposcopy images	

	to support diagnostic decision-making. The work will focus on the use of open-source tools, with a view to developing systems for others to use at zero cost. **Workpackage 2 – Off-grid electronic diaries to explore the temporal dynamics of FGS related symptoms** We will explore the use of "Badger 2040", a programmable badge with e-ink display, as a low-cost, low power, off-grid electronic data collection platform for longitudinal studies. This work will develop new hardware and software, whilst answering key questions about the temporal dynamics of FGS and to monitor genital symptomatology across multiple menstrual cycles. A subset of women already recruited to the Schista! study will be invited to participate in exploratory longitudinal FGS research using these devices. **Techniques to be used** • Machine Learning • Convolutional Neural Networks • Python programming • Epidemiology **Confirmed availability of any required databases or specialist materials** • Access to training colposcope imagery is confirmed from two large studies in Zambia (PI ABustinduy) • Badger 2040 devices are available off-the-shelf **Potential risks to the project and plans for their mitigation** • We may need more training data than is available • Prof. Bustinduy has a rich professional network of FGS experts, some of which could be approached to share their colposcopy image libraries • We will need to apply for some funding for travel to Zambia • CNN may not work as well as we hope • Other methods exist, many clinical outcomes of interest that can also be explored **Other useful information** • This project requires a high level of computer literacy and will suit candidates with computer sciences backgrounds • Experience working with Python and R, Keras, & Tensorflow will be beneficial.
Further reading (Relevant preprints and/or open access articles)	https://doi.org/10.1016/bs.apar.2021.12.003 https://github.com/chrissyhroberts/Badger2040_PH

Additional information from the supervisory team	The supervisory team has provided a recording for prospective applicants who are interested in their project. This recording should be watched before any discussions begin with the supervisory team. To access the recording please see MRC LID Project – Roberts & Bustinduy.