A Google Earth-based database management for schistosomiasis control in Zanzibar

Ming-Zhen He,1,2 Wei Li,1,3 Saleh Juma,4 Fatma Kabole,5 Da-Cheng Xu,5 Xin-Yao Wang,1,3 Jian He,1,3 Tao Jiang,6 Robert Bergquist,7 Kun Yang1,3

1Jiangsu Institute of Parasitic Diseases, Key Laboratory of National Health Commission on Parasitic Disease Control and Prevention-Jiangsu Provincial Key Laboratory on Parasite and Vector Control Technology, Wuxi, Jiangsu, China; 2Center for Disease Control and Prevention of Changzhou City, Changzhou, Jiangsu, China; 3College of Medicine, Jiangnan University, Wuxi, Jiangsu, China; 4Ministry of Technology, Wuxi, Jiangsu, China; 2 Center for Disease Control and Prevention of Changzhou City, Changzhou, Jiangsu, China; 6Center for Disease Control and Prevention of Danyang City, Zhenjiang, Jiangsu, China; 7Ingerod, Brastad, Sweden

Abstract

Schistosomiasis remains a serious health problem in Africa. Although a strong, coordinated agenda for research on this disease has been in place for the last 50 years in Zanzibar, data storage, retrieval of survey data and management remain problem areas. We investigated the use of Google Earth (GE) in conjunction with a hand-held, global positioning system as a pilot project for managing schistosomiasis control. In this way, risk areas can be surveyed and followed up by visualizing both the distribution of human infections and that of the intermediate snail host together with environmental information. A platform with three spatial databases was created: i) distribution of infected humans; ii) distribution of the intermediate snail host in ponds (infected and not infected specimens); iii) distribution of the intermediate snail host in streams (infected and non-infected specimens). The GE spatial database increased the efficiency of follow-up case treatment as well as snail control and contributed also to the discovery of previously unknown areas in need of snail control. We conclude that this platform is advantageous not only by being useful for management and visualization of spatial data, but also because it is easy to operate and available free of charge.

Introduction

Schistosomiasis, one of the poverty-related, neglected tropical diseases (NTDs), is caused by trematode parasites of the genus Schistosoma, which reside in the abdominal capillaries of the infected human definitive host. Different kinds of snails play the role of intermediate host depending on the species of the parasite. So far, many areas in sub-Saharan Africa still have a high schistosomiasis prevalence with many people presenting with high intensities of disease due to constant reinfection (Olsen et al., 2015). The World Health Organization (WHO) estimates that more than 90% of all humans requiring treatment for this disease currently live in Africa (WHO, 2018). In Zanzibar (including the islands of Unguja and Pemba), urogenital schistosomiasis is still a considerable public health problem with high prevalence (Knopp et al., 2013; Pennance et al., 2016). Schistosomiasis on these islands is urogenital which is caused by Schistosoma haematobium relying on the intermediate host snail Bulinus globosus (Allan et al., 2013).

In 2012, WHO set the goal to interrupt transmission of schistosomiasis in selected countries of the African region by 2025 (WHO, 2013). Much work and research have been carried out to reach this goal in Zanzibar. Prevention and control on the island of Pemba was once the focus of strong control efforts (Savioli et al., 1989; Savioli and Mott, 1989; Lwambo et al., 1997; Guidi et
al., 2010); however, although disease burdens have been allevi-ated through preventative chemotherapy together with additional control interventions, transmission has not been interrupted on a large scale (Knopp et al., 2012; 2013; 2016). Regrettably, collect-ed data are dispersed and scattered and cannot be effectively uti-lized since the local schistosomiasis control institution, the NTD office, has neither access to easily available information about the distribution of B. globosus nor to environmental data associ-ated with this infection. At the World Health Assembly on May 21st 2014, the Government of Zanzibar, WHO and China jointly signed a memorandum of understanding to cooperate in the control of schistosomiasis in Zanzibar. This multilateral cooperation resulted in the decision to conduct a pilot project with the aim of eliminat-ing schistosomiasis in this part of Tanzania. In August 2016, Jiangsu Institute of Parasitic Disease in Wuxi, China officially took on this China-aid project for schistosomiasis elimination in Zanzibar. The project was stationed on Pemba Island and a systematic disease investigation and snail survey was initiated together with the local NTD office. It was felt that the work should focus on the spatial component to establish effective data collection and management for schistosomiasis control using geographical information systems (GIS) based on the Google Earth (GE) platform (Google Inc., Mountain View, CA, USA), which is a free software displaying geographic data.

Elliott and Wartenberg (2004) pointed out the importance of spatial epidemiology and a review by Bergquist and Rinaldi (2010) noted the usefulness of remote sensing (RS) for health-related research on endemic diseases. GE integrates the functions of RS, GIS and global positioning (GPS) technologies to present users with satellite-generated aerial imagery that varies from the global scale to the local (Lisle, 2006). Combined with GIS software, e.g., ArcGIS (ESRI, Redlands, CA, USA), and network technology, GE is an excellent approach to spatial epidemiology as various files used for the collection of data, e.g., Excel and shapefiles. Such files can easily be reformatted and incorporated as keyhole markup language (KML) files, a commonly used format pioneered by Keyhole Inc., which was acquired by Google in 2004. The shapefile is a popular geospatial vector data format developed by ESRI for interoperability between various software products. Human residences as well as streams and ponds are spatially represented as points, lines and polygons, respectively. A large number of different annotations describing these vectors, such as name, environmental variables, infection status, etc., can be added as annotations.

The GE virtual globe approach incorporates aerial images, digital terrain and a large amount of basic geographic information, the utilization of which is simple and convenient. Collected data can be accurately recorded and zooming in to present geographically limited areas as flat maps where the landscape scene can be intuitively understood with distances and areas derived in a straightforward way. The spatial data management methods applicable to basic schistosomiasis control work on Pemba were explored by adopting the GE platform to establish a spatial database for the demonstration of the distribution of human schistosomiasis infections and the presence of the intermediate snail host in the various water bodies available on the island. The specific aim of this work was to improve management of schistosomiasis control in Zanzibar by investigating the spatial relation between infected humans and B. globosus habitats.

Materials and Methods

Time and area under study

S. haematobium prevalence and snail status on Pemba Island was surveyed from February 2017 to May 2018. The island is divided into four districts that are further divided into about 120 smaller administrative areas, referred to as shehias, which can contain several villages varying in population size and household number. The total area of Pemba Island is 984 km², and the average area of a shehia is 8.2 km².

Data records and generation of spatial coordinates

The data consisted of geographic information, such as the whereabouts of human residences, ponds and streams. All findings regarding these entities, such as diagnostic results and snail data, etc., were named attribute data and given unique identification numbers (ID), for example for the human cases the record included name, house location, laboratory test results and treatment (Shehia no./Village no./Household no./Person no.).

Human data

Schistosomiasis-related data included the infection rate for each shehia and individual case investigation of positive cases. There are around 3,000 people in each shehia on Pemba Island. The whole population were surveyed in four pilot shehias, i.e., Mtangani, Kiyuya, Wingwi and Uwandani. In other shehias, about 200 residents (including at least 30 children aged 9-12 years) from 50 households were randomly selected and surveyed. Urine samples were collected and 10 ml of each sample was filtered and subjected to microscopy by experienced laboratory technicians who recorded all S. haematobium eggs present for each individual. All this information was recorded in an Excel database.

Snail data

Snail survey was conducted after the rainy season. All water bodies in the project area were investigated. The snail data included the distribution of B. globosus in the environment inhabited by humans and included the number of surveyed sites (ponds and streams), the total number of snails and the number of infected ones. The ponds were encoded as Shehia no./Village no./Pond no./Serial no./Pond type, while the streams were only labelled with its serial no. according to order of investigation. The sites were surveyed for snails at distance gaps of 30 and 100 m, respectively. At each identified site, presence of B. globosus was searched for 15 min over a maximum 15 m² area. The borders of the water bodies and their vegetation, where snails are most likely to be found, were more intensively searched than other areas. B. globosus was preliminary identified upon shell morphology, placed in screw top plastic containers with freshwater, and taken to the laboratory for microscopic examination of cercariae after crushing. A handheld Garmin etrex 30x device (Kansas City, KS, USA) was used to specify the coordinates of all findings. All data were saved as KML files through BaseCamp software, version 4.5.2 (Garmin).

Annotating attribute data

Positive human cases were recorded as a point on the map and the total number counted, while the length of the streams, given as lines, and the surface area of the ponds, given as polygons, were
measured using the GE properties window and recorded in the Excel database. Water bodies without snails were given blue colour, those with presence of non-infected snails yellow and those with presence of infected snails red. These spatial, qualitative pieces of information, each identified by its ID, were saved as shapefiles, reformatted as KML files and subsequently imported into GE Pro, version 7.1.8.3036 (Google).

The spatial database

The spatial, qualitative pieces of information, each identified by its ID, were collected in a GIS database, reformatted by ArcGIS, version 10.2.2 (ESRI) as KML files and subsequently imported into GE Pro, version 7.1.8.3036 (Google) to display the distribution of positive human cases in their environment and the snails (infected and uninfected) in theirs (ponds and streams). In this way, a GIS database containing the coordinates of all human positive cases found, including the status of the surrounding ponds and streams was established. Digital maps and shapefiles of Pemba Island were obtained from the NTD office and associated with the collected schistosomiasis data with reference to the shehias so that the GIS database could tell the schistosomiasis infection rate in all the different shehias on the island. Managed this way, all geographic and attribute data would be available and possible to visualize by using the software and clicking on the object in question.

Training

The Chinese team trained the local NTD staff on the operation of GE and how to use a handheld GPS device to specify the coordinates of all findings, save tracks and mark points, lines and polygons, how to import GPS data into GE as well as how to measure distances and how to calculate areas. The staff was also trained to use ArcGIS software manipulating ArcGIS shapefile vector features of interest such as points, lines and polygons, here representing human infections, streams and ponds, respectively, and converting this format into KML files to be fed into the GE software. They were also trained in associating attribute data related to the spatial data collected so as to create spatial databases.

Figure 1. Prevalence of *S. haematobium* in 38 surveyed shehias on Pemba Island.
Results

Surveys

From Feb 2017 to May 2018, a total of 38 shehias were investigated for *S. haematobium* infection. A total of 17,674 urine samples were collected and 546 were found to be positive corresponding to an average positive rate of 3.1%. Figure 1 shows the prevalence rates in the surveyed 38 shehias on Pemba Island.

Spatial databases

Three spatial databases were created by means of GE and they included distribution of infected humans (191 persons in 8 shehias); distribution of the intermediate snail host in ponds (all the ponds surveyed in 20 shehias); and distribution of the intermediate snail host in streams (all the streams surveyed in 38 shehias) (Figure 2).

Infected humans

The diagnostic examination showed that 191 persons in the 8 shehias were infected. The coordinates of these residences were available in the database that had been established and the ID and name, householder name, egg counts and treatment were recorded in the attribute data for each positive case. They could then be visualized and marked on GE as shown in Figure 3. The attribute information of each case could be displayed by clicking on the point in question.

Snail distribution

A total of 117 ponds in the 20 shehias were surveyed for *B. globosus* (Figure 4). The total area of the ponds surveyed was 2,052,965 m². Although we found 36 ponds with *B. globosus* covering an area of 1,133,724 m² only 14 (808,194 m²) were harboured infected *B. globosus*. Figure 5 shows ponds without snails, ponds with non-infected snails and ponds with infected snails. The basic information and snail information for each pond could be displayed by clicking on that pond shown in GE.

In the shehia of Uwandani, four infected ponds were found in August 2017, and nearly 200 positive cases were found in the whole population survey. The location of human cases and infected waterbodies were displayed on GE, and we found there were no infected ponds nearby some human cases. Then the waterbodies and snail situation were investigated again, and another two infected ponds surrounding the human cases were found (Figure 6).

A total of 72 streams in the 38 shehias were surveyed as shown in Figure 4. The total length of the streams surveyed was 183 km. We found 44 streams with *B. globosus*, out of which 6 harboured infected snails. Figure 7 shows streams without snails, streams with snails and streams with infected snails. The basic information and snail information for each stream could be displayed by clicking on that stream in the software.

Training

After training and work practice, all NTD staff easily grasped the use of GPS and GE. They could utilize GPS devices to specify coordinates and save tracks and use GE to import GPS data. They
Figure 3. Distribution of positive human cases and display of annotated information of cases.

Figure 4. Distribution of *B. globosus* in ponds and streams surveyed in 38 shehias on Pemba Island.
could also measure distances and areas and use ArcGIS to convert shapefile vector files to KML files, create spatial databases and make maps.

Discussion
Many diseases show spatial clustering (Elliott and Wartenberg, 2004; Kirby et al., 2017), including parasitic diseases (Barbosa et al., 2014; Hundessa et al., 2016), other infectious diseases (Smith et al., 2015; Lai et al., 2018) and also non-communicable diseases like cancer (Goodman et al., 2014; Sun et al., 2015).

Schistosomiasis transmission is focal in nature (Manyangadze et al., 2016) and schistosomiasis cases are consequently also spatially clustered as shown in many studies (Brooker et al., 2006; Clements et al., 2006; Simoonga et al., 2008). Also, the intermediate snail host has the tendency to appear in cluster (Cheng et al., 2016; Gao et al., 2014). It follows that spatial data collection are important for schistosomiasis control. Indeed, this helped us to identify four infected ponds in Uwandani that had not been found until we located the human cases near these ponds.

On Pemba, *B. globosus* is the only intermediate host of *S. haematobium*, so the geographical distribution of this snail species determines the spatial distribution of schistosomiasis on the island. Clennon et al. (2004) has already shown that households with high infection intensities of *S. haematobium* were significantly clustered around a water body with infected snails near the southern coast of Kenya. Following this lead, we established a spatial database of distribution of positive cases, together with the distribution of *B. globosus* in ponds and streams based on GPS and GE, which helped following up cases as well as guide us to other key areas in need of snail control as mentioned above.

GE is increasingly applied to the study of infectious diseases (Escamilla et al., 2014; Simonsen et al., 2016). For example, it was used to map HCV infections by home, social, and sexual neighbourhoods among community-recruited men who have sex with men in New York City (Tieu et al., 2018). In another study, Google Earth imagery and geographical analysis software was used to develop a sampling frame by digitizing and assigning coordinates for each household within the catchment area (Escamilla et al., 2014). The GE platform is particularly useful for research on schistosomiasis (Clennon et al., 2004; Clements et al., 2006; Simoonga et al., 2008, 2009; Yang et al., 2012; Sun et al. 2014; Gao et al., 2014; Wang et al., 2014; Manyangadze et al., 2015; Xiao et al., 2016a, 2016b). A great advantage is that GE users need not be computer specialists since its operation is straightforward and the software easy to apply, e.g., for map creation (Lefer et al., 2008) which is the first step. As long as the computer is connected to Internet, any map can be rapidly displayed. The application of GE in *grass-root* schistosomiasis control institutions is therefore a useful platform for the storage and display of spatial data (Wang et al., 2014; Xiao et al., 2016a). Combined with GIS software, such as ArcGIS, GE is also reliable for surveillance and forecasts (Xiao et al., 2016b) and can be implemented for epidemiological monitoring and the development of early warning systems (EWS) (Simonsen et al., 2016).

An on-site trial with GE as the operative platform achieved good results. Indeed, already after a few weeks of simple training and some practice, the NTD staff mastered all the different steps from recording results to displaying them. The spatial data were

Figure 5. Distribution of *B. globosus* in ponds and display of annotated information of ponds.
first generated using the coordinates taken from the GPS device adding attribute information for human cases and snails previously recorded in Excel. Secondly, both spatial and attribute data were organically integrated via ArcGIS establishing the GIS database, which made it possible to effectively manage and utilize the large number of epidemiological data collected. ArcGIS permits the provision of various thematic maps of the distribution of both human cases and positive snails. This database can also be used for infor-
mation query and statistical analysis. After the establishment of the GIS database, ArcGIS provided the link to GE through the conversion of shapefiles to KML files and thus clearly visualize the geographic position of positive cases as well as ponds and streams with infected snails, while attribute data, such as case information or snail distribution information, were only a click away.

According to the needs, a certain type of data can be selected for display on the map, or multiple different types of data can be displayed at the same time, easily visualizing the degree of severity of the endemic situation. We could also explore the spatial relation between infected human cases and *B. globosus* habitats by analysing the location of human cases and water bodies.

There are some limits in this study. For snail classification, we just conducted preliminary identification using shell morphology, while it is advisable to apply molecular diagnosis based on DNA detection for further identification. Secondly, the location of households with infected cases is not a proxy of where transmission takes place. In the next step, the proximity of a household to the nearest water body could be used to identify clusters of infected cases which might point towards a transmission source.

It should be borne in mind that, in contrast to complete eradication of a disease, elimination means controlling it so well that it no longer is of public health importance. However, it should also be remembered that keeping a disease eliminated in the face of ongoing transmission can be a tall order in the absence of strong economic development. A surveillance tool like GE that can connect all information needed and also play the role of EWS would be most useful for the rapid deployment of an effective response. As such, GE would not only be useful on Pemba but also in other countries on the African continent and elsewhere.

### Conclusions

It might seem possible to also eradicate schistosomiasis on an island like Pemba as the authorities there should be able to introduce visitor controls isolating it from influx of the infection from surrounding endemic areas. As a free software, the technique of GE proved applicable and useful for spatial data management for schistosomiasis control. Based on this platform, the severity of schistosomiasis distribution and snail situation can be judged intuitively and timely, supporting the identification and location of transmission sites as well as communities at high risk. Therefore, it can support planning and contribute to timely application of control measures by the health services. Our experience with training of staff without previous knowledge of advanced use of computers and GPS instruments for the collection and display of epidemiological data was very positive. All NTD staff easily grasped the use of GPS and GE and quickly applied this new knowledge for the collection of data, adding coordinates and saving tracks. They could also quickly learn how to measure distances and areas and use ArcGIS to convert shapefile vector files to KML files, thereby creating spatial databases and produce the maps needed.

### References

Allan F, Dunn AM, Emery AM, Stothard JR, Johnston DA, Kane RA, Khamis AN, Mohammed KA, Rollinson D, 2013. Use of sentinel snails for the detection of *Schistosoma haematobium* transmission on Zanzibar and observations on transmission...


