

Appropriate Tools and Methods for Tropical Microepidemiology: a Case-study of Malaria Clustering in Ethiopia

Tedros A Ghebreyesus¹, Peter Byass², Karen H Witten¹, Asfaw Getachew¹, Mitiku Haile³, Mekonnen Yohannes³, Steven W Lindsay⁴

Abstract

Background: The importance of local variations in patterns of health and disease are increasingly recognised, but, particularly in the case of tropical infections, available methods and resources for characterising disease clusters in time and space are limited. Whilst the Global Positioning System (GPS) allows accurate and easy determination of latitude and longitude, sophisticated Geographical Information Systems (GIS) that can process the data may not be available and accessible where they are most needed.

Objective: To describe an appropriate procedure for interpreting GPS information.

Methods: An example of space-time clustering of malaria cases around a dam in Ethiopia (106 cases in 129.7 child-years-at-risk) is used to demonstrate that GPS data can be interpreted simply and cheaply in under-resourced health service settings to provide timely and appropriate epidemiological assessments.

Results: Malaria cases were clustered in time and space in the area surrounding a microdam.

Conclusion: Quickly identifying disease foci using appropriate procedures in this manner could lead to better informed control and treatment activities which would represent a better use of resources as well as improved health for the community. [*Ethiop.J.Health Dev.* 2003;17(1):1-8]

Introduction

Increasing attention is being paid to micro-level variations in patterns of health and disease (1), and increasingly sophisticated tools and

methods are becoming available for such investigations. One of the most significant factors has been the advent of the Global Positioning System (GPS) which enables relatively precise determination of latitude and longitude at any point on the earth's surface from triangulated satellite signals. At the same time, the technological advances in computer-based Geographical Information Systems (GIS) and the availability of large and comprehensive environmental data bases have opened up possibilities of extremely sophisticated investigations and analysis.

Unlike most manifestations of sophisticated technology, the GPS often works at its best in remote rural areas of the world, where man-made structures and electromagnetic interference are at a minimum. The current generation of GPS receivers are small and portable, requiring only battery power, very

¹Tigray Regional Health Bureau, Mekelle, Ethiopia, Tel 251 3 400222, Fax 251 3 413329, e-mail: planprog@telecom.net.et; ²School of Community Health Sciences, University of Nottingham, Nottingham, UK and Department of Epidemiology and Public Health, University of Umeå, Umeå, Sweden, Tel 44 1949 836016, Fax 44 7989 338460 e-mail: peterbyass@aol.com; ³Mekelle University, Mekelle, Ethiopia, Tel: 251 3 400512; ⁴Department of Biological Sciences, University of Durham, Durham, UK and Danish Bilharziasis Laboratory, Charlottenlund, Denmark, Tel: 44 191 374 3349, Fax: 44 191 374 2417, e-mail: S.W.Lindsay@durham.ac.uk

easy to use (Figure 1) and now cost as little as US \$100. GPS was originally developed by the United States Department of Defense, primarily for military purposes, with an associated commitment to make the satellite signals generally available on a worldwide basis. However, a technical feature known as selective availability (SA) was incorporated into the signals, reducing accuracy to around ± 50 metres, except for the US military who could decode SA. Non-military solutions to the SA problem were developed, such as differential GPS (DGPS), or averaging many readings, but which involved much greater complexity of use. However, on 1 May 2000, President Clinton announced that SA would be immediately discontinued, with the result that more accurate GPS data are now easily available to epidemiologists throughout the world with an accuracy of around ± 5 metres.

Unfortunately, as far as developing countries are concerned, GIS are subject to the more typical constraints of high technology, including relatively high costs, and need stable power supplies and considerable human skills and resources to be used effectively. Whilst this may be quite feasible for centres of excellence such as universities in developing countries, GIS are generally beyond the scope of district and regional health services, even though the latter represent a key group in terms of their need to effectively assess the geographical epidemiology of outbreaks and changing patterns of disease.

This paper therefore sets out to bridge the gap between what is easily and appropriately accessible even in relatively poorly resourced environments (GPS data linked to other epidemiological information) and to explore appropriate methods for investigating and carrying out relatively simple analyses of these data. The basic criteria for any such approach are (a) requiring only a basic personal computer, (b) not having significant software

costs, and (c) not requiring extensive specialised training.

Methods

As a case-study, data are presented from a study of malaria epidemiology in relation to water resource development in Tigray, northern Ethiopia, undertaken during 1997. This study has been described elsewhere (2). Data from one dam area (Mai Temen) are presented here, which relate to a surveyed community including 497 children under the age of 10 years living in 247 households. The location (latitude and longitude to 5 decimal places of a degree) of each household was noted using a simple hand-held GPS receiver (GPS 38, Garmin International Inc., Olathe, Kansas), and a centre point for the dam located at 14.325 °N, 38.163 °E. During the course of 4 quarterly 30-day incidence surveys, 106 incident episodes of malaria were detected. Each quarterly survey was undertaken in two phases; firstly blood films were sought from all children under 10 years, and those with clinical malaria (defined as 1 or more days of fever, chills, sweating, headache) were treated with chloroquine (25mg/kg over 3 days) and another blood film taken after 14 days. Secondly, further blood films were sought from all children as close to 30 days as possible after their initial film. The data were handled using Epi-Info software (3) which is available on a public-domain basis. Figure 2 shows the location of the households with children under 10 years of age in relation to the dam, the filled dots representing households with at least one case of childhood malaria during the year. Although this figure was prepared for publication using mapping software, for field purposes these kind of data can equally well be drawn as a sketch map on squared paper.

• With the application of some high-school mathematics, it is possible to process GPS data in a number of ways. For example, the distance between two GPS points, or the compass bearing from one point to another, can easily be

calculated, as shown in the Appendix. These methods were used with these data to calculate, for each household, its distance and direction from the centre point of the dam. These distance data then enable household data to be grouped into concentric bands around the dam, as shown for example in the concentric circles of radius 1.5 and 3 km in Figure 2. Similarly the direction data can be used to group households radially around the centre point, for example into the quadrants shown in Figure 2.

Incidence rates were calculated within different sub-groups using the number of incident parasitaemia cases detected over the corresponding child-years observed. Confidence intervals and comparisons between such incidence rates can also readily be calculated (4), as detailed in the Appendix.

Although it is obvious from the sketch map in Figure 2 that malaria cases were unevenly distributed, establishing where and when clusters of cases occur requires further analysis. For this purpose further public-domain software, SaTScan (5), was applied to the data. This uses a replication method to create a spatial scan statistic by imposing a circular window which is then centred around each of several possible centroids positioned throughout the study region. For each centroid, the radius of the window varies continuously in size from zero to an upper limit, thus creating an infinite number of distinct geographical circles, each being a possible candidate for a cluster.

Ethical approval for this study was given by the Ethiopian Science and Technology Commission and the Tigray Regional Health Bureau and local consent obtained through community meetings.

Results

Overall, 106 incident cases of malaria were observed in a total of 129.7 child-years-at-risk (cyar), corresponding to an incidence rate of 0.82 cases/cyar. Table 1 shows incidence broken down by time, distance and direction from the dam. Significantly more cases occurred in the second half of the year (1.21 cf 0.44 cases/cyar, a rate ratio of 2.8 [95% CI 1.8 to 4.2]). Incidence was also significantly higher within 3 km of the dam (1.16 compared with 0.32 cases/cyar, a rate ratio of 3.6 [95% CI 2.1 to 6.0]).

SaTScan was applied to the data to determine clusters of cases in time and space. The most likely cluster to emerge after 9,999 replications was centred on 14.339 °N, 38.146 °E with a radius of 0.33 km during the 3rd quarter survey cycle. This cluster contained 13 households in which 15 cases occurred during that period, associated with 2.0 cyar and thus an incidence rate within the cluster of 7.5 cases/cyar. This cluster is represented in Figure 2 by the shaded circle to the north-west of the dam. The second most likely cluster was centred on 14.318 °N, 38.185 °E with a radius of 2.31 km during the latter half of the year. This cluster included 36 cases in 72 households, with an incidence rate of 2.7 cases/cyar, and is represented in Figure 2 by the large shaded circle to the south-east of the dam.

Table 1: Incidence of malaria per child-year-at-risk among 497 Ethiopian children under the age of 10 years living near a microdam, by time and space

	Group	Cases	Child-years-at-risk	Incidence (95% CI)
Overall incidence	all	106	129.7	0.82 (0.66 to 0.98)
Incidence by quarter during 1997	1 st	9	30.5	0.30 (0.11 to 0.49)
	2 nd	20	35.7	0.56 (0.31 to 0.81)
	3 rd	34	35.2	0.97 (0.65 to 1.29)
	4 th	43	28.3	1.52 (1.07 to 1.97)
Incidence by distance from the dam	<1.5 km	24	36.0	0.67 (0.40 to 0.94)
	1.5-2.99 km	65	41.0	1.59 (1.20 to 1.98)
	3 km and over	17	52.7	0.32 (0.17 to 0.47)
Incidence by direction from the dam	NE quadrant	27	32.2	0.84 (0.52 to 1.16)
	SE quadrant	37	15.3	2.42 (1.64 to 3.26)
	SW quadrant	6	46.3	0.13 (0.03 to 0.23)
	NW quadrant	36	35.9	1.00 (0.67 to 1.33)



Figure 1 Hand-held GPS receiver, in use at position 13.35967 °N and 39.52304 °E

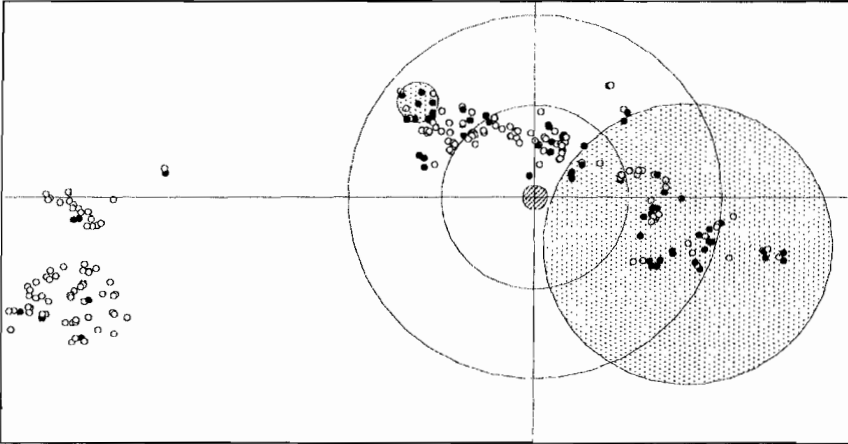


Figure 2: Sketch map of the dam area showing surrounding households with and without malaria cases. Shaded circles represent space-time clusters of cases identified from SaTScan analysis

Discussion

This case-study demonstrates the potential of analysing a combination of GPS and epidemiological data without needing to use complex GIS facilities, whilst revealing detailed microepidemiological patterns within the data. None of this analysis depends on the mapping process, although some kind of sketch map is a useful tool in understanding the data and results. However, there is no need for the mapping process to be computerised. Both Epi-Info and SaTScan require only relatively modest computing facilities that are now commonly available throughout the world, which, coupled with their free availability, make them extremely attractive for this type of investigation.

In the past GPS data as determined with the simplest receivers has been associated with the SA margin of error, although for many purposes it was quite sufficient to locate, for example, a household to within 50 m in absolute terms. Some authors have argued for

the use of more sophisticated differential GPS receivers for greater accuracy (6), but, with the discontinuance of SA, the additional expense and complexity would no longer seem to be justified for most epidemiological applications.

A variety of epidemiological investigations have been reported in which GPS data have been processed using GIS techniques. In some cases, for example in Thailand, existing, good quality maps have been used as the primary GIS data, with GPS providing locations of health facilities and other relevant features (7). Similarly in Bolivia map-based GIS approaches were enhanced with GPS fieldwork in analysing health care access (8). Elsewhere in Africa, GPS has been used at a basic level to locate villages in the Congo for larger scale epidemiology of onchocerciasis (9) and for more detailed mapping of tuberculosis treatment in South Africa (10). A malaria field study in Kenya using detailed mapping techniques for spatial analysis has been described (6).

The interpretation of our case-study results in terms of malaria epidemiology in this community remains to some extent speculative. The extremely small and intense focus of transmission early in the epidemic season is particularly interesting to observe. Although this particular analysis was carried out retrospectively, it illustrates the potential of using these methods for rapid assessment of disease patterns at the local level. A real-time approach to this analysis could well have influenced decisions on local strategies for residual spraying, case treatment and bednet distribution or re-dipping, particularly in relation to this small cluster detected early in the season. It is tempting to suggest that prompt control of this cluster might have influenced overall malaria transmission in the village later in the year, although we have no evidence to support this.

These results have been achieved within the criteria set (a basic personal computer, without significant software costs, and not requiring extensive specialised training), thus illustrating the feasibility of undertaking micro-epidemiological investigations from basic resources. Although Boelaert *et al.* (11) have questioned the appropriateness of GIS approaches at the district health level, Tanser and Wilkinson (10) have argued that "GIS/GPS is no longer exclusively a research tool", while Porter (12) stresses that "GIS and GPS need to be seen simply as tools" and not "displace simpler methods like drawing a map or a graph". Clarke *et al.* (13) conclude that while GIS "holds distinct promise as a tool" for public health, "it does not fit neatly into the health scientist's toolbox". Whilst GIS do, and will continue to, have an important role in many health-related applications, particularly for detailed research, we would wish to distinguish between GPS and GIS in terms of their routine applicability and appropriateness. We conclude that there are working environments in which GPS can be used as part of rapid and effective epidemiological

assessments, even though GIS hardware, software and expertise may not be available or appropriate.

Acknowledgements

This project was funded by SIDA/SAREC, Sweden, and supported by the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR). We are grateful for the contributions made by support staff at the institutions involved, the village-based enumerators and the villagers who agreed to participate in the study.

References

1. Greenwood BM (1989). The microepidemiology of malaria and its importance to malaria control. *Trans R Soc Trop Med Hyg.* 83:S25-29.
2. Ghebreyesus TA, Haile M, Witten KH, Getachew A, Yohannes AM, Yohannes M, Teklehaimanot HD, Lindsay SW & Byass P (1999). Incidence of malaria among children living near dams in northern Ethiopia: community-based incidence survey. *Br Med J* 319:663-666.
3. Dean AG (1995). *Epi Info, Version 6: A Word-Processing, Database, and Statistics Program for Public Health on IBM-compatible Microcomputers.* Atlanta, Georgia: The Division of Surveillance and Epidemiology, Epidemiology Program Office, Centers for Disease Control and Prevention (CDC). Available at Epidemiology Program Office, Centers for Disease Control and Prevention (CDC). Available at: <http://www.cdc.gov/epo/epi/epiinfo.htm>
4. Kirkwood B (1988). *Essentials of Medical Statistics.* Blackwell Scientific Publications, Oxford. pp128-130.
5. Kulldorff M, Rand K, Gherman G, Williams G & DeFrancesco D (1998). *SaTScan v 2.1: Software for the spatial and space-time scan statistics.* Bethesda, MD: National Cancer Institute. Available at <http://dcp.nci.nih.gov/bb/satscan.html>

6. Hightower AW, Ombok K, Otieno R, Odhiambo R, Oloo AJ, Lal AA, Nahlen BL & Hawley WA (1998). A Geographic Information System applied to a malaria field study in western Kenya. *Am J Trop Med Hyg* 58:266-282.
7. Entwistle B, Rindfuss RR, Walsh SJ, Evans TP & Curran SR (1997). Geographic Information Systems, Spatial Network Analysis, and Contraceptive Choice. *Demography* 34:171-187.
8. Perry B & Gesler W (2000). Physical access to primary health care in Andean Bolivia. *Soc Sci Med* 50:1177-1188.
9. Law PA, Ngandu ON, Crompton .P, Usungu O, Kosten D, Law JK & Burnham G, (1998). Prevalence of *Onchocerca volvulus* nodules in the Sankuru river valley, Democratic Republic of the Congo, and reliability of verbal assessment as a method for determining prevalence. *Am J Trop Med Hyg* 59:227-230.
10. Tanser F & Wilkinson D (1999). Spatial implications of the tuberculosis DOTS strategy in rural South Africa: a novel application of geographical information system and global positioning system technologies. *Trop Med Int Health* 4:634-638.
11. Boelaert M, Arbyn M & Van der Stuyft P (1998). Geographical Information Systems (GIS), gimmick or tool for health district management? *Trop Med Int Health* 3:163-165.
12. Porter JHD (1999). Geographical information systems (GIS) and the tuberculosis DOTS strategy. *Trop Med Int Health* 4:631-633.
13. Clarke KC, McLafferty SL & Tempalski BJ (1996). On Epidemiology and Geographic Information Systems: a Review and Discussion of Future Directions. *Emerg Infect Dis* 2:85-92.

Appendix

1. Calculating the distance between two GPS points

Although, due to the curvature of the earth, degrees of latitude and longitude are not absolutely related to distance, in the equatorial region is it reasonable to make the assumption that one degree of latitude or longitude is equivalent to 110 km. Using this assumption, the latitudinal distance between two points can be calculated as

$$\begin{aligned} \text{latitudinal distance} = \\ (\text{latitude}_1 - \text{latitude}_2) \times 110 \text{ km} \\ \text{e.g.} \\ (14.3258^\circ - 14.3041^\circ) \times 110 = 2.39 \text{ km} \end{aligned}$$

and the corresponding longitudinal distance as

$$\begin{aligned} \text{longitudinal distance} = \\ (\text{longitude}_1 - \text{longitude}_2) \times 110 \text{ km} \\ \text{e.g.} \\ (38.1209^\circ - 38.1157^\circ) \times 110 = 0.57 \text{ km} \end{aligned}$$

and then by Pythagoras' theorem the actual distance as

$$\begin{aligned} \text{actual distance} = \\ \{(\text{latitudinal dist})^2 + (\text{longitudinal dist})^2\}^{0.5} \\ \text{e.g.} \\ \{(2.387^2 + 0.572^2)\}^{0.5} = 2.46 \text{ km} \end{aligned}$$

2. Calculating the direction from one GPS point to another

Following from the distance calculation, the compass bearing from one point to another is given by

$$\begin{aligned} \text{compass bearing} = \\ \sin^{-1}(\text{longitudinal distance/actual distance}) \\ \text{e.g.} \\ \sin^{-1}(0.57/2.46) = 14^\circ \end{aligned}$$

3. *Calculating a 95% confidence interval for an incidence rate and the ratio between two incidence rates*

An incidence rate expresses the number of events (for example, new cases [n]) in terms of people's possible exposure leading to the event (often expressed in terms of person-years-at-risk [pyar]). The incidence rate λ is expressed as

$$\lambda = n/pyar$$

e.g. 106/129.7 = 0.82 cases/pyar

and the corresponding 95% confidence interval is given by

$$\lambda \pm 1.96 \times (\sqrt{n / pyar})$$

e.g. 0.82 ± [1.96 x (√106 / 129.7)] = 0.66 to 0.98 cases/pyar

Similarly a ratio between two incidence rates can be calculated for the purposes of comparison as

$$\lambda_1 / \lambda_2 = (n_1/pyar_1) / (n_2/pyar_2)$$

e.g. (29/66.2) / (77/63.5)
 $= 0.438 / 1.213$
 $= 0.361$

and a corresponding 95% confidence interval calculated for the ratio as

$$\exp\{ \ln(\lambda_1 / \lambda_2) \pm [1.96 \times \sqrt{(1/n_1 + 1/n_2)}] \}$$

e.g. exp{ ln[(29/66.2) / (77/63.5)] ± [1.96 X √(1/29 + 1/77)] }
 $= \exp\{ \ln(0.361) \pm 1.96 \times \sqrt{0.0475} \}$
 $= \exp\{ -1.019 \pm 0.427 \}$
 $= 0.235 \text{ to } 0.553$

This type of confidence interval for a ratio of rates is asymmetrical, and, if the value of 1.0 does not lie within the interval, the implication is that the two rates in the ratio differ significantly.