Multiple correspondence analysis as a tool for analysis of large health surveys in African settings

Dawit Ayele, Temesgen Zewotir, Henry Mwambi

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, Private Bag X01, Scottsville 3209, South Africa.

Abstract

Background: More than two thirds of the total population of Ethiopia is estimated to be at risk of malaria. Therefore, malaria is the leading public health problem in Ethiopia.

Objective: To investigate the determinants of malaria Rapid Diagnosis Test (RDT) result and the association between socio-economic, demographic and geographic factors.

Method: The study used data from household cluster malaria survey which was conducted from December 2006 to January 2007. A total of 224 clusters of about 25 households each were selected from the Amhara, Oromiya and Southern Nation Nationalities and People (SNNP) regions of Ethiopia. A multiple correspondence analysis was used to jointly analyse malaria RDT result, socio-economic, demographic and geographic factors.

Results: The result from multiple correspondence analysis shows that there is association between malaria RDT result and different socio-economic, demographic and geographic variables.

Conclusion: There is an indication that some socio-economic, demographic and geographic factors have joint effects. It is important to confirm the association between socio-economic, demographic and geographic factors using advanced statistical techniques.

Keywords: MCA, CA, malaria, RDT.

DOI: http://dx.doi.org/10.4314/ahs.v14i4.35

Introduction

While malaria has long been a cause of human suffering and mortality in Sub-Saharan Africa, in Ethiopia the problem is particularly severe[1].

Malaria is a leading cause of death amongst children in many African countries[2]. In Ethiopia, malaria is a major public health problem with 68% of the total population living in areas at risk of malaria[3,4]. From the total population of Ethiopia, more than 50 million people are at risk[5,6]. In highlands or highland fridge areas of Ethiopia, i.e. mainly areas 1,000-2,000 meters above sea level[7,8], epidemics of malaria are relatively high[6,9,10]. Notably this altitude covers 48% of the regions of Amhara, Oromiya and Southern Nations Nationalities and People regions of Ethiopia.

Corresponding author:
Ayele Dawit
University of KwaZulu-Natal, Pietermaritzburg,
Private Bag X01, Scottsville 3209, South Africa.
Phone Number: +270739720957
Fax Number: +27332605648
Postal Address: School of Mathematics,
Statistics and Computer Science,
E-mail: ejigmul@yahoo.com , ayele@ukzn.ac.za

Methods and materials

Study design

From December 2006 to January 2007, baseline household cluster malaria survey was conducted by the Carter center (TCC). The questionnaire was developed as a modification of the malaria indicator survey (MIS) household questionnaire. The questionnaire had two parts; the household interview and malaria parasite form. For this survey, multi-stage cluster random sampling was used. For the sampling purpose, the lowest measurement of malaria prevalence was used. Based on this, the sample size was estimated. In Amhara region, each zone was regarded as a separate domain, while in Oromiya and SNNP, the community-directed treatment with ivermectin (CDTI) areas combined were one domain. In Oromiya and SNNP, sampling was done directly at the kebele level. Kebele is the smallest administrative unit in Ethiopia. Therefore, the sampling frame was the rural populations of Amhara, Oromiya and SNNP regions, which is Kebele.

From the three regions, 5,708 households were included in the survey. From 5,708 households, Amhara, Oromiya and SNNP regions cover 4,101 (71.85%), 809 (14.17%) and 798 (13.98%) households respectively. To conduct the survey, first, 224 Kebeles were selected. From each Kebele, 12 households were selected for malaria tests. In the survey each room in the house was listed separately. Using the presence of mosquito nets, it was possible to ascertain the density of occupation per room as well as how many sleeping rooms were in or outside each house. In addition to the number of rooms and number of nets, the persons sleeping under each net were listed. The detailed sampling procedure for the base line household survey was discussed by different authors[11-13].

To obtain malaria parasite testing, consent from participants was obtained. To collect the blood sample, finger-prick blood samples was collected from participants for malaria Rapid Diagnostic Test. The test used was ParaScreen which is capable of detecting both Plasmodium falciparum and other Plasmodium species. Participants with positive rapid tests were immediately offered treatment according to national guidelines[14,15].

The socio-economic, demographic and geographic covariates comprised the baseline socio-economic status,demographic, and geographic variables that included gender, age, family size, region, altitude, main source of drinking water, time taken to collect water, toilet facilities, availability of electricity, radio and television, total number of rooms, main material of the room's wall, main material of the room's roof and main material of the room's floor. Malaria test RDT result, age and sex were collected at individual level. Altitude, main source of drinking water, time taken to collect water, toilet facilities, availability of electricity, radio, television, total number of rooms, main material of the room's walls, main material of the room's roof and main material of the room's floor were all collected at household level.

Statistical Methods

The cross-tabulation of categorical data is perhaps the most commonly encountered and simple form of analysis in research. Therefore, ordering things in time has been the interest of many researchers. Correspondence analysis is one of a wide range of alternative ways of handling and representing the relationships between categorical data. Correspondence analysis can suggest unexpected dimensions and relationships in the tradition of exploratory data analysis. The results of the correspondence analysis can be seen analytically and visually. This method first developed in France[16,17]. Different authors have proposed this method under various names. These methods are Dutch Homeneity Analysis[18], the Japanese Qualification Method[19], the Canadian Dual Scaling[20]. These methods have different theoretical foundations but all methods leads to equivalent solutions[21,22]. Correspondence analysis is thought of as principal component method for normal and contingency table data. It can be used to analyze cases-by-variable-categories matrices of non-negative data. Correspondence analysis is also a multivariate descriptive data analytic technique.

Even the most commonly used statistics for simplification of data may not be adequate for description or understanding of the data. The correspondence analysis results provide information which is similar to that produced by principal components or factor analysis[23]. Using the result, it is possible to explore the structure of the categorical variables included in the table. The simplified form data provides useful information about the data[24,25]. The relationship of the categories of rows and columns of the data can be represented using correspondence analysis. The graphical representation of the relationships between the row and column categories is in the same space which is also produced using correspondence analysis. In general, correspondence analysis simplifies complex data and provides a detailed description of practically every bit of information in the data, yielding a simple, yet exhaustive analysis [21, 26].

Correspondence analysis has several features that distinguish it from other techniques of data analysis. The multivariate treatment of the data through multiple categorical variables is an important feature of correspondence analysis. This multivariate nature has advantage to reveal relationships which could occur during a series of pair wise comparisons of variables[27]. Correspondence analysis works effectively for the large data matrix, if the variables are homogeneous, and the data matrix structure is either unknown or poorly understood. There are some advantages of correspondence analysis over other methods. This advantage is related to joint graphical displays. This graphical display produces two dual displays whose row and column geometries have similar interpretations. This facilitates the analysis to detect different relationships.

In other multivariate approaches to graphical data representation, this duality is not present[28].

Multiple correspondence analysis (MCA) which is part of a family of descriptive methods, is an extension of correspondence analysis (CA) and allows to investigate the pattern of relationships of several categorical dependent variables. It is the multivariate extension of CA to analyze tables containing three or more variables. In addition to this, MCA can be considered as a generalization of principal component analysis for categorical variables which reveal patterning in complex data sets.

MCA helps to describe patterns of relationships distinctively using geometrical methods by locating each variable/unit of analysis as a point in a low-dimensional space. MCA is useful to map both variables and individuals, so allowing the construction of complex visual maps whose structuring can be interpreted. Moreover, this technique offers the potential of linking both variable centred and case centred approaches.

Suppose there are n observations on p categorical variables. Assume q_j different values for variable j. Next define a matrix, G_j which is $n \times q_j$ matrix. This matrix is known as indicator matrix. The $n \times q_j$ matrix G, with g the sum of q_j can be obtained by concatenating the G_j's [17]. In general, MCA is defined as the application of weighted Principal component analysis (PCA) to the indicator matrix G[29]. Furthermore, G is divided by its grand total np to obtain the correspondence matrix $F = \frac{1}{np}G$, i.e. $1_n^t F 1_q = 1$ where is 1_i an $i \times 1$ vector of ones. The vectors $r = F1_q$ and $c = F^t 1_n$ are the row and column marginals respectively. These marginals are the vectors of row and column masses. Suppose the diagonal matrices of the masses are defined as $D_r = diag(r)$ and $D_c = diag(c)$. Note that, the i^{th} element of r is $f_{i.} = \frac{1}{n}$ and the s^{th} element of c is $f_{.s} = \frac{n_j}{np}$ where n_s is the frequency of category s[21].

MCA can be defined as the application of PCA to the centered matrix $D_r^{-1}(F - rc^t)$ with distances between profiles given by the chi-squared metric defined by D_c^{-1}. The n projected coordinate of the row profiles on the principal axes are called row principal coordinates. The $n \times k$ matrix X of row principal coordinates is defined by $X = D_r^{-1/2}\widetilde{F}V$

where $\widetilde{F} = D_r^{-1/2}(F - rc^t)D_c^{-1/2}$ and V_k is the q×k matrix of eigenvectors corresponding to the k largest eigenvalues $\lambda_1, \ldots, \lambda_k$ of the matrix $\widetilde{F}'\widetilde{F}$. The projected row profiles can be plotted in the different planes defined by these principal axes called row principal planes[21].

The categories for column profile can be described by the column profiles. The value can be calculated by dividing the columns of **F** by their column marginals. Interchanging rows with columns and all associated entities can be used for the dual analysis of columns profiles. This is done by transposing the matrix **F** and repeating all the steps. The metrics used to define the principal axes (weighted PCA) of the centered profiles matrix $D_r^{-1/2}(F - rc^t)^t$ are D^c and D_r^{-1}.

The q ×k matrix Y of columns principal coordinates is now defined by

$$Y = D_c^{-1/2}\widetilde{F}^t U_k,$$

where U_k is the n ×k matrix of eigenvectors corresponding to the k largest eigenvalues $\lambda_1, \ldots, \lambda_k$ of the matrix $\widetilde{F}\widetilde{F}^t$. To aid visualization and interpretation of the projected column profiles in the planes defined by principal axes, which are called column principal planes, can be plotted[26].

The absolute contribution of the variable j to the inertia of the column principal component α in the α^th column of Y is given by

$$c_{j\alpha} = \sum_{s \in M_j} s \in M_j f_{.s} y_{s\alpha}^2$$

where M_j is the set of categories of variable j. The relation between the absolute contribution $c_{j\alpha}$ and the correlation ratio between the variable j and the row standard component is given by

$$\eta_{j\alpha}^2 = \sum_{s \in M_j} \frac{n_s}{n} (\bar{x}^* - D)^2 = p \times c_{j\alpha},$$

Note that factor loadings for PCA are correlations between the variables and the components (the correlation ratios) are known as discrimination measures. More details for correspondence analysis can be found in different literatures[21,23-27].

Results

The application of multiple correspondence analysis is useful to visualize the associations between the socio-economic, demographic and geographic parameters and the malaria RDT result. Therefore, applying correspondence analysis helps to reduce the interaction parameters. Furthermore, the graphical interpretation of the data could be useful tool in an exploratory research and the reduction of the level of the associations between the investigated parameters.

For the applications of MCA, variables were divided into different subgroups that contain variables of similar types such as socio-economic, demographic and geographic variables. Variables analyzed with MCA generally are assumed to be categorical. This technique is described by Guitonneau and Roux[30]. To apply MCA to both continuous and discrete data, continuous variables could be categorized through a process of mutually exclusive and exhaustive discretization or coding [17]. Multiple correspondence analysis locates all the categories in a Euclidean space. To examine the associations among the categories, it is important to plot the first two dimensions of the Euclidean space. For the multiple correspondence analysis, malaria RDT result and the other socio-economic, demographic and geographic variables were considered. The demographic variables are sex, age and family size. For the multiple correspondence analysis, the continuous age and family size variables were recorded to be appropriate for the analysis.

The socio- economic variables are source of drinking water, time to collect water, toilet facility, availability of radio, television and telephone, construction material for room's floor, wall and roof, use of anti-mosquito spray, use of mosquito nets, total number of rooms in the house and total number of nets in the house. Besides the socio-economic and demographic variable, there were geographic variables included in the analysis. These variables are region and altitude. To be appropriate to MCA analysis, altitude has been recoded as categorical variable. Therefore, to perform the MCA analysis all socio- economic, demographic and geographic variables were included to the multiple correspondence analysis. The MCA analysis was performed using SAS 9.3 software.

In the MCA analysis, each principal inertia values expressed as a percentage of the total inertia. These values quantify the amount of variation accounted for by the corresponding principal dimension. In addition to this the principal inertia is decomposed into components for each of the rows and columns. The decomposed rows and columns provide the numerical contributions used to interpret the dimensions and the quality of display of each point in the reduced space. The parts which expressed as percentages are useful to explain the method of determination of the dimensions. The same parts of the dimensions can be expressed relative to the inertia of the corresponding points in the full space and this help to assess how close the individual points are to the dimension.

Table 1 presents inertia and Chi-Square decomposition for multiple correspondence analysis. Correspondence analysis employs chi-square distances to calculate the dissimilarity between the frequencies in each cell of a contingency table. The calculation of the chi-square distances is cell- independent. Pairs of cells whose observed and expected values are the same and can be considered to be independent of each other. Therefore, pairs of cells for observed and expected values are different. Table 1 suggested that the dimensions 1 and 2 account for 19.4% of the total association. The total chi-square statistic in Table 1, which is a measure of the association between the rows and columns in the full dimensions of the table, is 2,169,476 with degrees of freedom 2050. This chi-square represents all pairwise interactions among the factors. The maximum number of dimensions (or axes) is the minimum of the number of rows and columns, minus one.

Table 1: Inertia and Chi-Square Decomposition

Singular Value	Principal Inertia	Chi-Squar	Percent	Cumulative Percent	2 4 6 8 10 ----+----+----+----+----+---
0.42757	0.18282	232503	10.72	10.72	**************************
0.38438	0.14775	187901	8.66	19.38	**********************
0.3126	0.09772	124277	5.73	25.11	**************
0.29555	0.08735	111089	5.12	30.23	*************
0.28047	0.07866	100043	4.61	34.84	************
0.26462	0.07002	89054	4.1	38.94	**********
0.26193	0.06861	87250	4.02	42.97	**********
0.25503	0.06504	82716	3.81	46.78	**********
0.24806	0.06154	78259	3.61	50.39	*********
0.24591	0.06047	76909	3.55	53.93	*********
0.24557	0.06031	76696	3.54	57.47	*********
0.24356	0.05932	75444	3.48	60.94	*********
0.23959	0.0574	73005	3.37	64.31	********
0.23772	0.05651	71869	3.31	67.62	********
0.23474	0.0551	70079	3.23	70.85	********
0.23154	0.05361	68179	3.14	73.99	********
0.22675	0.05142	65388	3.01	77.01	********
0.22274	0.04961	63094	2.91	79.92	*******
0.21997	0.04839	61539	2.84	82.75	*******
0.21788	0.04747	60370	2.78	85.54	*******
0.2095	0.04389	55817	2.57	88.11	******
0.2031	0.04125	52458	2.42	90.53	******
0.1965	0.03861	49106	2.26	92.79	******
0.18357	0.0337	42856	1.98	94.76	*****
0.17417	0.03033	38578	1.78	96.54	****
0.16618	0.02761	35119	1.62	98.16	****
0.14744	0.02174	27646	1.27	99.44	***
0.08754	0.00766	9745	0.45	99.89	*
0.04423	0.00196	2488	0.11	100	
Total	1.70588	2169476	100		
Degrees of Freedom = 2025					

From Table 1, the singular value indicates the relative contribution of each dimension to an explanation of the inertia, or proportion of variation, in the participant and variable profiles. The singular values can be interpreted as the correlation between the rows and columns of the contingency table. As in principal components analysis, the first dimension explains as much variance as possible, the second dimension is orthogonal to the first and displays as much of the remaining variance as possible, and so on. Singular values of greater than 0.2 indicate that the dimension should be included in the analysis[25]. However, the proportion of variance explained by each dimension must be balanced with the cut-off point. The singular value and the inertia are directly related i.e., the inertia is an indicator of how much of the variation in the original data is retained in the dimensional solution[31]. Furthermore, the percentages of inertia accounted for by the first twelve axes are 10.7 per cent and 8.7 per cent, 5.73 per cent, 5.12 per cent, 4.61 per cent, 4.1 per cent, 4.02 per cent, 3.81 per cent, 3.61 per cent, 3.55 per cent, 3.54 per cent and 3.45 per cent, respectively (Table 1).

Based on this result, the first twelve axes as accounting for similar amounts of variance and would expect 39.1 per cent of the inertia to be accounted for by the remaining axes. As can be seen from the table, 93 per cent of the association can be represented well in twenty three dimensions. However, these data can be considered in just two dimensions. The first axis accounting for approximately 10.72 per cent of the inertia and the second axis accounts approximately 8.66 per cent. The percentages of inertia in MCA are low and tend to be close to one another and this latter fact might lead to an assumption that individual axes might be unstable.

Figure 1 presents the scree plot of singular values. One method to assess most appropriate number of dimensions for interpretation is using scree plot. The scree plot presents the proportions of variance explained[25]. As can be seen from the figure, the scree plot suggests that the proportion of variance explained drops faster up to 7th dimension and less rapidly up to dimension [26]. As discussed by Hair[25], 0.2 can be considered as a cut-off point as a first step. But, this cut-off point suggests that only 90.5 per cent variation can be explained by 22 dimensions. However, working 22 dimensions would not achieve the conceptual clarity for the use of correspondence analysis. But interpreting 22 dimensions is unnecessary. In literature for multidimensional scaling solutions, usually two or three dimensions are interpreted.

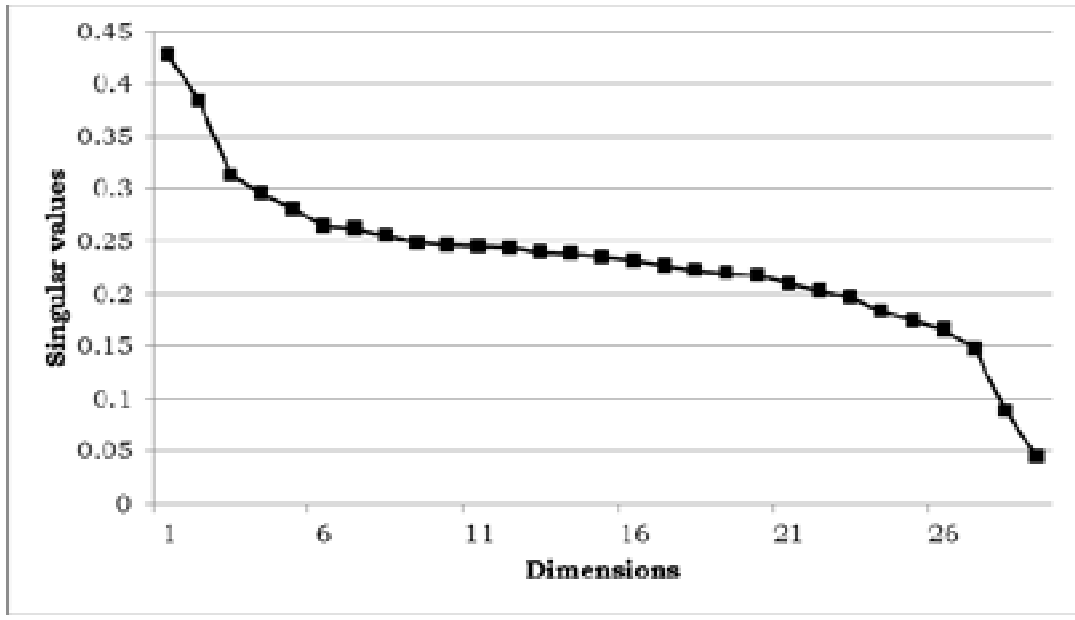
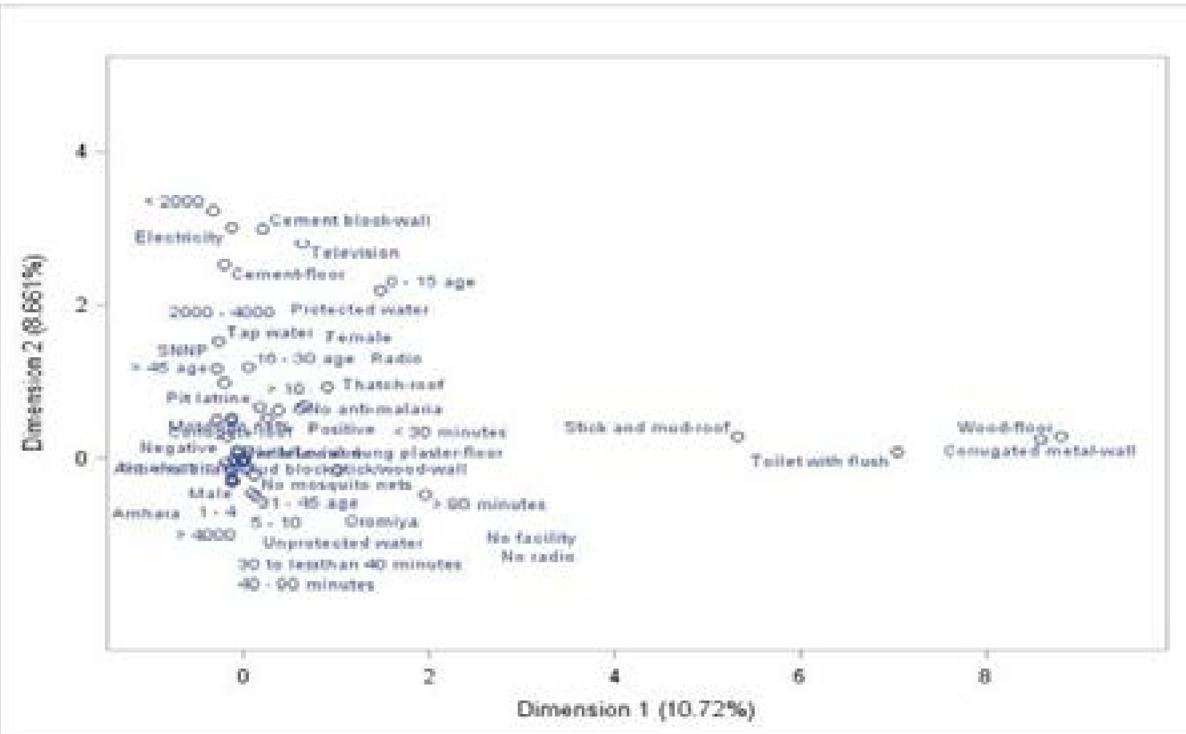


Figure 1: Scree plot of singular values

Figures 2a and 2b contain the multiple correspondence analysis scaling solution coordinates for the variables for twelve dimensions, with Dimension 1 on the horizontal axis and Dimension 2 on the vertical axis and so on.

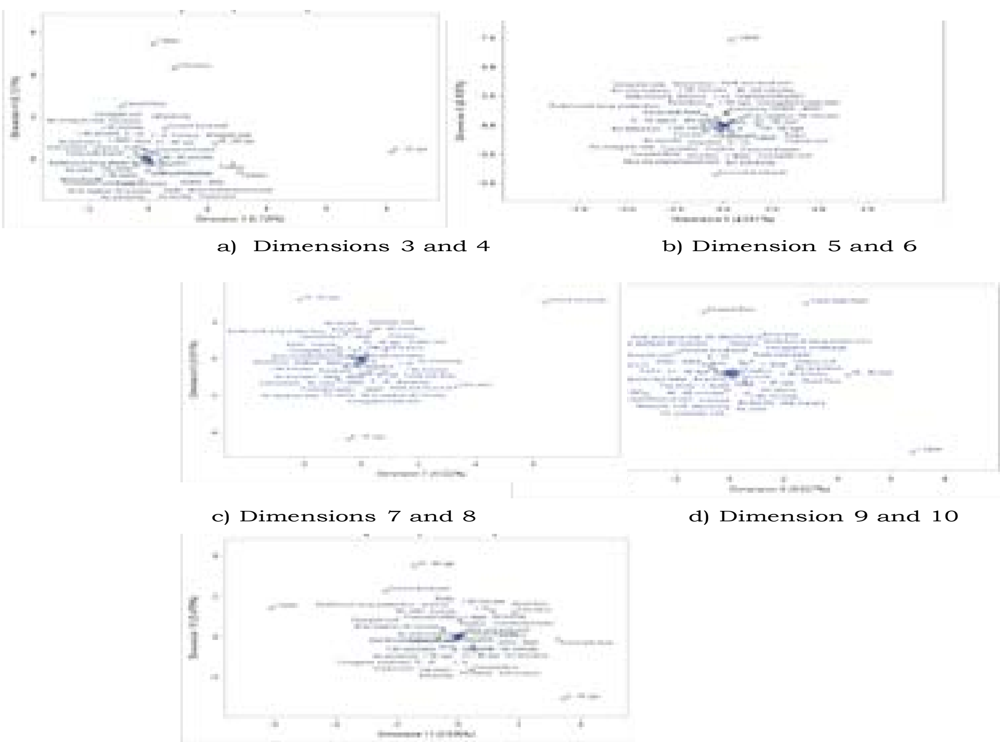
Figure 2a: Multiple correspondence analysis plot for dimensions 1 and 2



Multiple correspondence analysis locates all the categories in a Euclidean space. The first two dimensions of this space are plotted to examine the associations among the categories. Dimension 1 accounts for 10.72 per cent of the variance in the data and Dimension 2 accounts for 8.66 per cent of the variance (Figure 2a). The twelve dimensions totally accounts for 60.9 per cent of the variations. It can be seen that variable like stick and mud roof, toilet with flush, wood floor and corrugated metal wall appears separately in the right hand side of the chart. Therefore, these variables have to be included in the interpretation of dimension 1 and similarly for other dimensions.

It is important to note that this two-dimensional chart is part of the twenty two dimensional solutions. Interpreting of each dimension is considered as the contribution of variables to that dimension[32]. This is because a variable that appears on the two-dimensional chart might be a major contributor to another dimension but might not be located in the existing two-dimensional plane[33]. As can be seen in Figure 2a, the right quadrant of the plot (dimensions 1 and 2) shows that the categories stick and mud roof, toilet with flush, wood floor and corrugated metal wall are associated. To the top of the plot,altitude less than 2000 meter, use of electricity, cement block wall, cement floor, use of television, protected water, altitude between 2000–4000 meters are associated. On the other hand, positive malaria RDT result, not using anti-mosquito spray, thatch roof, earth or dung plaster floor are grouped together. Furthermore, negative malaria RDT result, use of anti-mosquito spray, use of malaria nets, pit latrine toilet and corrugated floor are associated. Similarly, unprotected water, 30–40 minutes walk to get water, no toilet facility and no radio are associated together. This interpretation of the plot is based on points found in approximately the same direction and in approximately the same region of the space.

Figure 2b: Multiple correspondence analysis plot for eleven dimensions



a) Dimensions 3 and 4 b) Dimension 5 and 6

c) Dimensions 7 and 8 d) Dimension 9 and 10

e) Dimensions 9 and 10

So far, the association between socio-economic, demographic, geographic variables and malaria RDT result was assessed based on dimension 1 and 2. As can be seen from Table 2 a, dimension 1 and 2 constitute 19.4 per cent of the variation. But, the other 20 dimensions all together constitute 71.2 per cent of the variation. Except the relationship between dimension 4 and 3, dimension 5 and 2, dimension 5 and 3 and dimension 7 and 1, the relationship between the variables for other combination of dimensions shows that it is located at the center of the graph. The relationship between variables shows similar relationship as of dimension 1 and 2.

Discussion and conclusion

In this study, multiple correspondence analysis was used as a way to graphically represent and interpret the relations between primary meanings in different malaria RDT result, socio-economic, demographic and geographic variables. Multiple correspondence analysis provides useful interpretative tools that can further the understanding of the conceptual context in which socio-economic, demographic and geographic variables by malaria RDT result occurs.

As it was discussed above, multiple correspondence analysis is a method for exploring associations between sets of categorical variables. Mathematically, it is a method for breaking down the value of the goodness-of-fit statistic into components due to the rows and columns of the contingency table. It can also be considered as a technique for assigned order to unordered categories. Therefore, the MCA approach involves defining a set of points, with associated masses, in a multidimensional space structured by Euclidean distance. Furthermore, the display is also thought of as a framework for reconstructing the original data as closely as possible. To display the relationship, the coordinate positions of the row and column points are used.

The association using MCA gives the relationship among coded variables and their associations. The technique allows the analysis of the relationships between the variables and different levels of one variable. Furthermore, the results of the analysis can be seen analytically and visually. This method of display gives detailed information of the relationship between variables and their associations. Therefore, the result from multiple correspondence analysis shows that there is association between malaria RDT result and different socio-economic, demographic and geographic variables. Moreover, there is an indication that some socio-economic, demographic and geographic factors have joint effects.

It is important to confirm the association between socio-economic, demographic and geographic factors using advanced statistical techniques. Therefore, future investigations need to be done to identify those variables that show significant relationships. By identifying those variables which could have joint effect, it is important to determine the principal axes and the identification of selection of variables to take forward for further analysis. In conclusion, the aim of the multiple correspondence analyses was to summarise the multidimensional data into an interpretable smaller dimensional factor and to reveal some association between different types of respondents. But, this reduction was not suitably achieved. This can be put down to either (i) all the factors being too scattered to be summarized in a smaller dimension, and/or (ii) the number of observations obtained in the cross tabulation being too small for all possible pairs of levels in the study.

Acknowledgements

We thank, with deep appreciation, The Carter Center of Ethiopia, for providing and giving permission to use the data for this study.

Ethical clearance
The ethical protocol received approval from the Emory University Institutional Review Board (IRB 1816) and Amhara, Oromiya and SNNP regional health bureaux. Informed consent was sought in accordance with the tenets of the declaration of Helsinki.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
DGA acquired the data, performed the analysis and drafted the manuscript. TTZ and HGM designed the research. All authors discussed the results and implications and commented on the manuscript at all stages. All authors contributed extensively to the work presented in this paper. All authors read and approved the final manuscript. All authors contributed extensively to the work presented in this paper.

References

1.WHO.Malaria:http://www.who.int/mediacentre/factsheets/fs094/en/ 2012 [cited 2012 May].
2. Snow, R.W., N. Peshu, and D. Forster, Environmental and entomological risk factors for the development of clinical malaria among children on the Kenyan coast. Transactions of the royal society of tropical medicine and hygiene, 1998. 92: p. 381 - 385.
3. Adugna, A., Malaria in Ethiopia http://www.ethiodemographyandhealth.org/ MedVectoredDiseasesMalaria.pdf. Accessed Nov. 14, 20112011, Addis Ababa, Ethiopia: Ethiopian Demography and Health.
4. Federal Ministry of Health (FMH), Malaria and Other Vector-borne DiseasesControl Unit, 1999: Addis Ababa, Ethiopia.
5. FMH, Federal Ministry of Health: Malaria: Diagnosis and Treatment Guidelines for Health Workers in Ethiopia. Addis Ababa, Ethiopia, Federal democratic Republic of Ethiopia, Ministry of Health, 2004.
6. FMH, National five-year strategic plan for malaria prevention and control in Ethiopia 2006 – 2010, 2006.
7. WHO, World Health Organization: Systems for the early detection of malaria epidemics in Africa: an analysis of current practices and future priorities, country experience. Geneva, Switzerland, World Health Organization, 2006.
8. Zhou, G., et al., Association between climate variability and malaria epidemics in the East African highlands. Proceedings of the National Academy of Sciences 2004. 101: p. 2375-2380.
9. Adhanom, T., et al., Malaria. In The Eipdemiology and Ecology of Health and Disease in Ethiopia 1st edition, in Edited by: Berhane Y, Hailemariam D, Kloos H and Shama PLC. Addis Ababa, Ethiopia 2006. p. 556-576.
10. Tulu, N.A., Malaria. In The Ecology of Health and Disease in Ethiopia 2nd edition. Edited by: Kloos H and Zein AZ. Boulder, USA, Westview Press Inc.1*993: p. 341-352*.
11. Ayele, D.G., T. Zewotir, and H. Mwambi, Prevalence and risk factors of malaria in Ethiopia. *Malaria Journal,* 2012. 11:195 p. doi:10.1186/1475-2875-11-195.
12. Emerson, P.M., et al., Integrating an NTD with One of "The Big Three": Combined Malaria and Trachoma Survey in Amhara Region of Ethiopia. PloS Neglected tropical diseases, 2008. 2(8).
13. Shargie, E.B., et al., Malaria prevalence and mosquito net coverage in Oromia and SNNP regions of Ethiopia. *BMC Public Health*, 2008. 8:321.
14. WHO, Malaria Rapid Diagnostic Test Performance, 2009: Geneva, Switzerland.
15. Wongsrichanalai, C., et al., A Review of Malaria Diagnostic Tools: Microscopy and Rapid Diagnostic Test (RDT). *The American Journal of Tropical Medicine and Hygiene,* 2007. 77 (Suppl 6): p. 119-127.
16. Benzécri, J.-P., Analyse des Donnees Vols 1 and 2 1973 Dunod.
17. Greenacre, M.J., Theory and Applications of Correspondence Analysis 1984: Academic Press.
18. Gifi, A., Nonlinear Multivariate Analysis1990, New York: John Wiley & Sons.
19. Hayashi, C., Multidimensional quantication{with applications to analysis of social phenomena. Annals of the Institute of Statistical Mathematics, 1954. 5(2) p. 121-143.
20. Nishisato, S., Analysis of categorical data: Dual Scaling and its applications1980, Toronto: University of Toronto Press.
21.Greenacre, M.J. and J. Blasius, Multiple correspondence analysis and related methods 2006, Boca Raton: Chapman & Hall/CRC.
22. Tenenhaus, M. and F.W. Young, An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. Psychometrika, 1985. 50: p. 91-119.
23. Hill, M.O., Correspondence Analysis: A Neglected Multivariate Method. J*ournal of the Royal Statistical Society,* 1974. 23 (3): p. 340-354.
24. Van der Heijden, P.G.M. and J. de Leeuw, Correspondence analysis used complementary to loglinear analysis. Psychometrika, 1985. 50 (4) p. 429-447.
25. Hair, J.F., et al., Multivariate Data Analysis. Vol. 4. 1995, Upper Saddle River, NJ: Prentice Hall.
26. Johnson, R.A. and D.W. Wichern, Applied multivariate statistical analysis.Vol. 6. 2007, New Jersey: Pearson/Prentice Hall.
27. Tian, D., S. Sorooshian, and D.E. Myers, Correspondence Analysis with MATLAB. Computers & Geosciences 1993. 19 (7): p. 1007-1022.
28. Askell-Williams, H. and M.J. Lawson, A Correspondence Analysis of Child- Care Students' and Medical Students' Knowledge about Teaching and Learning. International Education Journal, 2004. 5 (2).
29. Benzécri, J.-P., Analyse des Donnkes, Tome 2: Analyse de Correspondences1973, Paris: Dunod.
30. Guitonneau, G.G. and M. Roux, Sur la Taxinomie du Genre Erodium. Cahiers de l'Analyse des Donnees, 1977. 2 (1): p. 97-113.
31. Bendixen, M. A practical guide to the use of correspondence analysis in marketing research. Marketing Research On-Line, 2003, 16-38. http://mro.massey.ac.nz/ca.html. 1996 [cited 2011 Oct 03].
32. Clausen, S.-E., Applied Correspondence Analysis. An Introduction1998, Thousand Oaks, CA: Sage.
33. Nishisato, S., Elements of dual scaling: An introduction to practical data analysis 1994, Hillsdale, NJ: Erlbaum.