Asian Journal of Probability and Statistics

2(3): 1-13, 2018; Article no.AJPAS.46007

Assessment of Under Five Child Mortality in Tanzania Mainland Based on Principal Component Analysis

Privatus Christopher[1*]

[1] *Department of ICT and Mathematics, College of Business Education, P.O.Box 1968, Dar es salaam, Tanzania.*

Author's contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/AJPAS/2018/v2i328788

Editor(s):

(1) Dr. S. M. Aqil Burney, Department of Computer Science, University of Karachi, Pakistan.

(2) Dr. Jiteng Jia, School of Mathematics and Statistics, Xidian University, China.

Reviewers:

(1) Nalini V. Vaidya, G. H. Raisoni College of Engineering Nagpur, India.

(2) Aliyu Bhar Kisabo, Nigeria.

(3) Madalina Ecaterina Popescu, The Bucharest University of Economic Studies, Romania.

Complete Peer review History: http://www.sciencedomain.org/review-history/28108

Received: 04 October 2018

Accepted: 22 December 2018

Published: 03 January 2019

Original Research Article

Abstract

Deaths of children younger than 5 years has been a global problem for long time. This study is focused on evaluating diseases that caused under five child mortality in Tanzania in 2013. Diseases that causes child mortality were collected from 25 regions and analysed for 42 disease variables. The data obtained were standardized and subjected to principal component analysis (PCA) to define the diseases responsible for the variability in child mortality. PCA produced seven significant main components that explain 73.40% of total variance of the original data set. The results reveal that Thyroid Diseases, Snake and Insect Bites, Vitamin A Deficiency /Xerophthalmia, Eye Infections, Schistosomiasis (SS), Intestinal Worms, Ear Infections, Haematological Diseases, Diabetes Mellitus, Ill Defined Symptoms no Diagnosis, Poisoning, Anaemia, HIV/AIDS, Burns, Rheumatic Fever, Bronchial Asthma, Peri-natal conditions and Urinary tract infection are most significant diseases in assessing under five child mortality in Tanzania mainland. This study suggest that PCA technique is useful tool for identification of important diseases that causes death of children less than five years.

Keywords: Principle component analysis; variance; child mortality; disease.

**Corresponding author: E-mail: p.ntare@cbe.ac.tz*

1 Introduction

1.1 Background

Reduction of child mortality has been the global focus for many years. It started more than three decades ago where substantial political, donor and country focus on improvement of child survival[1]. According to [2] child mortality has been declining globally due to socialeconomic development and implementation of child survival interventions even though 8.8 million children younger than 5 years die annually. This indicate that child mortality is still a global challenge. In year 2000 the United Nation (UN) established eight international development goals and among others the fourth goal (MDG4) captured the intention of high level leaders since it targeted the reduction of child mortality by two third from 1990 to 2015 [3]. Worldwide under five mortality rate has declined more than half from 90 to 43 death per $1,000$ live birth between 1995 and 2015, this implies that there is a decline from 12.7 million death in 1990 to 6 million death in 2015 [4].

There are many global and local effort on the improvement of child survival. The global campaign such as child survival revolution, heath for all by the year 2000, and Jim Grant pioneering work at UNICEF and child intervention are example of international effort on improvement of child survival [5, 1]. It is widely accepted that global community have failed to achieve MDG4 since many countries especially from south Asia and Sub-Saharan Africa are not on track to meet the target [6, 1]. In Africa six million children younger than 5 years die every year due to infectious diseases [7]. Children die before discharge or after discharge (post discharge) and because of limited data available especially in developing countries, it is possible that many death are not recorded.

In order to provide the important input to global, national and regional debate on child mortality is important to have clear evaluation of global and local trend of child mortality. In Tanzania the largest gain in child survival occurred between the year 2000 and 2012 when the mortality of children under 5 years decrease at average rate of 8.5% per year [8]. This decrease enables Tanzania to achieve millennium development goal(MDG4) for child survival. In 2014, as the reference goals for the international development community for the period of 2015 to 2030, United Nations member states proposed the new set of Sustainable Development Goal(SDGs) which will achieve the Millennium Development goals (MDGs) [9].

Total of 17 SDGs were proposed but Goal 3 which is focusing on ensuring healthy lives and promote the well being for all at all ages still captured the intention of many Nations. The second objective of Goal 3 is by the end of 2030 to end preventable deaths of newborns and children under 5 years with the aim to reduce the neonatal mortality to at least 12 per 1,000 live births and under 5 mortality to at least as 25 per 1,000 births [10]. In order for Tanzania achieve the above objective and make coherent policies and strategies, it is important to understand diseases that contribute much to death of children under 5 years. This will provide the evidences and support in decision making about the disease priorities. Therefore this paper contributes to these accountability efforts and seeks to use the Principal Component Analysis to assess and analyse diseases that contribute to high mortality rate for children under 5 years in Tanzania.

1.2 Literature Review

In recent years, many studies have been done using different multivariate statistical techniques such as Principal Component Analysis (PCA), Cluster Analysis (CA), analysis of variance (ANOVA), discriminant analysis (DA), factor analysis (FA) and multiple regression analysis (MRA) in analysing and interpreting diseases that contribute to mortality rate for children under 5 years. [7] used univariate logistic regression to determine level of association between acute infectious diseases that

leads to post-discharge mortality.[11] employed the cox regression models to identify risk factors that lead to post discharge mortality and survival probability in children less that 15 years.[12] employed the descriptive statistics study the post- discharge mortality for children with severe malnutrition and pneumonia in Bangladesh. [13] employed the systematic review to identify all studies reporting post- discharge mortality in the children and estimate the likelihood and most risk factors for death. [14] employed the regression models to examine the relationship between travel time,cause -specific hospitalization rates and probability of death in hospital. Few mentioned literature have proved that different statistics techniques have been successfully applied to study mortality rate for children of different ages. Therefore this study employed the Principal Component Analysis to study diseases that contribute to high mortality rate for children under 5 years in Tanzania.

2 Methods and Data

2.1 Data Source and Software Used

Number and causes of death under 5 years data for 2013 were downloaded from the government open data portal for Tanzania National Bureau of Statistics (NBS) on May 12, 2018 [15]. The data set consists of 26 regions from Tanzania mainland which include Arusha, Dodoma, Dar es Salaam,Geita, Iringa, Kagera, Katavi, Kigoma, Kilimanjaro, Lindi, Manyara, Mara, Mbeya, Morogoro, Mtwara, Mwanza, Njombe, Pwani, Rukwa, Ruvuma, Shinyanga, Simiyu, Singida, Tabora and Tanga. The variables were 42 diseases which includes Acute Respiratory Infections (ARI), Diarrhoeal Diseases (DD), Intestinal Worms (IW), Malaria- Severe Complicated (MSC), Malaria-Uncomplicated (MU), Schistosomiasis (SS), Tuberculosis (TC), Severe Protein Energy Malnutrition (SPE) . Other Nutritional Disorders (OND), Sickle cell Disease (SCD), Anaemia (AN), Epilepsy (EP), Ear Infections (EI), Eye Infections (EY), Other Eye Diseases (OED), Vitamin A Deficiency /Xerophthalmia (VAX), Cardiac Failure (CF), Other Cardiovascular (Ocr), Rheumatic Fever (RF), Bronchial Asthma (BA), Pneumonia(PN), Respiratory Disease (RD), Non-Infectious Gastrointestinal Diseases (NGD), urinary tract Infections (UTI), Non-Infectious Kidney Diseases (NKD), Skin Infections (SI), Skin Diseases Non-Infectious(SDI), Joint Disorders (JD), Peri-natal Conditions (PNC), Snake and Insect Bites (SIB), Burns (BN), Poisoning (PO),HIV/AIDS (HA), Neoplasms (NP), Haematological Diseases (HDA) Osteomyelitis (OM), Congenital Diseases(CD), Fractures/ Dislocation(CD), Ill Defined Symptoms no Diagnosis(DSD), Thyroid Diseases(TD), Diabetes Mellitus (DM) and other diseases (OT). Results in this study were obtained using the statistical software environment R [16]. For clear visualization of results the package RColorBrewer [17] and FactoMineR [18] were used.

2.2 Principal Component Analysis

Principal Component Analysis is one of the multivariate methods widely used for extracting the linear relationship among a set of variables in the data set [19]. It extract low dimensional set of features from a high dimensional data set with the aim of capturing much information as possible. The basic idea of PCA is to describe the variation of a set of of uncorrelated variables in which each is a particular linear combination of the original variable [20]. Using PCA the decreasing order of importance derive a new variables, this imply that the maximum variation in the original dataset is much explained by the first principal component (PC1). The second principal component (PC2) will explain as much as possible the remaining variation under condition that it is uncorrelated with the PC1, and this procedure continues. The Principal components can be expressed using the following equation

$$Z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \cdots + a_{ip}x_{pj}, \tag{2.1}$$

where Z is the component score, a is the component loading, x is the measured value of a variable, i is the component number, j is the sample number, and p is the total number of variables.

Since the main objective of using the PCA is to find few components that can explain better the variation in the original data. There are many techniques used to choose the number of PCs to be retained but this study used the Kaiser's rule which retain the PCs with eigenvalues greater than 1. The final number of PCs retained were determined by scree plot. This reduces the number of variables in which each linear combination will correspond to a principal component (PC).

3 Results and Discussion

3.1 Data Description

The minimum (Min), first quartile (Q1), median(Med), Mean, third quartile(Q3), maximum (Max), skewness, Kurtosis and standard deviation (Sd) values of each disease under study are presented in Table 1.

Table 1. Descriptive statistics of the diseases under study

	Min	Q1	Med	Mean	Q3	Max	skewnes	kurtos	Sd
ARI	0.00	1.00	9.00	14.20	18.00	69.00	1.50	1.60	18.00
DD	8.00	14.00	26.00	32.00	42.00	104.00	1.30	1.30	23.80
IW	0.00	0.00	0.00	0.50	0.00	11.00	4.20	16.90	2.20
MSC	31.00	66.00	161.00	182.40	287.00	414.00	0.40	-1.20	124.50
MU	0.00	2.00	6.00	18.80	26.00	144.00	2.70	8.00	30.80
SS	0.00	0.00	0.00	3.10	0.00	66.00	4.30	17.30	13.30
TC	0.00	0.00	1.00	4.20	4.00	48.00	3.80	14.80	9.50
SPE	0.00	5.00	11.00	13.00	19.00	45.00	1.10	0.80	11.30
OND	0.00	0.00	0.00	2.80	7.00	13.00	1.00	-0.10	3.80
SCD	0.00	0.00	1.00	1.50	2.00	10.00	2.10	4.10	2.40
AN	8.00	25.00	68.00	89.60	135.00	287.00	0.90	-0.20	77.10
EP	0.00	0.00	0.00	0.10	0.00	1.00	2.90	6.80	0.30
EI	0.00	0.00	0.00	0.40	0.00	8.00	4.00	15.30	1.60
EY	0.00	0.00	0.00	0.40	0.00	9.00	4.40	18.20	1.80
OED	0.00	0.00	0.00	0.00	0.00	0.00			0.00
VAX	0.00	0.00	0.00	1.40	0.00	34.00	4.40	18.20	6.80
CF	0.00	0.00	0.00	5.40	1.00	112.00	4.40	18.00	22.30
Ocr	0.00	0.00	0.00	4.20	2.00	44.00	2.90	7.80	10.10
RF	0.00	0.00	0.00	0.10	0.00	2.00	4.40	18.20	0.40
BA	0.00	0.00	0.00	1.40	2.00	7.00	1.40	0.60	2.20
PN	20.00	49.00	74.00	92.60	124.00	237.00	1.00	0.00	58.50
RD	0.00	0.00	1.00	3.50	3.00	26.00	2.30	4.30	6.60
NGD	0.00	0.00	1.00	2.20	4.00	10.00	1.20	0.10	3.00
UTI	0.00	0.00	1.00	4.00	2.00	45.00	3.30	11.10	9.50
NKD	0.00	0.00	0.00	0.20	0.00	3.00	3.80	14.20	0.60
SI	0.00	0.00	0.00	0.40	0.00	6.00	3.40	11.40	1.30
SDI	0.00	0.00	0.00	0.00	0.00	0.00			0.00
JD	0.00	0.00	0.00	0.10	0.00	2.00	3.40	11.00	0.40
PNC	0.00	2.00	12.00	38.70	27.00	517.00	4.10	16.30	102.10
SIB	0.00	0.00	0.00	2.80	1.00	55.00	4.30	17.30	11.00
BN	0.00	3.00	6.00	6.00	9.00	20.00	1.00	1.40	4.50
PO	0.00	0.00	1.00	1.80	3.00	7.00	1.00	-0.20	2.10
HA	0.00	4.00	8.00	14.00	16.00	82.00	2.20	5.10	18.40
NP	0.00	0.00	0.00	0.20	0.00	3.00	3.80	14.20	0.60
HDA	0.00	0.00	0.00	0.00	0.00	1.00	4.40	18.20	0.20
OM	0.00	0.00	0.00	0.60	0.00	5.00	2.50	4.80	1.40
CD	0.00	0.00	0.00	1.20	1.00	12.00	2.70	6.70	2.80
FD	0.00	0.00	0.00	1.30	1.00	9.00	1.90	3.00	2.30
DSD	0.00	0.00	0.00	4.40	2.00	66.00	3.90	14.80	13.40
TD	0.00	0.00	0.00	0.20	0.00	4.00	4.10	15.80	0.80
DM	0.00	0.00	0.00	0.50	0.00	9.00	4.00	15.50	1.80
OT	0.00	10.00	16.00	36.50	37.00	264.00	2.70	7.80	56.40

3.2 Compositional Relation

The correlation between diseases can give more insight on the relationship between different diseases as shown in Fig. 1

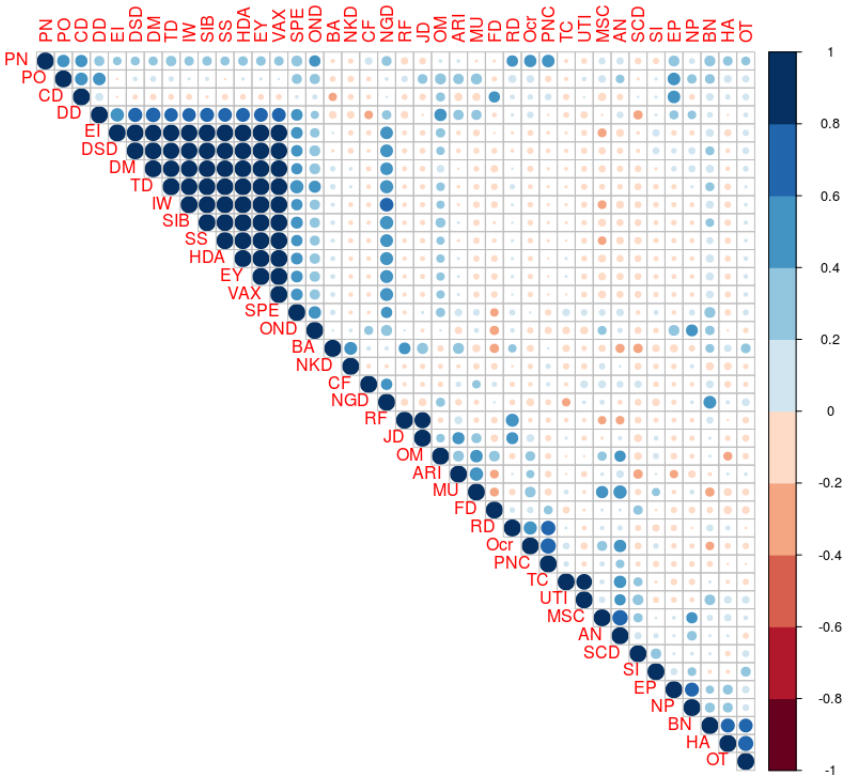

Fig. 1. Correlation matrix of diseases

The compositional relations in Table 1 show that the strong correlation (> 0.8) exist between (DSD,EI), DM with EI and DSD, TD with EID, SD and DM, IW with EI,DSD,DM and TD, SIB with EI, DSD, DM, TD and IW. SS with EI, DSD, DM, TD, IW and SIB. HAD with EI, DSD, DM, TD, IW, SIB and SS. EY with EI, DSD, DM, TD, IW, SIB, SS and HDA. VAX with EI, DSD, DM, TD, IW, SIB, SS,HDA and EY.

Also, moderate correlations (> 0.6) exist between (CD,PN),(CD,PO),DD with PO, EI, DSD, DM, TD, IW, SIB, SS, HDA and VAX, (EP,PO),(EP,CD),(PN,Ocr),(BN,NGD),(HA,BN)and (OT,HA). At this stage is difficult to group the diseases into components and attach the physical significance. Therefore correlation matrix in Fig. 1 has been applied to the principle component analysis.

3.3 Principal Component Analysis

PCA was applied to dataset with $n \times p$ matrix, where $n = 25$ is the number of regions in Tanzania mainland and $p = 42$ is the number of variables (diseases). This resulted to 25×42 matrix. Table

2 show the eigenvalue of each PC, percent and cumulative percent of the variance.

Taking into account the Kaiser's criteria, the PCA algorithms resulted into 12 PCs with eigenvalues greater than 1 as shown in Table 2. Then rotated component matrix of first 12 PCs are shown in Table 3.

Due to standardization, all principal components (PCs) have mean zero, the standard deviation is also given for each of the components and it is the square root of eigenvalue. The purpose is to find the correlation between the principal components and the original variables. The final number of PCs retained was determined by the analysis of Scree graph in Fig. 2.

From Fig. 2, we observe that the scree plot descends more rapidly towards the 8^{th} PC before levelling out. The first 7 PCs were retained for for subsequent analysis and accounted for the 73.4% of the total variation of children death variables. Ignoring of other PCs could results in discording of some important information but the aim of the study was to explain the maximum variation in the data. Variables correlating with each of the 7 PCs are indicated in Table 4.

Table 4 shows that PC1 was positively high correlated with ten variables. PC2, PC3 and PC4 were positively high correlated with two variables while PC5 and PC6 were highly correlated with one variable. Since PC1 was the component with high variance (28.05%), explaining more relevant information, it was plotted against the other six components to provide detailed information about the disease effects on death of children under 5 years in Tanzania.

Table 2. Eigenvalues, percentage variance and percentage cumulative variance for PCs

PCs	Eigenvalue	% variance	% cumulative variance
PC1	11.22	28.05	28.05
PC2	4.09	10.22	38.26
PC3	3.52	8.79	47.06
PC4	3.16	7.90	54.96
PC5	2.78	6.96	61.91
PC6	2.53	6.32	68.23
PC7	2.07	5.17	73.40
PC8	1.77	4.44	77.83
PC9	1.70	4.24	82.07
PC10	1.47	3.67	85.74
PC11	1.25	3.12	88.86
PC12	1.10	2.75	91.61
PC13	0.87	2.17	93.78
PC14	0.67	1.67	95.45
PC15	0.53	1.32	96.77
PC16	0.39	0.97	97.74
PC17	0.24	0.59	98.33
PC18	0.21	0.53	98.87
PC19	0.18	0.44	99.31
PC20	0.12	0.30	99.61
PC21	0.07	0.18	99.79
PC22	0.05	0.12	99.91
PC23	0.03	0.06	99.97
PC24	0.01	0.03	100.00
PC25	0.00	0.00	100.00

Table 3. Rotated component matrix of first 12 PCs

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
ARI	-0.0028	0.1464	-0.2744	-0.1838	-0.2430	-0.0320	-0.1189	-0.1109	-0.2733	-0.0541	0.1081	0.1369
DD	-0.2051	0.1977	-0.0771	0.0246	-0.0170	-0.2079	-0.1437	0.0624	-0.1126	-0.1453	0.0114	-0.1168
IW	-0.2903	-0.0651	-0.0300	0.0133	0.0130	0.0287	-0.0056	0.0222	0.0037	-0.0118	0.0498	-0.0155
MSC	0.0468	0.3049	-0.0263	0.1635	-0.1135	0.1266	0.2985	0.0349	-0.0387	0.1924	-0.1304	-0.2248
MU	0.0234	0.2766	-0.2857	0.1015	-0.2602	-0.1070	0.0419	-0.0235	-0.1102	0.0369	0.1878	0.0919
SS	-0.2897	-0.0650	-0.0404	0.0200	0.0245	0.0226	-0.0186	0.0372	0.0004	-0.0045	0.0248	-0.0415
TC	0.0172	0.0685	-0.0625	0.1918	-0.0704	0.4049	-0.2137	0.1557	0.1746	-0.3360	-0.0334	0.0922
SPE	-0.1578	0.1705	-0.0083	-0.0107	-0.1155	0.0958	0.0217	-0.0881	0.0974	-0.0899	-0.1213	0.2683
OND	-0.1268	0.1607	0.1292	-0.0552	-0.0743	-0.1076	0.3512	-0.0618	0.3259	-0.0375	0.0429	0.1086
SCD	0.0478	-0.0011	0.1271	0.2549	-0.0151	0.1718	-0.1483	0.0941	0.1936	0.3568	-0.1216	0.2430
AN	0.0391	0.3480	-0.1403	0.2309	-0.0771	0.2055	0.1061	0.0860	0.0013	-0.0242	-0.1109	-0.1416
EP	-0.0036	0.2086	0.2948	-0.0423	0.0323	-0.2886	-0.0088	0.1707	0.1927	-0.1245	-0.0020	0.0377
EI	-0.2855	-0.0627	-0.0200	0.0467	0.0211	0.0107	-0.0894	0.0453	-0.0129	0.1161	-0.0461	0.0727
EY	-0.2940	-0.0475	-0.0358	0.0120	0.0112	0.0360	-0.0114	0.0390	-0.0023	-0.0057	0.0182	-0.0104
VAX	-0.2936	-0.0484	-0.0373	0.0135	0.0096	0.0346	-0.0105	0.0412	-0.0053	-0.0078	0.0225	-0.0152
CF	0.0100	0.0493	0.0256	0.0637	-0.0530	0.1472	0.2486	-0.4204	0.1132	0.2282	0.2748	0.2491
Ocr	0.0236	0.2635	-0.2636	-0.0076	0.3230	0.0882	0.0528	0.1443	-0.1519	0.0569	0.0983	0.0567
RF	0.0193	-0.0388	-0.1363	-0.4124	-0.0705	0.0754	-0.1365	0.0361	0.3731	0.1755	-0.0066	-0.1166
BA	-0.0151	-0.0951	-0.0367	-0.4020	-0.1308	0.1560	0.1315	-0.0776	-0.1621	0.0028	-0.3378	0.1310
PN	-0.1023	0.2917	0.1133	-0.1048	0.3023	0.0394	0.0860	-0.0926	-0.0490	-0.1019	0.1194	0.2590
RD	-0.0046	0.1339	-0.1309	-0.3060	0.3200	0.2098	0.0265	0.1244	0.1681	0.1240	0.0084	-0.0602
NGD	-0.1845	0.0258	0.1186	0.0268	-0.0400	0.0815	0.1600	-0.4010	-0.0388	0.1429	0.0670	-0.0978
UTI	0.0313	0.0696	0.0987	0.1592	-0.1311	0.4457	-0.2423	-0.0745	0.1458	-0.2171	-0.1012	-0.0175
NKD	0.0175	-0.0828	-0.0426	-0.0835	-0.0207	0.0052	0.3427	0.0320	-0.1177	-0.1443	-0.6429	0.1832
SI	0.0112	0.0604	0.0343	0.1009	-0.1029	-0.1151	-0.1715	0.3108	-0.0786	0.4662	-0.0707	0.4179
JD	0.0223	0.0846	-0.2653	-0.3539	-0.1495	0.0258	-0.1662	0.0353	0.2481	0.1816	-0.0071	-0.0734
PNC	0.0111	0.1062	-0.1030	-0.0453	0.4873	0.2039	0.1264	0.1243	-0.1308	0.0265	0.0281	0.0365
SIB	-0.2942	-0.0383	-0.0352	0.0270	0.0219	0.0211	-0.0383	-0.0220	0.0018	0.0160	-0.0409	-0.0414
BN	-0.0821	0.1571	0.3261	-0.1823	-0.0921	0.2304	-0.0881	-0.1848	-0.0137	0.1086	-0.0031	-0.1343
PO	-0.0344	0.3701	0.0383	-0.1278	-0.0398	-0.1859	-0.1886	-0.0914	0.1244	-0.1308	-0.0839	0.1133
HA	0.0193	0.1318	0.3375	-0.1931	-0.0517	0.1311	-0.0892	0.0980	-0.3018	-0.0225	0.1281	-0.2743
NP	-0.0104	0.2183	0.2238	0.0038	-0.1256	-0.1344	0.2227	0.3100	0.1493	0.1061	-0.0971	-0.2587
HDA	-0.2940	-0.0475	-0.0358	0.0120	0.0112	0.0360	-0.0114	0.0390	-0.0023	-0.0057	0.0182	-0.0104
OM	-0.0775	0.2120	-0.2368	0.1161	-0.0657	-0.1386	-0.1868	-0.2580	-0.1122	0.0951	-0.2867	-0.2105
CD	0.0056	0.1539	0.1668	-0.0157	0.2867	-0.2189	-0.2822	-0.2552	0.0901	-0.1283	-0.2037	0.1663
FD	0.0338	-0.0183	-0.0263	0.1776	0.3055	-0.0676	-0.1706	-0.2625	0.0539	0.3059	-0.2709	-0.2770
DSD	-0.2874	-0.0152	0.0274	0.0432	-0.0185	-0.0158	-0.0237	0.1169	0.0151	0.1087	0.0004	-0.0022
TD	-0.2914	-0.0096	-0.0278	0.0058	0.0048	0.0427	0.0632	0.0176	0.0402	-0.0125	-0.0484	0.0090
DM	-0.2881	-0.0173	0.0030	0.0047	-0.0208	-0.0062	0.0304	0.1232	0.0000	-0.0138	-0.0380	-0.0347
OT	-0.0292	0.0943	0.3148	-0.1400	-0.0722	0.1425	-0.1672	0.0640	-0.4122	0.1860	-0.0424	0.1092
EV	11.22	4.09	3.52	3.16	2.78	2.53	2.07	1.77	1.70	1.47	1.25	1.10
%CV	28.05	10.22	8.79	7.90	6.96	6.32	5.17	4.44	4.24	3.67	3.12	2.75
C%V	28.05	38.26	47.06	54.96	61.91	68.23	73.40	77.83	82.07	85.74	88.86	91.61

Table 4. Variables correlating with each of the seven PCs

PC1	Cor	PC2	Cor	PC3	Cor	PC4	Cor	PC5	Cor	PC6	Cor	PC7	Cor
SIB	0.9855	PO	0.7482	HA	0.6330	RF	0.7329	PNC	0.8127	UTI	0.7086	CD	0.4056
EY	0.9847	AN	0.7035	BN	0.6116	BA	0.7145	Ocr	0.5387	TC	0.6438	MSC	-0.4290
HDA	0.9847	MSC	0.6163	OT	0.5903	JD	0.6289	RD	0.5337	EP	-0.4588	NKD	-0.4926
VAX	0.9835	PN	0.5896	EP	0.5529	RD	0.5437	FD	0.5095		OND	-0.5048	
TD	0.9758	MU	0.5590	NP	0.4197	AN	-0.4104	PN	0.5042				
IW	0.9722	Ocr	0.5327	OM	-0.4440	SCD	-0.4530	CD	0.4781				
SS	0.9702	NP	0.4412	Ocr	-0.4943			ARI	-0.4052				
DM	0.9651	OM	0.4285	JD	-0.4976			MU	-0.4339				
DSD	0.9296	EP	0.4216	ARI	-0.5146								
EI	0.9561	DD	0.3996	MU	-0.5359								
DD	0.6869												
NGD	0.6179												
SPE	0.5284												
OND	0.4246												

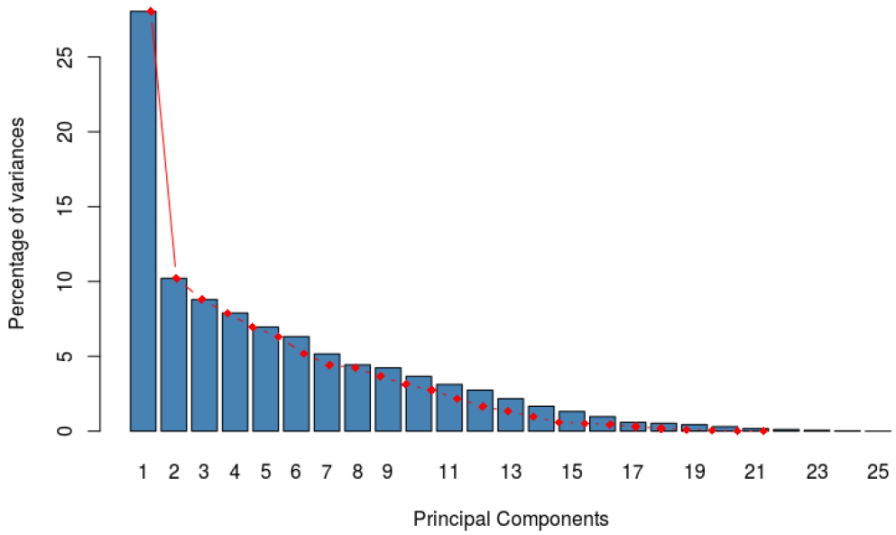

Fig. 2. Scree graph showing percentage of variance explained by each of the first 25 principal components

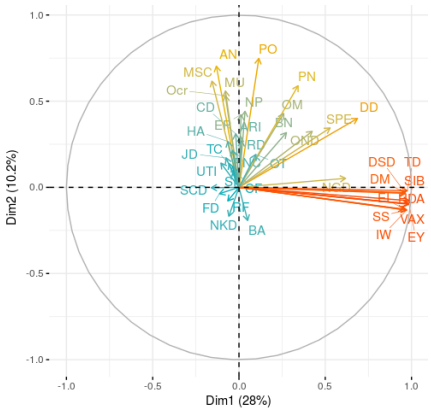

Fig. 3. Biplot relating PC1 with PC2

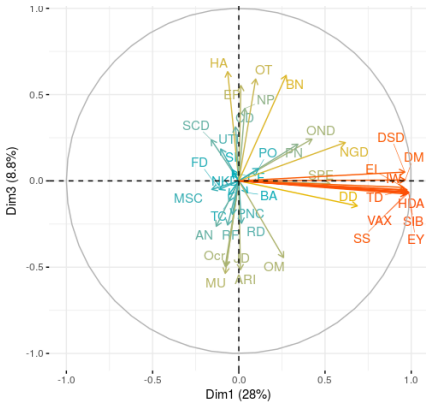

Fig. 4. Biplot relating PC1 with PC3

Biplot results of PC1 with PC2

The biplot presented in Fig. 3 shows the correlation between variables of PC1 and PC2. This interaction present some high correlated variables such as: (a) BN with OM and PN: this correlation, located in the first quadrant, indicates that when death caused by BN increases, OM and PN also increases. (b) DD with OND and SPE: this correlation is located in the first quadrant and indicates that child mortality due to DD was related with increased in OND and SPE. (c) PO with NP, RD and ARI: this correlation is located in the first quadrant and indicates that child mortality due to PO was related with NP, RD and ARI (6 and 24). (d) MU with Ocr, EP and CD: This interaction is located in the second quadrant of the biplot and suggests that an increase in MU is related with an increase in Ocr, EP and CD. (e) MSC with HA: The child mortality due to MSC is associated

with HA. (f) TD with SIB, VAX, EY, SS, IW, EI, HDA, DM and DSD: this correlation is located in the fourth quadrant and have a strong positive correlation with PC1, which indicates that higher child mortality is caused by these diseases.

In Fig. 3 one can also verify negative correlations between variables, located in opposite quadrants, such as: (a) PN and RF: this correlation indicates that increased child mortality due to PN was associated with lower death due to FR. (b)PO and NKD increased child mortality due to poisoning(PO) was associated to a lower non infectious kidney diseases. (c) MSC and BA: this correlation indicates that increased in children death due to MSC was associated with lower BA. (d) NGD and SCD: this correlation indicates that increased child mortality due to NGD was associated with lower mortality due SCD. It is noteworthy that PO and AN were positively high correlated with PC2. There were no variables which strong negatively correlated with either PC1 or PC2.

Biplot results of PC1 with PC3

The biplot presented in Fig. 4 shows the correlation between variables of PC1 and PC3. The main interactions between variables of this plot are: (a) OND with PO and PN: this correlation is located in the first quadrant and shows that when child mortality due to OND is high, PO and PN is also high. (b) HA with CD: this correlation is located in the second quadrant and indicates that HIV/AIDS is associated with congenital diseases.(c) DSD with DM: this correlation is located in the first quadrant and indicates that diabetes mellitus is associated with DSD. These variables were strongly related with PC1. (d) SCD with UTI and SI: this correlation is located in the second quadrant and indicates that sickle cell disease is associated with the UT1 and skin infection. (e) MU with JD and Ocr:this correlation is located in the third quadrant and shows that child died for uncomplicated malaria were also suffering from joint disorder and other cardiovascular diseases.

Fig. 4 also reveals some negative correlations between variables, located in opposite quadrants such as: (a) BN and AN: this correlation explains that child mortality associated with burns is opposite to Anaemia. (b) EP and ARI: this correlation shows that the decrease of Acute respiratory infection is associated with the increase of epilepsy disease for children.(c) OT with MU, OG and JD : this correlation explains that association between child mortality due to OT and uncomplicated malaria , joint disorder and other cardiovascular and is opposite. (d) SCD with OM : this correlation indicates that the increase death due to sickle cell disease is related to the decrease of the Osteomyelitis. One can verify the same high correlated variables for PC1 as in Fig. 3, and One can observe that HA and BN were positively high correlated with PC3. In contrast, ARI and MU were moderately negatively correlated with PC3 (Table 4).

Biplot results of PC1 with PC4

The biplot presented in Fig. 5 shows the positively high correlated variables in this interaction: (a) ARI with RD: this correlation is located in first quadrant and was related to the association between Acute respiratory infections and other respiratory diseases.This was expected since ARI is the type of RI. (b) JD with RF: this correlation is located in second quadrant and was related to the association between joint disorder and rheumatic fever. (c)NKD with PNC: this correlation is located in the second quadrant and relates NKD with PNC. (d) SCD with FD and UTI: this interaction is located in the third quadrant of the biplot and suggests an increase in SCD is in response with FD and UTI.

Some negative correlations were also identified: (a) PO with MSC: this correlation showed that child mortality due to MSC is opposite to PO. (b) OT with MU: this correlation indicates that OT decreases MU. one can verify the same high correlated variables for PC1 as in Fig. 3, and we can

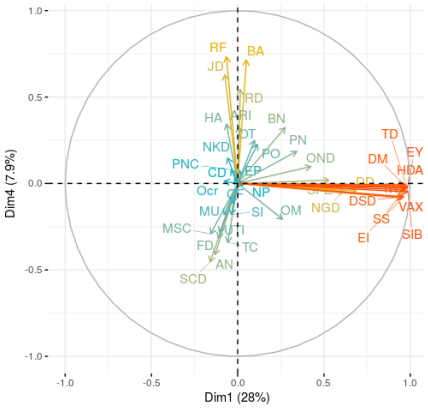

Fig. 5. Biplot relating PC1 with PC4

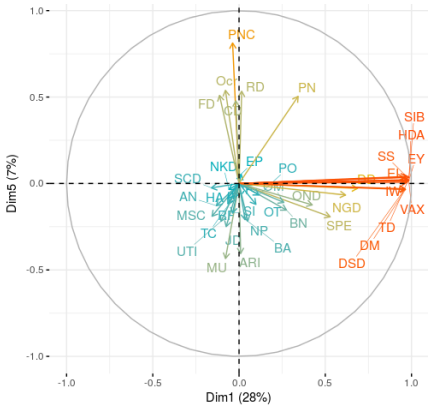

Fig. 6. Biplot relating PC1 with PC5

observe that RF and BA were positively high correlated with PC4. In contrast, AN and SDC were negatively weak correlated with PC4 (Table 4).

Biplot results of PC1 with PC5

The biplot presented in Fig. 6 shows the correlation between features of PC1 and PC5. The main interaction between variables of this plot are: (a) PNC with CD: this correlation is located in the second quadrant and shows that when child death due to PNC is higher, the death due to CD is also higher. (b) TC with UTI:this correlation is located in the third quadrant and explains the fact that child suffering from UTI was associated with suffering from tuberculosis. (c) NP with BA: this correlation is located in the fourth quadrant and explains the fact that child death due to neoplasm was associated with suffering from bronchial asthma.

Some negative correlations were also identified: (a) ARI with PNC and CD: this correlation shows that the decrease of ARI is associated with the increase of PNC and CD. (b) PN and MSC: this correlation indicates that child die from pneumonia disease, decreases the possibility of death from MSC. (c) FD with NP: this correlation indicates that child die from neoplasm, was not associated with fractures dislocation. Also one can verify the same high correlated variables for PC1 as in Fig. 4, and we can observe that PNC was positively high correlated with PC5. In contrast, ARI and MU were negatively weak correlated with PC5 (Table 4).

Biplot results of PC1 with PC6

The biplot presented in Fig. 7 shows the correlation between features of PC1 and PC6. The main interaction between variables of this plot are: (a) NKD with NGD: this correlation is located in the first quadrant and shows that when child die due to NKD,there is high possibility that was also suffering from NKD. (b) NP with EP: this correlation is located in the fourth quadrant and shows that children die due to NP ,were also associated with suffering from EP.

Some negative correlations were also identified: (a) OT with ARI: this correlation shows that the decrease of ARI is associated with the increase in OT. (b) MU and RD: this correlation indicates that child died from MU disease, had a lower possibility that was also suffering from RD. (c) PNC with PO: this correlation indicates that child die from PNC, was associated with low possibility of dying from poisoning. one can verify the same high correlated variables for PC1 as in Fig. 3, and

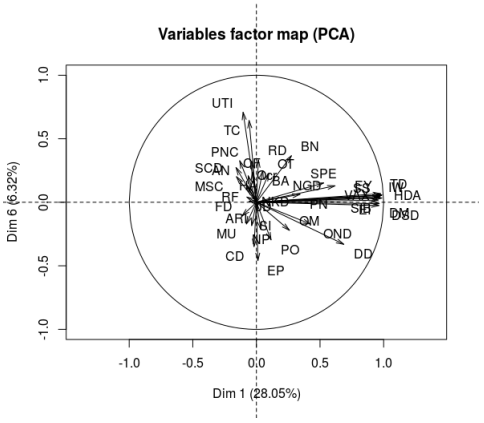

Fig. 7. Biplot relating PC1 with PC6

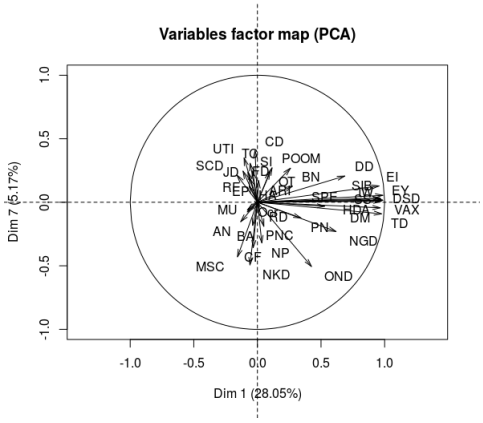

Fig. 8. Biplot relating PC1 with PC7

we can observe that UTI was positively high correlated with PC6. In contrast, EP was negatively weak correlated with PC6 (Table 4).

Biplot results of PC1 with PC7

The biplot presented in Fig. 8 shows the correlation between features of PC1 and PC6. The main interaction between variables of this plot are: (a) MU with AN: this correlation is located in the third quadrant and shows that when child die due to MU,there is high possibility that was also suffering from AN. (b) CF with BA: this correlation is located in the third quadrant and shows that children die due to CF ,were also associated with suffering from BA.(c) PN with NGD: this correlation is located in the fourth quadrant and shows that child death due to NGD is associated with suffering from PN.

In Fig. 8 one can also verify negative correlations between variables, located in opposite quadrants, such as: (a) JD and PNC: this correlation indicates that increased child mortality due to JD was associated with lower death from PNC. (b)OT and MU increased child mortality due to uncomplicated malaria was associated to a lower death due to other diseases. Also one can verify the same high correlated variables for PC1 as in Fig. 3, and we can observe that CD was positively weak correlated with PC7. In contrast, MSC, NKD and OND were negatively weak correlated with PC7 (Table 4).

4 Conclusion

In conclusion, it has been identified that from 25 components, the first seven components are sufficient to explain the diseases that causes child mortality for children under five years in Tanzania mainland. These components retained 73.40% of total variance of the original dataset. PCA results also show that TD, SIB, VAX, EY, SS, IW, EI, HDA, DM , DSD, PO, AN, HA, BN, RF,BA, PNC and UTI are significant diseases. It is worthwhile to stress that these important diseases should be observed more systematically and carefully. The outcome shows that there is a potential need for for improving the efficiency in preventing and controlling child mortality by reducing the number of diseases from 42 to 18. This reduction may result in significant cost in saving the child mortality control program without sacrificing important diseases. However it should be noted that in this

study the only one year (2013) were used. To make critical decision in assessing the diseases that lead to child mortality for children less than 5 years in Tanzania mainland, the longer time scale should be performed. Therefore, for Tanzania to attain the 2030 goal of 14 death per 1000 live birth, effective child survival interventions are required. There is the need to properly manage child mortality and monitor child diseases. This study may prove to be very useful in achieving this goal.

Acknowledgement

I would like to thank Almighty God for his protection, Tanzania National Bureau of Statistics, my mother Stephania Ntare and all friends for their encouragement throughout the work.

Competing Interests

Author has declared that no competing interests exist.

References

[1] Haidong Wang, Chelsea A. Liddell, Matthew M. Coates, Meghan D. Mooney, et al. Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990-2013: A systematic analysis for the global burden of disease study 2013. The Lancet. 2014;384(9947):0 957-979.

[2] Danzhen You, Tessa Wardlaw, Peter Salama, Gareth Jones. Levels and trends in under-5 mortality, 1990-2008. The Lancet. 2010;375(9709):100-103.

[3] Danzhen You, Lucia Hug, Simon Ejdemyr, Jan Beise. Levels and trends in child mortality. Report 2015. Estimates Developed by the un inter-agency group for child mortality estimation; 2015.

[4] United Nations. Department of Economic. The Millennium Development Goals Report 2015. United Nations Publications; 2015.

[5] Bellagio Study Group on Child Survival, et al. Knowledge into action for child survival. The Lancet. 2003;362(9380):323-327.

[6] Robert E. Black, Simon Cousens, Hope L. Johnson, Joy E. Lawn, et al. Global, regional, and national causes of child mortality in 2008: A systematic analysis. The Lancet. 2010;375(9730):1969-1987.

[7] Wiens MO, Kumbakumba E, Larson CP, Ansermino JM, et al. Postdischarge mortality in children with acute infectious diseases: Derivation of postdischarge mortality prediction models. BMJ Open. 2015;5(11):e009449.

[8] Hoviyeh Afnan-Holmes, Moke Magoma, Theopista John, Francis Levira, Georgina Msemo, Corinne E. Armstrong, et al. Tanzania's countdown to 2015: An analysis of two decades of progress and gaps for reproductive, maternal, newborn, and child health, to inform priorities for post-2015. The Lancet Global Health. 2015;3(7):e396–e409.

[9] David Le Blanc. Towards integration at last? The sustainable development goals as a network of targets. Sustainable Development. 2015;23(3):0 176-187.

[10] United Nations. Transforming our world: The 2030 agenda for sustainable development. Resolution adopted by the General Assembly; 2015.

[11] Jennifer C. Moïsi, Hellen Gatakaa, James A Berkley, Kathryn Maitland, et al. Excess child mortality after discharge from hospital in Kilifi, Kenya: A retrospective cohort analysis. Bulletin of the World Health Organization. 2011;89:725-732.

[12] Mohammod Jobayer Chisti, Stephen M Graham, Trevor Duke, Tahmeed Ahmed, Abu Syed Golam Faruque, et al. Post-discharge mortality in children with severe malnutrition and pneumonia in Bangladesh. PloS One. 2014;9(9):e107663.

[13] Matthew O. Wiens, Shane Pawluk, Niranjan Kissoon, Elias Kumbakumba, J Mark Ansermino, et al. Pediatric post-discharge mortality in resource poor countries: A systematic review. PLoS One. 2013;8(6):e66698.

[14] Jennifer C. Moïsi, D James Nokes, Hellen Gatakaa, Thomas N Williams, et al. Sensitivity of hospital-based surveillance for severe disease: A geographic information system analysis of access to care in Kilifi district, Kenya. Bulletin of the World Health Organization. 2011;89:102-111.

[15] Tanzania Open Data. Number and causes of death under 5 years - 2013. Tanzania National Bureau of Statistics, Dar es salaam, Tanzania; 2018.
Available:http://statistics.go.tz/dataset/b0e2ccbf-7350-478f-bb2d-7cc75d7e7936/resource/7fe2f901-70a3-4b85-a4b8-39eee44aca63/download/Death-Under-5-years—2013.csv

[16] Core Team R. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2018.
Available:https://www.R-project.org/

[17] Erich Neuwirth. RColor Brewer: Color Brewer Palettes; 2014. Available:https://CRAN.R-project.org/package=RColorBrewer

[18] Sébastien Lê, Julie Josse, François Husson. FactoMine R. A package for multivariate analysis. Journal of Statistical Software. 2008;25(1):1-18. DOI:10.18637/jss.v025.i01

[19] Adamu Mustapha, Ado Abdu. Application of principal component analysis & multiple regression models in surface water quality assessment. Journal of environment and earth science. 2012;2(2):16-23.

[20] Brian S. Everitt, Graham Dunn, et al. Applied multivariate data analysis. volume 2. Wiley Online Library; 2001.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sciencedomain.org/review-history/28108