Journal of Theoretical and Applied Information Technology
15th April 2026. Vol.104. No.7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# IDENTIFYING A TYPOLOGY OF MOROCCAN HEALTH PROVINCES IN URBAN AREAS BASED ON MATERNAL AND CHILD HEALTH INDICATORS USING A CLUSTERING APPROACH

**MERYEM CHAKKOUCH[1], HICHAM SADIKI [2], MEROUANE ERTEL [3], AZIZ MENGAD [4], SAID AMALI [5], ABDELMALEK AZZOUZI [6]**

[1,2,3] Informatics and Applications Laboratory (IA)-Faculty of Sciences, Moulay Ismail University Meknes, Morocco

[4] Centre for Doctoral Studies "Life and Health Sciences"-Drug Sciences Formation, Laboratory of Pharmacology and Toxicology (LPTR), Faculty of Medicine and Pharmacy of Rabat (FMPh), Impasse Souissi Rabat, Morocco

[5] Informatics and Applications Laboratory (IA), FSJES Moulay Ismail University, Meknes, Morocco

[6] Applied Economics Laboratory of Meknes Faculty of Legal, Economic and Social Sciences Moulay Ismail University, Meknes, Morocco

E-mail: [1] m.chakkouch@edu.umi.ac.ma, [2] h.sadiki@umi.ac.ma, [3] m.ertel@edu.umi.ac.ma, [4] mengad.aziz@hotmail.com, [5] s_amali@yahoo.com, [6] a.azzouzi@umi.ac.ma

**ABSTRACT**

Maternal and child health remains a public health priority, particularly given significant spatial disparities in healthcare accessibility and outcomes. It is essential to clearly identify these differences in order to refine intervention strategies and optimize resource allocation. In this article, we propose a clustering approach to group and identify typologies of Moroccan health provinces based on MCH indicators. We conducted a comparative analysis to evaluate five dimensionality reduction (DR) techniques, including PCA, ICA, and nonlinear methods (KPCA, LE, Isomap), combined with three clustering methods (K-Means, agglomeration clustering, GMM) on Moroccan urban areas data characterized by a high dimensionality, low sample size (HDLSS) structure on MCH indicators. The model selection was guided by internal validation indicators (Silhouette, Calinski-Harabasz, Davies-Bouldin). Laplacian Eigenmaps followed by K-means achieves superior performance measures reflecting the most robust and consistent clustering structure.

Three clusters of provinces were identified in this comparative study: moderate-performing provinces characterized by the presence of both public and private urban healthcare systems; high-performing provinces with robust urban public services; and lower-performing provinces due to rural dependence on urban facilities and resource constraints. The main contribution of this study is to identify disparities in maternal and child health across Moroccan provinces/prefectures using urban-area indicators and a clustering-based approach. It thus makes it possible to propose a territorial typology that can support public health interventions and improve the strategic allocation of resources.

**Keywords:** *Clustering; PCA, ICA, KPCA, LE, Isomap, K-Means, Agglomerative Clustering, GMM, MCH Indicators*

## 1. INTRODUCTION

Reflecting a society's level of development and the effectiveness of its healthcare system, maternal and child health (MCH) remains a major public health concern worldwide [1,2]. Preventable maternal and infant mortality and morbidity persist, despite significant progress in recent years, with large disparities between and within countries [3,4,5]. Several factors strongly influence transitions in maternal mortality, as recent global evidence indicates. Such factors include structural determinants (socio-economic development, fertility trends, education), health system capacity, and factors related to the quality of care, which interact differently across contexts [6,7]. Under these circumstances, WHO calls for accelerated, equity-oriented action on maternal health and child survival, emphasizing that national averages can mask

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

pockets of high burden and that better subnational intelligence is needed to guide targeted interventions and resource allocation [3,4,8].

In Morocco, considerable efforts have been made in this crucial area of MCH. The country has adopted the Global Strategy for Women's, Children's and Adolescents' Health 2016-2030, which aims to end preventable deaths among all women, newborns, children and adolescents, to substantially improve their health and well-being and to bring about the transformations that will create a better and more prosperous future [4]. These efforts have been reflected in an increase and positive development of the various indicators of MCH, , as evidenced by the results of the National Population and Family Health Survey (2018): - A 35% reduction in the maternal mortality ratio between 2010 (112 deaths per 100,000 live births) and 2018 (72.6 deaths per 100,000 live births); - A 27% reduction in the mortality rate among children under 5 years of age, between 2011 (30.5 deaths per 1,000 live births) and 2018 (22.16 deaths per 1,000 live births); The infant mortality rate fell from 28.8 deaths to 18.0 per 1,000 live births (a reduction of 38%) [9]. However, despite these appreciable achievements, Morocco also presents significant territorial disparities and inequalities in maternal and infant health indicators, particularly between rural and urban areas, where, for example, the maternal mortality rate is 111.1 per 100,000 live births in rural areas while it is 44.6 per 100,000 live births in urban areas [9].

Many research studies confirm this territorial approach. These studies show that MCH results and service coverage are not equal everywhere and tend to cluster spatially and socioeconomically. District-level analyses in India and multicentre data in sub-Saharan Africa reveal marked gradients according to wealth and development, with clear geographical patterns of "coverage gaps" that are concentrated in specific regions and persist over time[10,11]. Spatial epidemiology studies further document geographic hotspots and the role of social determinants in shaping maternal risks, for example, national-scale spatial clustering of COVID-19 maternal mortality in Brazil [12], district mortality hotspots in Pakistan using spatial autocorrelation and geographically weighted regression [13], and repeated low-uptake clusters of antenatal care in Ethiopia identified through spatial scan statistics and regression [14], alongside broader evidence of geographic differences in RMNCH utilization [15]. On the methodological level, recent advances increasingly combine geospatial analysis and machine learning to produce accurate and policy-relevant maps of service coverage and disparities (e.g., high-resolution mapping of essential maternal and child health services in Nigeria) [16,17]. In parallel, other machine learning studies use clustering and feature importance modelling to stratify local risk profiles and identify key drivers. [18,19,20]. Clustering has been used successfully to extract regional typologies from datasets of MCH indicators, such as the grouping of districts in Bangladesh using hierarchical clustering of MICS indicators [21]. It has also been applied to datasets on maternity and childbirth to compare algorithms (e.g., K-means and K-medoids) and to highlight the importance of preprocessing choices [22]. In this context, this work encourages the application of unsupervised clustering to Moroccan provincial indicators of MCH to produce interpretable typologies that support targeted planning and translate inequality measurement into concrete actions, especially in the Moroccan context. Most existing studies remain based on descriptive assessments or on isolated indicators, which limits their ability to capture the multidimensional nature of territorial disparities. This study addresses this gap through a clustering-based analysis of aggregated MCH data, using various DR techniques and clustering algorithms to identify disparities in MCH across Moroccan provinces/prefectures and to establish a territorial typology.

The article is structured as follows. First, the data collection and preprocessing methods are detailed. In particular, the study highlights the normalization and standardization procedures. Then, dimensionality reduction techniques are applied and compared. The effectiveness of different clustering algorithms is then evaluated using specific performance metrics. Finally, the results are presented and discussed, emphasizing practical implications for healthcare policy and resource allocation.

## 2. PROPOSED MODEL

In this article, we proceed for a comparative analysis of different models obtained from the coupling of DR techniques and clustering algorithms. We adopt a classical clustering approach, starting with detailing data collection and preprocessing methods such as standardization procedures. Then, DR techniques are applied, and clustering algorithms are evaluated using specific performance metrics. Finally, the results are presented and discussed, emphasizing practical implications for healthcare policy and resource allocation.

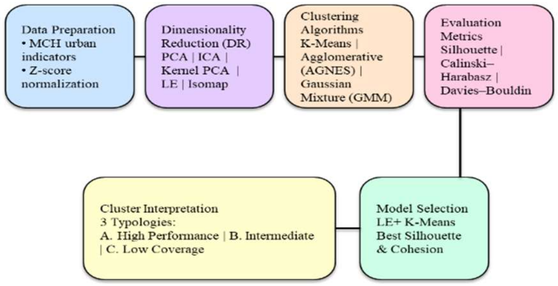


*Figure 1 : Clustering methodology*

## 3. MATERIALS AND METHODS

### 3.1 Data collection

MCH data is collected from the Health Information System for Reproductive Health, Child Health, and Curative Care (SREC), which includes important data on MCH such as prenatal and postnatal consultations, delivery assistance, breastfeeding practices, childhood illnesses, vaccination coverage, prevalence of fever, diarrhea, and cough, as well as the nutritional status of children under five years of age.

We note that the dataset is available on "Health in Figures" [23], consisting of an annual report that summarizes statistical information and data related to the national healthcare system. It outlines the availability of both public and private healthcare services, as well as the activities and outputs of public healthcare facilities at national, regional, and provincial/prefectural levels. This document is considered an official data source provided by the Ministry of Health and Social Protection. It serves as an important reference for healthcare administrators, medical professionals, researchers, students, and other individuals interested in health-related data [23].

In this study we collect 100 features about maternal and child health from 82 Province/prefecture. These features are organized into the following domains (Table 1) (See the appendix for the entire features list):

*Table 1 : List of study variables domains*

| | |
|---|---|
| **Prenatal consultation** | Prenatal consultation services |
| | Medicalized CPN and diagnosed high-risk pregnancies |
| **Childbirth and postpartum care** | Deliveries performed in public delivery facilities |
| | Births registered at public delivery facilities |
| | Diagnosed complications and number of malformations in newborns recorded in public delivery facilities |
| | Postnatal consultations for the mother |
| **Postnatal consultation** | Postnatal consultations for the newborn |
| **Immunization** | Children born protected and the number of children vaccinated by antigen (Hep.B1, BCG, VPO.O, DTP-Hib-HB (Penta), Rotavirus… ), |
| | Distribution of women of childbearing age who received the tetanus and diphtheria vaccine |
| | Distribution of pregnant women who received the tetanus and diphtheria vaccine |
| **Nutrition** | Children under 5 years of age who attended ESSPs according to their nutritional status |
| | Preventive vitamin A and vitamin D supplementation benefits for children under 5 years of age |
| | Curative vitamin A and zinc supplementation benefits for children under 5 years of age |
| | Preventive iron and vitamin D supplementation services for women in CPN and CPON |
| **Medical consultation for children under 5 years old** | Infants (0 to 2 months) sick depending on the illness |
| | Children (2 to 59 months) sick according to the type of illness |

### 3.2 Data Preprocessing

Every machine learning journey starts with data preprocessing. Feature scaling is one of the most important steps in preprocessing.

When working with datasets that include variables measured in different units or scales (Total prenatal consultations vs. Number of premature newborns), directly combining them in calculations can lead to skewed results. Variables with larger numerical ranges (like Total prenatal consultations) tend to overpower those with smaller ranges (like Number of premature newborns), even if both are equally important [24].

So, we first divide features by the population, and then we standardize the resulting per capita values.

As the features are numbers (for example number of women examined postpartum), and as it is necessary to take into consideration the demographic weight of each province to allow for fair comparisons across provinces, we divided all the variables for each province by the size of its urban population.

To normalize the data, we use Z-score standardization, a technique that centers features around a mean of 0 and scales them to a standard deviation of 1.

Given a X feature, the formula of the new feature standardized X' is defined as: $X' = \frac{X-\mu}{\sigma}$ ; $\mu$ and $\sigma$ are the sample mean and standard deviation of the feature X

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The final standardized values represent how many standard deviations each observation deviates from the mean, creating a unitless, comparable scale.

### 3.3 Feature Engineering: Dimensionality reduction techniques

We will apply five DR methods: Principal component analysis (PCA), Gaussian kernel Principal Component Analysis (Gaussian kPCA), Independent Component Analysis (ICA), Laplacian Eigenmaps (LE) and sometric Mapping (Isomap).

#### 3.3.1 Principal Component Analysis (PCA)

PCA is a linear DR technique that compresses high-dimensional data into a smaller set of uncorrelated variables, the principal components, while preserving most of the original variance. Mathematically, PCA solves an eigenvalue problem on the covariance (or correlation) matrix of mean-centred, usually standardized, features, yielding an orthogonal basis ordered by descending variance. Equivalently, PCA can be performed via the Singular Value Decomposition (SVD) of the data matrix, which provides the same principal directions and variances in a numerically stable way [24].

#### 3.3.2 Gaussian kernel Principal Component Analysis (Gaussian kPCA)

kPCA is a non-linear extension of PCA that leverages kernel methods to capture complex, non-linear patterns in data. The most used kernel is the Gaussian radial-basis function (RBF): $k_\gamma(\mathrm{x}, \mathrm{x}') = \exp(-\gamma \parallel \mathrm{x} - \mathrm{x}' \parallel^2)$;
$\gamma = \frac{1}{2\sigma^2} > 0$ & x and x′ are two data samples (row vectors) from the original input space $R^p$

#### 3.3.3 Independent Component Analysis (ICA)

ICA is a statistical technique used to separate a multivariate signal into additive, statistically independent components. Unlike PCA (which finds orthogonal directions of maximum variance) or kPCA (which handles non-linearity), ICA focuses on uncovering hidden factors or sources that are independent of each other. It is widely used in blind source separation (BSS) and signal processing. ICA models $X = AS$ where S: Matrix of independent source signals and A: Mixing matrix (unknown linear transformation).

#### 3.3.4 Laplacian Eigenmaps (LE)

LE is a nonlinear dimensionality reduction technique and a spectral manifold-learning method that preserves local geometric structure by leveraging the graph Laplacian. It assumes data lies on a low-dimensional manifold embedded in high-dimensional space. It assigns edge weights $W_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$, computes the graph Laplacian $L = D - W$ (where $D$ is the degree matrix with $D_{ii} = \sum_j W_{ij}$, and solves the generalized eigenvalue problem $Lv = \lambda Dv$. The low-dimensional embedding is derived from the eigenvectors corresponding to the smallest non-zero eigenvalues, minimizing $\sum_{i,j} W_{ij} ||y_i - y_j||^2$ under the constraint $Y^\top DY = I$ . Effective for visualizing manifolds, it prioritizes local connectivity but scales poorly and requires careful tuning of k and $\sigma$

#### 3.3.5 Isometric Mapping (Isomap)

Isomap is a nonlinear dimensionality reduction technique that preserves the geodesic distances (manifold-approximated shortest paths) between data points, unlike linear methods like PCA that use Euclidean distances. It extends Multidimensional Scaling (MDS) for manifolds by approximating the intrinsic geometry of the data. It constructs a $k$-nearest neighbour graph, computes pairwise geodesic distances $D_G(i,j)$ via Dijkstra's algorithm and applies MDS to the distance matrix $D_G$. The embedding $Y \epsilon R^p$ minimizes $\sum_{i,j}(D_G(i,j) - ||y_i - y_j||)^2$, achieved by eigen decomposition of the kernel matrix $K = -\frac{1}{2} H {D_G}^2 H$, where $H = I - \frac{1}{n} 11^\top$ centers the data. The solution $Y = \Lambda^{1/2} V^\top$ (with $\Lambda$ and V as top eigenvectors/values of K) preserves global manifold structure and is sensitive to noise or non-convex manifolds.

### 3.4 Machine Learning Models

In our study, we will apply three clustering algorithms to compare their performance and ability to reveal relevant structures in the data and subsequently retain the most consistent and interpretable partition. First, we will apply the famous K-means algorithms, then Gaussian Mixture Model (GMM), and Agglomerative clustering.

#### 3.4.1 K-means

K-means is one of the most used methods for partitioning a dataset into $k$ distinct, non-overlapping clusters. Its goal is to group observations so that points within the same cluster are as similar as possible (intra-cluster cohesion) and points in different clusters are as dissimilar as possible (inter-cluster separation). It aims to minimize the variance within each cluster by iteratively refining centroids (cluster centers) using a distance matrix like Euclidean distance. It consists of minimizing the sum of squared errors (SSE) within clusters:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$SSE = \sum_{i=1}^{k} \sum_{x \epsilon C_i} ||x - \mu_i||^2$ where $C_i$: Points in cluster $i$ and $\mu_i$: Centroid of cluster $i$.

### 3.4.2 Gaussian Mixture Model (GMM)

A GMM is a probabilistic unsupervised learning algorithm that assumes data points are generated from a mixture of $k$ Gaussian (normal) distributions. Unlike k-means (which uses hard clustering), GMM provides soft clustering by assigning probabilities to each data point for belonging to each cluster.

It is a powerful tool for modeling complex data distributions but requires careful tuning of $k$ and regularization (e.g., diagonal covariance) to avoid overfitting.

### 3.4.3 Agglomerative

AHC is a bottom-up approach that builds a hierarchy of clusters by iteratively merging the nearest pairs of data points or clusters until all points converge into a single cluster. Starting from each point as a separate cluster, it uses linkage criteria, such as simple linkage (minimum distance between pairs), full linkage (maximum distance), average linkage (average distance), or Ward's method (minimizing the variance of merged clusters), to calculate inter-cluster distances and guide the merging. The process generates a dendrogram to visualize the relationships between clusters and allows users to "cut" the tree at the desired height to extract clusters. AHC avoids predefining the number of clusters (a key advantage over the k-means algorithm) and captures hierarchical structures. Ideal for small and medium datasets, it is widely used in biology (gene clustering), social network analysis and taxonomy, where interpretable hierarchies are essential

## 3.5 Performance indicators

The evaluation of clustering results performances, we will apply three distinct internal validation metrics: the Silhouette Score, Calinski-Harabasz Score, and Davies-Bouldin Index.

### 3.5.1 Silhouette Score

The SI is metric that evaluates clustering quality through intra-class proximity (how closely data points are grouped within their clusters) and inter-class distance (how distinct they are from other clusters):

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$

where *a(i)* is the average intra-cluster distance and *b(i)* is the average distance to the nearest neighboring cluster, the score ranges from -1 to 1. Values near 1 indicate well-defined clusters, 0 suggests overlapping clusters, and negative values imply misassignment

### 3.5.2 Calinski-Harabasz Score

The CHI index is defined as the ratio of the weighted sum of between-cluster variance (separation) to within-cluster variance (compactness) :

$$CH = \frac{SS_B/(k-1)}{SS_W/(n-k)}$$

where:

$SS_B = \sum_{i=1}^{k} n_i . \|\mu_i - \mu\|^2$ : is the dispersion of the between-cluster

$SS_W = \sum_{i=1}^{k} \sum_{x\epsilon\ i} \|x - \mu_i\|^2$: is the dispersion of the within-cluster

$k$: number of clusters, $n$: total data points, $\mu_i$: cluster centroids, $\mu$: global centroid, $n_i$: cluster size

A higher score indicates better-defined clusters, as it reflects tightly grouped points within clusters and distinct separation between them.

### 3.5.3 Davies-Bouldin Index

DBI Index assesses the quality of the clustering by balancing intra-cluster compactness (average distance of points to their cluster centroid) and inter-cluster separation (distance between centroids). It measures, for each cluster, the ratio of its within-cluster scatter to its separation from the most "similar" other cluster, and then averaging these worst-case ratios:

$DBI = \frac{1}{k} \sum_{i=1}^{k} \max_{i \neq j} \left(\frac{S_i + S_j}{d(c_i, c_j)}\right)$, where:

$S_i = \frac{1}{|C_i|} \sum_{x\epsilon\ i} \|x - c_i\|$ : average intra-cluster distance for cluster $C_i$ (compactness).

$d(c_i, c_j)$ : Distance between centroids $c_i$ and $c_j$ (separation).

The lower the value, the better the clustering: with tight and well-separated clusters, with 0 being the ideal minimum.

# 4. RESULTS AND DISCUSSION

## 4.1 Descriptive statistics

The dataset contains a number of features about maternal and child healthcare monitoring registered at public health facilities (the private sector is not included) throat different sanitary provinces of Morocco in urban areas.

Our interest in urban environments is explained by the concentration of hospitals and health centers in them, while rural areas generally only have access to health centers.

Some key indicators from the dataset for the year 2022 are presented in Table 2 below. These

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

include 236,581 new prenatal care registrations and 408,981 follow-up prenatal consultations, reflecting respectively the number of pregnant women who started prenatal care during the year and the high continuity of prenatal care and access to routine examinations. 142,883 postnatal care registrations were recorded, demonstrating strong continuity of follow-up care after childbirth, a crucial period for maternal and neonate health.

166,577 live births were recorded, of which 153,078 being conducted in maternity hospitals.

Vaccination coverage is showing positive results demonstrating good alignment with national immunization schedules, with a total of 358,896 children having received the BCG vaccine (against tuberculosis), 350,396 having received the third dose of the pentavalent vaccine (DTP3 + Polio3 + Hib3 + HB3) and 344,721 children having received the first dose of MMR vaccine. In this context, maintaining high coverage in all districts remains essential for disease prevention.

*Table 2 : Some key indicators from the dataset*

| Indicators | 2022 |
|---|---|
| Number of new antenatal care registrations | 236581 |
| Number of returning antenatal care visits | 408981 |
| Number of women examined postnatally | 142883 |
| Total number of deliveries (in maternity hospitals) | 153078 |
| Total number of live births | 166577 |
| Total number of live births < 2500g | 5649 |
| BCG | 358896 |
| DTP3 + Polio3 + Hib3 + HB3 (3rd dose) | 350396 |
| MMR (1st dose) | 344721 |
| Tetanus-diphtheria vaccine (Td3) for pregnant women | 17814 |
| Number of vaginal deliveries | 138651 |
| Number of cesarean deliveries | 28022 |
| Total number of stillbirths | 2444 |

Statistics on childbirth indicate that 138,651 women gave birth vaginally, whereas 28,022 were performed caesarean sections. Finally, the data shows that 2,444 stillbirths were recorded which represents about 1.4% of all births. This underscores the need to keep working to improve care during childbirth and respond rapidly to any problems to reduce these cases.

Figure 2 below, which visualizes the correlation matrix, demonstrates the existence of strong linear relationships between MCH variables, particularly between variables grouped within the same domain (Table 1), such as maternal care or vaccination. For example, prenatal consultations and delivery outcomes are strongly correlated, as doses of BCG and DTP-Hib-HB.

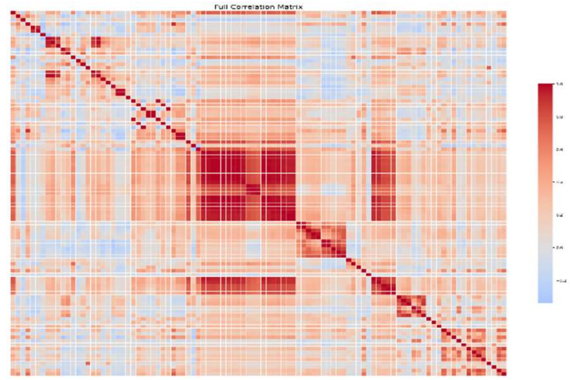


*Figure 2 : Correlation plot*

## 4.2 Clustering analysis

In this study, we based our decision regarding the number of clusters to choose on an examination of the elbow graph shown in Figure 3 below. This method graphically represents the within-cluster sum of squares (WCSS) as a function of increasing values of k: number of clusters, thus assessing the contribution of each new cluster to the overall structure. According to the resulting curve, a marked inflection point is observed at k=3, after which the WCSS curve flattens out, thus representing the optimal solution in terms of the number of clusters that ensures a good balance between intra-cluster compactness and inter-cluster separation. This three-cluster structure serves, for subsequent analyses, as a basis for typological exploration, allowing a coherent examination of the interaction between different DR techniques and clustering algorithms in order to reveal the underlying organization of the provincial dataset.

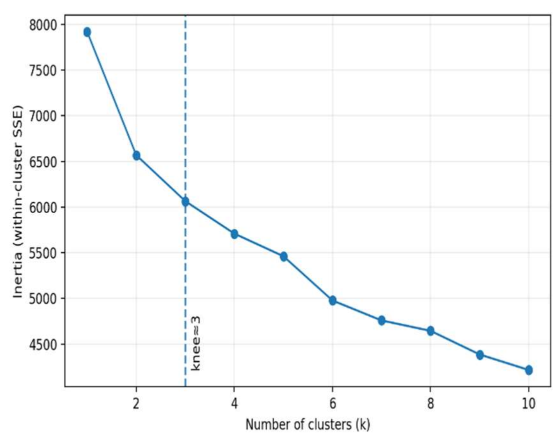


*Figure 3 : Optimal Number of Clusters: Elbow Method*

Figure 4 below presents the graphical representations of different models obtained from the interaction between DR techniques and clustering algorithms. In other words, it allows the

Journal of Theoretical and Applied Information Technology
15th April 2026. Vol.104. No.7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

description of resulting models that differ according to the choice of data representation technique and clustering methodology. Through which illustrates the cluster geometry that reflecting the choice of embedding and how each algorithm handles overlap, curvature, and group separation.

Each row corresponds to a clustering algorithm (K-means, GMM, AHC) and each column corresponds to a DR technique: two linear (PCA, ICA) and three nonlinear (kernel PCA, LE and Isomap). The three clustering algorithms partitioned Moroccan health provinces into three groups based on standardized per-capita MCH features. The first column of the visualisation demonstrates that clustering without a prior DR technique produces the weakest structure and clusters appear less distinct. While applying DR technique yields significant gains and improves cluster quality for these MCH indicators.

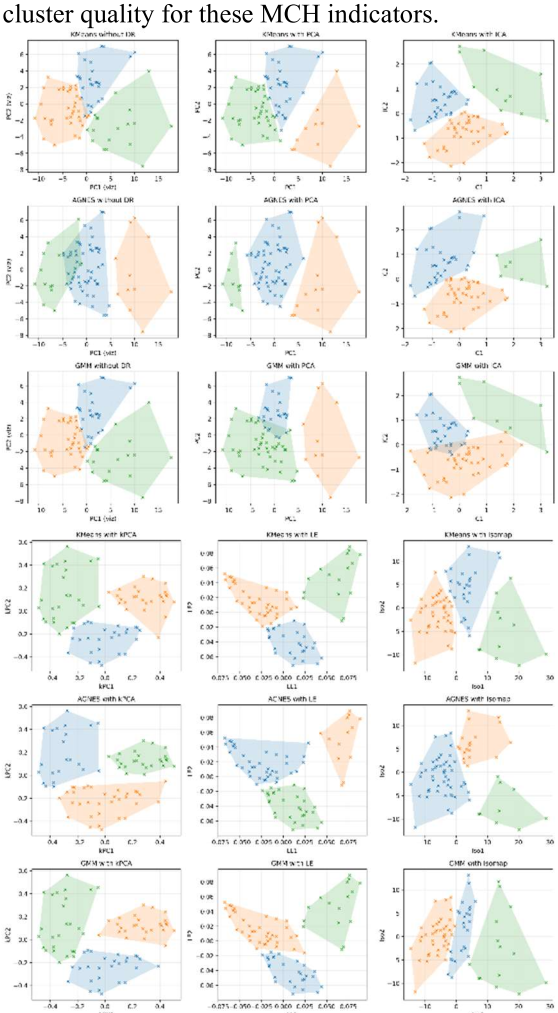

*Figure 4 : Visualization plots*

### 4.3 Performance Evaluation

The evaluation of clustering performance across the 18 combinations of DR techniques and algorithms based on the validation metrics shown in the table 2 below reveals distinct patterns in terms of effectiveness, shaped by the interaction between data structure, reduction methods, and clustering approaches.

The combination of Laplacian Eigenmaps (LE) and K-means stands out as the strongest and most consistent model by achieving the highest SI (0.469) and CHI (99.60), alongside the lowest DBI (0.696). These metrics collectively indicate well-separated, cohesive clusters with minimal overlap. The success of LE lies in its ability to preserve non-linear relationships within the data, which K-means leverages effectively due to its adaptability to large dataset including HDLSS data type. Additionally, the relatively balanced cluster sizes ((27, 37, 16)) suggest that the reduced data retains meaningful structure, enabling robust partitioning. The LE shows, in second degree, good performances with Agglomerative Clustering and GMM with closest metrics (SI 0.454, CHI 93.09, DBI 0.733 and SI 0.445, CHI 90.80, DBI 0.739 respectively). The second DR method that captures meaningful nonlinear structure while preserving a stable global geometry is kPCA. It offers a solid and stable performance across different algorithms, with balanced or quasi-balanced partitions: K-means + kPCA (Silhouette 0.424, CH 79.05, DB 0.808, (27, 27, 26)), GMM + kPCA (Silhouette 0.425, CH 76.93, DB 0.804, (27, 27, 26)), and Agglomerative Clustering + kPCA (Silhouette 0.420, CH 74.12, DB 0.823, (22, 33, 25)).

These results underscore the importance of selecting DR methods aligned with clustering objectives and show that non-linear DR approaches generally outperform linear ones, and that K-Means algorithm benefit more from these transformations. LE outperforms both linear methods (PCA, ICA) and other non-linear techniques (KPCA, Isomap), highlighting its superiority in capturing complex, non-linear data patterns. While PCA delivers moderate results, it often produces imbalanced clusters ((57, 13, 10) with Agglomerative Clustering), suggesting limitations in linear separability. Methods like KPCA and Isomap show partial success but fail to match LE's performance.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 3 : Performance Indicators*

| DR Technique | Clustering Algorithm | Silhouette | Calinski-Harabasz | Davies-Bouldin | Cluster Sizes |
|---|---|---|---|---|---|
| NoDR | K-means | 0,098 | 11,77 | 2,325 | (26, 36, 18) |
| | Gaussian Mixture Models | 0,098 | 11,77 | 2,325 | (26, 36, 18) |
| | Agglomerative Clustering | 0,136 | 12,11 | 2,049 | (56, 11, 13) |
| PCA | K-means | 0,328 | 56,91 | 0,963 | (31, 11, 38) |
| | Gaussian Mixture Models | 0,266 | 48,11 | 1,064 | (26, 11, 43) |
| | Agglomerative Clustering | 0,411 | 57,83 | 0,793 | (57, 13, 10) |
| ICA | K-means | 0,360 | 46,08 | 0,928 | (32, 37, 11) |
| | Gaussian Mixture Models | 0,301 | 33,59 | 1,052 | (33, 39, 8) |
| | Agglomerative Clustering | 0,356 | 46,17 | 0,888 | (33, 40, 7) |
| kPCA | K-means | 0,424 | 79,05 | 0,808 | (27, 27, 26) |
| | Gaussian Mixture Models | 0,425 | 76,93 | 0,804 | (27, 27, 26) |
| | Agglomerative Clustering | 0,420 | 74,12 | 0,823 | (22, 33, 25) |
| LE | **K-means** | **0,469** | **99,60** | **0,696** | **(27, 37, 16)** |
| | Gaussian Mixture Models | 0,445 | 90,80 | 0,739 | (27, 37, 16) |
| | Agglomerative Clustering | 0,454 | 93,09 | 0,733 | (38, 13, 29) |
| Isomap | K-means | 0,399 | 69,69 | 0,870 | (27, 42, 11) |
| | Gaussian Mixture Models | 0,305 | 52,18 | 1,176 | (25, 41, 14) |
| | Agglomerative Clustering | 0,426 | 66,57 | 0,779 | (52, 19, 9) |

The present findings are interpreted in light of both the methodological literature on clustering and the applied literature on territorial health inequalities. First, the results confirm that the quality of clustering depends strongly on data representation, as nonlinear dimensionality reduction methods, particularly Laplacian Eigenmaps and kernel PCA, produced more compact and better-separated clusters than linear approaches such as PCA and ICA. This suggests that disparities in MCH across Moroccan provinces/prefectures are structured by complex and potentially nonlinear relationships that cannot be fully captured through linear projections alone. Second, the identified clusters reveal that territorial inequalities are not limited to a simple gradient of service availability, but rather reflect distinct profiles combining differences in preventive care, delivery-related activity, postnatal follow-up, immunization, and child morbidity patterns. This interpretation is consistent with broader public health research showing that health-system performance is spatially heterogeneous and shaped by multiple interacting demographic, organizational, and service-utilization factors.

### 4.4 Sanitary provinces typology exploration

Three distinct profiles of Moroccan health provinces were obtained from the optimal model generated by the combination of LE followed by K-means. The interpretation of clustering analysis and the description of the three clusters is based on the different national programs related to MCH. These programs include: the Pregnancy and Childbirth Monitoring Program, the National Immunization Program, and the Health and Nutrition Program. Among the objectives of these programs, we mention: Achieving satisfactory prenatal and supervised delivery coverage; -Achieving and maintaining vaccination coverage of 95% or higher per antigen, per setting (urban and rural), and per level (national, regional, delegation, health district, and sector); -Reducing iron deficiency by one-third compared to 2000 levels; Eliminating vitamin A deficiency; -Reducing stunting in children under 5 years of age by 40%, among other specific objectives [25,26,27].

The first cluster present moderate performance regarding urban MCH indicators such as prenatal consultations, childhood vaccination coverage, and nutritional outcomes. It comprises 27 provinces including Casablanca, Rabat, and Agadir-Ida-Ou-Tanane. These provinces are characterized by a high rate of urbanization and the presence of private sector that offers same MCH services, which reduce the dependence to public facilities and explain this moderate performance.

The second cluster groups provinces showing good performance in MCH indicators. It is the largest cluster with 37 provinces, and it includes provinces like Tanger-Asilah, Meknes and Mohammadia. This profile reflects better-resourced urban health systems, with good implementation of public programs, greater availability of services and more efficient delivery mechanisms.

The third cluster with smallest size comprises 16 provinces, that are characterized by lowest performance according urban MCH indicators. These results are explained by the fact that the dataset focuses exclusively on urban indicators. With lower levels of urbanization, these provinces often serve large rural populations that rely on urban healthcare infrastructure, particularly hospitals, for childbirth and paediatric care. This demand puts pressure on relatively limited urban healthcare infrastructure, exacerbating service constraints and highlighting persistent territorial inequalities in access to care.

## 5. CONCLUSION

This study conducted a comparative analysis of regional health data relating to MCH indicators to extract the main spatial partitions. It focuses on urban areas that concentrate both hospitals and health centers. By comparing eighteen combinations of DR techniques and clustering algorithms, we selected the optimal model achieving high values of internal validation indexes (SI=0.469, CHI=99.60, DBI=0.696). The resulting provincial health typology from the LE followed by K-means reveals three distinct clusters: 1. Moderate Performance (27 provinces) characterized by a high rate of urbanization and significant private health sector presence resulting in moderate public health utilization services. 2. High Performing Provinces (37 regions) reflecting good outcomes of MCH indicators. 3. Low Performance (16 provinces) due to constraints related to the dependence of the rural population on urban infrastructure, particularly hospitals.

The results reveal that there are inequalities in achieving the objectives of the various programs related to MCH between the different provinces of Morocco, including unequal access to public services in areas with high private sector activity and resource constraints in mixed urban and rural contexts, guiding policymakers towards targeted investments in infrastructure and equitable healthcare planning.

This study addressed the lack of a data-driven territorial typology for examining maternal and child health disparities across Moroccan provinces/prefectures based on urban indicators. The results should be interpreted in light of certain limitations, including the restriction to urban data, the exclusion of the rural dimension, and the exploratory rather than causal nature of the analysis. Despite these limitations, the study is important from both a methodological and practical standpoint. Methodologically, it demonstrates the value of combining dimensionality reduction and clustering techniques to analyse complex public health data. From a public policy perspective, it proposes a territorial typology that could facilitate more targeted interventions and a more evidence-based allocation of health resources.

As a perspective to this work, the approach developed allows its results to be integrated into prediction models in order to anticipate future needs for medical equipment, personnel and infrastructure in the different provinces, thus promoting more proactive and equitable health planning.

## REFERENCES

[1] OECD, Health at a Glance 2025: OECD Indicators. in Health at a Glance. OECD Publishing, 2025. doi: 10.1787/8f9e3f98-en.

[2] World Health Organization. (2025, April 7). Maternal mortality ; https://www.who.int/news-room/fact-sheets/detail/maternal-mortality

[3] “Acceleration towards the Sustainable Development Goal targets for maternal health and child mortality”.

[4] “Global Strategy for Women’s, Children’s and Adolescents’ Health”.

[5] S. M. H. Jamaluddin, “Saving Lives, Building Futures: Importance of Maternal and Child Health in Achieving the Sustainable Development Goals,” Medtigo J Med, vol. 1, no. 1, pp. 1–5, Dec. 2024, doi: 10.63096/medtigo30622458.

[6] J. P. Souza et al., “A global analysis of the determinants of maternal health and transitions in maternal mortality,” The Lancet Global Health, vol. 12, no. 2, pp. e306–e316, Feb. 2024, doi: 10.1016/S2214-109X(23)00468-0.

[7] B. Madaj, H. Smith, M. Mathai, N. Roos, and N. Van Den Broek, “Developing global indicators for quality of maternal and newborn care: a feasibility assessment,” Bull. World Health Organ., vol. 95, no. 6, pp. 445-452I, Jun. 2017, doi: 10.2471/BLT.16.179531.

[8] J. Grove et al., “Maternal, newborn, and child health and the Sustainable Development Goals—a call for sustained and improved measurement,” The Lancet, vol. 386, no. 10003, pp. 1511–1514, Oct. 2015, doi: 10.1016/S0140-6736(15)00517-6.

[9] Enquête Nationale sur la Population et la Santé Familiale (ENPSF - 2018)

[10] B. K. Panda, G. Kumar, and A. Awasthi, “District level inequality in reproductive, maternal, neonatal and child health coverage in India,” BMC Public Health, vol. 20, no. 1, p. 58, Dec. 2020, doi: 10.1186/s12889-020-8151-9.

[11] Y. Lee, S. Bolongaita, R. Sato, J. B. Bump, and S. Verguet, “Evolution in key indicators of maternal and child health across the wealth gradient in 41 sub-Saharan African countries, 1986–2019,” BMC Med, vol. 22, no. 1, p. 21, Jan. 2024, doi: 10.1186/s12916-023-03183-0.

[12] T. S. Siqueira et al., “Spatial clusters, social determinants of health and risk of maternal mortality by COVID-19 in Brazil: a national population-based ecological study,” The Lancet Regional Health - Americas, vol. 3, p. 100076, Nov. 2021, doi: 10.1016/j.lana.2021.100076.

[13] F. S. Muhammad, S. M. Shahabudin, and M. B. A. Talib, “Measuring spatial inequalities in maternal and child mortalities in Pakistan: evidence from geographically weighted regression,” BMC Public Health, vol. 24, no. 1, p. 2229, Aug. 2024, doi: 10.1186/s12889-024-19682-5.

[14] A. Yeneneh, K. Alemu, A. F. Dadi, and A. Alamirrew, “Spatial distribution of antenatal care utilization and associated factors in Ethiopia: evidence from Ethiopian demographic health surveys,” BMC Pregnancy Childbirth, vol. 18, no. 1, p. 242, Dec. 2018, doi: 10.1186/s12884-018-1874-2.

[15] A. Defar, Y. B. Okwaraji, Z. Tigabu, L. Å. Persson, and K. Alemu, “Geographic differences in maternal and child health care utilization in four Ethiopian regions; a cross-sectional study,” Int J Equity Health, vol. 18, no. 1, p. 173, Dec. 2019, doi: 10.1186/s12939-019-1079-y.

[16] Y. Kawakatsu et al., “High-resolution mapping of essential maternal and child health service coverage in Nigeria: a machine learning approach,” BMJ Open, vol. 14, no. 6, p. e080135, Jun. 2024, doi: 10.1136/bmjopen-2023-080135.

[17] M. Emmanuel and A. Rajuroy, “Machine Learning-Based Predictive Models for Mapping Maternal and Child Health Disparities Using Geospatial Data”.

[18] B. Zaman, A. Sharma, J. Garg, C. Ram, R. Kushwah, and R. Muradia, “Analysis of factors influencing maternal mortality and newborn health—a machine learning approach,” J Med Artif Intell, vol. 7, pp. 1–1, Mar. 2024, doi: 10.21037/jmai-23-107.

[19] H. Bao et al., “How Can a High-Performance Screening Strategy Be Determined for Cervical Cancer Prevention? Evidence From a Hierarchical Clustering Analysis of a Multicentric Clinical Study,” Front. Oncol., vol. 12, p. 816789, Jan. 2022, doi: 10.3389/fonc.2022.816789.

[20] A. Victor, “The role of artificial intelligence in maternal and child health: Progress, controversies, and future directions,” PLOS Digit Health, vol. 4, no. 7, p. e0000938, Jul. 2025, doi: 10.1371/journal.pdig.0000938.

[21] M. Rahman, Md. F. Rabby, Md. M. H. Sakib, and Md. M. Hossain, “Examining regional disparities in maternal and child health in Bangladesh using cluster analysis of MICS 2019 data,” Sci Rep, vol. 15, no. 1, p. 32741, Sep. 2025, doi: 10.1038/s41598-025-03003-w.

[22] S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif, and S. Z. H. Kazmi, “K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir,” IEEE Access, vol. 8, pp. 151847–151855, 2020, doi: 10.1109/ACCESS.2020.3014021.

[23] MSPS, Santé en chiffre 2022

[24] Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. Introduction to Data Mining. New internat. edition. Always Learning. Harlow: Pearson, 2014

[25] Ministère de la Santé (Maroc), “Manuel Pratique sur la vaccination à l’intention des Professionnels de Santé,” 2016.

[26] H. Zerouali, “PROGRAMME DE SURVEILLANCE DE LA GROSSESSE ET DE L'ACCOUCHEMENT(PSGA),” 2013

[27] Ministère de la Santé (Maroc), “Programme National de Nutrition,” 2018, [Online]. Available: http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

**APPENDIX**

A: list of features

| Domain | Subdomain | Variables |
|---|---|---|
| Prenatal consultation | Prenatal consultation services | Total prenatal consultations (NI) |
| | | Total prenatal consultations (AI) |
| | | First trimester prenatal consultation (NI) |
| | Medicalized CPN and diagnosed high-risk pregnancies | Medicalized prenatal consultations |
| | | High-risk pregnancies (Diagnosed) |
| | | High-risk pregnancies (Supported) |
| Childbirth and postpartum care. | Deliveries performed in public delivery facilities | Deliveries performed in birthing centers |
| | | Deliveries performed in hospital maternity wards |
| | | Method of delivery (Vaginal delivery) |
| | | Mode of delivery (Caesarean section) |
| | | Direct obstetric complications (Diagnosed) |
| | | Direct obstetric complications (Supported) |
| | | Direct obstetric complications (Deceased) |
| | | Total obstetric complications (direct and indirect) (Diagnosed) |
| | | Total obstetric complications (direct and indirect) (Covered) |
| | | Total obstetric complications (direct and indirect) (Deceased) |
| | Births registered at public delivery facilities | Stillborn |
| | | Live births |
| | | Newborns weighing <2500g |
| | | Newborns who died < 24 hours |
| | | Number of premature newborns |
| | Diagnosed complications and number of malformations in newborns recorded in public delivery facilities | Complications in newborns diagnosed |
| | | Complications in the newborn Supported |
| | | Number of newborns with malformations |
| Postnatal consultation | Postnatal consultations for the mother | Number of women examined in post-natal consultations (early cpon) |
| | | Number of women examined in post-natal consultations (late cpon) |
| | | Number of women examined in post-natal consultations (post-natal consultations outside the time limits) (recommended) |
| | | Number of postpartum obstetric complications (Diagnosed) |
| | | Number of obstetric complications in postpartum (Supported) |
| | Postnatal consultations for the newborn | Number of newborns examined in postnatal consultations (early cpon) |
| | | Number of newborns examined in postnatal consultations (late cpon) |
| | | Number of newborns examined in postnatal consultations (postnatal consultations outside the recommended timeframes) |
| | | Postnatal consultations for the mother-newborn couple under medical supervision |
| | | Number of complications in postpartum newborns (Diagnosed) |
| | | Number of complications in postpartum newborns (Supported) |
| Immunization | Children born protected and the number of children vaccinated by antigen (Hep.B1, BCG and VPO.O), | Number of children born protected |
| | | Children vaccinated (Hep.B1) Done at birth |
| | | Vaccinated children (Hep.B1) Done with BCG |
| | | Children vaccinated with BCG |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| | | |
|---|---|---|
| | | Children vaccinated with OPV.1 |
| | Children vaccinated with antigen (OPV and DTP-Hib-HB (Penta)) | Children vaccinated with VPO 1st dose |
| | | Children vaccinated with VPO 2nd dose |
| | | Children vaccinated with VPO 3rd dose |
| | | Children vaccinated with DTP-Hib-HB (Penta) 1st dose |
| | | Children vaccinated with DTP-Hib-HB (Penta) 2nd dose |
| | | Children vaccinated with DTP-Hib-HB (Penta) 3rd dose |
| | Children vaccinated with antigen (IPV and Rotavirus), | Children vaccinated with IPV |
| | | Children vaccinated with Rotavirus 1st dose |
| | | Children vaccinated with Rotavirus 2nd dose |
| | | Children vaccinated with Rotavirus 3rd dose |
| | Children vaccinated with antigen (Pneumococcus and RR), | Children vaccinated against Pneumococcus 1st dose |
| | | Children vaccinated against Pneumococcus 2nd dose |
| | | Children vaccinated against Pneumococcus 3rd dose |
| | | Children vaccinated with RR 1st dose |
| | | Children vaccinated with RR 2nd dose |
| | Children vaccinated by antigen (OPV.O booster and DTP booster) | Children vaccinated with VPO booster.0 |
| | | Children vaccinated with DTP booster |
| | Distribution of women of childbearing age who received the tetanus and diphtheria vaccine | Women of childbearing age who received the tetanus and diphtheria Td1 vaccine |
| | | Women of childbearing age who received the tetanus and diphtheria Td2 vaccine |
| | | Women of childbearing age who received the tetanus and diphtheria Td3 vaccine |
| | | Women of childbearing age who received the tetanus and diphtheria Td4 vaccine |
| | | Women of childbearing age who received the tetanus and diphtheria Td5 vaccine |
| | Distribution of pregnant women who received the tetanus and diphtheria vaccine | Pregnant women who received the tetanus and diphtheria Td1 vaccine |
| | | Pregnant women who received the tetanus and diphtheria Td2 vaccine |
| | | Pregnant women who received the tetanus and diphtheria vaccine Td3 |
| | | Pregnant women who received the tetanus and diphtheria vaccine Td4 |
| | | Pregnant women who received the tetanus and diphtheria vaccine Td5 |
| Nutrition | Children under 5 years of age who attended ESSPs according to their nutritional status | Children under 5 years of age who attended ESSP according to their nutritional status: Underweight |
| | | Children under 5 years of age who attended ESSPs according to their nutritional status: Growth retardation |
| | | Children under 5 years of age who attended ESSPs according to their nutritional status: Malnutrition |
| | | Children under 5 years of age who attended ESSP according to their nutritional status: Overweight |
| | | Children under 5 years of age who attended ESSP according to their nutritional status: Obesity |
| | Preventive vitamin A and vitamin D supplementation benefits for children under 5 years of age | Vitamin D 1st Take for children under 5 years old |
| | | Vitamin D 2nd dose for children under 5 years old |
| | | Vitamin A 1st Take for children under 5 years old |
| | | Vitamin A 2nd dose for children under 5 years old |
| | | Vitamin A 3rd Take for children under 5 years old |
| | Curative vitamin A and zinc supplementation benefits for children under 5 years of age | Curative vitamin A supplement 100,000 1st dose for children under 5 years old |
| | | Curative vitamin A supplement 100,000 2nd dose for children under 5 years old |
| | | Curative vitamin A supplement 100,000 3rd intake for children under 5 years old |
| | | Curative vitamin supplement 200,000 1st dose for children under 5 years old |

| | | |
|---|---|---|
| | | Curative vitamin supplement 200,000 2nd dose for children under 5 years old |
| | | Curative vitamin supplement 200,0003rd Take for children under 5 years old |
| | | Children who received Zinc for children under 5 years old |
| | Preventive iron and vitamin D supplementation services for women in CPN and CPON | Preventive iron and vitamin D supplementation for women in CPN and CPON (Iron Supplementation) |
| | | Preventive iron and vitamin D supplementation for women in CPN and CPON (Vitamin D Supplementation) |
| Medical consultation for children under 5 years old | Infants (0 to 2 months) sick depending on the illness | Number of infants examined |
| | | Infants (0-2 months) sick: Bacterial infection |
| | | Infants (0 to 2 months) sick: Jaundice |
| | | Infants (0 to 2 months) sick: Diarrhea |
| | Children (2 to 59 months) sick according to the illness | Children (2 to 59 months) who are ill: Serious problem (Very serious illness) |
| | | Children (2 to 59 months) who are sick: Coughing problem |
| | | Children (2 to 59 months) sick: Diarrhea problem |
| | | Children (2 to 59 months) who are sick: Angina problem |
| | | Children (2 to 59 months) sick: Ear problem |
| | Children (2 to 59 months) sick according to the illness | Children (2-59 months) ill: Total number of children with clinically diagnosed meningitis |
| | | Children (2 to 59 months) sick: Measles |
| | | Children (2 to 59 months) who are sick: Anemia problem |
| | | Children (2 to 59 months) who are sick and are cared for |