Notater
Documents
2017/3

Statistics Norway

Statistisk sentralbyrå

John Åge Haugen and Dag Roll-Hansen

The health management Information system in Malawi

Assessment of data quality and methods for improvement

Documents 2017/3

John Åge Haugen and Dag Roll-Hansen

The health management Information system in Malawi

Assessment of data quality and methods for improvement

Statistisk sentralbyrå • Statistics Norway
Oslo–Kongsvinger

In the series Documents, documentation, method descriptions, model descriptions and standards are published.

Published 18 January 2017

ISBN 978-82-537-9475-4 (electronic)

Symbols in tables	**Symbol**
Category not applicable	.
Data not available	..
Data not yet available	…
Not for publication	:
Nil	-
Less than 0.5 of unit employed	0
Less than 0.05 of unit employed	0.0
Provisional or preliminary figure	*
Break in the homogeneity of a vertical series	—
Break in the homogeneity of a horizontal series	|
Decimal punctuation mark	.

Preface

Statistics Norway (SN) contributes to improvement of statistical systems in many developing countries. This report is based on Statistics Norway cooperation with the Malawian authorities on development of health statistics.

Information on available resources and service production in the health sector is crucial for any government to manage health services properly. Collecting information of high quality has however proven to be a challenging task. This document provides an assessment of data from the Health Management Information Systems (HMIS) in Malawi and suggests ways to improve it. Activities on this project were carried out in 2014/2015.

A special thanks to Geir Hjemås and Remy Bråthen at Statistics Norway, who have developed a general analytical tool for analysing data quality based on a statistical approach.

Thanks to the Norwegian Agency for Development Cooperation (NORAD) for funding this report.

Statistics Norway, 16 December 2016

Bjørnar Gundersen

Abstract

Information on available resource- and service production for the health sector is crucial for any government in order to manage it well. Collecting information of high quality on the health sector has however proven to be a challenging task.

This report is an assessment of data from the Health Management Information Systems (HMIS) in Malawi and suggests a way to improve data quality. The assessment is based on a "desk study" of the Malawian HMIS data. A statistical approach to improve data quality is presented, and it is used to identify possible data errors reported by the health facilities. Among other, the goal is to show a method for prioritizing the use of limited editing resources where they can improve quality the most.

This quality assessment of the Malawian HMIS data has 4 objectives:

1. Identify areas for initial improvement of data quality
2. Present general guidelines for data editing
3. Present use of statistical methods to improve data quality
4. Recommend future activity

This quality assessment is from 2014/2015 (desk study) and focuses on reported Malawian HMIS data for 2013, but for some analysis 2012 data is also included.

Missing information is a severe threat to the Malawian HMIS data. A first step should be to reduce the amount of missing information. Further, it is also recommended to prioritize identifying health facilities that have errors influencing the end result at aggregated level.

Improving quality by analysing already collected data to point out weaknesses in the data is an efficient approach. By doing this, a highly recommended process of working in a "circular" manner to improve quality will be established.

The Health Information Systems Programme (HISP) is a global network emerging from the Department of Informatics at the University of Oslo. HISP has developed the District Health Information system (DHIS), a free open source software specialized for HMIS. Today HMIS data in Malawi is collected using DHIS2. The methods for editing suggested in this report can be implemented in DHIS2.

The aim of the suggested approach for improving data quality is to: Avoid missing data from health facilities, correct major errors and improve the source.

Contents

1. Background

A health information system (HIS) is important for monitoring health, and for evaluating and improving the delivery of healthcare services and programs[1]. The World Health Organization (WHO) defines a Health Information System (HIS) *as a system that integrates data collection, processing, reporting and use of the information necessary for improving health service, health service effectiveness and efficiency through better management at all levels of health services*[2]. A Health Management Information System (HMIS) is generally a central part of this. Even though HMIS is important in developing a healthcare service, data in many developing countries struggle with quality problems resulting in incomplete, inaccurate and untimely information which is not useful for health decision-making[3]. Therefore, documenting the actual quality of HMIS data is an important first step to improve quality. The scope for this quality assessment is to evaluate the quality of data collected by the Malawian Government through their HMIS, and to demonstrate methods that will improve data quality that are easy to implement in the data production system.

2. Quality assessment

The quality assessment in this report consist of two approaches: In chapter 4 a practical approach will be presented; data will be analysed in terms of what can be seen as the most obvious data challenges and how these can be addressed. It will be argued that the first and most important step is to reduce the amount of missing information. In chapter 5 a statistical approach to improve data quality will be presented, and in particular identify facilities where registered information should be looked at closer to verify data quality, and if it is necessary, how to prioritize and improve their quality through editing.

2.1. Description of data

The Health Information Systems Programme (HISP) is a global network established, managed and coordinated by the Department of Informatics at the University of Oslo. HISP was started in the 1990s and today it involves a significant number of countries world-wide. HISP developed the District Health Information system (DHIS), a free open source software specialized for HMIS. Today it used by almost 50 countries around the world.

Potentially every health facility can report data using the online DHIS2 system. But in practice, in many countries including Malawi, paper forms are filled out at facility level. These forms are collected by the health districts where the data is registered in to the DHIS2 system.

Malawi registers a large number of health related variables covering many years on facility level. And as a consequence, there is a massive amount of data in the Malawian DHIS2 database. Therefore, to narrow down the scope of this quality assessment, the variables from HMIS 15 (health management information system) have been chosen. The HMIS 15 are a group of variables defined as important by the Ministry of Health in Malawi. Only three of these variables will be used as examples, these are: Total personnel currently, HMIS Bed capacity and HMIS Fully immunized children less than one year.

[1] *DHIS2: The Tool to Improve Health Data Demand and Use in Kenya*
[2] *Developing Health Management Information Systems: A practical guide for developing countries*
[3] *DHIS2: The Tool to Improve Health Data Demand and Use in Kenya*

Furthermore, figures on national level and from five health districts will be presented: Blantyre, Chiradzulu, Chitipa, Dedza and Ntchisi. For the health districts Chiradzulu, Dedza, and Blantyre figures will also be presented at facility level. The districts are selected to cover the different zones and both rural and urban settings.

This quality assessment was performed in 2014/2015 and focuses on registered HMIS data for 2013, but for some analysis 2012 data is also included. The DHIS2 system was rolled out in Malawi in 2010- 2012 it is therefore important to be aware of possible quality issues the first years after this system migration.

2.2. Objectives:

This quality assessment of the Malawian HMIS data has 4 objectives:

1. Identify areas for initial improvement of data quality
2. Present general guidelines for data editing
3. Present use of statistical methods to improve data quality
4. Recommend future activity

2.3. Main findings

Missing data is a severe threat to the Malawian HMIS reported for 2013. Even though completeness improved significantly from the year before, showing a positive trend.

A facility not registering data will, in most cases, have a negative impact on data quality. It is recommended to identify where limited resources ought to be spent to achieve cost-efficient improvements of quality. Improving completeness of Malawian HMIS data is an important first step.

2.4. Recommendations and future activity

This quality assessment does not address all challenges related to the HMIS data in Malawi. It does however point to important quality issues and suggest ways to improve on them. Improving and maintaining quality will often imply that the quality improvement process is an integrated routine in the reporting system.

SN suggests the following steps for efficient data quality improvements:

1. *Evaluate necessity of all variables reported.* If data do not serve an important purpose, they should not be collected. If the scope is too wide and, for instance, reporting is difficult and time consuming it will give a high respondent burden and could have a negative effect on data quality. It should be an aim to reduce the number of variables reported through DHIS2 in Malawi.
2. *Define report frequency and level of dissemination, both in terms of geography and time*. If reports are unnecessary time consuming because of a high report frequency it will have a negative effect on quality because resources to report data are limited. The level of data dissemination is important to assess at what level data needs to have acceptable quality.
3. *Implement statistical methods to prioritize use of feedback resources*. The aim is to focus resources on editing data from facilities that may influence the end result (aggregated level). Further to implement a system for feedback to facilities. This can for instance be done by a team of two to three staff members in the Ministry of Health, calling health facilities with missing data or suspected errors. The aim is to verify if the data is correct, and guide the facility in what should be reported. Quality improvement is dependent on a well-functioning method for communication between national level, districts and facilities, and involves editing data. Changes in data should generally be done in cooperation with the reporting unit. Statistical methods are very effective when used to point out where to focus limited editing resources.

4. *Statistical methods ought to be implemented in DHIS2.* By developing an editing/quality module, based on well-known statistical methods the whole DHIS2 community may have access to effective means to edit data and increase quality. The suggested module ought to be based on modern techniques for data editing and experiences from Malawi.
5. *Evaluate and improve paper forms for the facilities and clarify definitions*. HMIS data in Malawi is generally first registered on paper before it is registered into the DHIS2-system, therefore these reports are important for the quality of collected data. An evaluation and improvement of the paper reports using questionnaire methodology have, in similar cases as this, well documented quality gains. This work ought to build on experiences from data editing since this is where quality issues are identified.
6. *Use DHIS2 as a data edition tool and implement imputation methods*. It is advisable to develop a method for imputation of missing values. Imputation can never replace direct contact with facilities in ensuring good data quality, but it can work as a backup when missing data is difficult, expensive or too time consuming to collect. This will reduce the impact of missing data at district and national level, but facilities with imputated values may need to be followed up closely.

The aim of these suggestions is first and foremost to improve data quality by avoiding missing data and thereafter to correct major errors. But most importantly to establish a data collection system were data quality issues are analysed based on collected data. Then necessary action can be taken to improve data quality, either through editing or improving the whole data collection system.

2.5. Improving quality

The HMIS-data in Malawi is not unique in terms of struggling with quality issues; this can be said about almost every statistical/register-system. Improving data is most efficient when performed in a circular way, meaning to start with an analysis of the end product and identifying troublesome areas of the data. As a result, improving quality involves making changes in the production process based on analysis of the end product. When suggested changes are applied, the end product should again be analysed to evaluate these changes. This circular process is best thought of as a “never ending story” and step by step; data quality will improve.

In general, there are two ways to improve data:

1. Improve quality on input data
2. Editing data

These two approaches are dependent of each other, as the process of editing data will identify troublesome areas quality wise in the data, and give valuable input to how input data can be improved. It is difficult to make data with poor quality good trough editing, but editing is an important part of the process of increasing quality.

In statistics one usually defines two main approaches in analysing data quality:

1. Knowledge of field and context
2. Statistical methods

A mix of these approaches is the most beneficial. Knowledge of the field and context provides an understanding of the data which is undeniably important. But, it can be less systematic and dependent on individuals. Statistical approaches will typically be more systematic, but not necessarily better. Statistical methods work well to get an overview of the data, to point toward areas of the data that should be checked, and to provide a way of prioritizing limited resources for data improvement. Therefore, combining the knowledge of the field and context with statistical methods is an effective approach to improve data quality.

3. Practical approach

In this analysis of the HMIS 15 variables it will be argued that the first and most important step at this stage is to reduce the amount of missing data. First and foremost, focus will be at causes for fluctuating values of one variable from month to month. This can be a sign of quality issues; typically missing values or faulty data. Core data in the Malawian HMIS (HMIS15) is registered every month, and from there aggregated up to quarterly and annual figures. The method used here to uncover possible quality issues is to “drill down” from national level to, health district level and at last all the way down to facility level. Since the main focus here is data fluctuations, it is important to know that seasonal changes can be “natural” and not necessarily a sign of quality issues.

There are three variables from HMIS15 that will be examined:

1. Total personnel currently
2. Bed capacity
3. Fully immunized children less than one year (Fully immunized < 1 year)

3.1. Completeness

An important starting point quality wise for the Malawian HMIS, is to have as many of the health facilities as possible registering data into the DISH2-system. For a HMIS to be able to give valuable information about the health sector it is obvious that enough data on facility level is needed. Missing data, either a facility not registering any data or not registering completely, will have a negative impact on data quality, and potentially have profound effect when data is aggregated to health district and national level. It is advisable to identify where limited resources should be used to gain most profit in terms of quality improvements, and for the Malawian HMIS improving on completeness is an important first step.

In DHIS2 there is a tool for analysing response rate, and it displays a high response rate of 92 percent for Malawi in 2013 on the HMIS 15 variables. Compared to 2012 when the response rate was close to 51 percent, identifies that a strong improvement has taken place (see table 1).

Table 1 Response rate for HMIS 15 in DHIS2. Malawi, health districts 2012 and 2013

Health districts	2012 HMIS 15	2013 HMIS 15
MALAWI	**50,5**	**92,1**
Blantyre	43,8	96,7
Chiradzulu	29,8	79,2
Chitipa	47,5	84,2
Dedza	26,0	72,3
Ntchisi	99,2	100

Response rates, for both Malawi and five of the health districts (Table 1), are in general quite good, and it is also an impressive improvement from one year to another. But even though the response rate is high in 2013 it only tells part of the story. Because this figure only registers a delivered report, it does not say anything about the quality of the report, for instance how many variables that have missing information. This issue will be addressed later in this assessment.

As a starting point for increasing quality a 92 percent response rate is very good. This gives evidence that the DHIS2 system reaches out to facilities nationwide in Malawi, meaning the data collection system works.

3.2. Total personnel currently: Malawi

The first variable that will be evaluated is *Total personnel currently*. This is an important variable; it gives the Government an overview of man power in the health sector. Further it also indicates the size of the facility. In many cases large facility with many employees will have more activity than a small facility, and can therefore have a larger effect on aggregated figures.

By the end of 2013 almost 27 500 personnel were reported working in the Malawian health sector. In Figure 1 the *Total personnel currently* for 2013 is broken down on months. At first sight the number of health workers seems quite stable from month to month, but there are some fluctuations. For instance, there is a marked increase of more than 13 percent, or 3 227 health personnel, from May to November in 2013.

Figure 1 Total personnel currently on months. Malawi. 2013

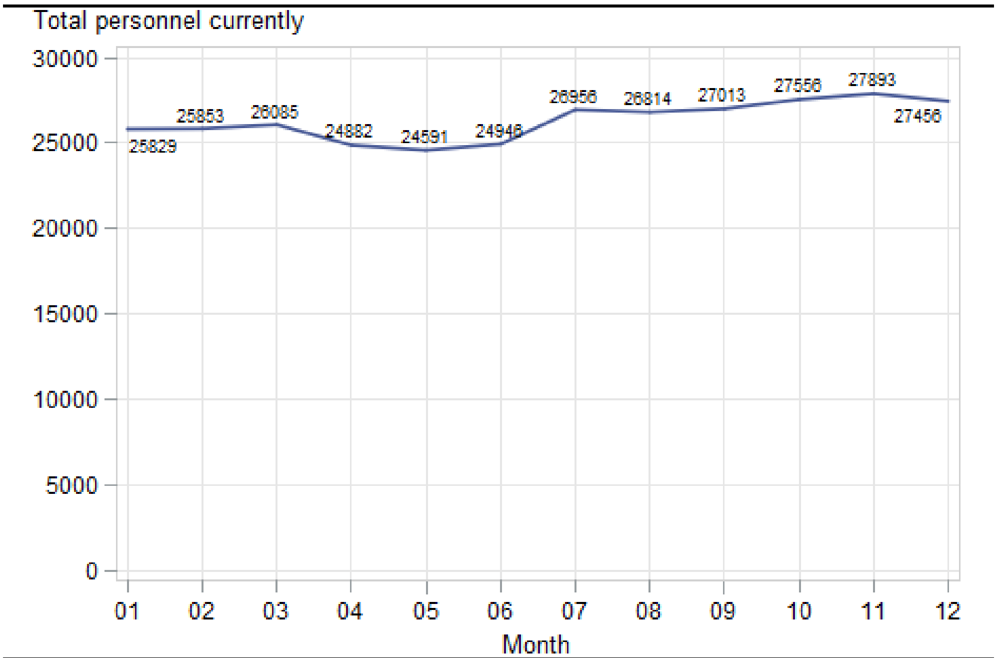
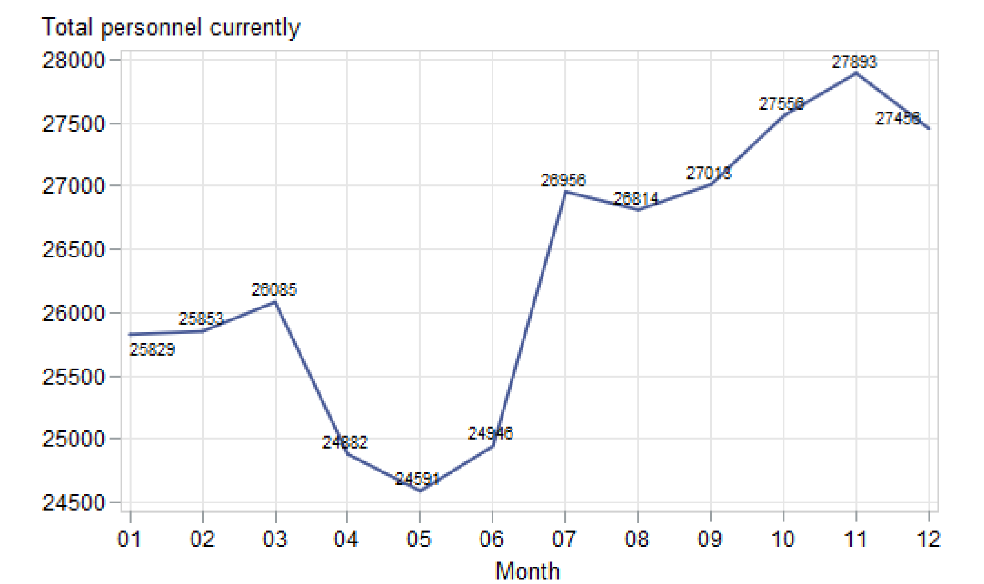


Figure 2 Total personnel currently on months. Malawi. 2013



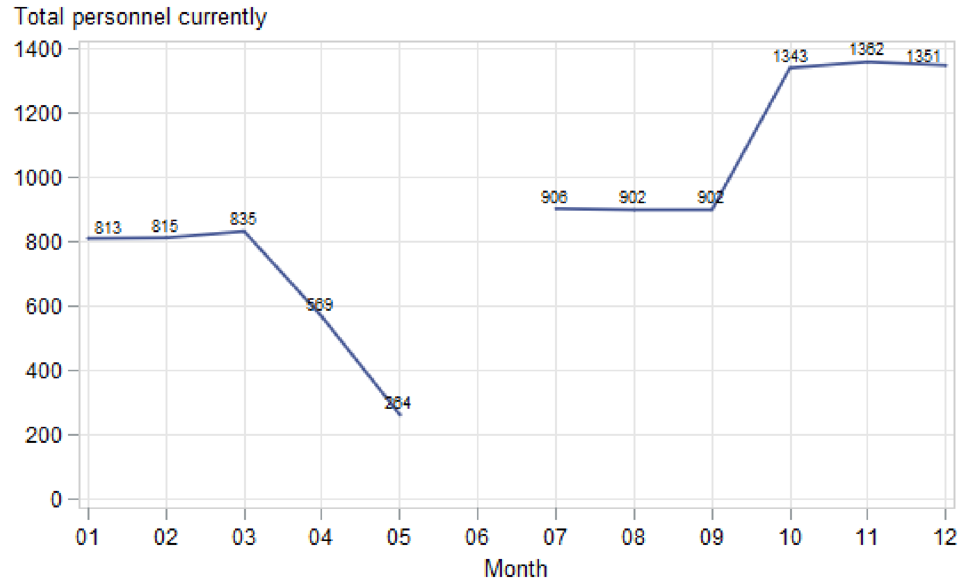
Zooming in at the monthly fluctuations for *Total personnel currently* (see figure 2) it becomes clear that there is a striking drop in the value of this variable from March to July compared to the rest of 2013.

The national figure of the total number of health personnel should, in normal circumstances, be quite stable from one month to another. But local context like holidays and understating of what actually should be registered can influence data. Here it turns out that some of the reason for the drop in values from March to July can be found at health district level, and Chiradzulu health district will be used as an example.

Chiradzulu health district

Chiradzulu health district has registered fluctuating values on the variable *Total personnel currently* in 2013 (see figure 3). The total number of registered health personnel increases almost by 1 100, or 415 percent, from May to November 2013. There is a fast decline in the registered health personnel from March to May, and there is no registered data for June 2013.

Figure 3 Total personnel currently on months. Chiradzulu health district. Malawi. 2013

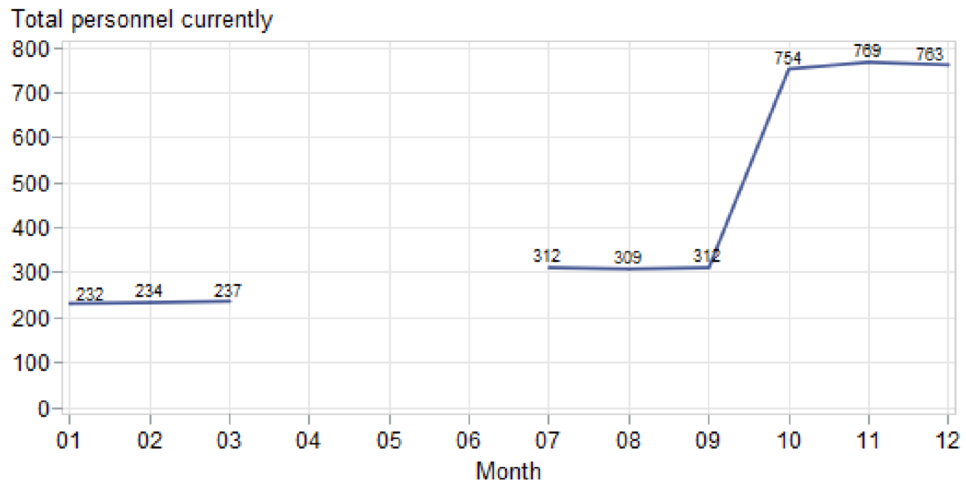


The Health district level is aggregated from the registered data from every health facility in that district. There will be small and large facilities, in terms of the figures they report on the different variables. A hospital will typically have a larger impact on data then a small countryside clinic.

In Chiradzulu health district the largest facility is Chiradzulu district hospital. Figures reported for this unit will have a strong influence on the figures for the whole district. Chiradzulu district hospital has not registered any values for the variable *Total personnel currently* from March to June 2013 (Figure 4). Figures for this hospital are also more than double from October to December compared to the rest of the year. Registered figures from Chiradzulu district hospital is a major weakness for the quality of data for the whole health district on the variable *total personnel currently*. By prioritizing use of resources in connection with feedback and editing, a data failure like this will be set to high priority and will in most cases be corrected.

In addition, the second largest facility in terms of *Total personnel currently* in the health district, Nguludi St. Joseph's Hospital, has missing values for May and June, weakening the quality of this variable even more.

Figure 4 Total personnel currently on month. Chiradzulu district hospital. Malawi. 2013



Completeness

To analyse completeness at district level the following method is used. All facilities that have reported zero or no value every month for all HMIS15 variables is removed from the data set. Chiradzulu health district then remains with 14 reporting facilities. Each of these facilities are supposed to report 12 times every year, adding up to a total of 168 reports.

As mentioned earlier, a response rate does not necessary say much about the quality of the report, just that the reports are registered. Evaluating completeness of data means that the focus is shifted to individual variables, and how complete the registered reports are. The first variable that will be evaluated here is *Total personnel currently* for Chiradzulu health district. In this district the number of reports with missing data on the variable *Total personnel currently* was 13 percent in 2013 (see Table 2). In total this means that 22 of 168 monthly reports have not reported values on this variable, and as described above it has a profound effect on the figures for the whole district (see figure 2)

Table 2 Response rate Total personnel currently. Chiradzulu health district. Malawi. 2013

Total personnel currently	Frequency	Percent
Missing value	22	13.10
Not missing value	146	86.90

Parts of the National decline of registered values for *Total personnel currently* in March to July can partly be explained by missing values in Chiradzulu Health districts report. But, at the same time there are also major drops in figures for health district Salima and Mwanza in the time period March to July

On the bright side it seems that the data quality on the national figures for *Total personnel currently* has increased since 2012. In 2012 there were only registered 2 000 personnel currently for the six first months, and the figures for 2013 for the same period was above 24 000 each month.

3.3. Bed capacity: Malawi

Bed capacity is a variable that, like *total personnel currently*, should be quite stable from one month to another. In terms of quality this variable can also be used to identify large units that should be followed up more closely. Registered *bed capacity* for Malawi in 2013 have some marked fluctuations (see figure 5). From the lowest nationally registered *Bed capacity* in April to the highest in September 2013 there is a 22 percent climb, a total of 2 645 beds. The registered *Bed capacity* is especially low in January and April (see figure 6).

Figure 5 **Bed capacity on months. Malawi. 2013**

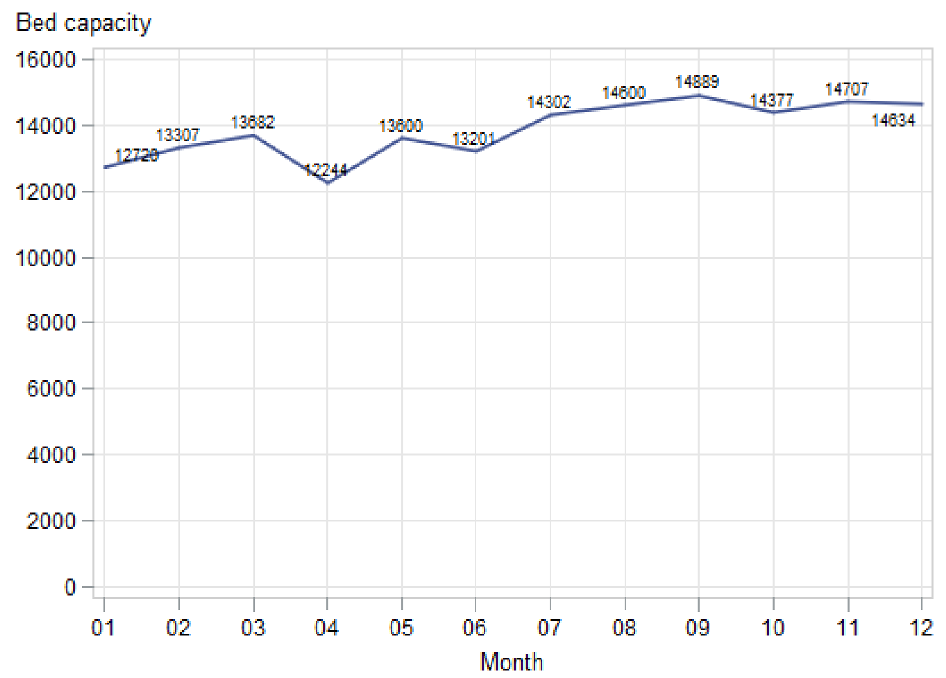
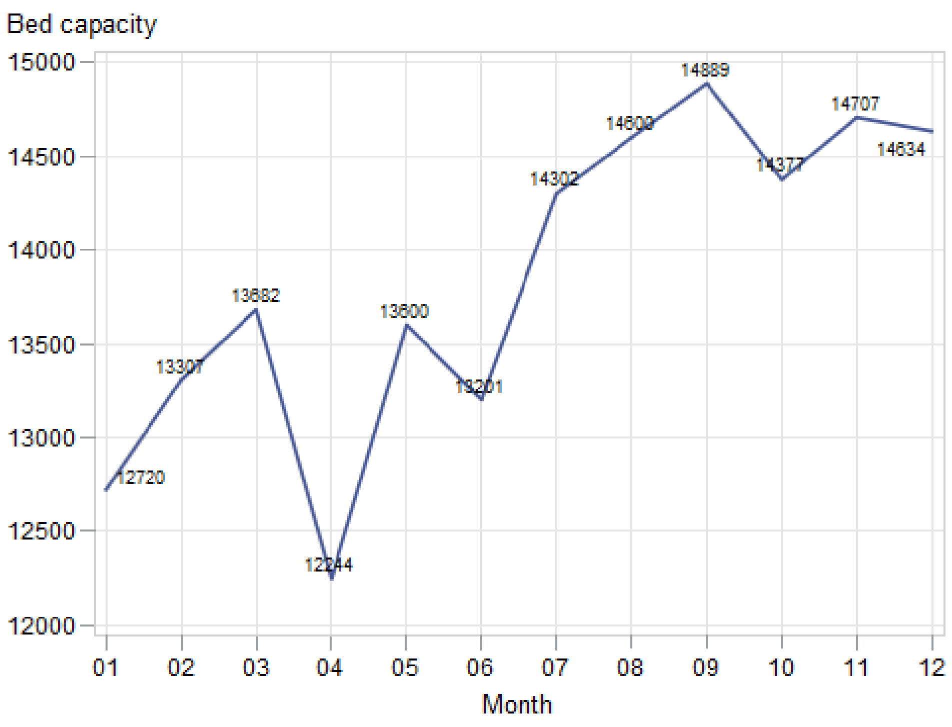


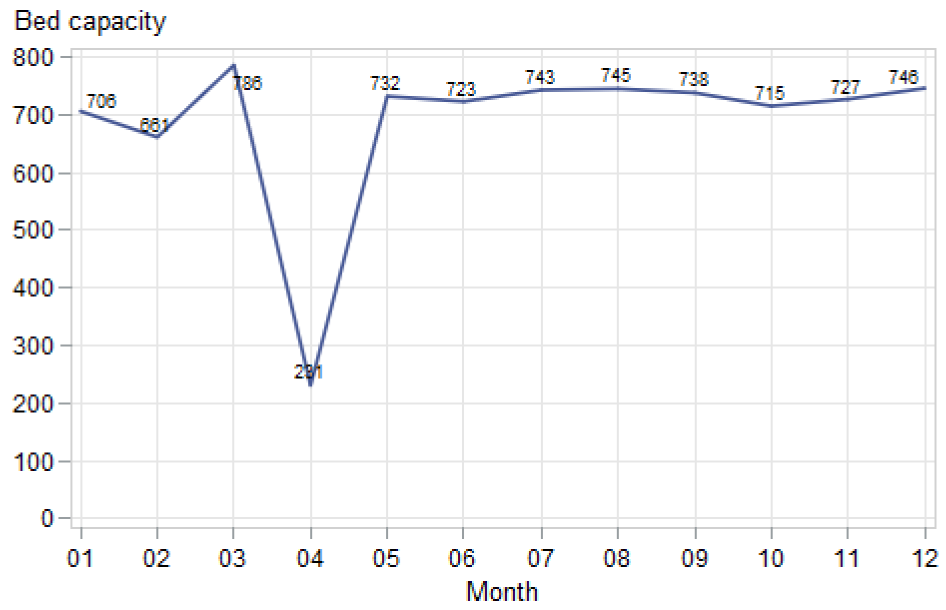
Figure 6 **Bed capacity on months. Malawi. 2013**



One way to get a better idea of the reasons for these variations is to drill one level down, from the national level to the health district level. There are some districts that show marked decreases in *Bed capacity* in April, two good examples are the health districts of Nkhotakota and Blantyre.

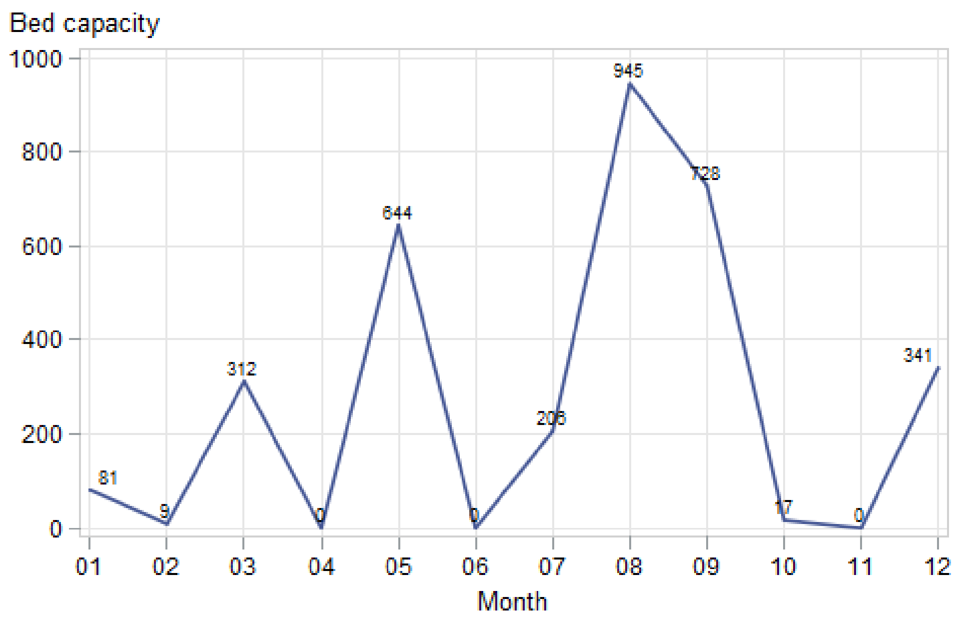
In Figure 7 the monthly *Bed capacity* of Nkhotakota health district is on display. As can be observed the monthly values are stable except for a major drop in April.

Figure 7 Bed capacity on months. Nkhotakota health district. Malawi. 2013



Blantyre health district on the other hand has reported very irregularly (see figure 8) on the variable *Bed capacity*. Figures from Blantyre vary from 0 to 945 from one-month no another, so here it is obvious that there are issues with the registered data.

Figure 8 Bed capacity on months. Blantyre health district. Malawi. 2013



In 2013 none of the facilities in Blantyre district had registered data every month on *Bed capacity*, and this will off course have a negative effect on data quality. If the facilities are large, the missing data will also have a profound effect on the district level figures.

Completeness

One of the main quality issues that have been identified for the variable *Bed capacity* at is that of missing data. It is therefore interesting to look at the response rate for the variable *Bed capacity* in 2013. The same method described in chapter 4.2.2 is used again, meaning that all facilities which have reported zero or no value for all HMIS 15 variables every month in 2013 are filtered out. In Blantyre health district, after this filtration, there are 29 reporting facilities, which should report 12 times each year to a total of 348 reports. In 2013 the total response rate for the variable *Bed capacity* was only 60 percent (see table 3). This means that a large portion of the facilities have not registered data for this variable each month, this has a negative impact on data quality resulting in uncertain figures on health district level, and it will also weaken the quality of the national figures for this variable.

Table 3 Response rate Bed capacity. Blantyre health district. Malawi. 2013

Bed capacity	Frequency	Percent
Missing value	138	39.66
Not missing value	210	60.34

Bed capacity 2012

Even though figures for Blantyre health district in 2013 seems to be lacklustre in terms of quality, compared to 2012 it the quality seems to be improving (see figure 9). For instance, the registered values in 2012 for *Bed capacity* are much lower compared to 2013, and there are no registered values at all from July to December in 2012.

Figure 9 Bed capacity on months. Blantyre health district. Malawi. 2012

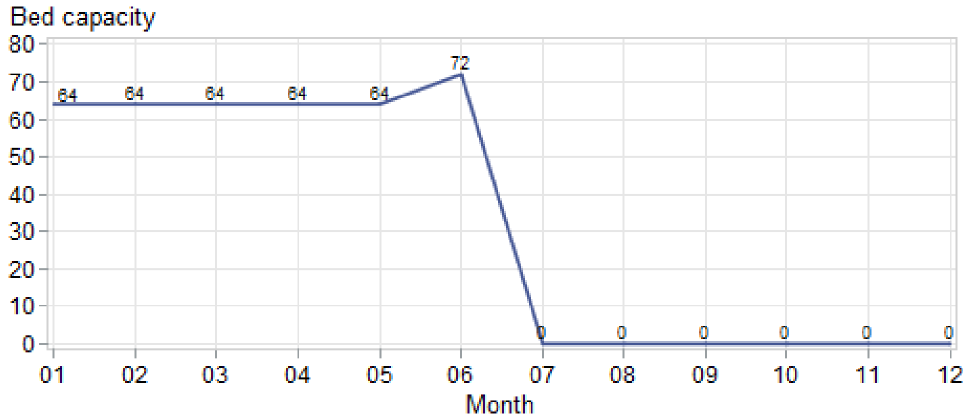
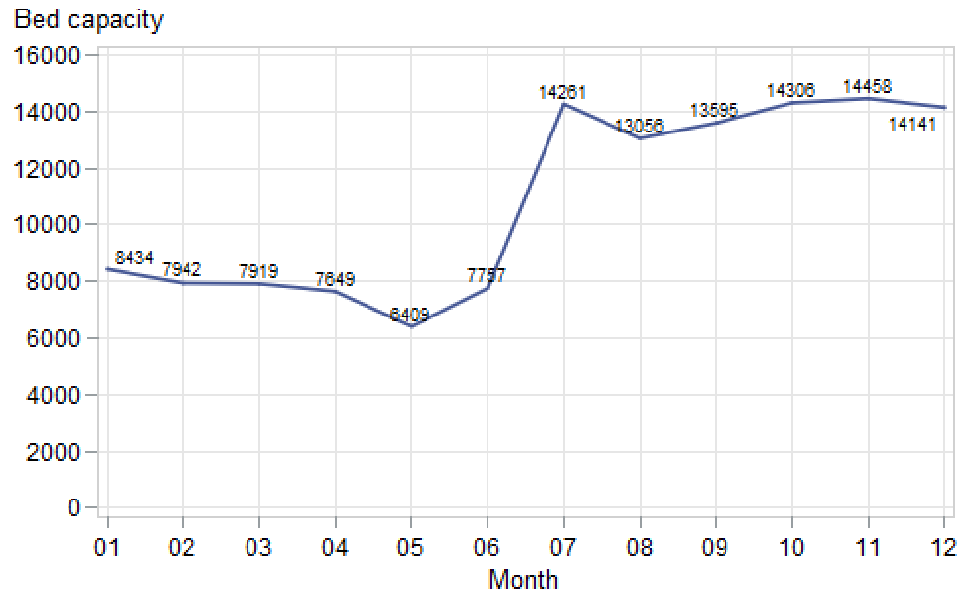


Figure 10 Bed capacity on months. Malawi. 2012



Another evidence that the data for *Bed capacity* is improving quality wise are the more stable figures reported for Malawi in 2013 compared to 2012 (see Figure 10 and 5).

3.4. Fully immunized children less than one year: Malawi

In the HMIS 15 data, the variable on *Fully immunized children less than one year (Fully immunized < 1 year)* gets a lot of attention, and it also provides important figures making it possible for the Government to monitor immunization. This is also a variable that should be quite stable during the year. In figure 11 and 12 the variable *for Fully immunized < 1 year* for Malawi in 2013 is on display. It can be observed that there are marked variations from month to month, and especially in the first half of 2013. From March to April there is a 15 percent drop or 7 241 children fully immunized. There is also a marked decline of 19 percent from July to December. Searching for reasons for these variations again leads us to the health district level. Two health districts, Dedza and Blantyre, will be used as examples to visualize how missing reports affect output data. Fluctuating data can for instance also be a result of stock out of certain vaccines and budget issues.

Figure 11 Fully immunized children less than one year on months. Malawi. 2013

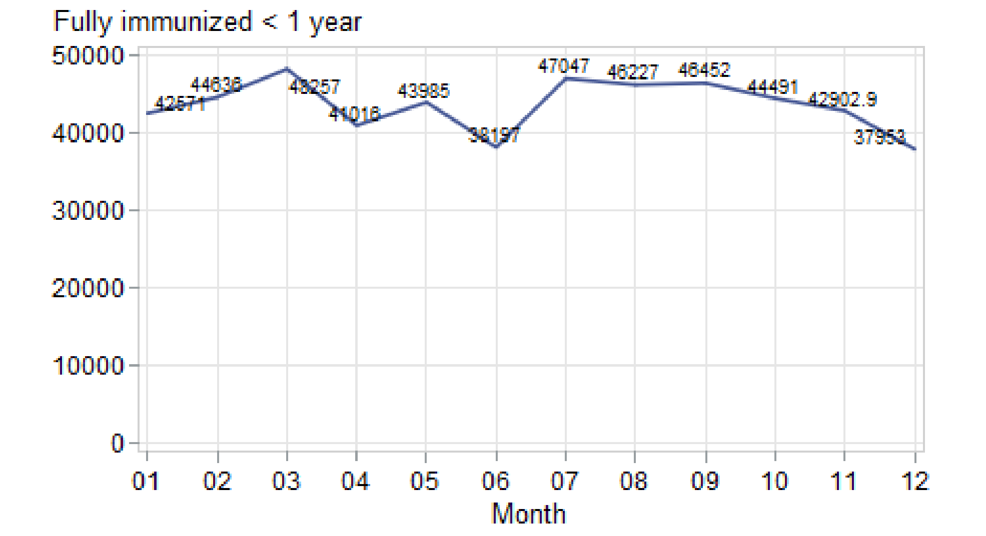
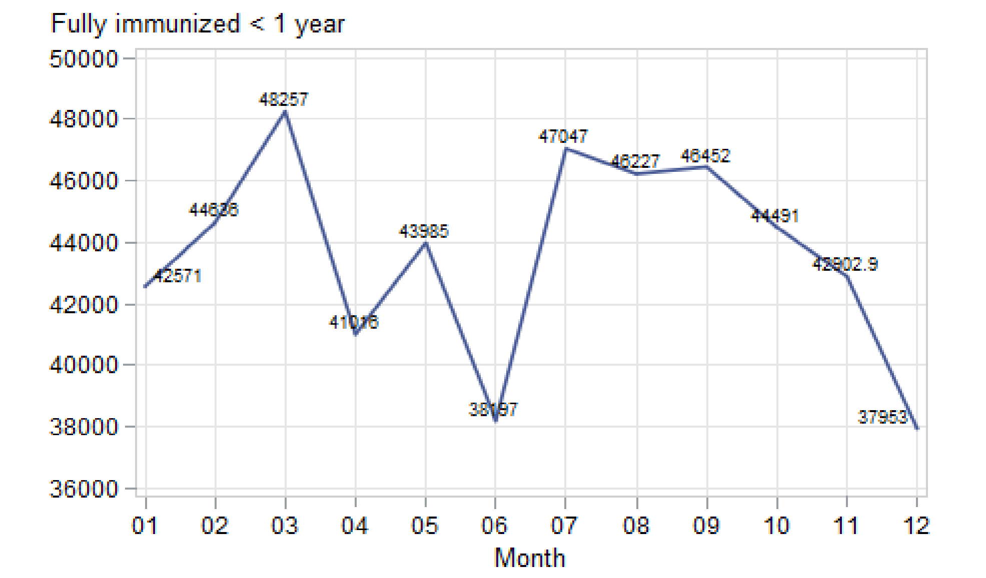


Figure 12 Fully immunized children less than one year on months. Malawi. 2013



Dedza health district

First out is Dedza health district, as can be observed in Figure 13 there are large variations from months to month, especially in the first half of 2013. From February to June the number of *Fully immunized < 1 year* decrease from 3 435 to 1 032 children, a 70 percent decrease. Then from June to July figures increases 177 percent, from 1 032 to 2 860 children. These are marked fluctuation at district level and some explanations for this are found at facility level.

Figure 13 Fully immunized children less than one year on months. Dedza health district. Malawi. 2013

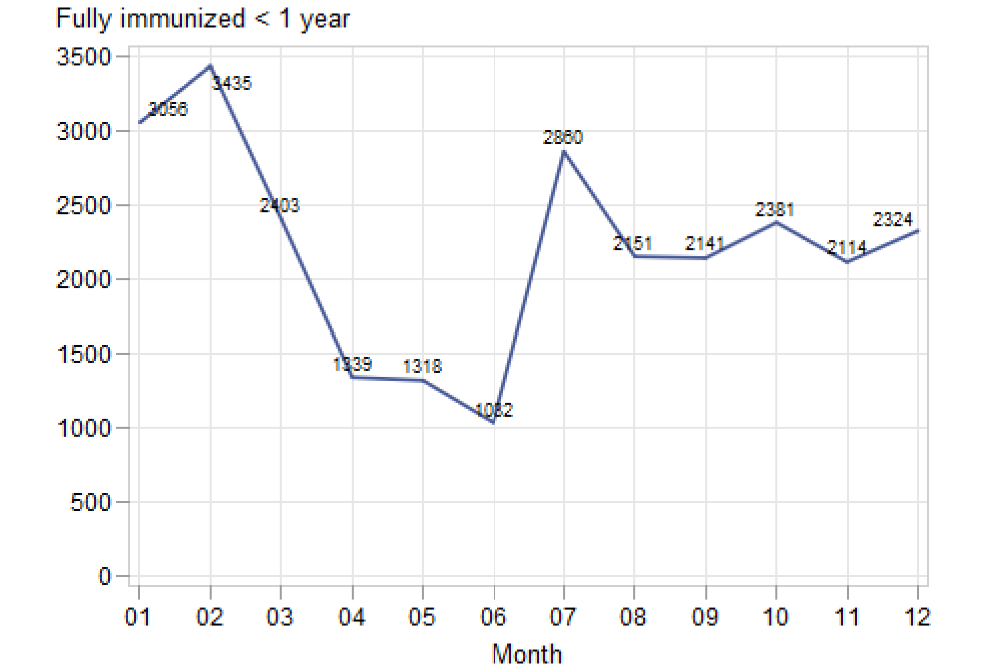
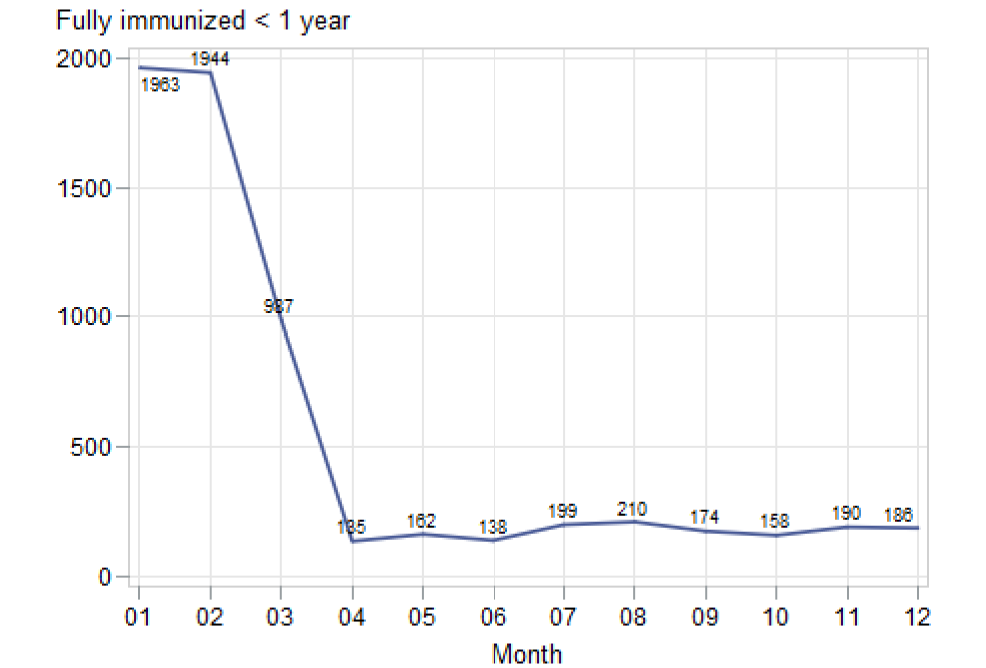


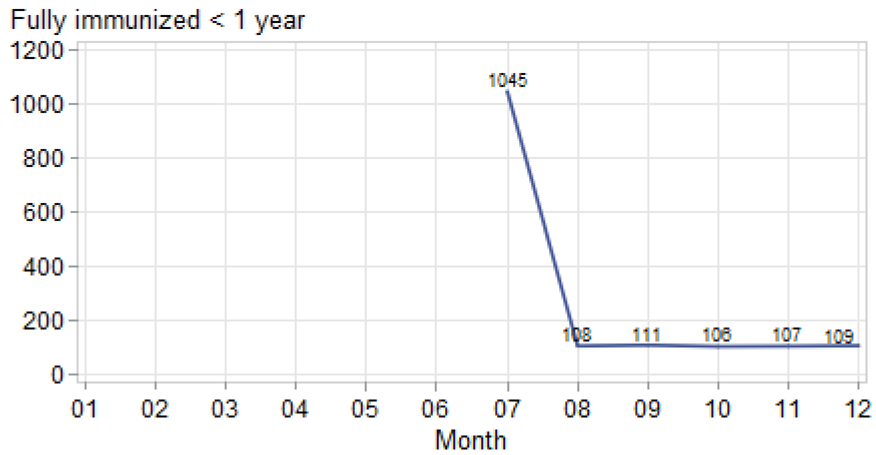
Figure 14 Fully immunized children less than one year on months. Dedza district hospital. Malawi. 2013



In Dedza health district the largest facility is Dedza district hospital. For this unit there is registered a profound decline in the total number of *Fully immunized < 1 year* from February to April in 2013 (see figure 14), a decline from 1 963 children

in January to 135 in April. This explains some of the decrease observed at district level.
The strong decline from February to June in figures for *Fully immunized < 1 year* from Dedza health district is followed by a marked rise again from June to July (see figure 13). Some of this rise can be explained by the reported data from Bembeke health centre in July 2013 (see figure 15). This facility had not reported any values in the time period from January to June 2013, but in July a total of 1 045 children less than one year are reported to have been fully immunized here. It can also be observed that after the July report, the figure sees a sharp decline to 108, and stays at this level to the end of the year.

Figure 15 Fully immunized children less than one year on months. Bembeke health centre. Malawi. 2013



Completeness

Again the method here is to filter out facilities with which has reported no data or zero every month of 2013 (se chapter 4.2.2 and 4.3.1). After the filtration 36 facilities are left in Dedza health district, each of these should report 12 times each month and in total 432 reports. Variation for the variable *for Fully immunized < 1 year* will also be dependent on the completeness of the registration. Of 36 facilities that have reported on *Fully immunized < 1 year* in 2013, only 15 facilities have reported each month in Dedza health district. The response rate for the variable *Fully immunized < 1 year* in Dedza health district for 2013 is 76 percent (See Table 4). When one fourth of the reports are missing, like here, it will lead to underreporting and in most circumstances to fluctuations in registered data. Again, increased completeness is therefore an important step to improve quality.

Table 4 Fully immunized children less than one year. Dedza health district. Malawi. 2013

Fully immunized < 1 year	Frequency	Percent
Missing value	104	24.07
Not missing value	328	75.93

Blantyre health district

Blantyre is the second health district that will be evaluated in terms of the variable *for Fully immunized < 1 year* in 2013. For this district a strong decline from January to December from 5 910 children per month to 2 208, or a decline of 63 percent (see Figure 16).

The decline throughout 2013 is based on declining figures from many of the largest facilities in Blantyre health district. An example is Bangwe Health Centre (see figure 16), which sees a decline from almost 900 to around 400 children fully immunized under 1 year.

Figure 16 **Fully immunized children less than one year on months. Blantyre health district. Malawi. 2013**

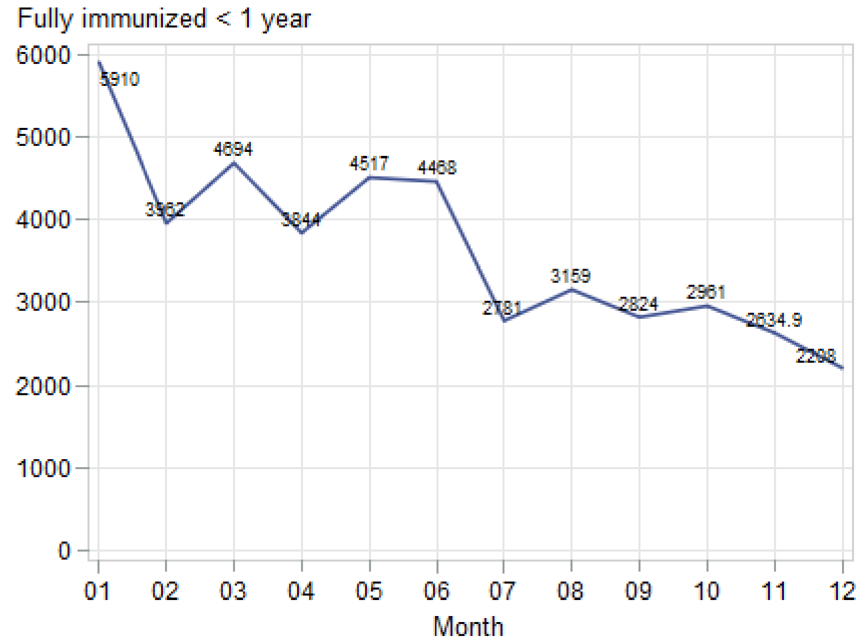
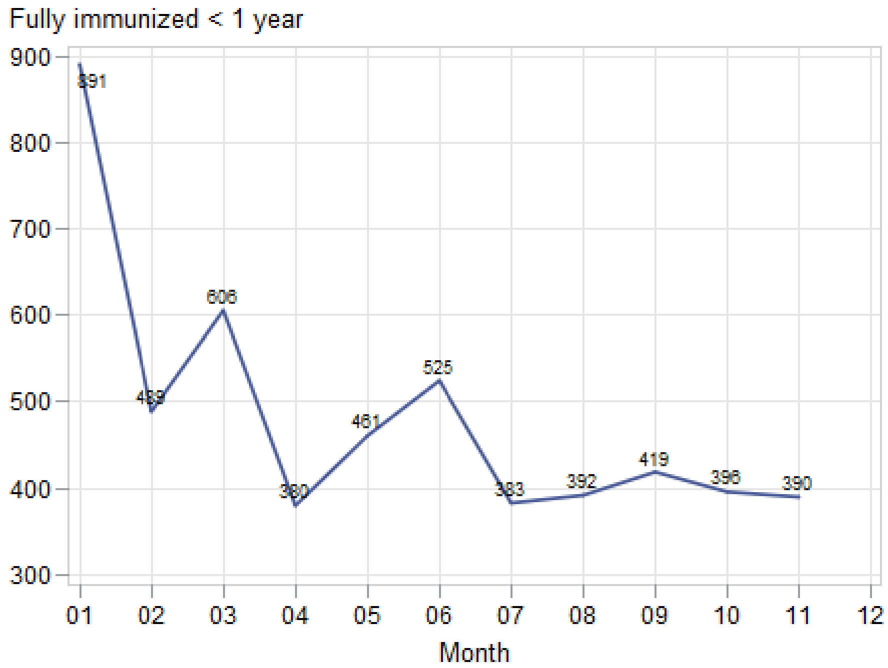


Figure 17 **Fully immunized children less than one year on months. Bangwe Health Centre. Malawi. 2013**



Completeness

There are 29 reporting facilities in Blantyre health district after all facilities that have reported zero or no value for all HMIS 15 variables every month in 2013 are filtered out. These facilities are supposed to report monthly each year, and in total this gives 348 reports. Of these reports 96 percent contain data on the variable *for Fully immunized < 1 year* is (see table 5). The high degree of completeness vastly reduces probability of negative impact on quality that missing reports have. In addition, an important parameter to review the quality is fulfilled with high degree of completeness and this opens up possibilities for more comprehensive quality analysis.

Table 5 Fully immunized children less than one year. Blantyre health district. Malawi. 2013

Fully immunized < 1 year	Frequency	Percent
Missing value	13	3.74
Not missing value	335	96.26

3.5. Conclusion and improvements
This part of the data quality assessment has been focused on missing and completeness of reports, and the examples chosen emphasize this. It is clear that data quality suffers when reports or variables are missing. When one health district does not report, it potentially has a critical effect on the national figures. This can also be said for missing reports for large facilities. Therefore, a first step should be to improve on completeness of the reports and as a starting point ensure that large facilities have reported data. There are many approaches to increase completeness of data reports. The most obvious is here is to work with the facilities, increasing their knowledge about the registration and why the report is important and give feedback on registered data.

4. Statistical approach

Statistical methods for improving data quality have a systematic approach and are powerful tools when used right. When implemented in a production system, these methods have the potential to vastly improve data quality. Still, it is important to acknowledge that statistical methods work best when combined with knowledge of the subject and context. There are a great number of statistical test and methods, and only a handful of them will be presented here. These are statistical tests Statistics Norway promotes as a first line of evaluating quality of reported data. Improving data quality has two main approaches; improve quality on input data and editing data. Statistical methods especially have an impact on the efficiency of data editing. The methods will help to prioritize where to focus limited resource by suggesting in what order units should be edited. It can also be used to evaluate automatic editing and imputation of missing values. There are three main approaches in the methods presented:

1. Comparing figures for two different years. This means to compare values for one or more variables between two years, thus identifying outliers. This is where the main focus will be in this quality assessment. There are different methods that can be used, and a few will be presented here.
2. Quartile methods compare the ratio between two variables within the same year. The logic here is that often there are connections between variables, and this can be used to identify quality issues in data. In practice this means to identify values that are acceptable but do not have a normal ratio compared to another variable in the data. An example of this will be presented.
3. Comparing edited data with raw data. Here data that have been through an editing process will be compared to data as they are when entered in to DHIS2. By comparing these two data sets the efficiency of the editing process can be evaluated. There is no available edited HMIS data for Malawi, but the method will be presented.

As mentioned earlier a process of improving quality relies on working in circle. This means not only to check the impact of performed editing, but also to improve on the causes making it necessary to edit data. By analysing causes and making changes to the production system, and again evaluating changes, input data will improve over time.

4.1. Raw data

One of the main principles when editing data is to keep a set of untouched raw data, meaning that editing is done on a copy of registered data. This is an important criterion because it makes it possible to keep track of and document changes that have been done. It is a key factor in evaluating the editing process in terms of influence on data, and also gives and understanding of how resources are managed.

4.2. Data overview

A preliminary overview of input data is an advisable place to start out an editing process. In table 6 the 2012 and 2013 data for the variable *Fully immunized < 1 year* are presented. The column called number of observations corresponds with the number of health districts in Malawi. As can be observed the figure is 29 both year. For Malawi there is reported an increase from 2012 to 2013 of 9 percent, or 43 557 children, on *Fully immunized < 1 year*. This can be a reasonable national figure. Comparing 2012 and 2013 data for the variable *Fully immunized < 1 year* for all 29 health districts display large differences, some districts increase and some decrease and to different degrees (see Table 7). The point here is to get an overview, and here it demonstrates that a large alteration in percent does not mean that it necessary will have a large impact on national figures. For instance, the 25 percent increase in Lilongwe represents more than 14 000 children, but the 60 percent increase in Mwanza only represent 1 323 children (see table 7). This is an issue that the Top down method addresses, which is presented in the next chapter.

Table 6 Fully immunized children less than one year. Malawi. 2012 and 2013

Number of observations 2012	fully immunized < 1 year 2012	Number of observations 2013	fully immunized < 1 year 2013	Change in absolute value of fully immunized < 1 year	Change in percent of fully immunized < 1 year
29	479 953	29	523 510	43 557	9.1

Table 7 Fully immunized children less than one year. Malawi, health districts. 2012 and 2013

Unit name	fully immunized < 1 year 2012	fully immunized < 1 year 2013	Change in absolute value of fully immunized < 1 year	Change in percent of fully immunized < 1 year
Blantyre-DHO	29 826	43 963	14 137	47.4
Lilongwe-DHO	56 775	70 889	14 114	24.9
Dedza-DHO	20 330	26 554	6 224	30.6
Nkhotakota-DHO	10 010	14 403	4 393	43.9
Chikwawa-DHO	18 079	22 058	3 979	22.0
Thyolo-DHO	16 964	20 431	3 467	20.4
Mzimba-North-DHO	16 253	13 817	-2 436	-15.0
Mulanje-DHO	23 322	20 961	-2 361	-10.1
Nsanje-DHO	12 531	10 208	-2 323	-18.5
Zomba-DHO	25 071	27 033	1 962	7.8
Dowa-DHO	24 762	22 863	-1 899	-7.7
Mangochi-DHO	31 178	32 754	1 576	5.1
Chiradzulu-DHO	10 036	11 398	1 362	13.6
Mwanza-DHO	2 205	3 528	1 323	60.0
Karonga-DHO	11 327	10 139	-1 188	-10.5
Kasungu-DHO	22 289	23 353	1 064	4.8
Nkhata-Bay-DHO	6 874	5 842	-1 032	-15.0
Balaka-DHO	12 331	13 359	1 028	8.3
Salima-DHO	14 868	13 863	-1 005	-6.8
Phalombe-DHO	11 819	12 732	913	7.7
Chitipa-DHO	5 780	6 447	667	11.5
Rumphi-DHO	6 849	6 266	-583	-8.5
Neno-DHO	4 211	4 754	543	12.9
Ntchisi-DHO	9 098	8 613	-485	-5.3
Machinga-DHO	22 209	21 931	-278	-1.3
Mzimba-South-DHO	15 969	16 214	245	1.5
Ntcheu-DHO	20 940	21 070	130	0.6
Mchinji-DHO	17 748	17 761	13	0.1
Likoma-DHO	299	306	7	2.3

4.3. Top Down method

It is time consuming to verify and control data, in most cases there will not be enough resources available to check each and every data entry. How much editing needed is also closely related to which level data will be disseminated. The essence of the Top down method is to start checking the units that are most important in terms of impact on data quality first, this is done by sorting data. Examples of sorting criteria are:

- Largest changes in absolute value
- Largest variance
- Largest weighed values

In table 8 the five health districts with the highest influence on national figures can be observed. This does not mean that there are quality issues with these data, but it is a natural point of departure for a more thorough analysis of the reporting units in the health district. Blantyre is high on this list and it is the health district that will be analysed further. This district has large changes in both absolute value and in percent when 2012 is compared with 2013 for the variable *for Fully immunized < 1 year* (see table 7).

Table 8 Units with largest impact on Fully immunized children less than one year. Malawi, health districts. 2013

Unit name	Total with unit	Total without unit	Change in absolute value of fully immunized < 1 year if left out	Change in percent of fully immunized < 1 year if left out
Lilongwe-DHO	523 510	452 621	70 889	-13.5
Blantyre-DHO	523 510	479 547	43 963	-8.4
Mangochi-DHO	523 510	490 756	32 754	-6.3
Zomba-DHO	523 510	496 477	27 033	-5.2
Dedza-DHO	523 510	496 956	26 554	-5.1

Blantyre health district reported a 47 percent increase from 2012 to 2013 for the variable *Fully immunized < 1 year*, in total figures this increase added up to more than 14 000 children (see Table 9).

Table 9 Fully immunized children less than one year. Blantyre health district. Malawi. 2012 and 2013

fully immunized < 1 year 2012	fully immunized < 1 year 2013	Change in absolute value of fully immunized < 1 year	Change in percent of fully immunized < 1 year
29 826	43 963	14 137	47.4

This is a large increase and under normal circumstances it should be checked. In Blantyre health district there are 5 facilities that have reported more than twice the value on *Fully immunized < 1 year* compared the other facilities in the district (see Table 10). This means these facilities can be flagged as important in this health district in terms of having a profound impact on the variable *Fully immunized < 1 year*.

Table 10 Units with largest impact on Fully immunized children less than one year. Blantyre health district. Malawi. 2013

Unit name	Total with unit	Total without unit	Change in absolute value of fully immunized < 1 year if left out	Impact on change in percent of fully immunized < 1 year if left out
Ndirande Urban Health Centre	43 963	38 353	5 610	-12.8
Bangwe Health Centre	43 963	38 631	5 332	-12.1
Zingwangwa Urban Health Centre	43 963	38 805	5 158	-11.7
Limbe Health Centre	43 963	39 363	4 600	-10.5
Chilomoni Health Centre	43 963	39 478	4 485	-10.2

Table 11 Units with largest impact on Fully immunized children less than one year. Blantyre health district. Malawi. 2012 and 2013

Unit name	fully immunized < 1 year 2012	fully immunized < 1 year 2013	Change in absolute value of fully immunized < 1 year	Change in percent of fully immunized < 1 year
Zingwangwa Urban Health Centre	2739	5158	2419	88.3
Chilomoni Health Centre	2473	4485	2012	81.4
Ndirande Urban Health Centre	3758	5610	1852	49.3
Chirimba Dispensary	137	1262	1125	821.2
Makhetha Dispensary	623	1662	1039	166.8

As stated earlier the increase in children *Fully immunized < 1 year* has increased by 14 000 cases from 2012 to 2013 for Blantyre health district. In table 11 the five facilities contributing the most to this increase are listed. Just by contacting the five units on this list more than half of the rise can be verified, and also some very unlikely increases in percentage can be checked out; for instance, Chirimba Dispensary which as reported 821 percent increase. Working this way will give priority to the facilities that influence the district figures the most and this is an efficient way to check and edit data. There are many ways to organize a top down approach. The main point is to make a prioritization and, as demonstrated, contacting only a few units that can have a large impact on the quality of the figures for one health district.

4.4. Thousand error

Common challenges with datasets are values which are entered with one, two or three extra zeroes and are often referred to as “thousand errors”. The result of this error is inflated figures. In many cases thousand errors can be corrected automatic. There are two main approaches to identify thousand errors:

- Counting number of digits
- Ratio between this year’s value compared to last year’s value

Searching for thousands error is important, this type of error can and in most cases will, have marked impact on outcome figures. In figure 18 a visual presentation of how to search for thousand errors is on display. The black dots represent facilities, the X-axis is the value that the facility has reported on the variable *Fully immunized < 1 year* and the Y-axis how many times higher or lower the 2013 values are than the 2012 values. The red lines are boarders for acceptable values, and here they are set to 0 and 4. A typical thousand error would have the value 1 000 in the Y-axis. For Blantyre there are no facilities that have this typical characteristic with three extra zeroes, but there is one facility than can by mistake have added a zero extra in the report, Chirimba Dispensary (see Figure 18 and black dot in top left corner). This method, even though it is set up to check thousand errors, can also as demonstrated here, be used to look at facilities that have a very high increase from one year to another. There are three facilities in Blantyre health district facilities that have reported a value for 2013 that is more than four times higher value compared to the year before (See Table 12).

Figure 18 Units with thousand error on Fully immunized children less than one year. Blantyre health district. Malawi. 2013

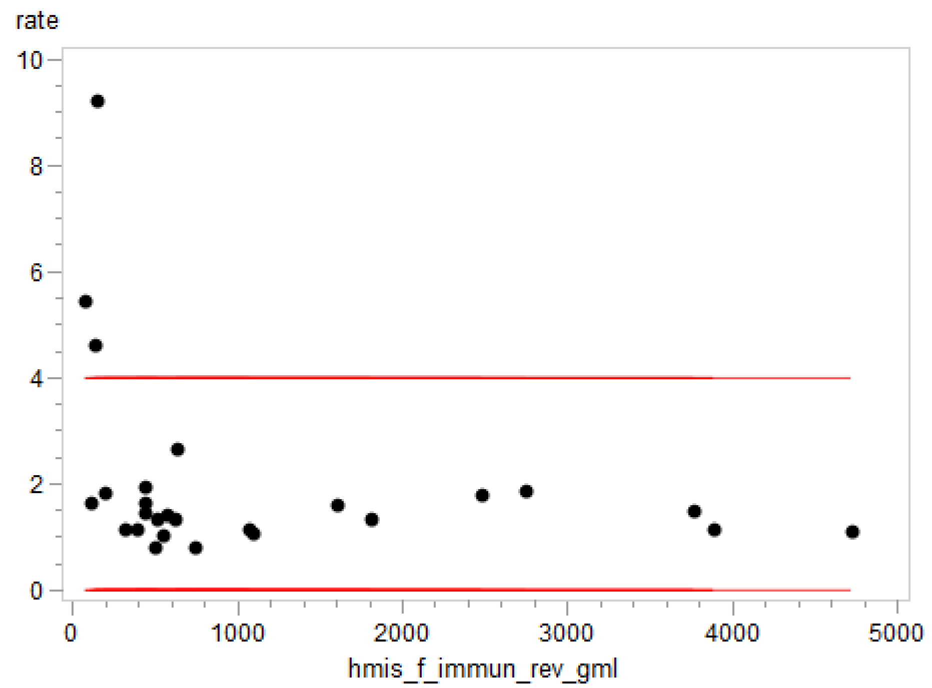


Table 12 Units with thousand error on Fully immunized children less than one year. Blantyre health district. Malawi. 2013

Unit name	fully immunized < 1 year 2012	fully immunized < 1 year 2013	Change in percent of fully immunized < 1 year
Pensulo Health Centre	72	394	447.2
Chirimba Dispensary	137	1262	821.2
Kadidi Dispensary	125	577	361.6

4.5. Quartile method

This method evaluates the ratio between two variables, and a limit value for an acceptable ratio is chosen. In figure 19 the quartile method is demonstrated on the variables *Bed capacity and Total personnel currently* for Malawi at health district level. These are two variables that in many cases influence each other. The black dots in the figure are health districts. The blue line is average ratio and the red line mark the area of acceptable ratio values. As can be observed in Figure 19 the ratio between the two variables looks quite stable, except for the three black dots to in the upper right area of the figure. Phalombe health district has a ratio of more than 10 for the two variables (See table 13), meaning that the district has reported very low figures on *Bed capacity* compared to the number *Total personnel currently* (or very high *total personnel* compared to *Bed capacity*). An advantage of this method is that it can identify inliers, meaning that both variables have acceptable values but an unacceptable ratio. Another strength is for instance when comparing values for 2012 and 2013, and there is a very high increase. Sometimes it can be difficult to know which year that is correct, using the quartile method this can in many cases be decided. If there exist a variable which has particular good quality (the gold variable), in many countries this will be population, it is especially efficient to use "the gold variable" and compare ratios between two years.

Figure 19 List of outliers/inliers using the quartile method on the variable Bed capacity and Total personnel currently. Malawi, health districts. 2013

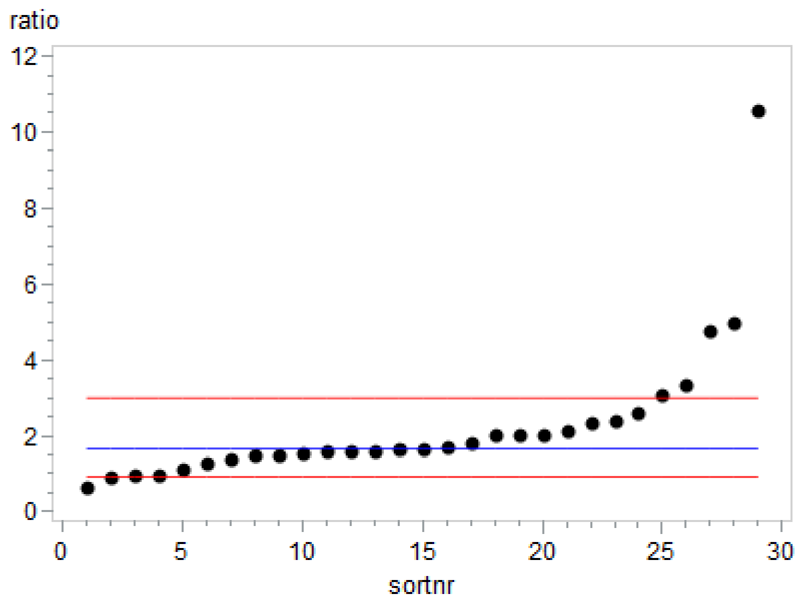


Table 13 Quartile method for the variable Bed capacity and Total personnel currently. Malawi, health districts. 2013

Health district	N-value	O-value	Bed capacity	Total personnel currently	ratio
Blantyre-DHO	0.92615	2.99258	3 283	15 679	4.7758
Chiradzulu-DHO	0.92615	2.99258	2 009	10 062	5.0085
Phalombe-DHO	0.92615	2.99258	921	9 741	10.5765

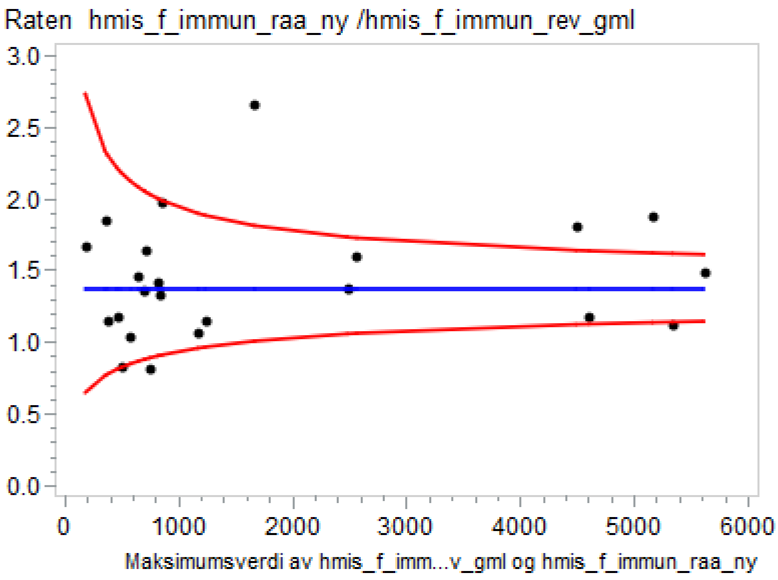
4.6. Hirdiroglou-Berhelot (HB) method

The Hirdiroglou-Berhelot (HB) method is a quartile method that especially takes in to account that normally it is easier to detect data errors with high values than errors with to low values. This method works best when thousand errors already have been edited.

There are three parameters that have to be established when using the HB method:
U= Emphasize on the size of the variable
C= Influences the width of the interval
A= Adjusting for very small quartile distances

These three parameters should be tested and decided by staff that work and have knowledge of the data ensuring that the test captures what it is indented to. Here the HB method is demonstrated on data for the variable *For Fully immunized < 1 year* for Blantyre health district in 2012 and 2013 (See Figure 20). In Figure 20 the black dots are facilities, the blue line represents the mean value and the red lines are boarders for acceptable values, thus identifying outliers. Here the parameters for the HB method are set to: U=0.5, A=0.05 and C=2.

Figure 20 HB method for the variable Fully immunized children less than one year. Blantyre health district. Malawi. 2012 and 2013



With these parameters the HB methods list out 5 outliers (See table 14). These are facilities that have a too high or too low ratio between the values on *Fully immunized < 1 year* when comparing 2012 and 2013. In addition, large facilities have a higher chance of being defined as an outlier (red lines are not straight), owing to the fact that they contribute more too aggregated figures.

Table 14 HB method outliers for Fully immunized children less than one year. Blantyre health district. Malawi. 2012 and 2013

Unit name	fully immunized < 1 year 2012	fully immunized < 1 year 2013	Change in percent of fully immunized < 1 year
Makhetha Dispensary	623	1662	166.8
Zingwangwa Urban Health Centre	2739	5158	88.3
Chilomoni Health Centre	2473	4485	81.4
Chileka Sda Health Centre	739	603	-18.4
Bangwe Health Centre	4715	5332	13.1

4.7. Imputation

When data from units/facilities are missing there is a possibility to imputate values, meaning creating a probable value based on a chosen method, for instance stochastic or “cold deck” method. The logic here is that a created value will be better than a missing value when figures are aggregated from facility to district or national level. Imputation can be looked on as controversial, but this is often based on examples of misuse or misunderstandings. Imputations do increases uncertainty but it can also have clear advantages and increase quality of published data. Missing values are a common challenge in statistics and imputation is used as a last resort, but is quite common. There are two important principles that should be followed when it comes to imputation. First, the method needs to be documented and public. Second, figures that actually have been imputated should be marked properly.

Evaluating the effect of different methods for imputations on the Malawian HMIS data will give an understanding of advantages and disadvantages of using these methods, and this can work as basis for a decision if imputation should be used. By implementing a possibility of imputation in DHIS2 it will be up to each country if they want to make use of this method or not. It is also possible to limit access to this method for instance only to key personnel.

5. Suggested way forward

It ought to be established an editing system in Malawi; aiming to avoid missing data from health facilities, to correct major errors, to improve the source data and the statistics produced. This can be done by setting up a team of two to three staff members in the Ministry of Health, calling health facilities with missing data or suspected errors. The aim is to verify if the data is correct, and guide the facility in what should be reported. The team can be small because the system will identify the facilities with the highest impact on the end result and this way help to prioritize which facilities to focus on.

The statistical methods presented in this data quality assessment are merely examples of techniques that can be used in the data editing process. If they are implemented in the production process and followed up by trained staff, they will potentially have a strong impact on the quality of data. The methods themselves are not the most important, but rather how they are followed up. By using statistical methods, like demonstrated for Blantyre health district, it is clear that contacting only a few facilities and having them to verify reported data will vastly reduce uncertainty. This feedback can be used to edit data, and again also improve registered data.

Implementing an editing module in DHIS2 will give all countries using this software the possibility to improve quality of their own data in a scientific and efficient manner. Currently DHIS2 is the preferred health management information system in 30 countries across four continents, in is also being used at various levels in 47 countries. Furthermore, it is beyond doubt that the methods presented here have the potential to contribute to increased quality for a large proportion of statistics and registers, and if used correct will have a distinct positive influence on HMIS data in Malawi.

References

Journal of Health Informatics in Developing Countries www.jhidc.org Vol. 8 No. 1, 2014 "*DHIS2: The Tool to Improve Health Data Demand and Use in Kenya"*. Josephine KARURI, Peter WAIGANJO, Daniel ORWA, Ayub MANYA. School of Computing & Informatics, University of Nairobi, Nairobi, Kenya, Ministry of Health, Nairobi, Kenya

WHO Library Cataloguing in Publication Data. *"Developing health management information systems: a practical guide for developing countries"* ISBN 92 9061 1650

Recommended reports/Documents

Statistics Norway. Documents 2017/4. *Manual for the DHIS2 quality tool: Understanding the basics of improving data quality*

Statistics Norway. Documents 2017/5. *Improving health data quality: Recommendations and guidelines. Based on the case of the Health Management Information System in Malawi and DHIS2*

List of Figures

List of Tables

Statistics Norway

Postal address:
PO Box 8131 Dept
NO-0033 Oslo

Office address:
Akersveien 26, Oslo
Oterveien 23, Kongsvinger

E-mail: ssb@ssb.no
Internet: www.ssb.no
Telephone: + 47 62 88 50 00

ISBN 978-82-537-9475-4 (electronic)

Statistisk sentralbyrå
Statistics Norway

Design: Siri Boquist