Improving public health information: a data quality intervention in KwaZulu-Natal, South Africa

W Mphatswe, a KS Mate, a B Bennett, b H Ngidi, c J Reddy, e PM Barker b & N Rollins c

Objective To evaluate the effect of an intervention to improve the quality of data used to monitor the prevention of mother-to-child transmission (PMTCT) of the human immunodeficiency virus in South Africa.

Methods The study involved 58 antenatal clinics and 20 delivery wards (37 urban, 21 rural and 20 semi-urban) in KwaZulu-Natal province that provided PMTCT services and reported data to the District Health Information System. The data improvement intervention, which was implemented between May 2008 and March 2009, involved training on data collection and feedback for health information personnel and programme managers, monthly data reviews and data audits at health-care facilities. Data on six data elements used to monitor PMTCT services and recorded in the information system were compared with source data from health facility registers before, during and after the intervention. Data completeness (i.e. their presence in the system) and accuracy (i.e. being within 10% of their true value) were evaluated.

Findings The level of data completeness increased from 26% before to 64% after the intervention. Similarly, the proportion of data in the information system considered accurate increased from 37% to 65% (P < 0.0001). Moreover, the correlation between data in the information system and those from facility registers rose from 0.54 to 0.92.

Conclusion A simple, practical data improvement intervention significantly increased the completeness and accuracy of the data used to monitor PMTCT services in South Africa.

Introduction

Reliable and accurate public health information is essential for monitoring health and for evaluating and improving the delivery of health-care services and programmes.1–6 As countries report their progress towards achieving the United Nations Millennium Development Goals, the need for high-quality data has never been greater.5,6 Furthermore, funding and support for public health activities, such as immunization programmes, remain contingent on demonstrating coverage using routine statistics.7 However, assuring the quality of health information systems remains a challenge.

Studies of public health information systems in resource-poor countries frequently document problems with data quality, such as incomplete records and untimely reporting.5,8 Yet these systems are often the only data sources available for the continuous, routine monitoring of health programmes.5,10 Efforts have been made to improve the quality and management of public health information systems in developing countries. Two examples are the Health Metrics Network, an international network that seeks to improve the quality of health information from various sources,11 and the Performance of Routine Information System Management (PRISM) framework, which was developed as a method for assessing the strengths and weaknesses of routine health information systems.12,13 Other initiatives, such as the Data Quality Audit, have been used by the GAVI Alliance to improve the monitoring of immunization coverage.14 However, the complex nature of health information systems and the demands placed upon them have complicated efforts to improve the quality of routine data.15

In South Africa, the effect of human immunodeficiency virus (HIV) infection on maternal and child health has raised considerable concern. The reported prevalence of HIV infection among women attending antenatal clinics in the province of KwaZulu-Natal, for example, is 38.7%, the highest in the country.16 This led South Africa to set a target of reducing the rate of mother-to-child HIV transmission to below 5% by 2011.17 Although HIV testing is essential for delivering interventions for the prevention of mother-to-child transmission (PMTCT) and thereby for reducing maternal and infant mortality,18 routinely available health information indicates that the HIV testing rate is highly variable.19,20 The questionable accuracy and reliability of these data compromise efforts to improve the health-care systems that provide PMTCT interventions.20

Routine health data from the primary health-care system in South Africa, which is organized in districts, are collected and stored in the District Health Information System (DHIS).21 Recent studies of this system, however, have reported that the quality of the data, including those used to track PMTCT care, is suboptimal and is hindering efforts to strengthen service delivery.20,22 As a result, in 2008 the KwaZulu-Natal Department of Health, the University of KwaZulu-Natal and the Institute for Health care Improvement launched a large-scale effort, entitled the 20000+ Partnership, to improve the quality of PMTCT services in three health districts in the province. This programme included an intervention to increase the completeness and accuracy of the public health data routinely recorded in the DHIS. The aim of this paper is to report on the effect of that intervention.

Abstracts in العربية, 中文, Français, Русский and Español at the end of each article.

---

a Department of Paediatrics and Child Health, University of KwaZulu-Natal, Durban, South Africa.
b Institute of Healthcare Improvement, 7th Floor, 20 University Road, Cambridge, Massachusetts, MA 02138, United States of America.
c Department of Maternal, Newborn, Child and Adolescent Health, World Health Organization, Geneva, Switzerland.

correspondence to KS Mate (e-mail: kmate@ih.org).

(Submitted: 29 June 2011 – Revised version received: 28 October 2011 – Accepted: 2 November 2011 – Published online: 5 December 2011)
Methods

Routine data collection for the DHIS in South Africa starts with information being collected in registers at each point where clinical care is provided. Every month the staff at the different primary health-care facilities collate the data and send monthly summaries on paper to a clinician supervisor who, on average, oversees six facilities. The monthly summaries are then converted into electronic format by an information officer based either at the facility or centrally within the district health office. Larger facilities such as community health centres and hospitals have dedicated facility information officers who themselves create electronic files containing the facility data. District information officers oversee the input of all data into the DHIS. In addition, district PMTCT coordinators work regularly with district information officers to ensure that all health-care workers receive training on data reporting requirements, data collection tools and any changes or updates to the definitions of the data elements used as indicators of the performance of PMTCT services.

Study design

The study involved data collected before and after the data quality improvement intervention at 58 antenatal clinics and 20 delivery wards in both urban and rural communities in KwaZulu-Natal. These 78 locations were randomly selected from 222 fixed facilities that provide PMTCT services in three of the 11 districts in the province: eThekwini, Umgungundlovu and Ugu districts. Across KwaZulu-Natal province, over 90% of pregnant women attend antenatal care facilities at least once during pregnancy and over 90% of deliveries take place in delivery wards at health-care facilities. The prevalence of HIV infection among expectant mothers in the three districts ranged from 40.3% to 45.7% in 2008.

The data quality improvement intervention took place between May and November 2008 and included three main components: (i) training days on data collection, (ii) monthly reviews of data and (iii) data audits at individual facilities.

Three data training days for health-care facility and district information officers and PMTCT programme managers took place in May, August and November 2008, respectively. Each one-day training session dealt with: standardizing data collection procedures; improving data quality; selecting data elements that were useful for monitoring the quality of data on PMTCT services; and providing clear definitions of those data elements. After training, each district information officer liaised on data quality with facilities in his or her district.

Monthly reviews of routine data on PMTCT services were carried out in each district by a PMTCT improvement task team, which comprised the district information officer, the PMTCT programme coordinator and the primary health-care coordinator. During these monthly meetings, the study team provided the task teams with information about the completeness and accuracy of the PMTCT data elements they were using to assess the PMTCT programme. Thereafter, each district task team provided feedback to the facilities on the quality of the data.

Data audits were carried out by study staff at the 78 randomly selected facilities three times: before, during and after the intervention. The original source data in facility registers covering the 3 months preceding the audits were collected and summarized in the same manner in which they were routinely dealt with at the end of each month. The accuracy of the process by which these data were collected at health-care facilities, transmitted to the district offices and entered into the DHIS was assessed by examining the degree of agreement between the data elements used for a particular clinic that appeared in the DHIS and the data elements recorded by the study team for the same time period. It is important to note that whether or not the data contained in the registers accurately represented the clinical activities taking place at each facility was not assessed, as resources did not allow for direct clinical observation. After each audit, the study team provided individual facilities with feedback on the quality of its data on PMTCT services.

Ethical approval for the study was obtained from the University of KwaZulu-Natal Biomedical Research Ethics Committee. In addition, consent was obtained from the KwaZulu-Natal Department of Health and the district managers of all three participating districts. The funding agency, the Centers for Disease Control and Prevention in Atlanta, United States of America (USA), also gave approval but did not influence the study design or data reporting.

Data collection and assessment

The study team visited the 78 randomly selected facilities before the intervention between September and November 2007, during the intervention between June and September 2008 and after the intervention in March 2009.

At each visit, the team collected information on the six data elements used for monitoring PMTCT services that were routinely recorded in registers on the number of: (i) pregnant women who were tested for HIV antenatally; (ii) pregnant women who tested positive for HIV; (iii) pregnant women in whom CD4+ lymphocyte quantitation was performed; (iv) pregnant women who received nevirapine in the delivery ward; (v) babies who were born to an HIV-positive woman and who underwent a polymerase chain reaction test for HIV; and (vi) babies who received nevirapine. These six data elements were selected from a total of 24 possible elements because of their clinical importance. The characteristics of each facility were recorded in a simple form that was completed by the facility manager during the site visit.

It was possible to obtain information on the completeness of the data reported between January 2007 and March 2009 to the DHIS for all 222 facilities providing PMTCT services in the study districts. Completeness was quantified by determining whether a value for a given data element from a specific facility was or was not present in the information system. Data completeness was analysed by data element and facility, as well as over time.

Data accuracy was determined by comparing the data recorded in the DHIS with the data obtained from individual facility registers during data audits. An individual data element value was defined as accurate if the value recorded in the information system lay within 10% of the corresponding value “reconstructed” by the study team from the original source data during the data audit. The 10% threshold was based on that used in previous studies. No assessment of the accuracy of a data element value was made if that value was missing from either the facility register or the DHIS or if the difference between the value reconstructed by the
The study team and that in the information system was greater than 1000%. Such large differences were regarded as “errors” rather than inaccuracies.

### Statistical analysis

The data analysis was carried out using Excel 2007 (Microsoft, Redmond, USA).

### Results

Of the 78 facilities randomly selected to participate in the study of the data quality intervention, 37 (47%) were urban sites, 21 (27%) were rural and 20 (26%) were semi-urban. The median number of clients seen daily at each site was 28 and the median number of health-care workers at each site was 6. All facilities had been offering PMTCT services for more than 2 years.

#### Data completeness

Overall data completeness improved from 26% before the intervention to 64% after: of the 1115 data element values that could have been completed in January 2007, 285 were completed, while of the 1010 that could have been completed in March 2009, 643 were completed. In the antenatal clinics, data completeness improved from 25% (i.e. 271 of 1065 relevant data element values were completed) to 63% (i.e. 600 of 960 relevant data element values were completed). In the delivery wards, data completeness improved from 28% (i.e. 14 of 50 relevant data element values were completed) to 86% (i.e. 43 of 50 relevant data element values were completed).

#### Data accuracy

Overall, 55% of data element values were excluded from the accuracy analysis before the intervention because of errors or missing data, compared with 37% after. The overall accuracy of the remaining data on all six data elements improved from 37% at the first data audit to 65% at the third data audit ($P < 0.0001$). In antenatal clinics, data accuracy improved from 41% to 65% ($P < 0.0001$); in delivery wards, it improved from 32% to 64% ($P < 0.0001$).

During the first data audit before the intervention, there was poor correlation between data that appeared in the DHIS and data obtained from facility registers ($r = 0.54$, Fig. 2). The correlation was much better for data obtained after the intervention during the third data audit ($r = 0.92$, Fig. 2).

#### Discussion

A data quality improvement intervention that involved specific training for health-care workers on the importance...
of public health information, monthly data reviews and feedback, and regular data audits was effective in significantly increasing the completeness and accuracy of the data used to monitor PMTCT services in South Africa.

Training principally involved the health-care facility and district information officers who were responsible for data collection and was intended to help them understand that high-quality data are essential for improving PMTCT programme performance. It was possible to use monthly district programme review meetings, which were often already under way, as an opportunity to reflect on data quality. In the study, data audits were carried out by a small team of four data clerks. Although we did not undertake a formal costing analysis, these activities are likely to be a relatively inexpensive way of improving data quality in resource-poor settings before more systemic or expensive solutions, such as computerized data management, can be implemented.

Although previous reports have made recommendations for improving public health information, few studies have gone further than describing the inadequacy of the information available.\textsuperscript{11, 26, 22, 27} One systematic review identified 12 studies of activities for improving data quality: all were conducted in well resourced health-care services and only one appeared to be an interventional study.\textsuperscript{26, 27} However, three other studies reported the outcomes of data quality improvement interventions in resource-limited settings, in Kenya, Malawi and Peru. In Peru, phone reminders to epidemiological surveillance teams and clinic visits were used to improve the timeliness and accuracy of data reported in an electronic surveillance system of infectious disease outbreaks used by the navy. Phone reminders but not clinic visits improved timeliness, whereas in some settings data accuracy was improved by visits but not phone reminders.\textsuperscript{26} In Malawi, the completeness and accuracy of public health information on HIV testing were reported, though not quantified, using a software package provided by a donor-funded HIV programme.\textsuperscript{26} In Kenya, investigators used on-site assessment and feedback on data completeness and validity to improve the quality of routine health information: the number of HIV tests recorded and the number of HIV positive patients identified both doubled.\textsuperscript{26} Finally, a recent study found that sequential data quality audits in six countries improved the quality and accuracy of data on immunization,\textsuperscript{27} which confirms our finding that regular audits can help improve data quality.

Despite the improvement in data quality observed during this study, the absence of a comparison study arm limits our ability to conclude that there was a strong cause-and-effect relationship with the intervention. Part of the improvement may have resulted from the stringent scrutiny that the South African PMTCT programme received from both programme donors and the department of health. However, we know of no specific efforts to improve data quality in the study facilities or districts at the time of the intervention. The well known inaccuracy of the DHIS has led many nongovernmental organizations, academic bodies and other development partners in South Africa to adopt parallel data collection systems to monitor PMTCT programme performance. These systems, while effective in serving the specific needs of these organizations, were not designed to improve the quality of the information system. An additional limitation of our study is that personnel at participating districts often retrospectively inserted missing data and corrected major errors in the DHIS for up to 1 year. These corrections could have made the data that were available at baseline appear to be better than they actually were and could have resulted in an underestimation of the effect of the intervention.

One unexpected complication that arose during the study could have reduced the effect of the data improvement intervention. The PMTCT programme in South Africa is relatively dynamic and the names and definitions of the data elements used for monitoring are frequently changed. Several challenging changes occurred during the study. For example, the data element used in the DHIS to record whether a baby had undergone a polymerase chain reaction test for HIV at 6 weeks was initially titled “HIV 1st test of baby born to HIV-positive woman” but was later changed to “HIV PCR test of baby born to HIV-positive woman at 6 weeks or later”. Such changes were made without the district offices providing definitions to the clinics. This could have caused considerable confusion at individual facilities and compromised the quality of reporting on that particular data element. To lower the chance of a similar occurrence we selected data elements whose descriptions had remained largely unchanged in recent years.

Despite these limitations, the improvement in PMTCT data quality observed in this study is encouraging, for it suggests that similar approaches could improve the quality of the data needed for decision-making and resource allocation in other public health programmes. The rationalization of data collection tools, clear definitions of data elements, continuous feedback on data quality and intermittent but regular data audits are effective ways of improving data quality. However, while this study shows that public health information can be improved, the final result falls short of what we should accept from our health information systems. The quality of public health data could be improved further. In addition, more work is needed to ensure that the data transmitted from bedside to the public health record are complete and accurate and that the original data collected in consultation rooms accurately reflect the interaction between patient and health-care provider.

In conclusion, this study reports a simple, practical approach to improving the quality of public health information, both locally and in a district health information system. Accurate information is vital for health-care planning and for evaluating progress towards specific health targets. Further research is needed to assess the effectiveness of similar interventions on a large scale, to estimate their cost implications and to clarify the connection between improved data quality and better health service delivery.

Acknowledgements

The authors thank Hlengiwe Gcaba, Thanda Goniwe, Sibongiseni Gumede, Sibongiseni Khambule, Lungisani Mhlungu, Nirvasha Moodley and personnel at eThekwini Municipality, Ugu, eThekwini and Umgungundlovu health districts and the KwaZulu-Natal Provincial Department of Health. In addition to their primary affiliations, KS Mate is also affiliated with the Department of Medicine, Weill Cornell Medical College, New York,
Research
Data quality intervention in KwaZulu-Natal

W Mphatswe et al.

USA, PM Barker is affiliated with the Department of Pediatrics, University of North Carolina, Chapel Hill, USA and N Rollins is affiliated with the Department of Paediatrics and Child Health, University of KwaZulu-Natal, Durban, South Africa.

Funding: This study was funded by the United Kingdom Department for International Development and the United States Centers for Disease Control and Prevention.

Competing interests: None declared.

The aim of this study was to evaluate the impact of a data quality intervention on the completeness and accuracy of data used in the monitoring of PMTCT services in KwaZulu-Natal, South Africa.

The intervention was implemented in 2008 and involved training on data collection and feedback to health information officers and program managers, monthly data reviews, and data audits in healthcare facilities.

The study found that the completeness of the data improved from 26% to 64% after the intervention, and the proportion of data considered accurate increased from 37% to 65% (P < 0.0001).

The correlation between the data from the healthcare facilities and the data recorded in the health information system increased from 0.54 to 0.92.

The study concluded that a simple and practical data improvement intervention significantly increased the completeness and accuracy of data used in the monitoring of PMTCT services in KwaZulu-Natal.
Улучшение медико-санитарной информации: мероприятия по улучшению качества данных в провинции Квазулу-Натал, Южная Африка

Цель
Оценить эффект мероприятий по улучшению качества данных, используемых для мониторинга и предотвращения передачи от матери ребенка (ППМР) вируса иммунодефицита человека в Южной Африке.

Методы
В исследовании приняли участие 58 женских консультаций и 20 родильных палат (37 городских, 21 сельских и 20 пригородных) в провинции Квазулу-Натал, которые предоставили услуги ППМР и отправляли данные в информационную систему здравоохранения района. Мероприятия по улучшению качества данных, проведенные в период с мая 2008 г. по март 2009 г., включали обучение персонала информационных отделов медицинских учреждений и руководителей программ сбору данных и организации обратной связи, ежемесячным обзором данных, а также аудиту данных в учреждениях здравоохранения. Данные по шести элементам данных, используемых для мониторинга услуг ППМР и записанных в информационной системе, были сопоставлены с источником данных из реестров медицинских учреждений до, во время и после проведения мероприятий. Были оценены полнота данных (т.е. их присутствие в системе) и точность (т.е. отклонение не более, чем на 10% от истинного значения).

Результаты
Уровень полноты данных увеличился с 26% до 64% после проведения мероприятий. Аналогичным образом, доля точных данных в информационной системе увеличилась с 37% до 65% (P < 0,0001). Более того, корреляция между данными в информационной системе и данными из реестров медицинских учреждений выросла от 0,54 до 0,92.

Вывод
Простые и практичные мероприятия по улучшению качества данных значительно увеличили полноту и точность данных, используемых для мониторинга услуг ППМР в Южной Африке.

Resumen
Mejorar la información acerca de la salud pública: intervención para mejorar la calidad de los datos en KwaZulu-Natal, Sudáfrica

Objetivo
Evaluar el efecto de una intervención para mejorar la calidad de los datos empleados para controlar la transmisión del virus de la inmunodeficiencia humana de la madre al hijo (TMT) en Sudáfrica.

Métodos
El estudio incluyó 58 clínicas prenatales y 20 paritorios (37 de ellos situados en áreas urbanas, 21 en áreas rurales y 20 en áreas parcialmente rurales) en la provincia de KwaZulu-Natal que proporcionaban servicios de prevención de transmisión de la madre al hijo y comunicaron sus datos al Sistema de Información Sanitaria del Distrito. La intervención para la mejora de los datos, que se aplicó entre mayo de 2008 y marzo de 2009, incluyó la formación sobre recopilación de datos y comunicación de resultados dirigida al personal sanitario encargado de la información así como a los gestores de programas, exámenes mensuales y auditorías de los datos en las instalaciones de asistencia sanitaria. Se compararon los datos basados en seis elementos de datos utilizados para controlar los servicios de prevención de TMT y archivados en el sistema de información con la fuente de datos de los registros de las instalaciones de asistencia sanitaria antes, durante y después de la intervención. Se evaluaron los datos en base a su exhaustividad (por ejemplo, su presencia en el sistema) y su exactitud (por ejemplo, estar situado dentro del 10% de su valor auténtico).

Resultados
El nivel de exhaustividad de los datos aumentó del 26% antes de la intervención al 64% después de la misma. De manera similar, la proporción de datos en el sistema de información considerados exactos aumentó del 37% al 65% (P < 0,0001). Además, la correlación entre los datos del sistema de información y aquellos de los registros de las instalaciones aumentó de 0,54 a 0,92.

Conclusión
Una intervención sencilla y práctica para mejorar los datos aumentó la exhaustividad y exactitud de los datos empleados para controlar los servicios de prevención de la TMT en Sudáfrica.

References
Research
Data quality intervention in KwaZulu-Natal
W Mphatswe et al.


