The use of routine case record data to evaluate quality of inpatient hospital care for newborns and children in Kenya

Gathara, D.N.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 5

Discussion
Discussion

Quality of care assessment is one of the ways of evaluating what the health system is providing and can allow monitoring and evaluation exercises to track progress and identify gaps. Availability of high quality data is central to these assessments, highlighting the need for integration of data collection activities for these processes into the routine health information systems to promote efficient use of resources especially in LIC where resources are scarce. The utility of data lies in its use, therefore, the availability of high quality patient-level data could consequently lead to data use for: i) health needs assessment and identification of gaps in service provision, ii) quality of care assessment, iii) health planning, iv) programme evaluation and v) monitoring of disease trends while generating a longitudinal dataset that can be used for research to evaluate progress and impact of interventions and guideline recommendations.

Each chapter in this thesis examined the utility of routine records in quality of care assessment briefly touching on some elements of structure (Chapters 2c and 4a), focusing predominantly on process measures (Chapters 2a, 2b, 2c and 4b) and exploring mortality as one obviously important outcome in two papers (Chapters 3a and 3b). The process of care examined in this thesis might best be considered as the technical content of clinical care. The approach adopted involves examining the process of care and comparing this against agreed standards.

The papers presented in this PhD work are united by the pragmatic approach to improving quality of care in real-life hospital conditions, the attempts to increase the utility of routinely collected information in LICs and the innovative approach to application of statistical methods rarely used in the analysis of routine data from LICs. Although the focus of this thesis is on the Kenyan setting, it will also be relevant to other LIC settings particularly those in Africa with similar context and challenges.

Use of routine data for quality of care reporting

The results in chapter 2 highlight the inadequacy of much routine case record data for quality of care assessment. In chapter 2a, at baseline it appeared almost impossible to evaluate neonatal care or care provided to severely malnourished children. For instance for newborn care, while the targeted number of records per hospital was 30 records from a period of 6 months in some hospitals the retrieval rate was as low as 7 records, a rate which was clearly inconsistent with prevailing hospital workloads. However, even where records were available, clinical documentation was poor with documentation of key essential signs and symptoms ranging from a score of 0 to 28. The work reported was conducted in
the context of an intervention trial aimed at improving paediatric care[1]. Although improvements were noted in neonatal care and care for the severely malnourished in the post-intervention period with implementation of standardized admission record forms, the adequacy of documentation remained sub-optimal with outcome data (whether they lived or died) still missing for 12% of the neonates. In chapter 2b, we present work from one tertiary hospital over a period of 5 years. In this work we demonstrate that where case records are relatively well documented it is possible to examine the impact of interventions aiming to change practices using process data. However, this was not without challenges inherent in the use of case record data. Firstly we observed differential documentation of disease severity over time where use of recommended syndromic (very severe, severe, non-severe) disease classifications for pneumonia for example were used more often as the study progressed. This therefore gave a false impression of increasing disease severity over time making it hard to examine changes in case-fatality within these clinical groupings over time. Secondly, still relatively poor documentation meant disease severity could not be linked to presenting signs and symptoms making it hard to judge true adherence to guidelines or classification of illnesses by clinicians. In chapter 2c, we report resource availability and uptake of guidelines (process measures) over a period of 11 years across studies that used similar methodologies. Although indicators to be assessed were similar over time, the indicators that could be evaluated at each time point were few due to inadequacy of quality data highlighting the challenges of tracking intervention adoption using routine data and their inability to provide insights into the process of care. Further, data on patient outcomes (mortality) were still often missing or unreliable.

These studies demonstrate that routine case record data may be inadequate to evaluate care except for basic quality indicators and that they are often too poor for a detailed analysis of quality gaps, tracking progress of new interventions or monitoring health outcomes. These findings are consistent with other studies that have reported difficulties using case record data for quality of care assessment in LICs[2]. with similar challenges in the use of routine records (EMRs) being reported in HICs[3]. However, this does not imply that routine case records cannot be used in evaluating health system performance. For instance the Kaiser Permanente division uses routine records from thousands of physicians stored in a clinical research data base that is used for research to undertake health services and population studies. Collaborations and learning health systems have been set up to undertake research using routinely collected data. Examples of such learning health systems include PEDSNet for paediatrics and the Vermont Oxford Network for neonates among others[4, 5].
These collaborative efforts, that often pave the way for wider health systems improvements, often begin with efforts to standardize data collection and improve clinical documentation. Work done by Mwakusya et al.[6], O’Driscoll[7] and in chapter 2a have shown that provision of standardised admission record forms significantly improves the level of documentation of key clinical signs and symptoms even in LIC. Such documentation of clinical features highlighted in national and WHO clinical guidelines subsequently provides the basis to evaluate care. Other approaches that improve documentation include critical care pathways[8] and computerised data collection (EMRs)[9–11]. These approaches not only improve quality of data available for process assessment but may be job aides in themselves linked to efforts to promote guideline adherence. Such forms developed by our group (http://idoc-africa.org) have been adopted by the Kenyan Ministry of Health and are being disseminated for use. Although some hospitals are struggling with adopting these new forms other hospitals have already taken them up and are using them routinely for up to 100% of their admissions. These hospitals therefore have relatively high quality and comprehensive data on paediatric and neonatal admissions. These routine data may be adequate to answer key research questions, track disease trends, evaluate quality of care, and progress of various interventions as illustrated in some of the work in this thesis if appropriate systems are developed for collating and analysing data. Demonstrating the value of these data may help to promote their collection at scale.

Use of routine records in examining associations with mortality

Quality of care has been shown to vary on weekends and weekdays with a higher mortality being reported in some studies of stroke[12, 13] and emergency medicine among others[14–17]. These differences in care have commonly been attributed to decreased levels of staffing, supervision and expertise and poorer access to diagnostics during weekends and night-time hours[16, 17]. In chapter 3a, we explored whether children admitted at weekends were at higher risk of death than those admitted at other times in a large Kenyan tertiary hospital. As a secondary hypothesis, we explored whether being admitted at the weekend with pneumonia or dehydration is associated with poorer quality of admission care, as a possible mechanism for higher mortality. We report no difference in mortality on weekends and weekdays. However, we speculate that our results may have been biased by our inability to classify night-time hours as out-of-office hours due to poor documentation of time of admission as the care provided during these night-time hours is likely to face similar challenges as that on weekends. The lack of a difference in mortality may then have been affected by this misclassification and further limited the power of the study to detect a small but clinically meaningful effect. However, we also found no
differences in process measures of quality of care between weekdays and weekends. This could be a result of the same misclassification as above or: i) a true finding because we studied a teaching hospital with resident doctors (trainee pediatricians) in theory being available for 24 hours each day (although anecdotal evidence suggests much reduced night-time staffing) or ii) due to poor staffing and resource availability on both weekdays and weekends. Inadequate staffing which is present routinely has been shown to affect quality of care and consequently mortality[18–21].

In chapter 3b, we use a large patient level, multi-site hospital dataset (the clinical information network) to illustrate variation in mortality across hospitals but also risk factors for mortality. Findings from this work indicate that mortality is highly variable with up to five-fold difference across settings even within a common age group. Further, all the risk factors identified and supported by evidence based guidelines were associated with mortality except for vomiting everything and jaundice (both rare findings). Our data supports the use of clinical risk factors drawn from guidelines in day-to-day use in prioritizing care and identifying children at the highest risk of death. Previous reports on mortality in LICs are commonly from single sites or are disease specific[22, 23][24]. We are not aware of similar work on all-cause mortality across multiple sites using contemporaneous routine data. These findings reiterate the utility of risk factors identified in guidelines and their potential usefulness in developing clinical risk scores applicable to LICs that might be used to develop risk adjusted mortality estimates across hospitals.

Basing such approaches on simple clinical features is more appropriate than using risk scores such as Paediatric Risk of Mortality Score (PRISM)[25], the Paediatric Early Warning System Score (PEWS)[26] or the Paediatric Index of Mortality (PIM)[27] developed in HICs as these require physiological or laboratory data not available in routine LIC hospitals.

The data used in this study, although routinely collected, is likely to be better quality data than would be found in many hospitals. However, given the limited resources provided to the CIN hospitals (a desktop and a clerk), it seems possible to achieve similar success in routine data collection at wider scale with appropriate, targeted investment and by building networks of partners. Thus hospitals improved their data collection over time (in most cases) and adopted the standard record form we recommended (but did not provide) that is endorsed by the Ministry of Health. Despite these measures the inherent problem of missing data was still persistent with less than 60% cases being available for complete case, multi-variable analysis using standard regression techniques that result in list-wise case deletion. However, improvements in computational abilities of standard computer hardware and the
development of statistical techniques such as multiple imputation in routine statistical software allowed use of all available data and strengthened the credibility of the results presented.

**Exploring variation in care using routine records**

Care has been shown to vary greatly across places. We explored this using a cross sectional dataset from 22 hospitals across Kenya. In chapter 4a, we demonstrated care varied greatly across places both for process indicators and for resources. In chapter 4b, we explore the sources of the variation in some process measures and illustrate that factors at the hospital level often contribute to most of the variability observed.

Variability in processes of care and outcomes has been reported widely in high-income settings and reported at geographic, hospital, physician groups and individual physician levels[28–30]. We believe the report we present is the most statistically sound examination of such variation in a LIC. The analyses also provide estimated intra-cluster correlation coefficients (ICCs) that are often lacking from pragmatic routine data sets to help inform sample size and power calculations and subsequently better study designs. Where ICCs are available in LICs they are from clinical trials and community settings therefore limiting their extrapolation to hospital based studies. Further, there is very little work from LICs that deals with clinician level clustering.

Why is variation important? Variability in care has been seen as an indication for need of quality improvement. There have been illustrations in literature where authors have used multi-level analysis to estimate the degree of variation across different levels to identify an optimum level for intervention[31][32, 33]. In the data that we present in chapter 4b, most of the variation was attributed to the hospital level. This therefore suggests that hospital directed quality improvement initiatives would be most effective, a contrast to the common practice in LICs where most interventions are targeted to individual clinicians. The ‘higher order’ influences of hospitals on care such as organisational policies or systems beyond the control of individuals make hospitals an appropriate target level of certain interventions. Further, some process indicators reflect team work rather than the effect of individuals. For example to ensure guideline adherence to malaria treatment (i.e only children with malaria positive test should be prescribed Artesunate), the laboratory needs to have capacity to test for malaria parasites, the pharmacist needs to ensure that Artesunate is available while the clinician will be required to prescribe the correct dose. However, interventions at the clinician level may be important for tasks that require cognitive skills like calculating weight-based dosages in children, an indicator
which demonstrated most of the variability at the clinician level in our data. Therefore, examining variability within levels might be a reasonable approach to identifying optimal strategies targeting different system levels for quality of care improvement approaches. However, interventions should not only be informed by the source and magnitude of variability at a level but also the absolute levels in performance.

**Statistical methods for routine data**

Reports on quality of care from LICs are mainly descriptive often reporting proportions, means or medians. However, advanced statistical methods provide an opportunity for better reporting and getting the most out of routine data. While detailed methods are provided in each of the manuscripts, I will briefly highlight the motivation behind the innovative application of each of the methodological approaches to analysis of routine data employed in this thesis. These type of methodological adaptations are rarely applied to data from LICs.

**Use of Propensity scores**

Routine observational data has often been used for reporting national, pre-specified indicators especially in HIC. However if of good quality, these data can be used for secondary data analyses to answer questions that are often not high priority for experimental research but are key in provision of care. Using observational data to answer such questions is often limited by uncertainty in attributing causality as non-random allocation of individuals to different forms of care may result in imbalances in risk factors between groups.

Propensity scores analysis provide a process for creating two comparable groups by adjusting for the observed covariates and therefore establishing a sub-classification (or stratification) that improves the ability to generate balanced groups and allow more robust comparisons[34, 35]. While propensity scores cannot adjust for unmeasured or unknown confounders, and the conclusions from an observational study may not be considered as strong as those from RCTs, they can provide useful data for providing evidence where RCTs are not available[36, 37]. For example it is not possible to allocate admissions to arrive on weekdays or weekends. Further, such data are relatively inexpensive to obtain and often represent the spectrum of medical practice better than patients actually included in randomized experiments.

**Funnel plots**
The most common form of graphical comparisons used in performance reports for quality of care are simple bar charts or league tables where indicators or institutions are ranked from ‘worst’ to ‘best’. These comparisons using league tables have often been criticised for the inability to allow for variation attributable to sampling error [38][39].

Funnel plots provide an alternative to these comparisons by allowing for chance variation by taking into account sample size in estimations, with places with small sample sizes being allowed (and expected) to vary more than places with estimates based on large sample sizes[39, 40]. They often demonstrate that most variation is as a result of chance and hence reduce mislabelling of institutions based on ranks. Further, they also demonstrate institutions that are performing out of control (good or bad) that are outside the set confidence limit that can then be followed-up for extra monitoring.

**Multi-level models**

Health care provision could be argued to be clustered within regions, hospitals and clinicians. While it is important to identify the good and poor performers within any health system it is equally important to identify and quantify the sources of variations in care with the aim of identifying priority areas for intervening in order to achieve harmonisation in care provided.

Multi-level modelling techniques[41] have been used in health care to explore and attribute variation at different levels of aggregation of health care data (geographic, hospital, and individual physician levels). An intra-class correlation coefficient is commonly used as a measure of the proportion of variability attributable to a level. The ICC is defined as the ratio of between-cluster variance to the total variance and has values between 0 and 1 with high ICCs suggesting that between cluster variance is more important than within cluster variance. When considering service improvement, levels that contribute only a small proportion of total variability may be interpreted as having less potential for quality improvement unless the total variability explained is low in a model in which case we are not sure where to intervene[42][43]. In chapter 3b we use ICCs for purposes of exploring sources of variation in care in Kenyan hospitals and illustrate the frequently powerful influence of hospital level effects on indicators of quality measured at the patient level.

**Imputation**

Missing data are consistently reported as a limitation to use of routine data and a potential source of bias. However with advancement in computing (better hardware) and statistical software, methods such as multiple imputation are now commonly available and much easier to implement. While the ultimate
aim is to have completely observed data, missing data are unfortunately unavoidable in epidemiological or clinical research in routine settings. Common approaches to dealing with missing data include: complete case analysis, mean substitution and last observation carried forward. However these are often criticised as providing biased estimates or standard errors[44]. More robust methods for handling missing data are the use of expectation maximization algorithms for maximum likelihood estimation[44] and multiple imputation[45][46].

In multiple imputation, missing values are predicted based on the observed data several times and pooled results are used that allow the use of all available data while providing more conservative standard errors around estimates. Although it is a relatively new method and only became available in statistical software recently we have demonstrated in chapter 4b the applicability and value of this method in routine data from LIC settings.

**Overall findings**

What do we learn from these datasets? Case record data is readily available and can with some basic enhancements be adequate to answer key questions on the quality of care, track disease trends and outcomes and could be a good resource for health services and operational research. We have demonstrated that with minimal investments in information systems and data capture systems, it is possible to routinely collect relatively good quality data to answer research questions and undertake detailed monitoring and evaluation exercises pragmatically. However, a more long term solution would be to integrate these data collection systems in routine health information systems. Although hospitals in LICs are rapidly adopting EMRs, recent evaluations have shown that these are mainly administrative and where clinical data are captured it is often in formats that cannot be easily analysed (text) with most systems also not being interoperable[47]. A potential approach to development of EMRs is to develop intermediary systems such as those reported in our work in Chapter 3b and those from disease specific systems like those developed for HIV by AMPATH in Kenya[48] and Baobab in Malawi among others[11, 49]. These systems typically focus initially on developing common data frameworks that may be incorporated into EMRs.

The clinical information network –comprising of 14 hospitals across Kenya collecting data on all admissions for paediatrics - provides an illustration of providing a large multi-site patient level dataset for quality of care assessment. Early work from this network has allowed reporting of all-cause inpatient
mortality and the variations and risk factors associated with it, and explores the utility of these kind of data for performance monitoring work which is rarely reported in LICs.

The over 40000 observations in this pragmatically collected data set illustrates that it may be possible to undertake research on mortality in LICs at even greater scale and to evaluate quality of care routinely. However the number of hospitals in CIN is small (14 hospitals), therefore larger data sets across a large number of sites will be required to better understand mortality and begin to use it to understand system performance. Such data might be used to develop standardized mortality ratios for institutional comparison although this approach is contentious[50–52].

Translating research work into policy and practice

Approaches described in this PhD thesis for routine performance monitoring might be used more widely in LICs. For instance, the paediatric protocols (guidelines) that formed the basis of the quality indicators used in the work reported in this thesis have been adopted in Uganda[53] and Rwanda[54] with discussions to implement them in Somaliland, Sierra Leone, and Zimbabwe. Therefore, Ministries of Health could adopt some of the available tools used for QoC assessment (http://idoc-africa.org). The data collection processes and analyses have been developed in non-proprietary or open source software (REDCap and R statistical software) which we believe could also be applied more widely to support evaluations of quality at scale and the effects of interventions to improve quality. In addition, this work prompted the Ministry of Health to initiate work geared towards formal routine performance monitoring. Firstly, based on work from chapter 2, the ministry of health undertook initiatives to help them understand health care provision in hospitals better. These efforts towards routine performance reporting focusing on ‘internship training centres (hospitals’)’ were started through the SIRCLE collaboration (Health Services Research and Clinical Excellence Collaboration). This aimed at developing approaches for routine monitoring and evaluation for inpatient services[55] – work which I helped develop and is reported in chapter 4 and elsewhere[56][57][58]. Secondly, a technical working group has been set up to identify quality of care indicators that feed into the routine monitoring across all levels of the health system (community, primary, hospital, county and national levels). Work I have been conducting is contributing to the development of these indicators. Thirdly, although still at an early stage, there is ongoing work on developing a national EMR with discussions on how to support these systems in performance monitoring with some of the lessons learnt from the analyses I report and from the CIN informing the design.
Limitations

Although we have reported limitations specific to each study in each of the papers, more generally, the work presented in this thesis needs to be interpreted in light of the following limitations. First, the data reported in all the papers is based on manual extraction from case records and therefore we have to assume that all care that was delivered was documented. Where documentation is poor, care might be interpreted as poor purely due to lack of data. However recent work by Mwaniki and colleagues demonstrated that retrospectively collected case record data are as good as prospectively collected case record data for quality of care assessment[59]. Secondly, although case record data are readily available and a good resource for performance monitoring and operational research, they are often not collected with a research question in mind. Therefore, researchers need to adequately interrogate the data for potential sources of bias. Common shortcomings of case record data identified in work reported in this thesis were: i) missing data linked to poor documentation, ii) differential documentation over time, iii) missing case records due to poor record archiving, and iv) limited use of diagnostic tests to inform patient diagnoses. Thus while some shortcomings in data collection may be reduced through development of careful electronic data capture such as that presented in chapter 3a that allow for data validation and cleaning during initial entry, primary errors in recording made by clinicians cannot be overcome. Finally, the selection of hospitals included in each of the studies was purposive and generally included a relatively small number of hospitals which might have led to selection bias. However efforts were made to ensure both epidemiological and geographical diversity of hospitals while the selection process was always guided by structured minimum criteria.

Implications for future research

To allow for routine performance monitoring, the use of electronic data capture mechanisms is not enough, there is need to ensure development of common data frameworks linked to commonly understood clinical and diagnostic clinical procedures, standardization of data formats, and development of database structures to allow inter-operability across institutions to ensure comparability of quality indicators across places. This would involve wide stakeholder engagement informed by agreements on appropriate quality of care indicators and schematically laying out the data capture processes and analytic algorithms to ensure consistency in data collection and analysis across places. Documenting the above processes for the CIN and any other networks and sharing these approaches will be useful in future to institutions aiming to implement similar networks or build routine analytics into EMRs.
We have demonstrated the use of various statistical methods in each of the chapters and described them briefly in the discussion. As earlier mentioned these methods are rarely used in data from LICs but have a huge benefit in allowing for better use of routine data for both performance monitoring and research. We suspect but cannot confirm that part of the reason they are not applied is due to limited investment in data collection and development of statistical skills among staff within academic institutions and Ministries of Health. Therefore, as we invest in developing better systems for collating data there will be need to develop analytic capacity concurrently for better reporting of observational data a position echoed by Whittaker and colleagues[60]. Further, as longitudinal datasets become increasingly available there is opportunity for using additional statistical process control methods to monitor care over time. While the methods we have used to explore variation in chapter 4 mainly use pooled data, these do not take into account changes that might be occurring over time during health care provision. Statistical process control methods evaluate the variation of care over time identifying time points when change is observed[40]. These methods and the wider suite of time-series methods are appropriate for evaluation of impact of interventions over time. Such methods might be used to evaluate at wider scale the effect of system related events such as the rotation of clinicians across departments, health worker strikes, the introduction of vaccines or even disease outbreaks[61–63].

Guidelines only reflect the knowledge of the day and need to be regularly updated. For instance, recently Maitland and colleagues demonstrated that guideline recommendations for fluid bolus for children with shock might actually be harmful[64]. While randomized clinical trials are considered as the gold standard for generating evidence they often focus on the efficacy of interventions (in ideal conditions), are expensive and time consuming and therefore difficult to generalize. Secondly, the cost of answering many clinical questions by collecting data is high. Therefore availability of routine datasets will provide an opportunity to create an observational evidence base while allowing research to be undertaken in a cost-effective manner. However for these to be effective there will be need for research and technical collaborations among institutions to create a platform for testing widely used interventions whose merits are unknown.

**Conclusion**

Quality of care assessment is one of the ways of evaluating what the health system is providing and can allow monitoring and evaluation exercises to track progress and identify gaps. Such monitoring however depends on an ability to measure quality. Care can only be measured through data, however in LIC data are very limited and where available are often of poor quality especially in routine settings. At present
most of the health information systems also only collect aggregate patient data that do not allow patient level quality of care assessment. However the quality of documentation can be greatly improved through use of standardised record forms whose consistent and wide scale use will allow for wider scale, more informative routine monitoring. Although these standardised records provide good quality data, such data are not analysable in their raw format and hence we need to electronically collect patient level data that can allow us to report on both process and outcome. Ultimately such data should be provided by contextually appropriate EMR development. However mortality data as a measure of quality will be hard to interpret and large high quality datasets linked to appropriate risk adjustment will be required to make any useful comparisons between places.

The availability of large datasets and sentinel systems such as the one illustrated in chapter 3b provide an opportunity for integrating research into routine clinical practice thereby accelerating the time from research to policy while providing context specific/relevant evidence. In addition, such a platform provides an opportunity for undertaking observational research and using statistical methodologies to allow for efficient use of routine data. For instance to undertake comparative effectiveness research and answering important questions that otherwise might not attract funding. While developing EMR systems that permit such evaluation will take some time a potential approach in LICs would be to develop intermediary electronic data capture systems which would not only provide data but also a learning platform on how best to develop and deploy electronic medical records in LICs.
References


5. Vermont Oxford Network – Improving care for infants and their families [https://public.vtoxford.org/]


28. Janjua NZ, Khan MI, Clemens JD: Estimates of intraclass correlation coefficient and design effect for surveys and cluster randomized trials on injection use in Pakistan and developing countries. *Tropical Medicine and International Health* 2006, **11**:1832–40.


53. Improving the quality of hospital care for sick children in East Africa through ETAT+ training [http://www.rcpch.ac.uk/what-we-do/rcpchinternational/volunteering-overseas/health-partnerships-scheme-grant-etat-east-africa]


