The Effect of an Improved Monitoring System on the Quality of Care at a Pediatric High-Dependency Unit in Malawi

A Quantitative Study Examining the Effectiveness of the IMPALA 2.0 System in Zomba Central Hospital using Propensity Score Matching

Name: Gabriela Riebeek

Student number: 2674217

Supervisors:

Wendy Janssens

School of Business and Economics of the VU

Ângela Ben Jornada

Department of Health Sciences of the Vrije Universiteit Amsterdam and the Amsterdam Institute for Global Health & Development.

Date: 29/06/2023

Master: Spatial, Transport, Regional & Environmental Economics

Abstract

Child mortality rates remain a global concern, with millions of children under the age of 5 dying annually, particularly in sub-Saharan African countries where child mortality rates are significantly higher. The IMPALA project aims to address this issue by introducing a monitoring system in low-resource settings to improve child survival. However, the effectiveness of such interventions can be affected by contextual factors.

This effectiveness is evaluated by comparing data from a period before the IMPALA project with data collected during the IMPALA 2.0 project using propensity score matching. The outcomes examined include the duration of HDU stay (1), the occurrence of Critical Illness Events (2), and the number of deaths (3). Nine covariates were used from one of the two categories: descriptive variables and vital signs.

After data cleaning, a total of 485 record IDs were included, with 303 children in the IMPALA 2.0 dataset and 182 children in the pre-IMPALA dataset. Propensity score matching reveals that there is a statistically significant decrease of 0.145 (95% CI ± 0.114) in the occurrence of critical illness events among children in the HDU during the IMPALA 2.0 period compared to the pre-IMPALA period. However, this significant difference was not observed when analyzing a smaller dataset that excluded records with missing data. No significant differences were found in the other two outcome variables 'duration of stay' and the 'number of deaths' between the two periods.

The results suggest that implementing a monitoring system, such as IMPALA, can contribute to improving healthcare quality in low-resource settings. The study anticipates that the new IMPALA 3.0 system, incorporating predictive capabilities, will have an even greater impact. Further research is necessary to evaluate the effects of the IMPALA 3.0 system and the role of well-trained staff in enhancing healthcare quality.

Key words:

Healthcare quality, High Dependency Units, IMPALA, Low resource settings, Malawi, Monitoring system, Pediatric care, Propensity score analysis, Sub Saharan Africa

1. Introduction

Global child mortality has declined substantially in the last three decades with it currently being at its all-time lowest level. The number of child deaths has more than halved, dropping from 12.5 million in 1990 to 5.2 million in 2019 worldwide (Roser et al., 2013). However, despite this progress, an alarming statistic reveals that approximately 5 million children under the age of 5 still lost their lives in 2021 (World Health Organization, 2022).

One of the Sustainable Development Goals, specifically Goal 3.2, aims to reduce the child mortality rate to a maximum of 2.5% in all countries by 2030. This ambitious target envisions that over 97.5% of newborns worldwide would survive their first five years, regardless of their place of birth (General Assembly Economic and Social Council - United Nations, 2023). However, countries are currently far from reaching this goal by 2030, especially in low- and middle-income countries (LMICs), emphasizing the need for intensified efforts to improve child survival rates (Roser et al., 2013).

Child mortality is not equally distributed across the globe, and children in sub-Saharan African countries are particularly vulnerable, facing a significantly higher risk of mortality. The likelihood of dying before the age of 5 in this region is nearly double the global average (Levels & Trends in Child Mortality: Report 2022 - World, 2023). Presently, in sub-Saharan Africa, some countries experience child mortality rates exceeding 10%, which implies that one in every ten children born in these countries does not survive to reach their 5th birthday (Kieny, et al. 2018).

Crucially, there is a growing body of evidence suggesting that the expansion of healthcare coverage does not always lead to improved outcomes, even for locations that are highly amendable for medical care (Powell et al, 2015). This highlights the detrimental impact of low-quality health systems on overall child mortality. These deaths arise from causes that 'should not occur in the presence of timely and effective healthcare' (Adhikari et al., 2010). Healthcare facilities in LMICs (including sub-Saharan Africa) often lack essential care units and high-quality healthcare (Kruk et al, 2010). This is particularly detrimental for critically ill children, as they depend on quality healthcare for their survival (Kayambankadzanja et al, 2020).

Vincent et al. (2014) found that there is a significant correlation between the risk of death of critically ill patients and the gross domestic product (GDP) of a country. While the GDP plays a significant role, various studies have found that several other factors influence patient mortality rates in healthcare facilities within sub-Saharan countries (Zaka et al, 2018; Katz et al, 2013; Kruk et al, 2018). The study by Kruk et al (2018) found that there is a correlation between GDP and child mortality, primarily due to GDP indirectly influencing the quality of healthcare through financial resources.

This finding aligns with the insights presented in Avedis Donabedian's seminal paper published in the Milbank Quarterly in 1966, wherein he stated that the foundation of a health system's success is the quality of a healthcare system. His framework on quality of care makes the distinction between the structure, the process, and the outcomes of healthcare provision. Donabedian argued that clinical care in low- and middle-income countries often falls short in all three parts. Despite nearly six decades since the publication of his paper, this argument remains relevant for critical care in low-income countries today, as the issue remains unresolved.

A significant factor related to the low quality of healthcare is the absence of specialized pediatric intensive care units (PICUs) in most low-income countries meaning that standard protocols for pediatric advanced life support are missing (Murthy et al, 2015). The lack of PICUs frequently leads to critically ill children being deprived of the necessary treatment they require (Black, 2022). This is

supported by evidence showing that the mere presence of dedicated pediatric ICUs can have a substantial impact on reducing child mortality rates (Slusher et al, 2018; Musa et al, 2018).

Another factor related to low-quality healthcare is the inadequate continuous monitoring of children and neonates by healthcare staff. This is a problem due to the fact that hospitals in Sub-Saharan Africa often suffer from an insufficient number of physicians and support staff (Kayombo, 2013). A prevalent attempt to deal with such a shortage is tasking parents with monitoring their own children's conditions (Kruk et al, 2022). Furthermore, the education and training levels of the healthcare personnel are frequently subpar, with many of them having never received the proper education and medical training. The study by Marie-Paule et al. (2018) states that healthcare workers in seven low- and middle-income African countries achieved accurate diagnoses only between one-third to three-quarters of the time, and clinical guidelines for common conditions were followed less than 45% of the time on average. This highlights a dual challenge of both insufficient healthcare staff and physicians with a very high workload and a deficiency in knowledge to provide high-quality care (Das et al., 2008).

The third factor related to the low quality of healthcare arises from the inadequacy or unavailability of healthcare facilities, equipment, and medicine in hospitals across sub-Saharan Africa, particularly when compared to facilities in high-income countries (Mills, 2014). Especially the availability of monitoring equipment can increase healthcare quality since proper surveillance of children in High Dependency Units (HDUs) and PICUs is crucial for delivering high-quality critical care. The reason for this is that a patient’s condition can deteriorate quickly, and life-threatening events can occur very quickly (Romare et al, 2022). This decreases the burden on the hospital staff.

However, monitoring systems commonly employed in high-income environments are not appropriate and, therefore, lacking in low-income settings due to their significant initial and ongoing high costs (Turner et al, 2019). Furthermore, these systems often lack the necessary flexibility to accommodate the socioeconomic context, working conditions, and workflows found in such settings (Jacobs et al, 2019; Kirigia et al, 2013).

To address the problem of limited monitoring in critical care units in low-resource settings, the IMPALA project has been designed in a collaborative effort between Malawian and European researchers. The focus of the IMPALA project is to aid healthcare workers by implementing a more suitable monitoring system for critically ill children in HDUs in Malawi, a country characterized by alarmingly high child mortality rates in sub-Saharan Africa (World Health Organization, 2022). Malawi. The project has been implemented since July 2022 and is projected to continue until April 2024. By integrating durable cutting-edge sensors and point-of-care biomarkers into predictive machine learning algorithms, the goals of the IMPALA project are to enable healthcare workers to timely detect critical illnesses and facilitate life-saving interventions. This can alleviate the workload of hospital staff, minimize avoidable errors, and thus improve the quality of healthcare (Kamuzu University of Health Sciences, 2022).

Currently, the first prototype of the IMPALA system has been used in two hospitals in Malawi. The purpose of this paper is to assess the effectiveness of the prototype of the IMPALA system on healthcare quality at Zomba Central Hospital. And with that, to find out what the effect of such a monitoring system is on the quality of healthcare in a low-resource setting.

It is important to recognize that the context in which health interventions are implemented can significantly impact their effectiveness. For example, even seemingly simple interventions, like the widespread distribution of malaria rapid diagnostic tests, can fall short of expected outcomes and

have unintended consequences (Romare et al, 2022).Therefore, it is important to evaluate the impact of prototypes before rolling them out at scale.

The research question of this study is as follows: "What is the effect of an improved monitoring system implemented in a pediatric High Dependency Unit in a low-resource setting?”

The paper is organized as follows: The methodology section begins by providing an overview of the study's setting and the IMPALA program. It then delves into the outcomes and control variables, followed by an explanation of the propensity score analysis. The subsequent section presents the results of variable cleaning and the propensity score analysis, along with various robustness checks conducted on the findings. Following the results, the paper proceeds with a discussion and concludes with the final remarks of the study.

1. Methods

Setting and study population

The IMPALA project has been implemented at Zomba Central Hospital's pediatric HDU in the district of Zomba, Southern Malawi. Zomba is largely a rural district with a population of approximately 746,000 (National Statistical Office, Preliminary Report, 2018). The Zomba Central Hospital is one of the four central hospitals in Malawi, functioning as a district hospital and a tertiary referral center for the other hospitals in Zomba. This study focuses specifically on the pediatric section of the hospital, which caters to children below 12 years of age. Critically ill neonates are, however, admitted to a separate neonatal ward. The pediatric HDU design consists of an admission room where children are assessed and triaged for placement in either medical or surgical wards, as well as a dedicated pediatric HDU with two rooms with eight beds. Both HDU rooms admit critically-ill children who require intensive monitoring (Mwale et al, 2023).

The study population comprises children aged 0 to 5 years old who require intensive monitoring in the pediatric HDU at Zomba Central Hospital. Data was gathered from both the Pre-IMPALA period (i.e., between the 1st of January and the 30th of April 2022) and the IMPALA 2.0 period (i.e., between the 1st of January and the 30th of April 2023). The comparison between Pre-IMPALA and IMPALA 2.0 data was restricted to the same four-month period (i.e., from January until April) to minimize the likelihood of substantial variations in the types of illnesses due to seasonal changes (World Health Organization, 2023).

The IMPALA monitoring system

Prior to this project, the HDU lacked continuous monitoring. In the HDU there was all sorts of hospital equipment such as blood pressure meters and thermometers. However, monitoring systems were often malfunctioning and/or insufficient in numbers. Because of this, the nurses often relied on manual monitoring of the children. This was carried out by a small, insufficient, group of nurses and was limited to four moments throughout the day (8 AM, 12 PM, 4 PM, and 8 PM). During these periods, the nurses would measure the oxygen saturation (SpO2) levels, heart rate, respiratory rate, temperature and blood pressure. They would also visually assess if any issues were present in the child's condition.

The IMPALA monitoring system is being designed to capture children's vital signs and biomarkers which will ultimately be processed through advanced predictive machine learning algorithms enabling the early detection of life-threatening situations. The new system will be able indicate the presence of critical illnesses hours before highly trained healthcare professionals would typically be able to recognize them.

The development of the IMPALA monitoring system includes three phases, namely: Pre-IMPALA, IMPALA 2.0, and IMPALA 3.0. The pre-IMPALA phase started in January 2022 and had as its main goal to assess different context-related factors that should be considered in the design of the IMPALA system's prototype. This prototype is launched during the IMPALA 2.0 phase that started in January 2023. During this phase, clinical implementation research is done as well as data collection for refining the usability of the system and developing the predictive algorithms that will serve as input for IMPALA 3.0.

This study focuses on evaluating the effect of the prototype of the IMPALA monitoring system (further referred to as IMPALA 2.0 system) compared to Pre-IMPALA. The IMPALA 2.0 system is a user-friendly, sturdy, low-cost monitoring system that measures vital signs (i.e., temperature, heart rate, respiratory rate, oxygen saturation, and blood pressure) and can send alarm signals, but it does

not include the machine learning algorithms to predict critical illness events yet. This prototype is used to help create the predictive algorithm of the eventual IMPALA 3.0 monitoring system.

The prototype does, however, display the oxygen saturation and ECGs of a child on a monitor and can alert the hospital through alarms when vital sign monitoring shows that vital thresholds are exceeded. This can suggests that a CIE will occur.

In the HDU there were two types of nurses: clinical nurses and nurses specifically assigned by the IMPALA program to oversee the functionality of the monitors. The clinical nurses were instructed not to make any changes to the monitors in order to avoid disrupting the study. However, this instruction was often interpret as not using the monitors at all and ignoring them[1]. Consequently, when alarms were triggered, they sometimes were ignored by the clinical nurses. They were, however, able to see the monitors, so the monitors could still impact the quality of healthcare in the HDU. The nurses assigned by the IMPALA study utilized the monitors, but primarily for monitoring their functionality rather than for clinical purposes.

The IMPALA system's prototype is used to help create the predictive algorithm of the eventual IMPALA 3.0 monitoring system.

Ethical statement

Ethical approval for the research (P.01/22/3552) was granted by the College of Medicine Research and Ethics Committee (COMREC) at the University of Malawi on the 6th of October 2022. The research methods followed the principles of the Declaration of Helsinki and COMREC guidelines on Health Research. Permission was obtained from the Zomba Central Hospital. Written informed consent was provided with a signature or thumbprint. Participants were assured that their confidentiality would be maintained, and no personal details would be given free. Their participation in the research was voluntary, and they had the right to withdraw at any time without facing any consequences.

Outcomes

In order to assess the impact of the IMPALA system's prototype on the quality of healthcare in Zomba Central Hospital's HDU, it is necessary to define measures for healthcare quality. According to Donabedian's framework (1966), healthcare quality can be defined by considering the structure, process, and outcomes of a healthcare system.

In this framework, 'structure' refers to the characteristics of the healthcare setting in which care is provided. The 'process' is defined as the actual delivery of care, so the utilization of resources and the actions involved in providing and receiving care. The 'outcome' describes the effects of healthcare on the health status of patients and a population.

This study focuses on the outcomes associated with the IMPALA system's prototype. Donabedian distinguishes between final outcomes, such as mortality, morbidity, disability, and quality of life, as well as intermediate outcomes, including blood pressure, body weight, personal well-being, functional ability, coping ability, and improved knowledge.

After examining the data from both datasets, three outcomes have been identified based on Donabedian's final outcomes: Duration of HDU stay, the Occurrence of a Critical Illness Event (CIE) during their stay, and the Number of Deaths.

[1] Lieke De Mare. Product developer of the IMPALA system at GOAL3. She has spent three months during IMPALA 2.0 in Malawi conducting research on the further advancement of the IMPALA system.

1. Duration of HDU stay: This study assesses whether there has been a decrease in the length of children's HDU stay, indicating a reduced workload on the staff. The duration of HDU stay is defined as the difference in hours between the moments of admission and discharge from the HDU. To avoid mistakes within the data from the data entry process, the times and dates from the recruitment form and discharge form are compared to the dates recorded in the daily sheets data.

2. Occurrence of a CIE: This study examines whether there has been a decrease in the occurrence of critical illness events following the implementation of the IMPALA 2.0 system. Critical illness events (CIEs) are defined as the occurrence of life-threatening events during HDU stay. These events are categorized into one of the following six categories informed by the healthcare staff: Respiratory, Circulatory, Neurological, Infectious and Other.
   The occurrence of a CIE gives insight into the morbidity of the critically-ill children, as well as possible future disabilities of the children and the quality of life since CIEs are often very invasive and recovering from them can have lifelong implications (Azoulay et al, 2017). A decline in CIEs would suggest the system's effectiveness in preventing or managing critical health issues. The occurrence of a CIE is measured by looking at the daily sheets data. When a child experienced a CIE during their stay at the HDU, the value '1' is assigned, if not the value '0'.
   For comparison, an examination is conducted to compare the occurrence of CIEs using the recruitment data and the daily sheets data. Moreover, an analysis is performed to determine the total number of CIEs, considering instances where a child encountered multiple CIEs. This analysis is carried out using both the recruitment and daily sheets data.

3. Number of deaths: The final outcome of Donabedian 'mortality' is investigated by looking at the number of deaths before and after IMPALA 2.0. This information was based on the daily sheets data informed by the healthcare staff.

To determine whether there was a change in these outcomes due to the IMPALA system's prototype, this study compares data from before the implementation of this system and after implementation. These datasets are called 'pre-IMPALA' for the data before implementation, and 'IMPALA 2.0' for the dataset after implementation. The IMPALA 2.0 data will be used as the treatment group and the pre-IMPALA dataset as the control group. Children in the IMPALA 2.0 dataset were labeled as 1 to indicate they have undergone treatment, while children in the pre-IMPALA dataset were labeled as 0 to indicate they have not undergone treatment.

Data source

Both Pre-IMPALA and IMPALA 2.0 data contain details of children admitted to the HDU in Zomba Central Hospital. The Pre-IMPALA data was obtained from paper-based medical record archives spanning from the 1st of January to the 30th of April 2022. The records were screened to only include children admitted to the HDU. The medical records contained a recruitment form, daily sheets with vital signs, CIEs and other medical information on a daily basis, and a discharge form.

The Pre-IMPALA data was retrospectively entered into the Research Electronic Data Capture (REDCap) system. Three research assistants from Kamuzu University of Health Sciences (KUHeS) were recruited and trained to collect retrospective baseline data from the paper-based medical records and entering it into REDCap. Sometimes the files of a child were put aside if that child has died during their stay at the HDU. Data quality control measures were implemented through randomized quality checks by a medical doctor on the research team.

As for the period during the implementation of the IMPALA-2.0, children were enrolled in the study between 1st of January and 30th of April 2023. Parents consented for their children to be participants in the study, and be monitored through the IMPALA 2.0 device. The data only includes children who were admitted to the HDU, based on the recruitment forms that were entered into REDCap by nurses. Nurses also directly entered the information from the daily sheets as well as the discharge forms directly into the REDCap tablets. Vital signs of the children were recorded through a the IMPALA 2.0 monitoring system directly attached to them. Following data collection, study statisticians performed data cleaning procedures.

Since the parents had to consent for their children to take part in the study, the dataset used in this study is a selective subsample of children admitted to HDU in 2023. This is one of the reasons matching is used in this study to compare both datasets.

Data cleaning and preparation

Prior to the analysis, Prior to the analyses, I cleaned and prepared the data in four steps, to ensure its quality and reliability. the data underwent a cleaning process to ensure its quality and reliability and included four steps. The initial step involved preparing both Pre-IMPALA and IMPALA 2.0 data for use in the StataSE-17.0, the statistical program used for this research. There were no issues with the IMPALA 2.0 data, as it had already been entered in a STATA file format. However, the Pre-IMPALA data needed to be converted from a csv file to a STATA file. This conversion involved manually adjusting the date and time variables in the csv file to conform to the format used in StataSE. This was done by changing these variables in Excel to the format “MM, YYYY, DD, hh, mm, ss", and then changing them back in STATA.

The second step of the data cleaning process involved addressing additional issues related to the date and time variables in both datasets. Specifically, it focused on resolving significant errors or inconsistencies that could not be fixed. These challenges included addressing issues such as duplicate entries, admission dates not aligning with the first daily sheet dates, substantial differences (exceeding a week) between various daily observations, and discharge dates not matching the last daily observed dates on the last daily sheet. As a consequence, certain dates had to could be manually corrected based on additional information from the paper-based files could be manually corrected, while in other cases, observations or record IDs had to be dropped.

The third step of the data cleaning and preparation procedure involved the creation of new variables based on observations from the data and involved the preparation of existing variables for the analysis. Several variables were generated, including three outcome variables: the duration of ICU stay, the number of critical illness events, and the number of deaths. Additionally, various covariates were generated, which were used to match the children from both Pre-IMPALA and IMPALA 2.0 data.

The final step of the data cleaning and preparation process involved dealing with missing observations in variables. Given that observations with missing values are automatically excluded from the propensity score model by default, mean imputation was used to handle missing values of covariates at admission (White & Thompson, 2005). This was done because analysis including only complete cases is inefficient as it reduces study power (White & Thompson, 2005). Thus, to assess the robustness of the results obtained from imputed data, a "complete case analysis (CCA)" was conducted only including children with completed covariate values at admission. The same propensity score method was used to compare the CCA results to the results of the imputed data.

Covariates – matching variables

To analyze the difference in healthcare quality between the period before the implementation and during the implementation of the IMPALA 2.0 system, propensity score matching was used (Blundell & Costa Dias, 2000). To this end, children admitted during IMPALA 2.0 were matched with children admitted pre-IMPALA based on their so-called propensity score.

Propensity scores estimate the likelihood of a patient receiving treatment (i.e., the implementation of the IMPALA 2.0 system) given a set of children's characteristics further referred to as covariates (Blundell & Costa Dias, 2000). The covariates were used to match children from both Pre-IMPALA and IMPALA 2.0 data, creating a control group and a treatment group with comparable characteristics and minimizing confounding variables. The covariates used in this study are categorized into two groups: (1) Descriptive variables pertaining to the children, such as age and sex, and (2) vital signs that were recorded during recruitment.

Descriptive variables

This study incorporated various descriptive variables concerning the children, including age, sex, weight, HIV status, and month of recruitment. The variables 'age' and 'weight' were treated as continuous variables. The variables 'sex' (1 = male; 2 = female) and HIV status (0 = no HIV, 1 = HIV) were considered binary variables. The variable 'month of recruitment' was categorized as 1 (January), 2 (February), 3 (March), and 4 (April). Appendix, Table 1 provides a more detailed overview of these covariates, including the input variables used to create them. Although additional descriptive variables related to the children's parents and previous health conditions were available for the IMPALA 2.0 dataset, they were not present in the Pre-IMPALA data and thus were excluded from the analysis.

Vital signs

The recruitment data and daily sheets contain information on five vital signs that are monitored at the recruitment of the children and throughout their stay in the HDU. These vital signs include temperature, heart rate, respiratory rate, oxygen saturation, and blood pressure. Since the variable 'blood pressure' has no entered values for both data, this variable was excluded from the analysis.

To create the vital signs variables, continuous values from the recruitment forms were first imputed and then transformed into categorical variables. A value of 1 or 0 is assigned based on whether a child has reached a vital threshold or not. These threshold values are determined by the World Health Organization (2013) guidelines. The threshold values are as follows:

1. Oxygen Saturation: Any measurement below 92% was considered life-threatening (i.e., = 1).
2. Temperature: Any measurement exceeding 38 degrees Celsius is considered abnormal (i.e., = 1), regardless of the child's age.
3. Heart Rate: The threshold for heart rate varies based on the child's age. If the child is between 0 and 1 year old, the normal heart rate should be between 100 and 160 beats per minute (bpm). For children between 1 and 3 years old, the range is between 90 and 150 bpm, and for children between 3 and 6 years old, between 80 and 140 bpm. Any heat rate outside those limits was considered abnormal (i.e., = 1).
4. Respiratory Rate: The threshold for respiratory rate also depends on the child's age. For infants less than 2 months old, the normal rate should be between 40 and 60 breaths per minute. For infants between 2 and 11 months old, it should range from 25 to 50 breaths per minute. For children between 1 and 5 years old, the range is between 20 and 40 breaths per minute. Any respiratory rate outside those limits was considered abnormal (i.e., = 1).

Propensity score analysis

After cleaning the covariates and harmonizing them between data from Pre-IMPALA and IMPALA 2.0, the IMPALA 2.0 group was labeled as 1 (i.e., treatment group) and the Pre-IMPALA group as 0 (i.e., control group) to indicate which record IDs are from the treatment group and which from the control group. Thereafter, both data were appended into one dataset to estimate the propensity scores.

A probit models was used to estimate propensity scores for each observation. In this model, the 'D' stands for the dependent variable (i.e., the group receive treatment or not). The 'x' indicates the independent pre-treatment characteristics (the covariates). The probit model gives the probability of receiving a treatment given pre-treatment characteristics x as indicated in Equation 1 (Schafer & Kang, 2008).

$$p(x) = prob(D = 1|x) = E(D|x) \quad (1)$$

After estimation of the propensity scores, the record IDs from the treated group and the created control group are matched based on these propensity scores. For this, one of three matching methods was used: nearest neighbor, radius, or kernel. The decision on which of these preferred matching methods to use was decided based on whether the propensity score successfully balances the data (Rosenbaum & Rubin 1985). This balancing is tested using the pstest function in StataSE. The remaining two not chosen methods were used for robustness checks.

After matching, the average treatment effect on the treated (ATET) was calculated (Formula Equation 2) and empirically estimated (Formula Equation 3). Theoretically, ATET shows the difference between the outcomes of the treated and the outcomes of the treated observations had they have not been treated. p(x) indicates the matched variables based on the propensity scores.

$$ATET = E(\Delta|p(x), D = 1) = E(y1|p(x), D = 1) - E(y0|p(x), D = 0) \quad (2)$$

$$ATET = \frac{1}{n1}\sum_{i(D=1)}(y_{1,i} - \sum_j w(i,j)y_{0,i}) \quad (3)$$

These steps were carried out in StataSE 17.0 using the implementation PSmatch2 (version st0026_2) (Leuven and Sijanesij, 2003). This package uses full Mahalanobis matching and multiple propensity score matching techniques to account for observable differences before treatment between a treated group and an untreated group (Leuven and Sjanesj, 2003). The decision which propensity score method to use was tested using the pstest function in StataSE. This function shows the difference between the treated and untreated for each covariate before and after matching. If the matching has been successful, the standardized bias is between -10% and +10 % (Rosenbaum & Rubin 1985).

To match the datasets better, the function 'common support' is used to omit propensity score ranges where there is insufficient overlap between the treatment and control groups. This allows for more meaningful matching and analysis. If there is limited common support, it may indicate difficulties in finding suitable matches or potential bias in the estimation of treatment effects (Garrido et al, 2014).

The minimum sample size

After propensity score matching, the sample size that is needed to detect a meaningful effect without wasting resources, is calculated. This is done using a power analysis for a one-sample mean test in Stata. This function estimates the required sample size, the power, and effect size for a test comparing a mean to a reference value. The values used are the difference in means and the

standard errors calculated during propensity score matching. The default levels are chosen for the significance level and power, these are respectively: 0.05 and 0.8.

The null hypothesis H0: $\mu = 0$ and the alternative hypothesis H1: $\mu \neq 0$.

2. Results

Data cleaning

The study sample includes 525 children who were admitted to the HDU between 1st of January and 30th of April of 2022 and 2023. After data cleaning, 485 record IDs of children were left in. Among them, 303 children (62%) were from the IMPALA 2.0 dataset, while 182 children (38%) belonged to the Pre-IMPALA dataset. Missing observations of the covariates were replaced with the means of these covariates for both the Pre-IMPALA and IMPALA 2.0 datasets. There were no missing values for the outcome variables.

Table 2 presents the summary statistics of the variables after completing the data cleaning process. The values in this table show that the children in the Pre-IMPALA period were, on average, younger compared to those of IMPALA 2.0. This difference is, however, not statistically significant on a 95% confidence interval. Additionally, the Pre-IMPALA children had lower average recruitment weight (Difference of means (DM)=-2.5 kg; 95% CI ± 0.3kg) and heart rate (DM=0.09; 95% CI ± 0.04). There is also a statistical significance between the recruitment month (DM=0.7; 95% CI ± 0.1). The values show that in the Pre-IMPALA group there were relatively more children in the months March and April, while in the IMPALA 2.0 group there were more children in the month January and February.

Furthermore, it was observed that in the Pre-IMPALA group, a higher proportion of children had a heart rate above the vital threshold compared to the IMPALA 2.0 group (DM=0.09; 95% CI ± 0.04). No significant differences were found in the vital signs of oxygen saturation and respiratory rate and in the variables 'sex' and 'HIV exposure'. The vital sign 'temperature' was omitted due to a potential bias caused by all children in the IMPALA 2.0 group having levels above the vital threshold.

Table 1: Baseline characteristics of the children

	Before propensity score matching		**After propensity score matching**	
	Pre-IMPALA (N=182)	IMPALA 2.0 (N=303)	Pre-IMPALA (N=182)	IMPALA 2.0 (N=295)
Descriptive variables				
Age in months (Mean, Std. dev.)	18 (13)	20 (16)	18 (13)	20 (16)
Sex (Male; Female; percentage))	82; 55 (60%; 40%)	163; 127 (56%; 44%)	82; 55 (60%; 40%)	159; 123 (56%; 44%)
Weight in kg (Mean, Std. dev.)	6.5 (3.6)	9.1 (3.4)	6.5 (3.6)	8.9 (3.3)
Recruitment month (Amount per month; percentage)	Jan: 14, Feb: 37; Mar: 63; Apr: 68 (8%, 20%, 35%, 37%)	Jan: 86, Feb: 85; Mar: 71; Apr: 58 (29%, 28%, 24%, 19%)	Jan: 14, Feb: 37; Mar: 63; Apr: 68 (8%, 20%, 35%, 37%)	Jan: 79, Feb: 84; Mar: 71; Apr: 58 (27%, 29%, 24%, 20%)
HIV exposure (Number of children; percentage)	4 (2.2%)	16 (5.3%)	4 (2.2%)	12 (4.1%)
Vital signs				
Oxygen saturation (Number of children; percentage)	25 (17%)	68 (24%)	25 (17%)	63 (23%)
Temperature (Number of children; percentage)	18 (13%)	303 (100%)	*Omitted due to perfect collinearity*	*Omitted due to perfect collinearity*
Heart rate (Number of children; percentage)	61 (42%)	95 (33%)	61 (42%)	92 (33%)
Respiratory rate (number of children)	13 (37%)	103 (40%)	13 (37%)	103 (40%)

**Percentages in this study are calculated using the number of available observations for a particular variable. The missing observations are therefore not taken into account when calculating these percentages.*

Propensity score matching

For the propensity score matching, the kernel matching method is was used. The reason for this is was that this method. The kernel matching demonstrated the most effective balancing of the matched variables, as evidenced by the %bias observed (Table 1 & Appendix, Table 2). Radius matching performed significantly worse than kernel and nearest neighbour matching. The majority of variables have %biases ranging from -10% to +10% which shows successful matching (Rosenbaum & Rubin 1985). While the nearest neighbor matching method also achieved favorable balancing, it was not selected due to the kernel-based matching method generally performing better than one-to-one propensity score matching (nearest neighbour matching) when there is an insufficient number of potential controls available (Berg, 2011).

After propensity score matching, a total of 303 (100%) children from the IMPALA 2.0 group were successfully matched. Figure 1 shows that the control group and the treatment group are much more similar after propensity score matching than before.

Utilizing the common support options, only 5 observations fell outside the range of support. The matching procedure effectively reduced many of the disparities between the two groups, as evidenced in Table 1 under the column labeled 'after propensity score matching'. Most of the baseline characteristics were balanced after the matching process. However, the variables 'age' and 'recruitment month' are the only ones that remain outside the acceptable range of -10% to +10%.

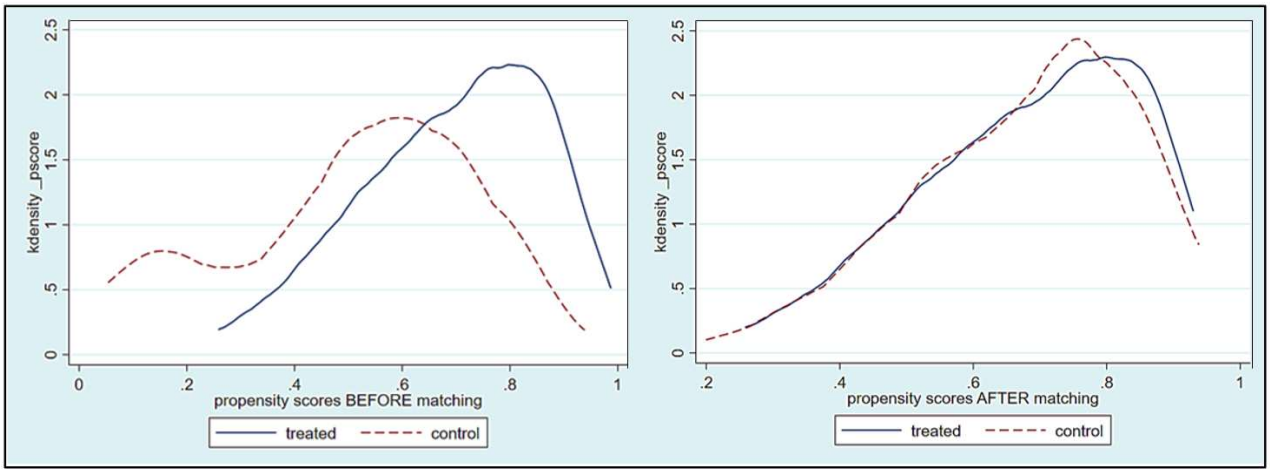

Figure 1: Comparison of the treated and control group before and after propensity score matching. The density of the kernel matching is on the y-axis, the propensity scores are on the x-axis.

Table 2 shows the results of each outcome variable using kernel propensity score matching. It shows the mean in the treatment and control group (columns 1 and 2) as well as the difference, its standard error in parentheses and the 95% confidence interval, both before and after matching. The difference in Column 3 after matching is the coefficient of interest.

After matching, the mean duration of stay in the HDU in the treatment group is 4.2 hours lower compared to the control group (decreasing from a mean in the control group of 69.4 hours). However, this decrease is not statistically significant.

The coefficient of the mean for the variable 'Occurrence of a CIE' in the treatment group is 0.027 lower than for the control group (95% CI ± 0.037). This decrease is statistically significant and can be

interpreted as children in the treatment group experiencing 36% less CIEs during their stay at the HDU.

The mean of the 'number of Deaths' increases from the control group to the treatment group by a coefficient of 0.028 . This increase is not statistically significant.

Table 2: Outcomes propensity score analysis (kernel matching) for all 3 outcome variables

	Treated group	**Control group**	**Difference (SE)**	**95% confidence interval**
Duration of HDU stay (hours)				
Before propensity score matching	64.828	73.407	-8.578 (5.732)	-19.813; 2.657
After propensity score matching	65.258	69.434	-4.176 (5.696)	-15.340; 6.988
Occurrence of a CIE (0,1)				
Before propensity score matching	0.261	0.346	-0.085 (0.043)	-0.043; 0.000*
After propensity score matching	0.258	0.403	-0.145 (0.058)	-0.259; -0.031*
Number of Deaths				
Before propensity score matching	0.040	0.027	0.012 (0.017)	-0.021; 0.045
After propensity score matching	0.043	0.013	0.027 (0.019)	-0.064 ; 0.010

**Statistically significant (0.05)*

Robustness of the results

To evaluate the robustness of the results, two additional methods of propensity score analysis, namely nearest neighbor matching and radius matching, were used. Three additional datasets were generated for this purpose. The first dataset excluded a greater number of record IDs compared to the original dataset. The second dataset incorporated recruitment data into the variable 'Occurrence of a CIE,' while the third dataset examined the variable 'Number of CIEs.'

Nearest neighbour matching and radius matching

By using nearest neighbor matching (refer to Appendix, Table 3), the analysis revealed a reduction in the duration of stay at the HDU by 7.8 hours (95% CI ± 14.9 hrs). The variable 'Number of CIEs' showed a coefficient decrease of 0.104 (95% CI ± 0.135), while the variable 'Number of Deaths' exhibited a coefficient increase of 0.024 (95% CI ± 0.065). However, none of these findings are statistically significant.

Applying radius matching, the duration of stay at the HDU showed a decrease of 7.7 hours (95% CI ± 8.5 hrs). The variable 'Number of CIEs' displayed a coefficient decrease of 0.088 (95% CI ± 0.057), and the variable 'Number of Deaths' demonstrated a coefficient increase of 0.013 (95% CI ± 0.024). The decrease in the 'Number of CIEs' was statistically significant, while the differences observed in the other two outcomes were not statistically significant.

The complete case analysis

The first newly created dataset is referred to as the 'complete case analysis'. This dataset included only those record IDs that had no issues or missing data. In total, the dataset comprises 254 record

IDs, with 31 (13%) from the pre-IMPALA dataset and 233 (87%) from the IMPALA 2.0 dataset. Appendix Table 4 presents the summary statistics of the variables within this dataset.

Using kernel propensity score matching on this dataset (see Appendix, Table 5), it was observed that the duration of stay at the HDU exhibited an average reduction of 12.2 hours between the control group and the treatment group. The variable 'number of CIEs' displayed a coefficient decrease of 0.062, while the variable 'number of deaths' showed a coefficient increase of 0.047. However, none of these findings were statistically significant at a 95% confidence level.

Occurrence of a CIE with the recruitment data

In order to evaluate the occurrence of a CIE with the recruitment data, an additional propensity score analysis was conducted. This analysis focused not only on the on the daily sheets data but also on the recruitment dataset. The inclusion of the recruitment data was important as it could potentially influence the daily sheets data. Although this approach may yield less 'pure' results and introduce some bias, it is still crucial to consider this aspect.

Appendix Table 6 presents the results of the propensity score analysis for this variable using all three matching methods. Kernel matching revealed a significant decrease in the coefficient of 0.137 (95% CI ± 0.073). Similarly, both nearest neighbor matching and radius matching demonstrated statistically significant decreases in the coefficient between the control group and the treatment group.

The total number of CIEs

Next to estimating the effect of the occurrence of CIEs, propensity score matching is also used for the total number of CIEs in both datasets. Appendix Table 7 indicates that with the results of kernel propensity score matching, a decrease of 0.164 (95% CI ± 0.202) is found for this variable. Similarly, both nearest neighbor matching and radius matching demonstrate coefficient decreases, with a mean of 0.242 (95% CI ± 0.245) for nearest neighbor matching and a mean of 0.093 (95% CI ± 0.104) for radius matching. However, none of these coefficients are statistically significant.

The minimum sample size

Using Stata, the minimum required sample size and effect size were calculated. As shown in Appendix Table 8, for the variable 'Duration of stay' to have a statistically significant difference between the means of the treatment and control groups, a minimum sample size of 4,354 is needed. For the variable 'Occurrence of a CIE,' a sample size of 376 is sufficient, while the variable 'Number of Deaths' requires a sample size of 1,161. Since the current sample size is 485, it indicates that there is only a sufficiently large sample size for the variable 'Occurrence of a CIE.'

3. Discussion

Main findings

In order to estimate the impact of the IMPALA's system prototype on healthcare quality in Zomba Central Hospital's HDU in in Malawi, a comparison was made between the pre- IMPALA control group and the IMPALA 2.0 treatment group. Three outcome variables; the duration of stay in the HDU, the number occurrence of a critical illness events, and the number probability of deaths, were used to evaluate healthcare quality.

Through the utilization of kernel propensity score matching, this study identified a statistically significant decrease of 0.145 (95% CI ± 0.114) in the variable 'Occurrence of a CIE' between the pre-IMPALA and IMPALA 2.0 time periods. This implies that children in the HDU who received treatment using the IMPALA 2.0 system's prototype experienced a 36% reduction in CIEs compared to children during the pre-IMPALA period. This finding is not only statistically significant but also holds practical significance. Similar results were obtained with 'radius matching', while 'nearest neighbor matching' showed a decrease that was not statistically significant.

The additional two outcome variables, 'Duration of Stay' and 'Number of Deaths', revealed no significant difference between the two time periods.

The variable 'Occurrence of a CIE' was further analyzed by including the recruitment dataset. In this version, a significant decrease between the two time periods was also observed. However, when examining the variable 'Number of CIEs', no statistically significant decrease was observed.

The 'complete case analysis' also showed no statistically significant findings for any of the three outcome variables.

Explanation of the main findings and comparison with the literature

The lack of a significant difference in the results of the 'complete case analysis' could be attributed to the limited sample size of the dataset, which consisted of only 254 record IDs, particularly with regards to the pre-IMPALA data, which only comprised 31 record IDs. Such a small sample size may not provide enough statistical power to detect significant differences between the two time periods.

Deaths were probably not always recorded in the paper-based medical records. That would mean that there may be under-reporting of deaths in the pre-IMPALA dataset which could be a possible explanation for the slight (albeit insignificant) increase in mortality (Number of Deaths).

The lack of a significant difference in the 'Duration of Stay' variable may be attributed to the relatively small sample size. Combining both datasets, there were only 525 children, and after data cleaning, the sample size reduced to 485. However, it is recommended to have a larger sample size of at least 1,000 to 1,500 for matching in order to minimize sampling errors (Shadish et al., 2013). Based on calculations, an even higher sample size (at least 4,350) would be suggested for accurate results in relation to the duration of stay.

The significant difference observed between the pre-IMPALA and IMPALA 2.0 periods in terms of the occurrence of a CIE can be explained by an exploratory qualitative analysis conducted by Mwale (2023) for the pre-IMPALA phase, as well as a personal interview with De Mare in June 2023[2] regarding the IMPALA 2.0 phase.

[2]Lieke De Mare. Product developer of the IMPALA system.

Mwale (2023) describes that monitoring during the Pre-IMPALA was intermittent rather than continuous, and suffered from various issues. His study involved in-depth interviews with selected healthcare workers in the HDU. He shows that there are still numerous challenges in the monitoring of critically-ill children within the HDU. He found that the main challenges involved the design of the HDU, insufficient staff and hospital caregivers and lack of monitoring devices as well as limited training on how to use existing monitoring tools.

The prototype may therefore still be an advancement compared to the previous situation: Before the prototype, vital signs monitoring occurred four times per day (unless patients were deteriorating) and there was no continuous monitoring. With the monitoring system, this has become more continuous by displaying the oxygen saturation and ECGs on a monitor and alerting the nurses when thresholds were exceeded. The system was however not used very effectively and often nurses ignored alarms that went off.

Nevertheless, the fact that the IMPALA system's prototype still decreased the number of children having a CIE during their stay shows that even when the system is not consistently used yet, it still has a positive effect.

This is in line with the findings of other studies that researched the effect of a simple monitoring devices on healthcare quality in low resource settings. Ryu et al. (2021) demonstrated that monitoring devices specifically designed for laboring women and their fetuses, incorporating advanced measurements such as continuous blood pressure measurement, uterine electrohysterography, and automated body position classification, were effective in improving healthcare quality in low resource settings.

Similarly, Hao et al. (2021) showcased the benefits of easy-to-use wearable monitoring devices in Vietnamese intensive care units. They further demonstrated that leveraging the monitoring data for machine learning purposes, particularly for risk prediction through training systems with observed data, significantly enhanced patient outcomes.

The IMPALA 3.0 program aims to achieve a similar objective through the utilization of risk prediction based on vital sign monitoring and machine learning. Therefore, it can be expected that this version of the system will have a more positive impact on healthcare quality. However, further research is necessary to evaluate the effects once the IMPALA 3.0 phase has started.

Strengths and limitations

This study uses propensity score analysis to compare the Pre-IMPALA period to the IMPALA 2.0 period. As this is a nonexperimental research design, it is important to acknowledge potential endogeneity issues (Lobmeier, 2010) that may arise, such as differences in outcome variables due to disparities between the two years. Propensity score matching addresses this concern by matching the children from both periods based on 9 different covariates. This helped reduce the differences between the groups, making them more comparable.

One of the limitations of propensity score matching is that it depends largely on the selected covariates. Biases can arise if important covariates are not included in the matching process. In this study, it would have been beneficial to include certain covariates from the IMPALA 2.0 dataset in the Pre-IMPALA dataset for a more robust matching process. These covariates, such as parental information and children's medical history, could have provided valuable insights. Unfortunately, incorporating these covariates was not feasible due to data unavailability in the Pre-IMPALA dataset.

In addition, a general drawback of matching methods is that by definition it is not possible to control for unobserved differences between the treatment and comparison groups. If those differences are systematic (i.e. not random) than that can introduce a bias in the results.

The datasets used in this study, have more treatment variables than control variables, which differs from the typical scenario in propensity score matching where there are usually more control variables than treatment variables. Nevertheless, this discrepancy should not be a significant issue as the goal of matching is not solely focused on finding pairs of record IDs. Instead, the primary objective is to effectively pair down the data (Ho et al., 2007), which has been done relatively well.

Another limitation of the study pertains to the data collection process. These data-related concerns highlight the possibility of unobserved biases within the dataset used for this study. The Pre-IMPALA dataset has, for example, numerous missing observations, evident from the small number of record IDs remaining in the 'complete case analysis'. This can create a bias in the results.

Additionally, both datasets had several issues related to the time variables. There was a notable time gap observed between different observations in the daily sheet data and there were recruitment and discharge dates that did not correspond to the first and last daily sheet dates. These inconsistencies raise concerns regarding the accuracy and reliability of the temporal information within the datasets.

Furthermore, based on observations of De Mare, there was probably a higher number of deaths in both the Pre-IMPALA period as well as during the IMPALA 2.0 period. This is probably due to files of children that died being set apart for M&M discussions. This probably created a bias in the results of the outcome variable ‘Number of Deaths’. Therefore, further research should look into another way to take mortality into account when assessing the quality of healthcare.

Implications for clinical practice and future research

Since the IMPALA 3.0 system wants to improve on the IMPALA 2.0 system by adding a predictive part, further research should look at the effectiveness of this monitoring system on the quality of healthcare in the HDU.

In addition to the three outcome variables examined in this research, it is essential to consider other significant measures of healthcare quality. These may include non-technical variables that assess both clinical quality and the perceived quality of care (Hanefeld et al., 2017). To capture these dimensions, one approach is to quantify the outcomes of qualitative research through the use of surveys. This allows for the systematic assessment of qualitative data and incorporates these qualitative insights into the evaluation of healthcare quality.

Further research should also explore the impact of a monitoring system in combination with effectively trained staff. It is important to investigate how the utilization of monitors in a more optimal manner can potentially yield more significant improvements in healthcare quality. By ensuring that healthcare professionals are adequately trained to operate the monitoring system effectively, the potential benefits and impact on healthcare outcomes can be further improved.

4. Conclusion

This study showed that children treated in the HDU using the IMPALA system's prototype experienced a significant reduction in critical illness events compared to the period before implementing this system. This significant difference was observed across two different propensity score matching methods. However, when a smaller dataset excluding records with missing data was used, this significant difference was not found. There was no significant difference for the duration the children that stayed in the HDU and the number of deaths of children. However, the observed results of this particular variable can be attributed to an underrepresentation of deaths in the files. This underrepresentation may be a consequence of files being set aside prior to their use in research.

These findings combined with previous published literature suggest that implementing a monitoring system can have a notable impact on healthcare quality. It is anticipated that this impact may be even greater when the data from the monitoring system is utilized for predictive purposes. Consequently, further research is required to evaluate the effects of the new IMPALA 3.0 system after its implementation, particularly in combination with well-trained hospital staff, to assess its impact on healthcare quality.

References

- Adhikari, N.K.J. Fowler, R.A., Bhagwanjee S., et al. (2010). Critical care and the global burden of critical illness in adults. The Lancet; 376:1339–46.

-Azoulay, E., Vincent, JL., Angus, D.C. et al. (2017). Recovery after critical illness: putting the puzzle together—a consensus of 29. Crit Care 21, 296. https://doi.org/10.1186/s13054-017-1887-7

- Black, R. E. (2022, 30 april). Health and development from preconception to 20 years of age and human capital. The Lancet. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)02533-2/fulltext

- Blundell, R., Dias, M. C. (2005). Evaluation Methods For Non-experimental Data. Fiscal Studies, 4(21), 427-468. https://doi.org/10.1111/j.1475-5890.2000.tb00031.x

- Das, Jishnu, Jeffrey Hammer, and Kenneth Leonard. (2008). "The Quality of Medical Advice in Low-Income Countries." Journal of Economic Perspectives, 22 (2): 93-114. DOI: 10.1257/jep.22.2.93

-Donabedian A. (2005). Evaluating the quality of medical care. 1966. The Milbank quarterly, 83(4), 691–729. https://doi.org/10.1111/j.1468-0009.2005.00397.x

- Garrido, M. M., Kelley, A. S., Paris, J., Roza, K., Meier, D. E., Morrison, R. S., & Aldridge, M. D. (2014). Methods for constructing and assessing propensity scores. Health services research, 49(5), 1701–1720. https://doi.org/10.1111/1475-6773.12182

- General Assembly Economic and Social Council - United Nations (2023). Progress towards the Sustainable Development Goals: Towards a Rescue Plan for People and Planet Report of the Secretary-General (Special Edition). ADVANCE UNEDITED VERSION.

-Gregory D. Berg (2011) An application of kernel-based versus one-to-one propensity score matching for a nonexperimental causal study: example from a disease management program evaluation, Applied Economics Letters, 18:5, 439-447, DOI: 10.1080/13504851003689692

- Hanefeld, J., Powell-Jackson, T., & Balabanova, D. (2017). Understanding and measuring quality of care: dealing with complexity. Bulletin of the World Health Organization, 95(5), 368–374. https://doi.org/10.2471/BLT.16.179309

- Ho, D. Imai, K., King, G. and Stuart, E.. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Political Analysis, 15, Pp. 199–236. Retrieved from: https://tinyurl.com/y4xtv32s

- Jacobs, B., Hui, K., Lo, V. et al. (2019). Costing for universal health coverage: insight into essential economic data from three provinces in Cambodia. Health Econ Rev 9, 29 https://doi.org/10.1186/s13561-019-0246-6

- Kayombo, E. J. (2013). Impact of Training traditional birth attendants on maternal mortality and morbidity in Sub-Saharan African countries. Tanzania Journal of Health Research, 15(2). https://doi.org/10.4314/thrb.v15i2.7

- Kieny, M. P., Evans, T. G., Scarpetta, S., Kelley, E. T., Klazinga, N., Forde, I., ... & Kim, S. M. (2018). Delivering quality health services: a global imperative for universal health coverage. Washington, DC: World Bank Group.

- Kirigia, J.M., Asbu, E.Z. (2013). Technical and scale efficiency of public community hospitals in Eritrea: an exploratory study. Health Econ Rev 3, 6

- Kruk ME, Freedman LP, Anglin GA, Waldman RJ. (2010). Rebuilding health systems to improve health and promote statebuilding in post-conflict countries: a theoretical framework and research agenda. Soc Sci Med; 70: 89–97.

- Kruk, M. E. (2018, 17 november). Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries. The Lancet. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)31668-4/fulltext

- Leuven, E. and Sianesi, B. (2003). "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing". Retrieved from: http://ideas.repec.org/c/boc/bocode/s432001.html This version. st0026_2

- Levels & Trends in Child Mortality: Report 2022 - World. (2023, January 9). ReliefWeb. https://reliefweb.int/report/world/levels-trends-child-mortality-report-2022

-Kamuzu University of Health Sciences (2022). Study protocol for ethical review: A predictive monitoring system for critically ill children in Malawi: a mixed-methods usability and implementation research (IMPALA-UIR).

-Katz J, ACC L, Kozuki N, Lawn JE, Cousens S, Blencowe H, et al. (2013). Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis. Lancet. 382:417–25.

-Kayambankadzanja, R. K., Likaka, A., Mndolo, S. K., Chatsika, G. M., Umar, E., & Baker, T. (2020). Emergency and critical care services in Malawi: Findings from a nationwide survey of health facilities.

-Kieny,Marie-Paule; Evans,Timothy Grant; Scarpetta,Stefano; Kelley,Edward T.; Klazinga,Niek; Forde,Ian; Veillard,Jeremy Henri Maurice; Leatherman,Sheila; Syed,Shamsuzzoha; Kim,Sun Mean; Nejad,Sepideh Bagheri; Donaldson,Liam. (2018). Delivering quality health services: a global imperative for universal health coverage (English). Washington, D.C. : World Bank Group.

-Kruk, M. E., Lewis, T. P., Arsenault, C., Bhutta, Z. A., Irimu, G., Jeong, J., Lassi, Z.S., Sawyer, S.M., Vaivada T., Waiswa, P. & Yousafzai, A. K. (2022). Improving health and social systems for all children in LMICs: structural innovations to deliver high-quality services. The Lancet.

-Max Roser, Hannah Ritchie and Bernadeta Dadonaite (2013) - "Child and Infant Mortality". Published online at OurWorldInData.org. Retrieved from: https://ourworldindata.org/child-mortality

- Mills, A. (2014). Reflections on the development of health economics in low- and middle-income countries. Proceedings of the Royal Society B: Biological Sciences, 281(1789), 20140451. https://doi.org/10.1098/rspb.2014.0451

- Murthy, S., Leligdowicz, A., & Adhikari, N. K. (2015). Intensive care unit capacity in low-income countries: a systematic review. PloS one, 10(1), e0116949. https://doi.org/10.1371/journal.pone.0116949

- Musa NL, Dursun O, Phan PH, et al. Pediatric Critical Care in Resource-Limited Settings—Overview and Lessons Learned. 6. Epub ahead of print 2018. DOI: 10.3389/fped.2018.00049.

- Mwale, D. Manda-Taylor, L. Langton, J. Likumbo, A. van Hensbroek, M.B. Calis, J. Janssens, W. Pell, C. (2023). The role of healthcare workers and caregivers in monitoring critically ill children: a qualitative study in Malawi.

- National Statistical Office. Preliminary Report. (2018). Malawi Population and Housing Census, www.nsomalawi.mw

- Powell-Jackson, T., Mazumdar, S., & Mills, A. (2015). Financial incentives in health: New evidence from India's Janani Suraksha Yojana. Journal of health economics, 43, 154–169. https://doi.org/10.1016/j.jhealeco.2015.07.001

- Romare C, Anderberg P, Sanmartin Berglund J, et al. (2022). Burden of care related to monitoring patient vital signs during intensive care; a descriptive retrospective database study. Intensive Crit Care Nurs; 71: 103213.

- Ryu, D., Kim, D., Price, J. T., Lee, J. Y., Chung, H. U., Allen, E., … & Xu, S. (2021). Comprehensive Pregnancy Monitoring With a Network Of Wireless, Soft, And Flexible Sensors In High- And Low-resource Health Settings. Proc. Natl. Acad. Sci. U.S.A., 20(118). https://doi.org/10.1073/pnas.2100466118

-Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: a practical guide and simulated example. Psychological methods, 13(4), 279–313. https://doi.org/10.1037/a0014268

-Shadish, W. R., Rindskopf, D. M., Hedges, L. V., & Sullivan, K. J. (2013). Bayesian estimates of autocorrelations in single-case designs. Behavior Research Methods, 45(3), 813–821. https://doi-org.vu-nl.idm.oclc.org/10.3758/s13428-012-0282-1

- Slusher TM, Kiragu AW, Day LT, Bjorklund AR, Shirk A, Johannsen C, et al. (2018). Pediatric critical care in resource-limited settings-overview and lessons learned. Front Pediatr. 6:49. doi: 10.3389/fped.2018.00049

- Statology (2020). Two Sample t-test: Definition, Formula, and Example. Retrieved from: https://www.statology.org/two-sample-t-test/

- The World Bank (n.d.).Propensity Score Matching. Retrieved from: https://dimewiki.worldbank.org/Propensity_Score_Matching#:~:text=Propensity%20score%20matching%20(PSM)%20is,the%20impact%20of%20an%20intervention.

- The World Health Organization (2023) Influenza (Seasonal) overview. Factsheet. Retrieved from: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)

- Turner HC, Hao N Van, Yacoub S, et al. (2019). Achieving affordable critical care in low-income and middle-income countries. BMJ Glob Health; 4: e001675.

- Van HMT, Hao NV, Phan Nguyen Quoc K (2021). On behalf of the Vietnam ICU Translational Applications Laboratory (VITAL) Investigators, et al. Vital sign monitoring using wearable devices in a Vietnamese intensive care unit. BMJ Innovations;7:s7-s11.

- Vincent, J. L., Marshall, J. C., Namendys-Silva, S. A., François, B., Martin-Loeches, I., Lipman, J., Reinhart, K., Antonelli, M., Pickkers, P., Njimi, H., Jimenez, E., Sakr, Y., & ICON investigators (2014). Assessment of the worldwide burden of critical illness: the intensive care over nations (ICON) audit. The Lancet. Respiratory medicine, 2(5), 380–386. https://doi.org/10.1016/S2213-2600(14)70061-X

- World Health Organization (2013). Pocket Book of Hospital Care for Children: Guidelines for the Management of Common Childhood Illnesses. ISBN-13: 978-92-4-154837-3

- World Health Organization (2022). Child Mortality Report World Health Organisation. Geneva, Switzerland: WHO

- World Health Organization (2022). Child Mortality Report World Health Organisation. Geneva, Switzerland: WHO

-White IR, Thompson SG. (2005). Adjusting for partially missing baseline measurements in randomized trials. Stat Med.24(7):993–1007.

-World Health Organization (n.d.). Child mortality and causes of death. Retrieved from: https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/child-mortality-and-causes-of-death#:~:text=Since%201990%2C%20the%20global%20under,to%20

Appendix

Table 1: Summary of the inputs and the types of the variables used for the analysis

Variables	**Input variables Pre-IMPALA**	**Input variables IMPALA 2.0**	**Variable type**
Outcome variables			
Duration	Recruitment HDU admission date, daily time	Recruitment interview date, daily time	Continuous
Amount of CIEs	Recruitment CIE number, daily new CIE number	Daily new CIE number	Continuous
Deaths	CIE type, discharge data outcome	Daily type CIE, discharge data outcome	Binary: 1 yes, 0 no
Descriptive variables			
Age (months)	Date of birth, HDU admission date	Date of birth, recruitment date	Continuous
Sex	Recruitment sex	Recruitment sex	Binary: 1 male, 2 female
Weight (kg)	Recruitment weight	Recruitment weight	Continuous
Recruitment month	Recruitment HDU admission date, daily time	Recruitment interview date, daily time	Categorical: 1 Jan, 2 Feb, 3 Mar, 4 Apr
HIV exposure	Recruitment HIV status	Recruitment HIV status	Binary: 1 yes, 0 no
Vital signs			
Oxygen saturation	Oxygen saturation (continuous variable)	Oxygen saturation (continuous variable)	Binary: 1 vital, 0 non vital
Temperature	Temperature (continuous variable)	Temperature (continuous variable)	Binary: 1 vital, 0 non vital
Heart rate	Heart rate (continuous variable)	Heart rate (continuous variable)	Binary: 1 vital, 0 non vital
Respiratory rate	Respiratory rate (continuous variable)	Respiratory rate (continuous variable)	Binary: 1 vital, 0 non vital

Table 2: Balance testing of the propensity score analysis for kernel matching, radius matching and nearest neighbour matching.

Covariate	Pre-IMPALA system	IMPALA 2.0 system	%bias	P score
Nearest neighbour matching				
Age (months)	19.84	15.0	33.1	0.000
Sex	1.44	1.3	22.9	0.006
Weight (kg)	8.94	9.1	-3.5	0.623
Recruitment month	2.40	2.6	-19.7	0.021
HIV exposure	0.04	0.03	7.2	0.364
Oxygen saturation	0.23	0.23	-1.6	0.862
Heart rate	0.33	0.36	-6.5	0.443
Respiratory rate	0.38	0.40	-2.2	0.814
Radius matching				
Age (months)	19.84	17.83	13.6	0.242
Sex	1.44	1.40	7.6	0.507
Weight (kg)	8.94	6.53	68.6	0.000
Recruitment month	2.40	3.02	-60.0	0.000
HIV exposure	0.04	0.02	9.9	0.366
Oxygen saturation	0.23	0.17	14.9	0.196
Heart rate	0.33	0.41	-19.4	0.085
Respiratory rate	0.38	0.37	5.1	0.683
Kernel matching				
Age (months)	19.84	17.94	12.9	0.106
Sex	1.44	1.44	-0.2	0.984
Weight (kg)	8.94	8.97	-0.9	0.904
Recruitment month	2.40	2.51	-11.1	0.196
HIV exposure	0.04	0.01	8.1	0.297
Oxygen saturation	0.23	0.26	-8.0	0.368
Heart rate	0.33	0.34	-3.6	0.659
Respiratory rate	0.38	0.43	-10.1	0.251

Table 3: Outcomes propensity score analysis (nearest neighbor & radius matching) for all 3 outcome variables

	Treated group	**Control group**	**Difference (SE)**	**95% confidence interval**
Duration				
Before propensity score matching	64.828	73.407	-8.578 (5.732)	-19.813; 2.657
After propensity score matching (nearest neighbour)	65.742	73.539	-7.797 (7.579)	-22.652; 7.058
After propensity score matching (radius)	65.742	73.407	-7.664 (4.360)	-16.210; 0.882
Occurrence of a CIE				
Before propensity score matching	0.201	0.346	-0.085 (0.043)	-0.043; 0.000*
After propensity score matching (nearest neighbour)	0.258	0.362	-0.104 (0.069)	-0.239; 0.031
After propensity score matching (radius)	0.258	0.346	-0.088 (0.029)	-0.145; -0.031*
Deaths				
Before propensity score matching	0.040	0.027	0.012 (0.017)	-0.021; 0.045
After propensity score matching (nearest neighbour)	0.041	0.017	0.024 (0.033)	-0.041; 0.089
After propensity score matching (radius)	0.041	0.027	0.013 (0.012)	-0.011; 0.037

**Statistically significant (0.05)*

Table 4: Baseline characteristics of the children of the 'complete case analysis'

	Before propensity score matching		After propensity score matching	
	Pre-IMPALA (N=32)	IMPALA 2.0 (N=254)	Pre-IMPALA (N=31)	IMPALA 2.0 (N=223)
Descriptive variables				
Age (months)	19 (14)	20 (17)	18 (14)	20 (17)
Sex (male; female)	17; 15 (53%; 47%)	147; 107 (58%; 42%)	17; 14 (55%; 45%)	130; 93 (58%; 42%)
Weight (kg)	6.8 (4.4)	9.1 (3.6)	6.7 (4.4)	9.1 (3.6)
Recruitment month	Jan: 5, Feb: 10; Mar: 9; Apr: 8 (16%, 31%, 28%, 25%)	Jan: 81, Feb: 69; Mar: 56; Apr: 48 (32%, 27%, 22%, 19%)	Jan: 5, Feb: 8; Mar: 10; Apr: 8 (16%, 26%, 32%, 26%)	Jan: 74, Feb: 58; Mar: 51; Apr: 40 (33%, 26%, 23%, 18%)
HIV exposure (number of children)	2 (6%)	15 (6%)	2 (6%)	13 (6%)
Vital signs				
Oxygen saturation (number of children)	8 (25%)	63 (25%)	9 (29%)	55 (25%)
Temperature (number of children)	4 (13%)	254 (100%)	*Omitted due to perfect collinearity*	*Omitted due to perfect collinearity*
Heart rate (number of children)	10 (31%)	83 (33%)	12 (39%)	73 (33%)
Respiratory rate (number of children)	13 (41%)	104 (41%)	13 (42%)	90 (40%)

Table 5: Outcomes propensity score analysis 'complete case analysis' (kernel matching) for all 3 outcome variables

	Treated group	Control group	Difference (SE)	95% confidence interval
Duration				
Before propensity score matching	59.974	72.906	-12.933 (11.914)	-36.284; 10.418
After propensity score matching	59.767	71.942	-12.175 (8.432)	-28.702; 4.352
Occurrence of a CIE				
Before propensity score matching	0.248	0.343	-0.095 (0.080)	-0.252; 0.062
After propensity score matching	0.245	0.308	-0.062 (0.103)	-0.264; 0.140
Deaths				
Before propensity score matching	0.044	0	0.044 (0.036)	-0.027; 0.115
After propensity score matching	0.047	0	0.047 (0.010)	0.027; 0.067*

**Statistically significant (0.05)*

Table 6: Occurrence of CIEs without the recruitment data

	Treated group	Control group	Difference (SE)	95% confidence interval
Occurrence of a CIE (0,1)				
Before propensity score matching	0.802	0.918	-0.116 (0.034)	-0.183; -0.049*
After propensity score matching (kernel)	0.807	0.944	-0.137 (0.037)	-0.210; -0.065*
After propensity score matching (nearest neighbour)	0.807	0.946	-0.139 (0.046)	-0.229; -0.049*
After propensity score matching (radius)	0.807	0.918	-0.111 (0.024)	-0.158; -0.064*

**Statistically significant (0.05)*

Table 7: Outcomes kernel propensity score analysis for variable 'Number of CIEs'

	Treated group	Control group	Difference (SE)	95% confidence interval
Number of CIEs				
Before propensity score matching	0.399	0.485	-0.086 (0.080)	-0.243; 0.071
After kernel propensity score matching	0.393	0.557	-0.164 (0.103)	-0.366; 0.038
After nearest neighbour propensity score matching	0.393	0.634	-0.242 (0.125)	-0.487; 0.003
After radius propensity score matching	0.393	0.485	-0.093 (0.053)	-0.197; 0.011

**Statistically significant (0.05)*

Table 8: Minimum sample sizes

	Difference in means	Standard deviation	Current sample size	Minimum sample size
Duration of HDU stay (hours)	-4.176	98.328	485	4,354
Occurrence of a CIE (0,1)	-0.145	1.271	485	376
Number of Deaths	0.027	0.328	485	1,161