



## RESEARCH ARTICLE

# Linking health facility data from young adults aged 18-24 years to longitudinal demographic data: Experience from The Kilifi Health and Demographic Surveillance System [version 1; referees: 1 approved, 1 approved with reservations]

Christopher Nyundo <sup>1,2</sup>, Aoife M. Doyle<sup>3</sup>, David Walumbe<sup>1,2</sup>, Mark Otiende<sup>1,2</sup>, Michael Kinuthia<sup>1</sup>, David Amadi<sup>1</sup>, Boniface Jibendi<sup>1</sup>, George Mochamah<sup>1,2</sup>, Norbert Kihuha<sup>1</sup>, Thomas N. Williams<sup>1,2,4</sup>, David A. Ross<sup>3</sup>, Evasius Bauni<sup>1,2</sup>

<sup>1</sup>KEMRI/Wellcome Trust Research Programme, Kilifi, Kenya

<sup>2</sup>INDEPTH, Accra, Ghana

<sup>3</sup>London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK

<sup>4</sup>Imperial College, London, SW7 2AZ, UK

**v1** First published: 17 Jul 2017, 2:51 (doi: [10.12688/wellcomeopenres.11302.1](https://doi.org/10.12688/wellcomeopenres.11302.1))  
Latest published: 17 Jul 2017, 2:51 (doi: [10.12688/wellcomeopenres.11302.1](https://doi.org/10.12688/wellcomeopenres.11302.1))

## Abstract

**Background:** In 2014, a pilot study was conducted to test the feasibility of linking clinic attendance data for young adults at two health facilities to the population register of the Kilifi Health and Demographic Surveillance System (KHDSS). This was part of a cross-sectional survey of health problems of young people, and we tested the feasibility of using the KHDSS platform for the monitoring of future interventions.

**Methods:** Two facilities were used for this study. Clinical data from consenting participants aged 18-24 years were matched to KHDSS records. Data matching was achieved using national identity card numbers or otherwise using a matching algorithm based on names, sex, date of birth, location of residence and the names of other homestead members. A study form was administered to all matched patients to capture reasons for their visits and time taken to access the services. Distance to health facility from a participants' homestead was also computed.

**Results:** 628 participated in the study: 386 (61%) at Matsangoni Health Centre, and 242 (39%) at Pingilikani Dispensary. 610 (97%) records were matched to the KHDSS register. Most records (605; 96%) were matched within these health facilities, while 5 (1%) were matched during homestead follow-up visits. 463 (75.9%) of those matched were women. Antenatal care (25%), family planning (13%), respiratory infections (9%) and malaria (9%) were the main reasons for seeking care. Antenatal clinic visits (n=175) and malaria (n=27) were the commonest reasons among women and men, respectively. Participants took 1-1.5 hours to access the services; 490 (81.0%) participants lived within 5 kilometres of a facility.

**Conclusions:** With a full-time research clerk at each health facility, linking health-facility attendance data to a longitudinal HDSS platform was feasible and could be used to monitor and evaluate the impact of health interventions on health care outcomes among young people.

## Open Peer Review

Referee Status:

	Invited Referees	
	1	2
version 1 published 17 Jul 2017	 report	 report

- 1 **Philippe Bocquier** , Université Catholique de Louvain, Belgium  
University of the Witwatersrand, South Africa
- 2 **Fiona J Stanley**, University of Western Australia, Australia  
**Hannah C. Moore**, University of Western Australia, Australia

## Discuss this article

Comments (0)

**Keywords**

Record linkage, Demographic Surveillance, Adolescent Health, Kenya



This article is included in the [KEMRI | Wellcome Trust](#) gateway.

**Corresponding author:** Christopher Nyundo ([cnyundo@kemri-wellcome.org](mailto:cnyundo@kemri-wellcome.org))

**Competing interests:** No competing interests were disclosed.

**How to cite this article:** Nyundo C, Doyle AM, Walumbe D *et al.* **Linking health facility data from young adults aged 18-24 years to longitudinal demographic data: Experience from The Kilifi Health and Demographic Surveillance System [version 1; referees: 1 approved, 1 approved with reservations]** Wellcome Open Research 2017, 2:51 (doi: [10.12688/wellcomeopenres.11302.1](https://doi.org/10.12688/wellcomeopenres.11302.1))

**Copyright:** © 2017 Nyundo C *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Grant information:** This work was supported by the Wellcome Trust [084535; 091758]; and the Hewlett Foundation through a grant awarded to INDEPTH [IND/0107].

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**First published:** 17 Jul 2017, 2:51 (doi: [10.12688/wellcomeopenres.11302.1](https://doi.org/10.12688/wellcomeopenres.11302.1))

## Abbreviations

HDSS: Health and Demographic Surveillance System

ID: Identification

INDEPTH: International Network of field sites with continuous Demographic Evaluation of Populations and Their Health in developing countries

## Introduction

Health and Demographic Surveillance Systems (HDSS) provide longitudinal information on populations living within geographically-defined areas, including data on fertility, mortality and migration<sup>1-3</sup>. These data can be useful to public health policy makers both locally and internationally<sup>2-5</sup>. When HDSS data are combined with data from health services, they can be used to monitor and evaluate outcomes of research and health care programmes<sup>1-3</sup>.

Data linkage is a process of pairing records from two data sources or bringing together information from two records that relate to the same individual or entity<sup>6-8</sup>. This linkage process frequently involves the use of basic socio-demographic indices that uniquely identify an individual across two or more datasets<sup>8</sup>. Although there is growing interest in integrating HDSS and health service data<sup>3</sup>, examples of how this can be performed are relatively rare<sup>5,9-11</sup>.

The Kilifi Health and Demographic Surveillance System (KHDSS) was established in 2000 to monitor births, deaths, pregnancies and migration events within a sub-population of Kilifi County on the coast of Kenya<sup>1</sup>. The KHDSS covers an area of 891km<sup>2</sup> with a resident population of approximately 280,000 in 2016. The area is served by 1 referral hospital, three health centres, 14 dispensaries and numerous private health service providers. KHDSS data include basic details of all homesteads and the names, dates of birth, sex, national identity (ID) card numbers and ethnicity of all homestead members, and are updated with births, deaths and in- and out-migration events three times a year. A homestead comprises of one or more houses or dwelling units with people, also referred as residents, who have one person they refer to as the head. Geographic coordinates for dwelling units and health facilities are routinely collected using global positioning system technology (GPS). The KHDSS population register has been linked to a surveillance system of children admitted to Kilifi County Hospital since 2001, and linkage was expanded to cover the maternity and adult wards admissions in subsequent years<sup>1</sup>. Finally, data on childhood immunizations administered at 30 government and private not-for-profit health facilities have been linked to the register since 2008<sup>12</sup>. However, data linkage for other age groups attending these peripheral health facilities has not been done. In this study, we linked young adults (18–24 year-olds) attending health facilities within the KHDSS area and asked their reasons for visiting. This exercise was part of the International Network of field sites with continuous Demographic Evaluation of Populations and Their Health (INDEPTH) Healthy Transitions to Adulthood Study (IHTAS), which also aimed to describe the nutritional and health problems of young people aged 13–24 years in Kilifi and Dodowa HDSS, both members of INDEPTH

Linking data relating to adolescents and young adults who attend such facilities was anticipated to provide a useful platform to monitor and evaluate interventions to improve adolescent health<sup>13-15</sup>.

## Methods

### Study area

The study was conducted within the area served by the KHDSS on the coast of Kenya (Figure 1) during a 3 month period from mid-August 2014 to end of October, 2014. Six health facilities within the area were evaluated with a view to include two within our pilot study on the basis of the availability of research space, youth-friendly services, a reliable power supply and a willingness to participate in the study. We characterised youth-friendly health care services as the availability of health workers who were trained to deal with and handle respectfully and confidentially the heterogeneous health issues of young people<sup>15,16</sup>. Two facilities were chosen: Matsangoni Health Centre in the northern part of our study area and Pingilikani Dispensary in the south. Young adults aged 18–24 years who sought health services from the facilities were the focus of our study.

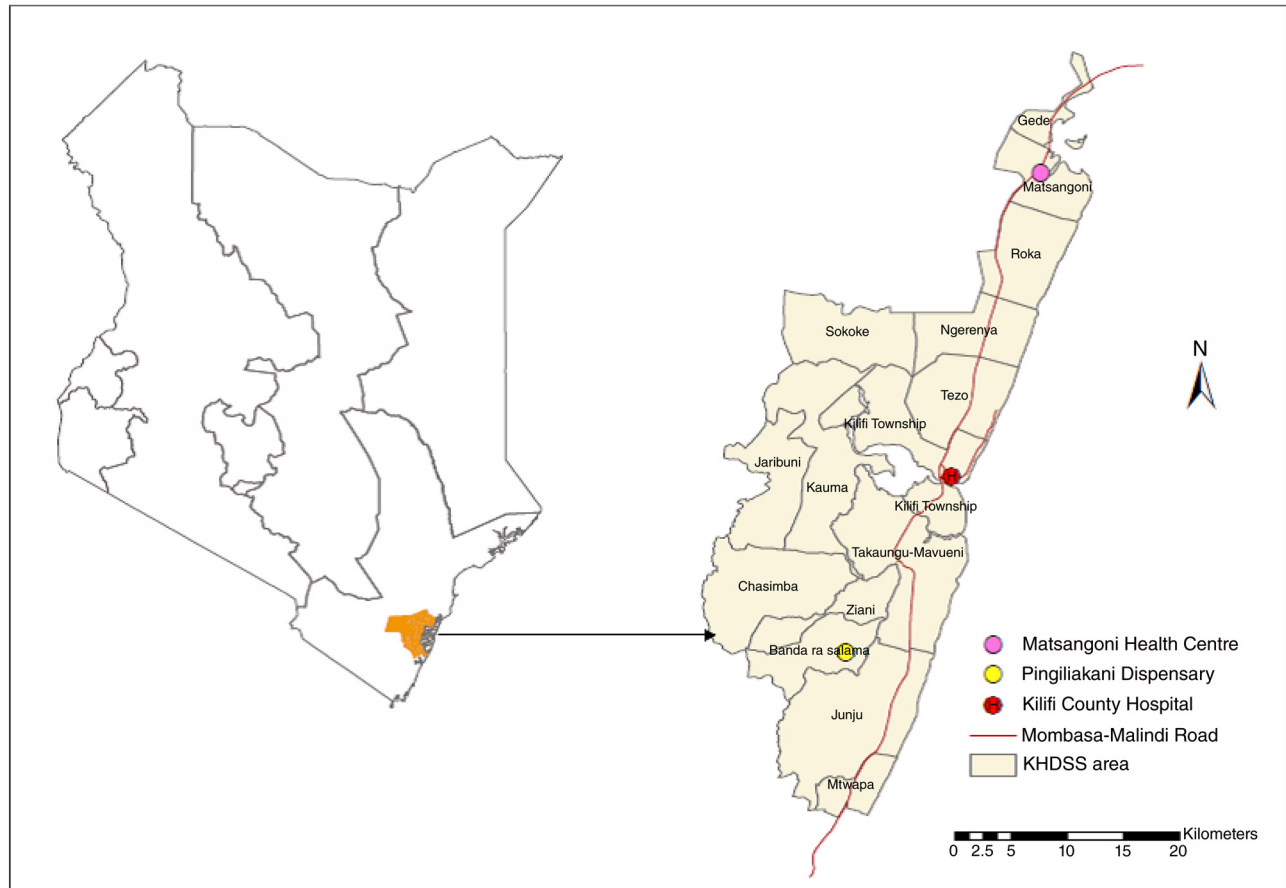
### Participant recruitment

Data clerks were trained on consenting, matching, and the administration of the study questionnaire (Supplementary File 1). We used laptop computers that were up-loaded weekly with the most recent version of the KHDSS database to identify and match resident young adults who visited these health facilities. All consenting patients aged 18–24 years were eligible to participate in the study.

Young people were identified at the health facility reception desk and referred to the study data clerk for assessment of eligibility, consenting, matching, linkage to the population register and collection of study-specific information. After obtaining consent from the participant, the data clerk searched the participant's records in the KHDSS records. Then those who successfully matched were issued with a study form on which the data clerk recorded the facility name, date and time of arrival. The patient proceeded with the form to the attending clinician who recorded the purpose of the visit. After completing facility procedures, including the collection of prescribed medications, the patient returned the form to the data clerk who recorded the time of departure. For those not found on the KHDSS register, details were recorded in a separate form (Supplementary File 2) for further matching by more experienced experts and to facilitate home visit follow-up. During the home visits, experienced DSS staff identified the homestead with its members, and with the help of other homestead members confirmed whether the participant's records existed in the KHDSS database.

### Matching

The matching and data-linkage process commenced after written consent had been obtained. We used ID numbers where available, or an algorithm based on first and second names, date of birth, gender, and ethnicity, and the homestead name, location and sub-location. The KHDSS database algorithm is a search engine that retrieves a set of records that meet the set criteria. The correct match was obtained by narrowing down using the names of the homestead head, the list of other homestead members, or varying the first 3 letters of names to account for spelling variations. Uncertainties



**Figure 1. Study area.**

regarding dates of birth were overcome by using a range of options of “+/- 1 year”, “+/-3 years” of the date of birth and “DO NOT KNOW” on instances the person does not know the date of birth

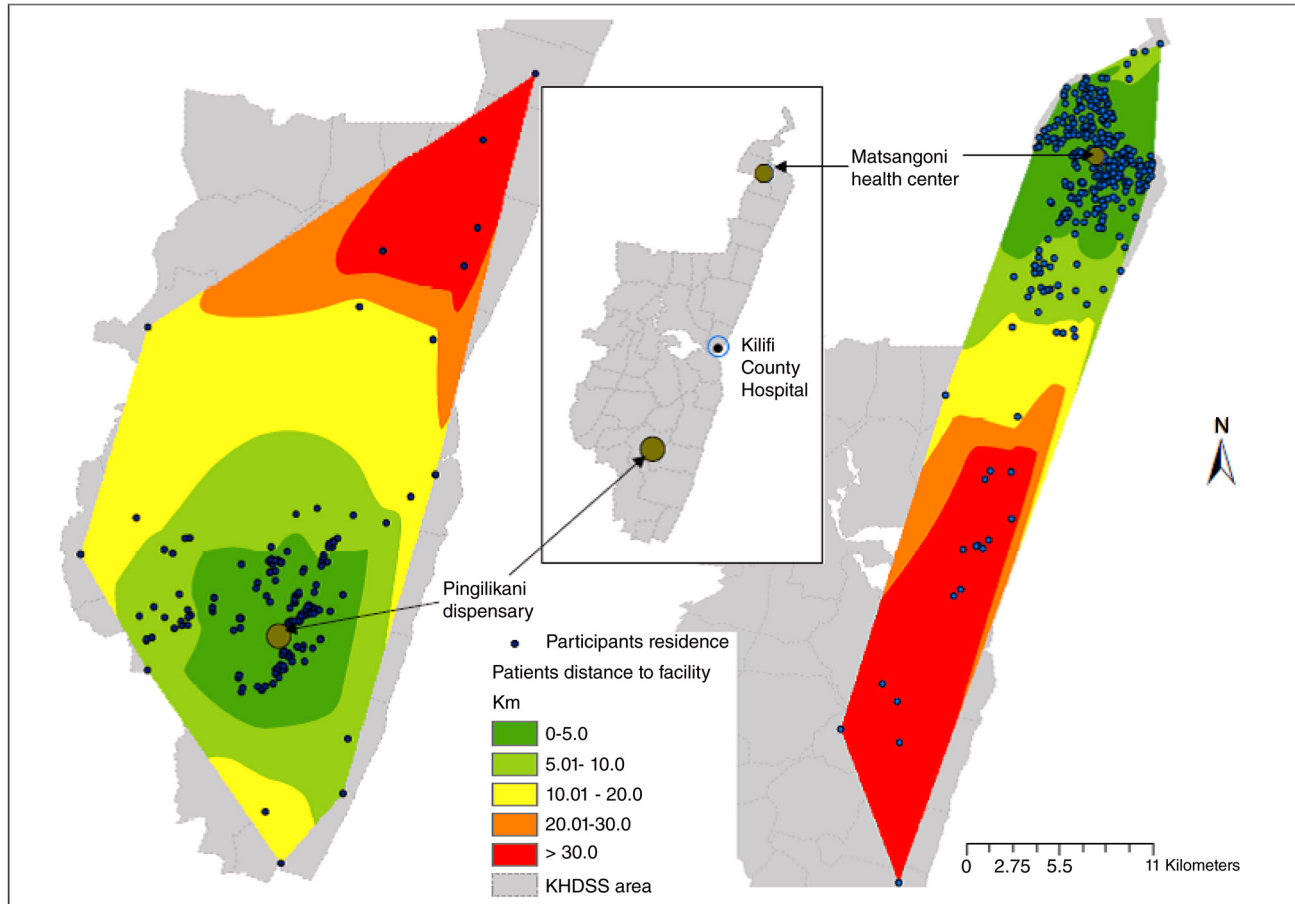
Distance to health facility from residence of consenting participants was computed using ArcGIS 10.1 desktop software (ESRI, Redlands, CA, USA) based on pairs of geographic coordinates for the facility and location of participants’ homestead.

## Results

All (n=628) participants who were approached to participate consented, 386 (61.5%) were recruited at Matsangoni Health Centre and 242 (38.5%) at Pingilikani Dispensary. A total of 605/628 participants (96.3%) were successfully matched at the health facilities, a process that took a minimum of 1 minute to a maximum of 3 minutes based on experience of matching patients in the wards and at the vaccine clinics. We failed to match 23 (3.7%) participants at the facility level, who were later followed in the field by more experienced staff and another search was performed with the help of other homestead members. We discovered

that some of these participants existed in the database with names different from those they used at the clinic. This process was successful for a further five patients, but remained unsuccessful for 10, while eight patients could not be found in the community.

Three of the unmatched participants were seen at Pingilikani Dispensary, which is relatively centrally placed, far from the border of the KHDSS area, while the other 15 were seen at Matsangoni Health Centre, which is situated close to the border of the KHDSS area. The homestead visits revealed that none of the 18 participants we failed to match were in the KHDSS population register. Three resided outside the KHDSS area, one was a high school teacher who had recently in-migrated, had given wrong information for the homesteads they said they came from and were not known in these homesteads, another 10 could not be verified because the data collection forms containing details of where they came from were misplaced. However all who did not match at the facility were excluded from the analysis. The spatial distribution of study participants is shown in [Figure 2](#). 81% of participants lived within 5km or less of the health facilities, and only 3.4% were more than 20 km ([Table 1](#)).



**Figure 2.** Participants' distance to health facility from their residence.

**Table 1.** Distance between participants' residence and health facilities.

Distance (km)	Matsangoni Health Centre (N=372)	Pingilikani Dispensary (N=233)	Total (N=605)	%
0-4	306	184	490	81.0
5-9	43	35	78	13.0
10-19	7	9	16	2.6
≥20	16	5	21	3.4

A total of 715 visits to the health facilities were recorded among the 605 matched participants since some had multiple visits during the study period (Table 2). Reasons for facility visit differed by gender and there were many more visits by women than men (548/715; 76.6%). The main reasons for visit by women were antenatal clinic (175; 32%) and family planning services (95; 17%) (Table 2).

Malaria (27; 16%) and skin infections (26; 16%) were the commonest reasons for men.

Matsangoni Health Centre had more visits (n=429) than Pingilikani Dispensary (n= 286) (Table 3). The main reasons for visits differed between the two facilities. The main reasons for visits in Matsangoni Health Centre were antenatal clinic (n=147) or family planning services (n=75), while the main reasons at Pingilikani Dispensary were malaria (n=59), non-infectious conditions (n=36), respiratory infections (n=35), antenatal clinic (n=31) and family planning (n=21).

The month of September had the highest number of visits (317; 44.3%), followed by October (293; 40.9%). The study was done for a half month in August and the results are not a true reflection for a month's observation to be compared with the other months. Antenatal care and family planning were the main reasons for health facility visit in October with 80 and 50 visits, respectively, while malaria, non-infectious, respiratory and skin infections were the main reasons in September. Participants took 10 to 480 minutes to be served at the facility, with a median of 80 minutes.

**Table 2. Overall reasons for health facility visit by gender.**

Reasons	Women, N (%)	Men, N (%)	Total, N (%)
Antenatal care clinic	175 (31.9)	1 (0.6)	176 (24.6)
Family planning	95 (17.3)	1 (0.6)	96 (13.4)
Respiratory infection	47 (8.6)	19 (11.4)	66 (9.2)
Malaria	37 (6.8)	27 (16.2)	64 (9.0)
Other non-infectious	39 (7.1)	18 (10.8)	57 (8.0)
Skin infection	21 (3.8)	26 (15.6)	47 (6.6)
Sexually transmitted infection	25 (4.6)	7 (4.2)	32 (4.5)
Gastro-intestinal disease	18 (3.3)	12 (7.2)	30 (4.2)
Urinary tract infection	21 (3.8)	9 (5.4)	30 (4.2)
Ear, Nose and Throat	15 (2.7)	8 (4.8)	23 (3.2)
Injuries	6 (1.1)	15 (9.0)	21 (2.9)
Chest infection	9 (1.6)	7 (4.2)	16 (2.2)
Asthma	6 (1.1)	3 (1.8)	9 (1.3)
Gynaecological disease	8 (1.5)	1 (0.6)	9 (1.3)
Other infections	5 (0.9)	3 (1.8)	8 (1.1)
HIV	11 (2.0)	1 (0.6)	9 (1.3)
Worms	0 (0.0)	6 (3.6)	6 (0.8)
Anaemia	4 (0.7)	1 (0.6)	5 (0.7)
Epilepsy	0 (0.0)	2 (1.2)	2 (0.3)
Abortion	1 (0.2)	0 (0.0)	1 (0.1)
Eclampsia	1 (0.2)	0 (0.0)	1 (0.1)
Mother Child Health	1 (0.2)	0 (0.0)	1 (0.1)
Pregnancy test	1 (0.2)	0 (0.0)	1 (0.1)
Missing	2 (0.4)	0 (0.0)	2 (0.3)
Total	548 (100.0)	167(100.0)	715 (100.0)

**Table 3. Reasons for visits by health facility.**

Reasons	Pingilikani, N (%)		Matsangoni, N (%)		Total, N (%)
	Women	Men	Women	Men	
Antenatal care clinic	31 (16.7)	0 (0.0)	147 (40.4)	0 (0.0)	178 (24.9)
Abortion	1 (0.5)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
Anaemia	2 (1.1)	1 (1.0)	2 (0.5)	0 (0.0)	5 (0.7)
Asthma	1 (0.5)	3 (3.0)	5 (1.4)	0 (0.0)	9 (1.3)
Ear, Nose and Throat	5 (2.7)	5 (5.0)	10 (2.7)	3 (4.6)	23 (3.2)
Epilepsy	0 (0.0)	1 (1.0)	0 (0.0)	1 (1.5)	2 (0.3)
Family Planning	21 (11.3)	0 (0.0)	75 (20.6)	0 (0.0)	96 (13.4)
Gastro-intestinal disease	8 (4.3)	3 (3.0)	10 (2.7)	9 (13.8)	30 (4.2)
HIV	3 (1.6)	0 (0.0)	8 (2.2)	1 (1.5)	12 (1.7)
Injuries	1 (0.5)	11 (11.0)	5 (1.4)	4 (6.2)	21 (2.9)
Mother child health	1 (0.5)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
Other infections	0 (0.0)	2 (2.0)	6 (1.6)	1 (1.5)	9 (1.3)
Pregnancy test	1 (0.5)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)
Respiratory infection	25 (13.4)	10 (10.0)	22 (6.0)	9 (13.8)	66 (9.2)
Sexually transmitted infection	4 (2.2)	1 (1.0)	21 (5.8)	6 (9.2)	32 (4.5)
Skin infection	8 (4.3)	15 (15.0)	13 (3.6)	11 (16.9)	47 (6.6)
Urinary tract infection	9 (4.8)	7 (7.0)	12 (3.3)	2 (3.1)	30 (4.2)
Chest infection	3 (1.6)	2 (2.0)	6 (1.6)	5 (7.7)	16 (2.2)
Gynecological disease	3 (1.6)	0 (0.0)	5 (1.4)	1 (1.5)	9 (1.3)
Malaria	34 (18.3)	25 (25.0)	3 (0.8)	2 (3.1)	64 (9.0)
Other non-infectious	25 (13.4)	11 (11.0)	14 (3.8)	7 (10.8)	57 (8.0)
Worms	0 (0.0)	3 (3.0)	0 (0.0)	3 (4.6)	6 (0.8)
Total	186 (100.0)	100 (100.0)	364 (100.0)	65 (100.0)	715 (100.0)

### Experience from The Kilifi Health and Demographic Surveillance System: Raw data for linked participants and unmatched participants

2 Data Files

<http://dx.doi.org/10.6084/m9.figshare.5202850>

**Dataset File 1: Linkage data.** Age\_group: age of participant on the day they visited the clinic, three missing values; facility: Matsangoni or Pingilikani; visit\_date: date when participant visited clinic, three missing values; service\_time\_minutes: duration (in minutes) it took for participant to get served, 29 missing values; visit\_reason: reason for visit, four missing values; exclusion: variable indicating the records that were included in the 'clinic visit analysis'; visit\_no: some participants made more than one visit during the study period - this variable indicates the number of visits made at any one instance; distance\_to\_facility: distance (in KM) from the homestead to health facility.

**Dataset File 2: Unmatched participant data.** Data clerk: Study staff who did the consenting, searched and linked study participant's demographic data with the clinic visit data; Hmname: Homestead name. Name a by which a homestead is known and referred to; HmHead: Homestead head. The person who heads a home, and makes important decisions for the family; Location: Administrative unit in the government structure headed by a chief; Sublocation: A sub-unit of a location headed by an assistant chief.

## Discussion

Through this study, we demonstrate that it is possible to link health data collected from the vast majority (97%) of young adults attending two peripheral health facilities to the population register of the KHDSS, using an algorithm with a range of routinely collected personal data. A total of 15 of the unmatched patients had sought services in Matsangoni Health Centre, which is located close to the KHDSS border on the Mombasa-Malindi highway. While national ID card numbers facilitated rapid matching and linkage for a few participants, the majority either did not have ID cards with them, or their ID numbers were not on KHDSS database. Although the speed, accuracy and potential of linking individuals using national ID numbers is consistent with similar observations from other studies<sup>7</sup>, there were particular challenges for this age group. While legally residents are supposed to register for a national ID card on reaching their 18<sup>th</sup> birthday, in practice this is not followed universally. Second, capturing these numbers on the KHDSS database is not immediate as update rounds are only conducted three times a year. Third, not everybody carries their ID cards when seeking health services.

The presence of the participants during the matching process was an advantage because they could quickly identify and confirm other members of their homestead. This speeded up the matching process. On average, in the absence of a national ID card number, it took 2 to 3 minutes to search, match and link an individual. This was reasonably fast and reflects a decade of learning and experience trying to perfect demographic data capture and real-time linkage to clinical data for research in Kilifi, particularly at the Kilifi County Hospital, and also in matching internal migrants. We have also been matching children and mothers in more than 30 facilities

that provide vaccinations for the last 8 years. The search engine used two names only but efficiency, speed and accuracy may improve if more and full names are used in the future<sup>4,9,17,18</sup>. We used a combination of other variables, which include location of residence, the name of the homestead head and the list of other homestead members for further confirmation. The greater speed and accuracy of matching using ID numbers reinforces the importance of recording these numbers when available, and of encouraging young people to get an ID card after reaching the statutory age of 18 years and carrying it when seeking health services.

Most of the successful matching was done using personal details of names, date of birth and ethnicity in combination with homestead name, name of homestead head and location and sub-location. A sub-location is the smallest administrative unit and several of them constitute a location. Failure to match was partly due to the use of different names from those on the population database, as was confirmed for the 5 who were matched during the home visit.

This study confirmed that it was possible to link individual level health data at peripheral health facilities with the longitudinal population register for young adult residents of KHDSS. This will enhance monitoring of health interventions in the future to inform health policy related to improving the health of young adults within our catchment population. With technological advances, increased accuracy and speed in record linkage may be possible with the adoption of a fingerprint biometric database system, although this is not without its own challenges<sup>19</sup>.

### Reasons for facility visit

There were substantial differences in the main reasons for clinic visits in young adults by sex, with antenatal care and family planning being the two commonest reasons for women, and malaria and skin infections for men. Differences by sex are likely to be minimal in young children, but diverge after puberty. The substantial differences by health facility partly reflect differences in malaria prevalence across the KHDSS area, with more malaria in the south where Pingilikani dispensary is situated<sup>20-22</sup>.

The present findings can be useful to inform and guide adolescent and young adult health programming<sup>23,24</sup>. The epidemiological transition is evident in the clinic visit data, where there has been a partial shift away from infectious diseases and undernutrition, with sexually transmitted infections, reproductive health problems and injuries becoming relatively more prominent<sup>13</sup>. The findings confirm previous work by Bauni and colleagues, who have shown that injuries and pregnancy-related conditions are now major causes of death among young men and women, respectively, living within the KHDSS area<sup>25</sup>. Although injury and HIV were not among the main reasons for visiting the health facility, other studies of the same community have shown injury and HIV as the main cause of death for young men and women, respectively<sup>25</sup>. Furthermore, Etyang and colleagues have reported that injuries were the most common cause of admission to Kilifi County Hospital among men, while infectious and parasitic diseases, and pregnancy-related complications were the most common among women<sup>26</sup>. Similarly, a review of medical causes of admission to hospital among adults in Africa reported

injuries to be a leading cause of admission for men, while HIV and pregnancy-related disorders were highest among women<sup>27</sup>.

The current Kenya health sector strategy recommends that, by 2016, at least 90% of households in Kenya should be within 5km or 1 hour travel time to a public health facility<sup>28</sup>. This study has shown that 81% of study participants were within 5km or less to the health facility and 6% were more than 10km from the health facility. A high proportion of those with longer distances had temporarily relocated. For instance, some married young women had temporarily moved back to their parents' homes for care during the postnatal period, while others had temporarily moved for work-related reasons or were students in boarding schools.

It took between half an hour and an hour to receive services. Longer times were usually due to long queues or the need for laboratory tests to confirm the clinical diagnosis. For example, all malaria cases were laboratory-test confirmed.

Strengths of the study included the fact that linking and matching of patient records was done in the presence of the participants who were usually able to confirm that the matching was correct. We also attempted to visit the probable homestead for those who could not be matched at the facility. During the home visits, we found that some people existed in the database with different names, while others did not exist in the KHDSS population register. Matching could be done very quickly (<1 minute) when the participant had their ID card with them and the ID number was in the KHDSS database, but even when this was not the case, the time taken for matching was a maximum of 3 minutes.

However, the study had limitations. Although health facility and KHDSS data were successfully linked for the great majority (97%) of young adults (18–24 years), it was limited to this age group and a small catchment area of two peripheral health facilities over a short period of two and a half months.

## Conclusions

This is the first study to link outpatient health facility data and KHDSS population data for young adults in peripheral health facilities. The study demonstrated that it was possible to link individual health data on young adults attending the two peripheral health facilities to a longitudinal population register. The main reasons for visiting health facilities identified in this study will inform policy makers on key areas to target for interventions.

These findings show that clinic/KHDSS data linkage is feasible in this context. The combination of the population-based fertility and mortality data from the demographic surveillance system, and the linked health facility data could be the basis for monitoring and evaluation of health care outcomes, demand for health services, and the effectiveness of public health interventions among young adults in this population.

## Ethical approval

Ethical approval for this study was given by the KEMRI/National Ethical Review Committee (Protocol SSC No. 2823: The Health

and Health Behaviours of young people (13–24) in Kilifi (version 2.1 dated July 17, 2014)), and by the Research Ethics Committee of the London School of Hygiene & Tropical Medicine (No. 8660). Written consent was obtained from all participants.

## Data availability

*Figshare:*

**Experience from The Kilifi Health and Demographic Surveillance System: Raw data for linked participants and unmatched participants.**

**Dataset File 1: Linkage data.** Age\_group: age of participant on the day they visited the clinic, three missing values; facility: Matsangoni or Pingilikani; visit\_date: date when participant visited clinic, three missing values; service\_time\_minutes: duration (in minutes) it took for participant to get served, 29 missing values; visit\_reason: reason for visit, four missing values; exclusion: variable indicating the records that were included in the 'clinic visit analysis'; visit\_no: some participants made more than one visit during the study period - this variable indicates the number of visits made at any one instance; distance\_to\_facility: distance (in KM) from the homestead to health facility.

**Dataset File 2: Unmatched participant data.** Data clerk: Study staff who did the consenting, searched and linked study participant's demographic data with the clinic visit data; Hmname: Homestead name. Name a by which a homestead is known and referred to; HmHead: Homestead head. The person who heads a home, and makes important decisions for the family; Location: Administrative unit in the government structure headed by a chief; Sublocation: A sub-unit of a location headed by an assistant chief.

<http://dx.doi.org/10.6084/m9.figshare.5202850><sup>29</sup>

Note that some information has been removed/amended in the datasets, in order to maintain the anonymity of the participants.

---

## Author contributions

CN prepared and drafted the manuscript, and developed the maps. AD, DR, TW, EB contributed to the conceptualization of the study and in the review and revision of the manuscript. NK, BJ and MK participated in the database design and development. DW, CN, GM, DA and MO participated in data collection, data management, analysis and interpretation. All authors read and approved the final manuscript.

## Competing interests

No competing interests were disclosed.

## Grant information

This work was supported by the Wellcome Trust [084535; 091758]; and the Hewlett Foundation through a grant awarded to INDEPTH [IND/0107].

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*



## Acknowledgements

The authors thank members of the Epidemiology and Demography Department at the KEMRI/Wellcome Trust Research Programme, the INDEPTH Secretariat and the Kilifi County Department of Health, including staff at the Pingilikani and Matsangoni health facilities, and the study's data clerks for their support and

contributions. We thank the young people who participated in the study, and the members of the KEMRI/Wellcome Trust Community Liaison Group for their help with community engagement.

This paper is published with the permission of the Director of KEMRI.

## Supplementary material

### Supplementary File 1. IHTAS Sub-Study 2: Health Facility Data Linkage data collection form.

[Click here to access the data.](#)

### Supplementary File 2. Unmatched participants demographic data form.

[Click here to access the data.](#)

## References

- Scott JA, Bauni E, Moisi JC, *et al.*: **Profile: The Kilifi Health and Demographic Surveillance System (KHDS).** *Int J Epidemiol.* 2012; 41(3): 650–7. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hoon-Leahy CE, Newbury J, Kitson A, *et al.*: **The LINKIN Health Census process: design and implementation.** *BMC Health Serv Res.* 2012; 12: 321. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ye Y, Wamukoya M, Ezeh A, *et al.*: **Health and demographic surveillance systems: a step towards full civil registration and vital statistics system in sub-Saharan Africa?** *BMC Public Health.* 2012; 12: 741. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kabudula CW, Clark BD, Gómez-Olivé FX, *et al.*: **The promise of record linkage for assessing the uptake of health services in resource constrained settings: a pilot study from South Africa.** *BMC Med Res Methodol.* 2014; 14: 71. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kelman CW, Bass AJ, Holman CD: **Research use of linked health data—a best practice protocol.** *Aust N Z J Public Health.* 2002; 26(3): 251–5. [PubMed Abstract](#) | [Publisher Full Text](#)
- Machado CJ: **A literature review of record linkage procedures focusing on infant health outcomes.** *Cad Saude Publica.* 2004; 20(2): 362–71. [PubMed Abstract](#) | [Publisher Full Text](#)
- Newcombe HB, Smith ME, Howe GR, *et al.*: **Reliability of computerized versus manual death searches in a study of the health of Eldorado uranium workers.** *Comput Biol Med.* 1983; 13(3): 157–69. [PubMed Abstract](#) | [Publisher Full Text](#)
- Holman CD, Bass AJ, Rosman DL, *et al.*: **A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system.** *Aust Health Rev.* 2008; 32(4): 766–77. [PubMed Abstract](#) | [Publisher Full Text](#)
- MacKian S, Elliott H, Busby H, *et al.*: **'Everywhere and nowhere': locating and understanding the 'new' public health.** *Health Place.* 2003; 9(3): 219–29. [PubMed Abstract](#) | [Publisher Full Text](#)
- Jutte DP, Roos LL, Brownell MD: **Administrative record linkage as a tool for public health research.** *Annu Rev Public Health.* 2011; 32: 91–108. [PubMed Abstract](#) | [Publisher Full Text](#)
- Li B, Quan H, Fong A, *et al.*: **Assessing record linkage between health care and Vital Statistics databases using deterministic methods.** *BMC Health Serv Res.* 2006; 6: 48. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Adetifa IM, Bwanaali T, Wafula J, *et al.*: **Cohort Profile: The Kilifi Vaccine Monitoring Study.** *Int J Epidemiol.* (in press), 2016; pii: dyw202. [PubMed Abstract](#) | [Publisher Full Text](#)
- Sawyer SM, Afifi RA, Bearinger LH, *et al.*: **Adolescence: a foundation for future health.** *Lancet.* 2012; 379(9826): 1630–40. [PubMed Abstract](#) | [Publisher Full Text](#)
- Patton GC, Coffey C, Cappa C, *et al.*: **Health of the world's adolescents: a synthesis of internationally comparable data.** *Lancet.* 2012; 379(9826): 1665–75. [PubMed Abstract](#) | [Publisher Full Text](#)
- Brittain AW, Williams JR, Zapata LB, *et al.*: **Youth-Friendly Family Planning Services for Young People: A Systematic Review.** *Am J Prev Med.* 2015; 49(2 Suppl 1): S73–84. [PubMed Abstract](#) | [Publisher Full Text](#)
- Chandra-Mouli V, Mapella E, John T, *et al.*: **Standardizing and scaling up quality adolescent friendly health services in Tanzania.** *BMC Public Health.* 2013; 13: 579. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- DuVall SL, Fraser AM, Rowe K, *et al.*: **Evaluation of record linkage between a large healthcare provider and the Utah Population Database.** *J Am Med Inform Assoc.* 2012; 19(e1): e54–9. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sariyar M, Borg A, Pommerening K: **Evaluation of record linkage methods for iterative insertions.** *Methods Inf Med.* 2009; 48(5): 429–37. [PubMed Abstract](#) | [Publisher Full Text](#)
- Serwaa-Bonsu A, Herbst AJ, Reniers G, *et al.*: **First experiences in the implementation of biometric technology to link data from Health and Demographic Surveillance Systems with health facility data.** *Glob Health Action.* 2010; 3(1). [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bejon P, Williams TN, Lijander A, *et al.*: **Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya.** *PLoS Med.* 2010; 7(7): e1000304. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bejon P, Williams TN, Nyundo C, *et al.*: **A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots.** *eLife.* 2014; 3: e02130. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ndungu FM, Marsh K, Fegan G, *et al.*: **Identifying children with excess malaria episodes after adjusting for variation in exposure: identification from a longitudinal study using statistical count models.** *BMC Med.* 2015; 13: 183. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

23. Mokdad AH, Forouzanfar MH, Daoud F, *et al.*: **Global burden of diseases, injuries, and risk factors for young people's health during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013.** *Lancet.* 2016; **387**(10036): 2383–401.  
[PubMed Abstract](#) | [Publisher Full Text](#)
24. Global Burden of Disease Pediatrics Collaboration, Kyu HH, Pinho C, *et al.*: **Global and National Burden of Diseases and Injuries Among Children and Adolescents Between 1990 and 2013: Findings From the Global Burden of Disease 2013 Study.** *JAMA Pediatr.* 2016; **170**(3): 267–87.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Bauni E, Ndila C, Mochamah G, *et al.*: **Validating physician-certified verbal autopsy and probabilistic modeling (InterVA) approaches to verbal autopsy interpretation using hospital causes of adult deaths.** *Popul Health Metr.* 2011; **9**: 49.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Etyang AO, Munge K, Bunyasi EW, *et al.*: **Burden of disease in adults admitted to hospital in a rural region of coastal Kenya: an analysis of data from linked clinical and demographic surveillance systems.** *Lancet Glob Health.* 2014; **2**(4): e216–24.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Etyang AO, Scott JA: **Medical causes of admissions to hospital among adults in Africa: a systematic review.** *Glob Health Action.* 2013; **6**: 1–14.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Health MO: **TRANSFORMING HEALTH: Accelerating Attainment of Universal Health Coverage.** *THE KENYA HEALTH SECTOR STRATEGIC AND INVESTMENT PLAN.* (accessed 14 October, 2016), 2013.  
[Reference Source](#)
29. Nyundo C, Doyle AM, Walumbe D, *et al.*: **Dataset Files 1 and 2: Experience from The Kilifi Health and Demographic Surveillance System: Raw data for linked participants and unmatched participants.** *figshare.* 2017.  
[Data Source](#)

# Open Peer Review

Current Referee Status:



Version 1

Referee Report 13 October 2017

doi:[10.21956/wellcomeopenres.12195.r26015](https://doi.org/10.21956/wellcomeopenres.12195.r26015)



**Fiona J Stanley**<sup>1</sup>, **Hannah C. Moore**<sup>2</sup>

<sup>1</sup> Telethon Kids Institute, University of Western Australia, Perth, WA, Australia

<sup>2</sup> Wesfarmers Centre of Vaccines and Infectious Diseases, Telethon Kids Institute, University of Western Australia, Crawley, WA, Australia

- Authors should be commended for this work which demonstrates the feasibility of conducting linkage work in this context and a basis for future research.
- It would be great for the authors to explain in more detail the selection of the 2 health facilities that were included in the study. The linkage rate of 96.3% is very high which is commendable, however would the linkage rate be similar if other health facilities were chosen instead? This would add to the study.
- A small clarification to add to the results would be the distance and whether this is measured by travel by road, therefore reflect access to services, or by overall distance.
- These next 2 suggestions are really extensions of the work that the authors could consider
  1. Was there any scope to record more than 1 reason for visiting the health care facility? Many of the reasons for clinic visits could simultaneously occur (ie HIV and chest infection). An ability to record more than 1 reason for a visit could greatly strengthen the work as less severe conditions could be captured (ie if someone presents for family planning reasons but also has a current skin infection, which condition would be recorded?) I'm unsure as to feasibility of recording this extra information in the study.
  2. Additionally, it would be good to have a measure if possible using the KHDSS dataset of how many young adults were registered and did not have any health facility visits. Again, I'm unsure of whether this would be possible to look at but it would help put into context the use of health care facilities in these areas.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Referee Report 12 September 2017

doi:[10.21956/wellcomeopenres.12195.r25438](https://doi.org/10.21956/wellcomeopenres.12195.r25438)



**Philippe Bocquier**  <sup>1,2</sup>

<sup>1</sup> Centre de Recherche En Démographie, Université Catholique de Louvain, LouvainlaNeuve, Belgium

<sup>2</sup> School of Public Health, University of the Witwatersrand, Johannesburg, South Africa

This article is quite relevant to anyone who is working with HDSS data that are collected in areas served with health facilities (most HDSS are actually initiated in collaboration with these facilities). It is important for the reader to note (as mentioned by the authors) that the team build on previous experiences of HDSS-health facilities matching with children records. I would have liked the authors to reflect on a non-random selection of the two facilities chosen "on the basis of the availability of research space, youth-friendly services, a reliable power supply and a willingness to participate in the study" (p3). What would be the outcome in the other, less-welcoming health facilities? How can we evaluate this selection (probably negative) bias? This is important for generalisation of results. Another source of bias is at the individual level, when young people are identified at the reception desk. The "assessment of eligibility" may have biased the result positively. What were the eligibility criteria? What if for example a young adult declares that s/he is not living in the HDSS (or not the right age) though s/he is? I would rather not impose eligibility criteria or adopt large criteria (e.g. extending the catchment area or the age group) and then apply the algorithm on all in order to minimize the chance of false negatives. With the procedures in place, maybe only false positives can be identified, hence maybe the impressive matching success rate. The two maps in figure 2 are not easy to understand as they are no explanations attached to these maps. Also the definition of distance is not given: is it bird distance or road distance? Are people travelling by foot or by bike or car or public transport? In other words, is there a cost attached to distance? Wouldn't time to health facility be more appropriate? The analysis and discussion on reasons for visit is somewhat disconnected from the main aim of the paper. I would have preferred an analysis that shows how matching records with HDSS data can help to better understand the purpose of the visits. The paper could actually do without the analysis of reasons for visit. The paper remains interesting but precisions on the potential bias and the consequence on generalisation are needed.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

***Competing Interests:*** No competing interests were disclosed.

***Referee Expertise:*** Migration and health; longitudinal data management and analysis;

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

---