

Longitudinal record linkage in sub-Saharan Africa countries: recommendations for healthcare research

Innocent B. Mboya^{1,2,3}, Stanley Teleka⁴, Sophia Kagoye⁵, Neema R. Mosh^{5,6}, Wende C. Safari⁷

¹ Department of Translational Medicine, Lund University, Malmö, Sweden. ² Institute of Public Health, Department of Epidemiology and Biostatistics, Kilimanjaro Christian Medical University College, Moshi, Tanzania. ³ Africa Academy for Public Health, Dar es Salaam, Tanzania. ⁴ Department of Clinical Sciences, Malmö, Lund University, Malmö, Sweden.

⁵ National Institute for Medical Research, Mwanza, Tanzania. ⁶ Ludwig-Maximilians University, Munich, Germany. ⁷ Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, United Kingdom.

Extended abstract

Background: The use of unique national personal identification numbers (PIN) for linkage of medical records across health facilities and population-based studies is limited in sub-Saharan African (SSA) countries. The disintegration of healthcare research with different participant identifiers creates methodological challenges in linking data from multiple sources to answer a diverse range of policy-relevant, clinical, administrative, and research questions. This commentary aims to provide recommendations for improved linkage of health services data in SSA for healthcare research.

Main text: Linkage of health services data is a complex process that requires health system thinking, multisectoral collaboration between the public and private sectors, and long-term investments. We recommend transitioning from paper-based healthcare data collection and storage systems, at health facility and population levels, to the use of electronic computer-assisted systems and a focus on improved data quality. SSA countries implementing this transition will benefit from additional technical and financial support. At the health-facility level, change from collecting aggregated to individual-level data. The use of PIN across multiple data sources is paramount, in addition to clear and transparent legal and ethical data protection and sharing guidelines, centralized data linkage by the governments, and partnerships with the private sector.

Conclusion: High-quality linked data in SSA are scarce. SSA countries should prioritize establishing a robust foundation for high-quality data collection and future linkage.

Keywords: Healthcare data, longitudinal record linkage, secondary data analysis, SSA.

Recommendations

1. Despite the limited resources, the transition from paper-based data collection and storage systems to the use of electronic health data records, across all health facilities, and where possible, at all levels of care. Most SSA countries rely on specially designed registers/forms to mostly collect aggregated and seldom patient-level data, which has been proven to hamper the quality. Even where patient-level data is collected, challenges exist in linkage to other data sources. Electronic data collection systems are a prerequisite to improving data quality [1]. SSA countries implementing this transition will benefit from additional technical and financial support. At the health-facility level, strengthen the collection of individual-level apart from aggregated data often used to monitor key indicators. Efforts should also be made to improve the quality of electronic medical records for future data linkage.
2. Capture the PIN across all healthcare data sources in the country to enable future linkage of multiple data sources. Medical records and health insurance numbers are useful but often are unique across healthcare facilities, hence impossible to link individual medical records, while population-based surveys and clinical trials use their identification systems. With proper legal and ethical guidelines, the national PIN is a potential form of identification, which should be scaled up for all individuals, and used in all healthcare data sources countrywide, as a common identifier [2, 3]. All other identifiers, such as health insurance, social security, and medical record number should complement PIN and not vice versa. Although conventional personal identifiers can be used [4], without a common identifier, the use of deterministic and probabilistic matching methods will pose a methodological difficulty or prove impossible [3-5]. Hence, independent researchers and organizations should be encouraged to capture the PIN, if available, across all studies and follow transparent data-sharing protocol/ guidelines. Other important non-health-related data sources such as census, migration, and registrations of birth and death should also be electronic and must capture PIN.
3. The protection of privacy and client/participant data should be taken into the highest possible consideration. The legal and ethical regulations in SSA should be enhanced following local and international standards to provide clear guidelines on collecting the PIN and other identifiable information, procedures for data anonymization, data access, data sharing, and protection of the intellectual property rights regarding data ownership [2, 5]. In addition, the local ethical review boards and data-sharing committees must be strengthened to deal with these sensitive ethical issues.
4. Centralized pooling, merging, and sharing of pseudo-anonymized data from multiple sources is highly recommended and should be by a dedicated government entity such as the Ministry of Health, National Bureaus of Statistics, or Commissions of Science and Technology [2, 3]. Researchers should not be allowed to access certain identifying patient information [3]. Procedures must be in place on when to update data linkage. To achieve this, the SSA governments should work with the private sector and international partners to secure the necessary resources, such as dedicated servers, software, and technical capabilities necessary for data linkage, anonymization, sharing, and storage.
5. The legal and ethical regulations must state when these linked data must be destroyed or are not available for research, e.g., once the research is completed by the individual or group of investigators, and should follow the national and international regulations [2, 3].
6. Recognizing a bulk of resources invested in collecting, linkage, and storage of this data, the research institutions and the government should provide a transparent and reasonable amount of money to be paid for the linked and pseudo-anonymized data access and to what entity, apart from the ethical clearance fees. Such information is essential to aid researchers in carefully planning their studies and include data access fees in their budgets.
7. Public-private partnerships in the healthcare sector are encouraged in creating and using data systems and removing fragmentation. As noted elsewhere [3], *“Without these partnerships, timely and comprehensive health information would not be available from*

private organizations and their patient populations to answer pressing health services and policy research questions”.

8. In the Western countries where the national PIN was implemented early, its importance was immediately recognized but has been appreciated even more over time. In contrast, the SSA countries have functioned without the use of PIN. For stakeholders to appreciate the importance of healthcare record linkage and using PIN, training frontline providers in collecting patient-level data is paramount [2]. The educational curriculum in medicine, nursing, and pharmacy, amongst others, should put a strong emphasis on the collection and utility of quality healthcare data beyond patient care. This would encourage these crucial stakeholders to appreciate its importance and motivate them to do their part.

References

1. Abdul-Rahman T, Ghosh S, Lukman L, Bamigbade GB, Oladipo OV, Amarachi OR, Olanrewaju OF, Toluwalashe S, Awuah WA, Aborode AT *et al*: **Inaccessibility and low maintenance of medical data archive in low-middle income countries: Mystery behind public health statistics and measures.** *J Infect Public Health* 2023, **16**(10):1556-1561.
2. Boulle A, Heekes A, Tiffin N, Smith M, Mutemaringa T, Zinyakatira N, Phelanyane F, Pienaar C, Buddiga K, Coetzee E *et al*: **Data Centre Profile: The Provincial Health Data Centre of the Western Cape Province, South Africa.** *Int J Popul Data Sci* 2019, **4**(2):1143.
3. Bradley CJ, Penberthy L, Devers KJ, Holden DJ: **Health services research and data linkages: issues, methods, and directions for the future.** *Health Serv Res* 2010, **45**(5 Pt 2):1468-1488.
4. Kabudula CW, Clark BD, Gomez-Olive FX, Tollman S, Menken J, Reniers G: **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.
5. Harron K, Dibben C, Boyd J, Hjern A, Azimaee M, Barreto ML, Goldstein H: **Challenges in administrative data linkage for research.** *Big Data Soc* 2017, **4**(2):2053951717745678.