Who uses health services? Record linkage between health facility and demographic surveillance data in rural South Africa.

Chodziwadziwa Kabudula, F. Xavier Gómez-Olivé Georges Reniers, Benjamin Clark Samuel Clark, Stephen Tollman

INDEPTH Network
Background

- Health and Demographic Surveillance Systems (HDSS) have progressed enormously in documenting the burden of diseases in communities with weak information systems.

- Several types of socioeconomic data are used as explanatory variables for observed differentials.

- HDSS, however, have provided limited information on utilisation of health services where they have existed in isolation from health facility data.
Question

- How feasible it is to produce representative datasets on patterns of health care utilisation in communities with weak registration systems in Primary Health Care by linking health facility and HDSS data?
Overview of record linkage

- Record linkage between two data sources A and B attempts to identify records that belong to the same individual

- Two types of approaches:
  - Deterministic → Record pairs linked based on agreement on specific identifiers or unique identifier.
  - Probabilistic → Record pairs linked based on probabilities of agreement on a set of identifiers.

INDEPTH Network
Use of record linkage for population–based research

- Common in developed countries (both approaches)
- Rare in developing countries.
  - Lack of unique identification system e.g. social security number
  - Inaccurate reporting of identifying information due to high illiteracy

INDEPTH Network
Objectives

- To assess the representativeness of datasets generated by record linkage between health facility data and health demographic surveillance system in rural northeast South Africa.

INDEPTH Network
Study area: Agincourt HDSS
Methods

- Construction of gold standard dataset

Fingerprints collection from males and females aged 18+ years from 2 villages in AHDSS

Fingerprints collection from all patients at Agincourt Health centre

Individuals linked by matching fingerprints
Methods (Cont’d…)

- Deterministic record linkage

All individuals in AHDSS database

Individuals in gold standard dataset from Agincourt Health Centre

Individuals linked by matching identifiers

Mobile phone number & JaroWinkler(First name) >= 0.8 & Gender

OR

Mobile phone number & DOB & Gender

JaroWinkler(First Name) >= 0.8 & JaroWinkler(Surname) >= 0.8 & Gender & Age difference <= 1 & Village & Clinic visit date before end of residence episode in AHDSS
Methods (Cont’d…)

- Comparison of fingerprint matched versus identifier matched:
  - Gender distribution
  - Age distribution
  - Ethnicity
  - Education status
  - Union status

INDEPTH Network
Results: Fingerprint linkage

Fingerprints collected in AHDSS villages
- 2895 individuals
- 1930 female
- 965 male

Fingerprints collected at Agincourt Health centre
- 4190 individuals
- 2734 female
- 1456 male

Linked by matching fingerprints
- 623 individuals
- 511 female
- 112 male

INDEPTH Network
## Results: Deterministic record linkage

<table>
<thead>
<tr>
<th>Identifiers</th>
<th>Finger prints</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Match</td>
<td>Non-match</td>
<td>Total</td>
</tr>
<tr>
<td>Match</td>
<td>399</td>
<td>19</td>
<td>418</td>
</tr>
<tr>
<td>Non-match</td>
<td>224</td>
<td>192653</td>
<td>192877</td>
</tr>
<tr>
<td>Total</td>
<td>623</td>
<td>192672</td>
<td>193295</td>
</tr>
</tbody>
</table>

Sensitivity = 64.04%

Specificity = 99.99%

Positive Predictive Value = 95.45%

INDEPTH Network
Results: Fingerprint vs Identifier linked by gender

P value = 0.505

INDEPTH Network
Results: Fingerprint vs Identifier linked

<table>
<thead>
<tr>
<th></th>
<th>Females (Chi sq p-values)</th>
<th>Males (Chi sq p-values)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age distribution</td>
<td>0.098</td>
<td>0.692</td>
</tr>
<tr>
<td>Ethnic group</td>
<td>0.59</td>
<td>0.448</td>
</tr>
<tr>
<td>Marital status</td>
<td>0.23</td>
<td>0.982</td>
</tr>
<tr>
<td>Education status</td>
<td>0.057</td>
<td>0.846</td>
</tr>
</tbody>
</table>
Conclusion

- Results suggest that record linkage between health facilities and AHDSS data can yield representative information on the characteristics of users of health services despite low linkage sensitivity.
- Further validation need on a larger dataset with known match status

INDEPTH Network
Thank You!