Methodology for monitoring data quality in cluster studies: application to the *TIMCI* project

Hélène Langet[1], Jean Augustin D. Tine, Samwel Lwambura, Silvia Cicconi, Charles Festo, Fenella Beynon, Anmol Jacob, Francis Njiri, Fabian Schär, Abdallah Mkopi, Kevin Ngari, Gillian Levine, Tracy R. Glass, for the TIMCI consortium

[1] helene.langet@swisstph.ch

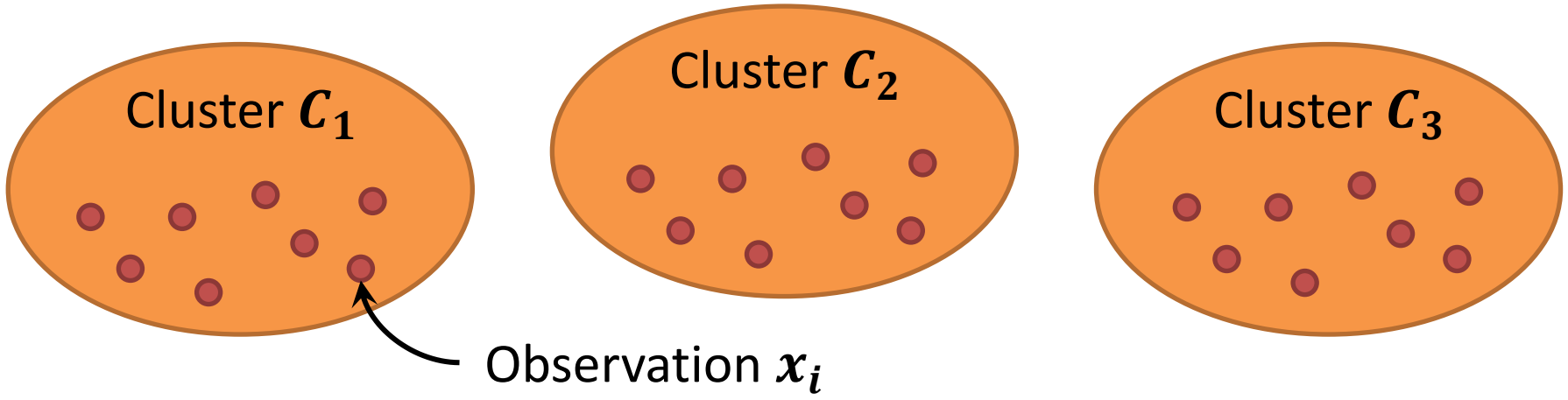
WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Presentation outline

- Introduction
- Methodology
- TIMCI project
- Results (application to TIMCI)
- Conclusion

Cluster studies

- individual observations **nested in a hierarchical structure** (*clusters*) [1]



- very common in **implementation research**, including **pragmatic studies**[2], to evaluate interventions in **real world** settings

[1]Galbraith S, Daniel JA, Vissel B. "A study of clustered data and approaches to its analysis." *J Neurosci. 2010 Aug 11;30(32):10601-8*. https://doi.org/10.1523/JNEUROSCI.0362-10.2010
[2]Weinfurt K. "What Is a Pragmatic Clinical Trial?: Why Are We Talking about Pragmatic Trials?" *Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials 2024. Bethesda, MD: NIH Pragmatic Trials Collaboratory*. https://doi.org/10.28929/089.

Data quality in cluster studies

- essential to ensure valid and reliable research findings
- **variability across clusters**
  e.g., natural epidemiological variations, human factors (individual practices)
- specific challenges associated with **real-world data sources**, including **adaptation** of data collectors to practices within their cluster based on their understanding

Need for a **practical methodology** for **monitoring data quality** in **cluster studies**

Cluster- time partitioning

Full dataset $\mathcal{D} = \begin{bmatrix} C_1 & t_1 & x_1 \\ \cdots & \cdots & \cdots \\ C_K & t_M & x_N \end{bmatrix}$

N observations $x_{i=1,\ldots,N}$
K clusters $C_{k=1,\ldots,K}$
M time periods $t_{m=1,\ldots,M}$

Partition $\mathcal{D}_{C_k,t_m} = \{x_i | \ \mathbf{cluster}(x_i) \in C_k \cap \mathbf{time}(x_i) \in t_m\}$

	t_1	t_2	$\cdots$	t_M
C_1	$\mathcal{D}_{C_1,t_1}$	$\mathcal{D}_{C_1,t_2}$	$\cdots$	$\mathcal{D}_{C_1,t_M}$
C_2	$\mathcal{D}_{C_2,t_1}$	$\mathcal{D}_{C_2,t_2}$	$\cdots$	$\mathcal{D}_{C_2,t_M}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
C_K	$\mathcal{D}_{C_K,t_1}$	$\mathcal{D}_{C_K,t_2}$	$\cdots$	$\mathcal{D}_{C_K,t_M}$

Statistical dispersion

Quality as a measure of **dispersion** of **clustered summary statistics**

Absolute modified z-scores

$$z_{C_k,t_m} = 0.6745 \times \frac{\left|\mathrm{X}_{C_k,t_m} - \tilde{\mathrm{X}}_{t_m}\right|}{\mathrm{MAD}_{t_m}}$$

X_{C_k,t_m} summary statistics for observations x_i in partition $\mathcal{D}_{C_k,t_m}$

$\tilde{\mathrm{X}}_{t_m}$ median of summary statistics X_{C_k,t_m} across all clusters for time period t_m

MAD_{t_m} median of absolute deviation $\left|X_{C_k,t_m} - \tilde{\mathrm{X}}_{t_m}\right|$ across all clusters for time period t_m

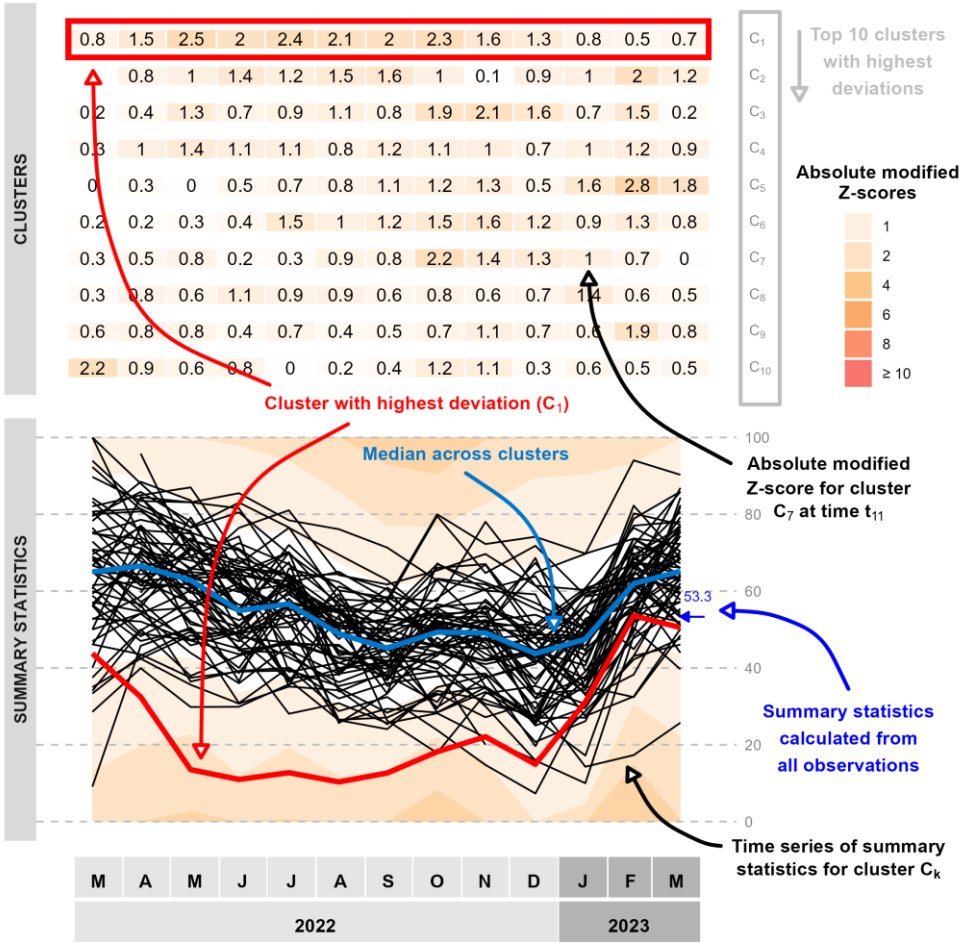
Iglewicz B, Hoaglin D. "How to Detect and Handle Outliers". *The ASQC Basic References in Quality Control: Statistical Techniques. 1993. In: Mykytka, E.F., Eds., How to Detect and Handle Outliers, ASQC Quality Press, Milwaukee, Vol. 16*.

Dispersion visualisation

- Heatmap of 10 **clusters with highest modified z-scores**

  outliers ⇒ $z_{C_k,t_m} \geq 3.5$

- Time series of **patterns of summary statistics** by individual clusters

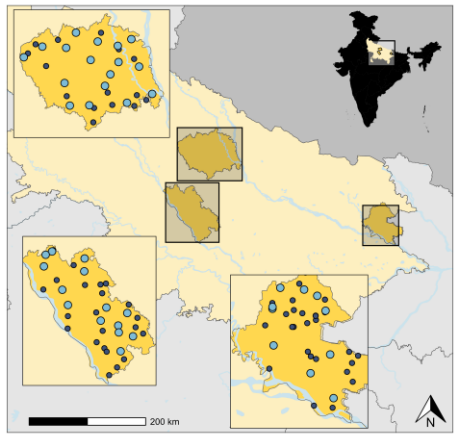


Iglewicz B, Hoaglin D. "How to Detect and Handle Outliers". *The ASQC Basic References in Quality Control: Statistical Techniques. 1993. In: Mykytka, E.F., Eds., How to Detect and Handle Outliers, ASQC Quality Press, Milwaukee, Vol. 16*.

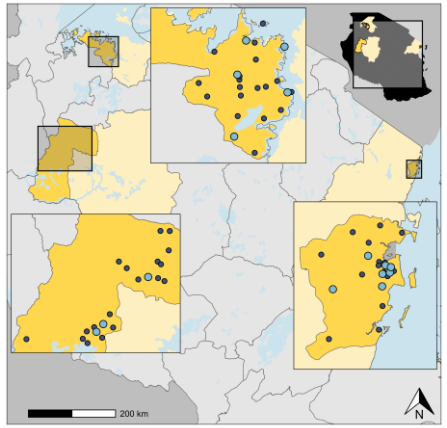
Project overview

TIMCI
Tools for Integrated Management of Childhood Illness

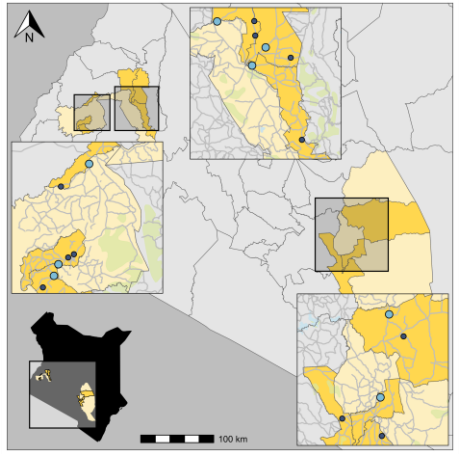
INDIA



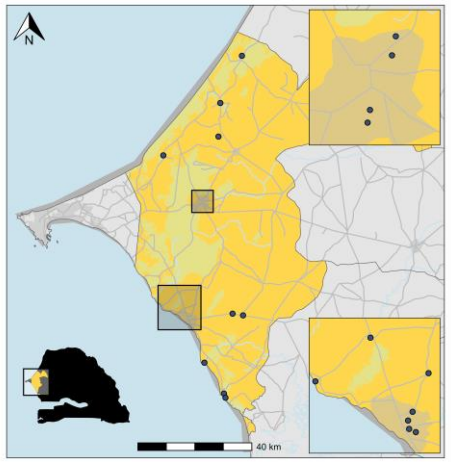
TANZANIA



KENYA



SENEGAL



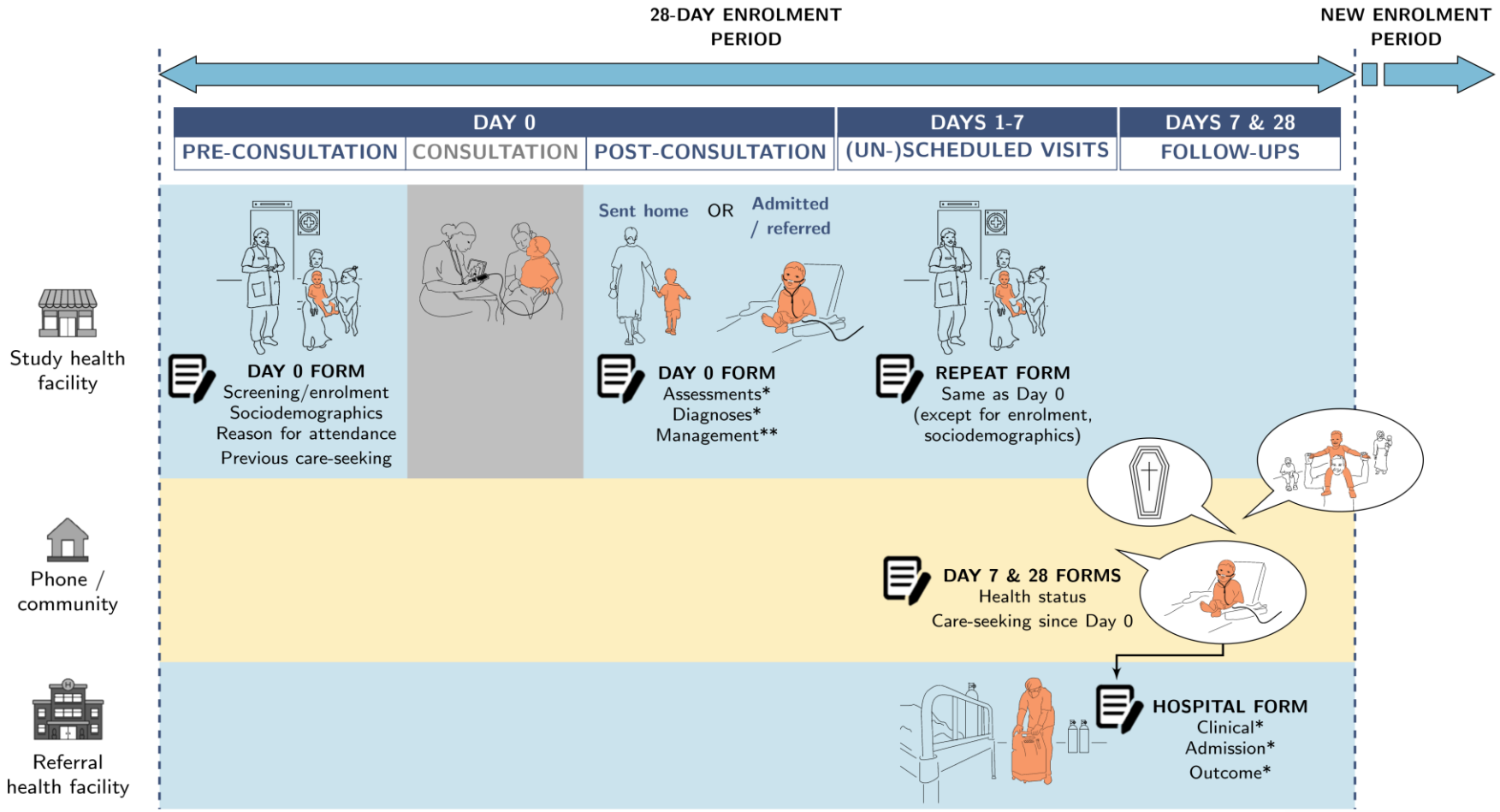
210 clusters (facilities)
209'269 children 0-59 months

- A **pragmatic cluster RCT**
  - NCT04910750
  - 106 facilities in India (9 months)
  - 66 facilities in Tanzania (12 months)
- A **quasi-experimental pre-post** study
  - NCT05065320
  - 19 facilities in Kenya (15 months)
  - 20 facilities in Senegal (18 months)

Beynon F, Langet H, Bohle LF et al. "The Tools for Integrated Management of Childhood Illness (TIMCI) study protocol: a multi-country mixed-method evaluation of pulse oximetry and clinical decision support algorithms." *Glob Health Action. 2024 Dec 31;17(1):2326253*. https://doi.org/10.1080/16549716.2024.2326253

Data collection flow

TIMCI
Tools for Integrated Management of Childhood Illness

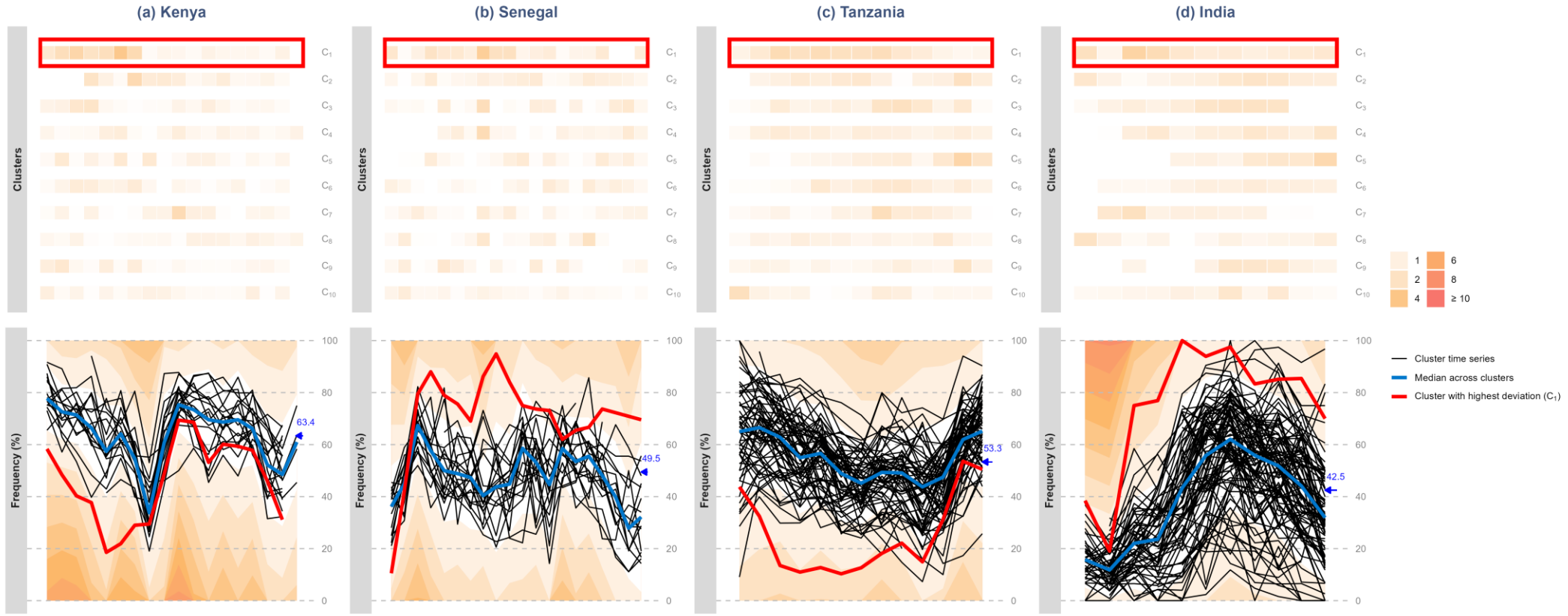


WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Results (application to TIMCI)

Review of a common event with natural epidemiological variations

Pre-consultation
Variable: **cough**
Source: **caregiver**
Statistics: **frequency**

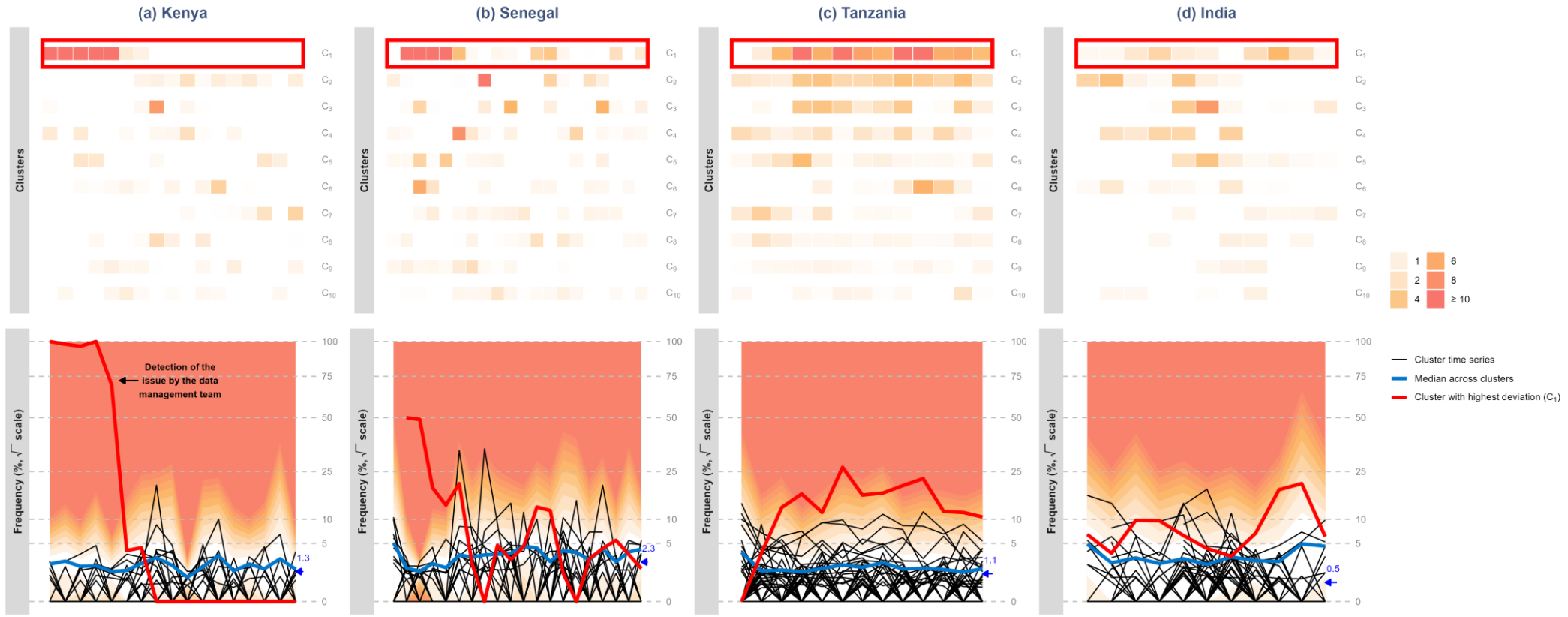


WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Results (application to TIMCI)

Review of a rare event

Post-consultation
Variable: **referral**
Source: **caregiver**
Statistics: **frequency**

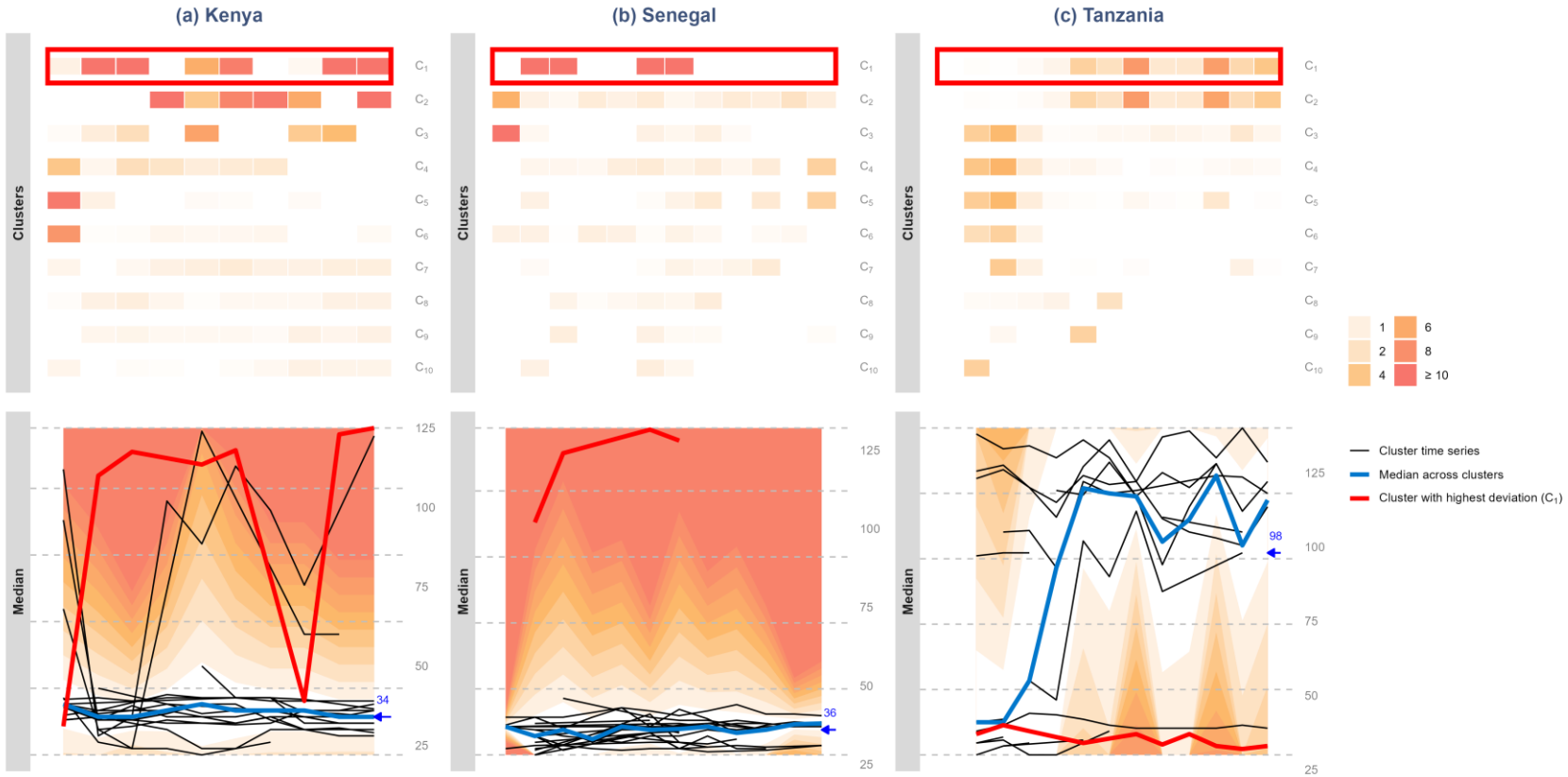


WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Results (application to TIMCI)

Review of a measurement

Post-consultation
Variable: **respiratory rate**
Source: **register**
Statistics: **median**



WCE WORLD CONGRESS OF EPIDEMIOLOGY 2024

Conclusion

- Targeted **identification** and **investigation** of potential data quality issues
- **Optimization** of the quality by adopting the approach early on and proactively implementing corrective measures throughout the **data acquisition process**
- Generation of **new knowledge**: findings can further inform the analysis and interpretation
- Can be adapted to integrate **other statistics**

Funded by Unitaid
2019-35-TIMCI

Acknowledgements

The TIMCI consortium thanks all the research assistants, participants, caregivers, healthcare providers, and Ministries of Health. This work was only possible through the collaboration between all institutions and individuals listed below (alphabetical order):

Ifakara Health Institute:
PI: Prof. Honorati Masanja
Research Lead: Dr. Grace Mhalu
Charles Festo, Samwel Lwambura, John Maiba, Susan Makawia, Irene Masanja†, Naomi Masanja, Abdallah Mkopi, Robert Moshiro, Ibrahim Mtebene

King George Medical University:
PI: Prof. Shally Awasthi
Research Lead: Dr. Divas Kumar
Girdhar Agarwal, Vineela Bandi, Anmol Jacob, Tuhina Rastogi, Mansi

PATH:
Project Lead: Mike Ruffo
Maymouna Ba, Method Bulo, Zach Clemence, Ray Cummings, Mira Emanuel-Fabula, Kanishka Gupta, Claudia Harner-Jay, Bonnie Keith, Tanya Lalwani, Andolo Miheso, Deusdedit Mjungu, Olgah Odek, Dickens Oviedo, Elena Pantjushenko, Manjari Quintanar Solares, Julie Rajaratnam, Kovid Sharma, Janet Shauri, Megan Shawcross, Lisa Smith, Helen Storey

Swiss TPH:
PI: Prof. Kaspar Wyss
Fenella Beynon, Leah F. Bohle, Silvia Cicconi, Lisa Cleveley, Tracy Glass, Kristina Keitel, Gaurav Kumar, Hélène Langet, Gillian Levine, Lena Matata, Martin Norris, Vânia Oliveira, Anja Orschulko, Elisabeth Reus, Fabian Schär

Unisanté:
PI: Prof Valérie D'Acremont
Vincent Faivre, Gregory Martin, Sylvain Schaufelberger, Rainer Tan, Julian Thabard, Alan Vonlanthen

Université Cheikh Anta Diop:
PI: Prof. Ousmane Ndiaye
Research Lead: Prof. Papa Moctar Faye
Ndèye Marème Sougou, Aliou Thiongane, Jean-Augustin Diégane Tine

University of Nairobi:
PI: Prof. James Machoki
Research Lead: Francis Njiri
Peter Arimi, Rose Kosgei, Mercy Mugo, Kevin Ngari, Mariah Ngutu

University of Waterloo / Cost & Cost-effectiveness:
PI: Prof. Susan Horton
Michael Onah