

Going global, redeeming the local: an innovative approach to implement the OMOP CDM in two countries of the Global South

Valentina Martufi¹, Emma Kalk², Enny S. Cruz¹, Juliana Araújo Prata de Faria¹, Adalton do Anjos Fonseca¹, Maurício L. Barreto¹, Maria Yury Travassos Ichihara¹, Jessica Gammon², Nicki Tiffin³, Chris Fourie², Danilo Luis Cerqueira Dias¹, Denise Moraes Pimenta¹, Tsaone Tamuhla³, Andrew Boulle², Themba Mutemaringa², Juan-Paul Hynek², Muzzammil Ismail², Julio Barbour Oliveira⁴, Ricardo Felix Monteiro Neto⁴, Júlia Pescarini¹, Fernanda Revoredo de Sousa¹, Marianne Costa e Silva Lage¹, Adam Loff², Melvin Moodley², Elzo Pereira Pinto Junior¹

¹CIDACS-IGM/FIOCRUZ-BA, ²WC-PHDC, ³SANBI-UWC, ⁴Precision Data

Background

Accelerated globalization has transformed health phenomena and health research into an international issue, with impact beyond local boundaries. Since its inception in 2008, the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) has facilitated the elaboration of federated analyses across different countries, expanding the knowledge base for the improvement of health for all populations globally. However, thus far, uptake of the approach has predominantly been in the Global North, and Global South countries remain significantly underrepresented in the Observational Health Data Sciences and Informatics (OHDSI) open-source development community. The OMOP CDM, therefore, has been strongly modelled to the socio-economic and health systems context of Global North countries, limiting its application more broadly, especially with respect to context-specific variables. We describe the experience of a collaborative endeavor, between the Center for Data and Knowledge Integration for Health (CIDACS) in Brazil and the Western Cape Provincial Health Data Centre (PHDC) in South Africa to apply a Common Data Model to investigate in the two different data environment the risk factors and birth outcomes of two infectious diseases: gestational syphilis and tuberculosis during pregnancy.

Methods

In Brazil variables are being mapped from the following databases: the live-births information system (SINASC), the mortality information system (SIM), the single registry for social welfare programs (CadÚnico), the gestational syphilis and tuberculosis compulsory notification systems, and the hospitalizations information system (SIH). In South Africa, the PHDC is a health information exchange that consolidates multiple electronic health data sources (laboratory, administrative, pharmacy, disease-specific registers) to generate an individual-level population database. Understanding that there are differences between the variables held by the two collaborating institutions, as well as those contemplated by the OMOP CDM – and due to a preoccupation with losing relevant context-specific information in the analyses to be carried out – we propose an alternative to conventional approaches for federated analyses comprising three parallel analyses: 1. The main, analysis will contemplate the variables included in the OMOP CDM, with data from both institutions; 2. will apply only to CIDACS data, and add to the previous analysis significant variables that characterize Brazil-specific nuances, such as participation in social welfare programs (e.g., the Bolsa Família conditional-cash transfer policy) and race – a fundamental variable to be considered when studying any health-related question in Brazil; and 3. will apply only to PHDC data, and add to the first analysis relevant variables available to this institution (and not to CIDACS), such as contact with health service providers and laboratory results (see Figures 1 and 4 for variables being mapped for Gestational Syphilis and Tuberculosis in pregnancy, respectively). Non-academic stakeholders are being included in this discussion and will contribute to the recommendations

elaborated.

Results

The results from the main analysis will provide enriched understanding of these two globally relevant infectious diseases, contextualizing their behavior in two Global South countries that hold many similarities – including global economic standing, colonial history, and stark levels of racial, socioeconomic and health inequalities – and many differences – such as health systems organization, post-colonial dynamics, and epidemiological profile. Furthermore, this three-tiered approach will allow us to: a. assess and measure the relevance of the additional variables from the second and third analyses, and b. provide evidence to make the case to include such variables in the OMOP CDM, so as to make it more inclusive and truly globally relevant.

Variable	Common variables	Common tentative variables	PHDC only	CIDACS only
Type of diagnosis	X			
Treatment received		X		
Prenatal care (at least 4)	X		6-8 = adequate	
First PNC visit within 20 weeks*	X		*parameter from SA	
Age of mother (w Date of birth)	X			
Birthweight	X			
Stillbirth		X		
Neonatal death 1	X			
Congenital syphilis (in her child)				
Congenital anomalies		X		
HIV Status			X	consider using SINAN-HIV integration
Geography (SocioEconomic characterization)		X	they could have an index for socioeconomic context	
Race				X
Mother's education				X
Maternal marital status				X
Maternal obstetric history (previous live birth and fetal lost)		X		X
Type of birth		X		
Partners treatment for syphilis (congenital form only)		X		
Rural vs urban residence				X

Figure 1. Three-tiered approach to applying a Common Data Model in two similar but different countries of the Global South: Gestational syphilis variables being mapped to the OMOP CDM

Variable	Common variables	Common tentative variables	PHDC only	CIDACS only
Type of diagnosis	X			
Time of diagnosis (within pregnancy or after birth)		X		
Treatment received	X			
Type of establishment where diagnosis occurred		X		
Prenatal care (at least 4)	X			
First PNC visit within 20 weeks*	X		*parameter from SA	
Age of mother (w Date of birth)	X			
Birthweight	X			
Stillbirth		X		
Neonatal deaths	X			
Infection in child		X		
Congenital anomalies		X		
HIV Status			X	consider using SINAN-HIV integration
Geography (SocioEconomic characterization)	X		they could have an index for socioeconomic context	
Race				X
Type of TB (pulmonar)		X		
Mother's education				X
Maternal marital status				X
Maternal obstetric history (previous live birth and fetal lost)				X
Type of birth		X		
Rural vs urban residence				X
Crowd (CADU only population)				X

Figure 4: Three-tiered approach to applying a Common Data Model in two similar but different countries of the Global South: Tuberculosis during pregnancy variables being mapped to the OMOP CDM

Conclusion

We look forward to presenting the results of this analytical approach to applying the OMOP CDM and discuss their significance with the OHDSI community. We hope this will lead to an expansion to

incorporate social programs and the creation of archetypes describing what each of these program disciplines would be, within their own domain table. This should provide an enrichment of socioeconomic information, including employment and income, education, etc., as well as an incentive to apply the OMOP CDM to other countries in the Global South.

The results from this experience could also support the case with decision-makers in both countries – and elsewhere – regarding the importance of collecting comprehensive data to facilitate the study and understanding of health phenomena that affect the population’s health.

References

1. Reinecke I et al. The Usage of OHDSI OMOP– A Scoping Review. Volume 483: German Medical Data Sciences 4041: Digital Medicine: Recognize – Understand – Heal. <https://doi.org/10.3433/SHTI410546>
2. Junior EPP, Normando P, Flores-Ortiz R, Afzal MU, Jamil MA, Bertolin SF, Oliveira VA, Martufi V, Sousa F, Bashir A, Burn E, Ichihara MY, Barreto ML, Salles TD, Prieto-Alhambra D, Hafeez H, Khalid S. Integrating real-world data from Brazil and Pakistan into the OMOP common data model and standardized health analytics framework to characterize COVID-19 in the Global South. Journal of the American Medical Informatics Association, Volume 30, Issue 4, April 2023, Pages 643–655, <https://doi.org/10.1093/jamia/ocac180>
3. Bhattacharjee T, Kiyuwa-Muyingo S, Kanjala C, Maoyi ML, Amadi D, Ochola M, Kadengye D, Gregory A, Kiragga A, Taylor A, Greenfield J, Slaymaker E, Todd J and INSPIRE Network. INSPIRE datahub: a pan African integrated suite of services for harmonising longitudinal population health data using OHDSI tools. Front. Digit. Health (2023) 6:1349630. <https://doi.org/10.3389/fdgth.2023.1349630>
4. KiyuwaMuyingo S, Todd J, Bhattacharjee T, Taylor A and Greenfield J. Enabling data sharing and utilization for African population health data using OHDSI tools with an OMOP common data model. Front. Public Health Volume 11 – 2023 <https://doi.org/10.3389/fpubh.2023.1116684>