FAIR CASE STUDY
SURVEY HARMONIZATION

UHC Summer Institute- June 26-30, 2023

Ana Ortigoza, MD PhD

DREXEL UNIVERSITY
Urban Health Collaborative
Dornsife School of Public Health

SALURBAL
Salud Urbana en América Latina

WorldFAIR

SALURBAL DATA

Health	Built Environments	Social Equity
• **Deaths and causes of death** • **Life expectancy** • Health risk factors • Health-related behaviors • Violence	• Land use and urban form • Transit options • Traffic congestion • Air pollution • Walkability • Green space • Water and sanitation • Housing	• Poverty • Income inequality • Housing conditions • Education • Employment

MOTIVATION FOR HARMONIZING SURVEY HEALTH DATA

Level 1: "City"
This level refers to urban agglomerations. While some only include a single municipality, others include various jurisdictions (e.g. Greater Bogotá). The "cities" or Level 1 units can be defined in various ways: as a collection of municipalities or similar units (L1AD), based on country-specific designations of metropolitan areas (L1MA), or based on the built-up area (or urban extent) identified using satellite and quantitative methods (L1UX).

Level 2: "Subcity"
These units are defined by the smaller administrative areas (e.g. municipio, comuna, distrito) that compose Level 1 cities.

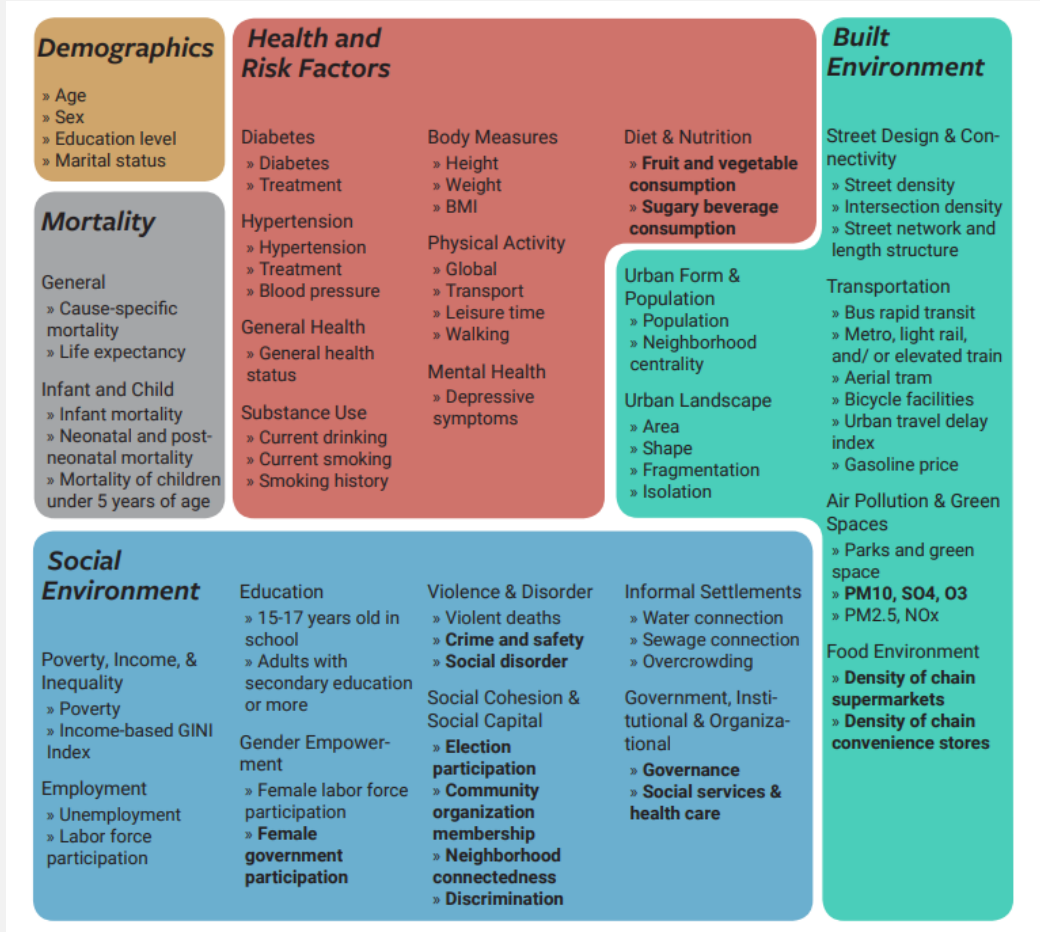
Level 3: "Neighborhood"
These are the smallest units for which data is available, defined by the censuses of each country (e.g. census tract, sector censal).

Example of a hypothetical city. The green line represents the boundaries of the city (L1AD) defined as a collection of two Level 2 units (municipalities separated by the blue line). In turn, the municipalities are divided into neighborhoods (L3) indicated by the red lines. The orange fill represents the city as defined by the urban extent (L1UX), which will not always coincide with the city defined by a collection of Level 2 units (L1AD).

SALURBAL cities and definitions of level 2 and 3 units by country

Country	Cities	Level 2 Unit	Level 3 Unit[b]
Argentina	33	Departamento/Partido/Comuna[a]	Radio Censal
Brazil	152	Municípios	Setor Censitário
Chile	21	Comuna	Zona Censal
Colombia	35	Municipio	Sector Urbano
Costa Rica	1	Canton	Unidad Geoestadistica Basica
El Salvador	3	Municipio	Sector Censal
Guatemala	3	Municipio	Sector Censal
Mexico	92	Area Geoestadistica Municipal	Area Geoestadistica Basica
Nicaragua	5	Municipio	Sector Censal
Panama	3	Corregimiento	Barrio
Peru	23	Distrito	Zona Censal

a Comunas in City of B.A., Partidos Province of B.A., Departamentos elsewhere. **b** As defined for country-designated urban areas.



FAIR → **Interoperability**

SURVEYS USED FOR HARMONIZATION BY COUNTRIES

- 27 surveys from 11 countries
- Spanning years 2000-2018
- ~250K adults (aged 18-100)
- ~70K children (aged 0-17)

Country	Survey name	Age	Survey year
Argentina	Encuesta Nacional de Factores de Riesgo, ENFR (National Risk Factors Survey)	≥18 years	2005, 2009, 2013
Brasil	Pesquisa Nacional de Saúde, PNS (National Health Survey)	≥18 years	2013, 2019
Chile	Encuesta Nacional de Salud, ENS (National Health Survey)	≥15 years	2003, 2010, 2017
	Encuesta Longitudinal de Primera Infancia (ELPI) (Longitudinal Survey of Early Childhood)	1-12 years	2017-2018
Colombia	Encuesta Nacional de Salud, ENS (National Health Survey)	18-69 years	2007
	Encuesta Nacional de la Situation Nutricional en Colombia, ENSIN (National Nutritional Situation in Colombia)	0-69 years	2005, 2010, 2015
Costa Rica	Encuesta Multinacional de Diabetes mellitus y Factores de Riesgo, CAMDI (Multinational Survey of Diabetes Mellitus & Risk Factors, Central American Diabetes Initiative)	≥20 years	2005
Guatemala	CAMDI (See Costa Rica)	≥20 years	2002-2003
	Demographic and Health Survey (DHS)	Females 18-49 years Children <5 years	2014-2015
Nicaragua	CAMDI (See Costa Rica)	≥20 years	2003
Mexico	Encuesta National de Salud, ENSA (National Health Survey)	Adults ≥18 Children <5 years	2000
	Encuesta Nacional de Salud y Nutricion, ENSANUT (National Survey for Health and Nutrition)	All ages	2006, 2012, 2016, 2018
Panama	Encuesta Nacional de Salud y Calidad de Vida ENSCAVI (National Survey of Health and Quality of Life)	≥18 years	2007
Peru	Encuesta Nacional de Demografia y Salud, ENDES (National Survey of Demographics and Health)	Adults ≥15 years Children <5 years	2016
El Salvador	CAMDI (see Costa Rica)	≥20 years	2004
	Encuesta Nacional de Salud Familiar (National Family Health Survey)	Females 18-49 years Children <5 years	2008
	Encuesta Nacional de Enfermedades Cronicas no transmisibles en Poblacion Adulta de El Salvador ENECA (National Survey of Non-communicable Chronic Diseases in the Adult Population of El Salvador)	≥20 years	2014-2015

STRATEGIES USED FOR HARMONIZATION

1. Use **existing** national **health survey data** administered by agencies within each country.
2. Use only surveys that could be **linked to** country geographical administrative IDs, corresponding to SALURBAL **sub-city level** (L2)*
3. Prioritize surveys with information on **non-communicable health** behaviors and risk factors.
4. Use harmonization **approaches that are rigorous but flexible** to accommodate differences across surveys.

STEPS FOR THE SURVEY HARMONIZATION

1. **Identifying and collating survey questions and responses by domain**
   - with attention to response patterns in the questionnaires asked on the survey and respondent universe.
2. **Reviewing surveys conducted by other institutions* for standard variable definitions and harmonization approaches.**
3. **Proposing harmonized variable definitions and response categories**
   - with attention to differences in wording across countries.
4. **Applying the harmonization approaches** that included
   - Creation of multiple versions due to country differences that did not allow a single harmonized variable (e.g., diabetes, hypertension).
   - Unit conversion (e.g., height, weight).
   - Collapsing categories (e.g., education, self-rated health).
5. **Revising the protocol as needed**, based on descriptive statistics of initial harmonized variables.

*CDC, NIH, WHO

SUMMARY OF HARMONIZED SURVEY DATA AVAILABLE BY DOMAIN

Domain	Number of core harmonized variables	% of NHS with all harmonized variables in the domain	% of NHS with at least one harmonized variable in the domain
		Adults	
Demographics	6	64%	100%
Socioeconomic status	15	67%	100%
Alcohol	6	33%	81%
Tobacco Use	8	26%	92%
Anthropometry	6	54%	96%
Diet	9	50%	75%
Physical Activity	14	34%	62%
Diabetes	3	73%	81%
Hypertension	7	64%	81%
Depressive symptoms	23	12%	50%
Self-reported health	3	22%	65%
Health care	4	41%	46%
Violence	3	13%	19%

NHS = National Health Surveys

SUMMARY OF HARMONIZED SURVEY DATA AVAILABLE BY DOMAIN

Physical activity & depressive symptoms were more feasible to harmonize →**standardized questionnaires**

NHS = National Health Surveys

Domain	Number of core harmonized variables	% of NHS with all harmonized variables in the domain	% of NHS with at least one harmonized variable in the domain
		Adults	
Demographics	6	64%	100%
Socioeconomic status	15	67%	100%
Alcohol	6	33%	81%
Tobacco Use	8	26%	92%
Anthropometry	6	54%	96%
Diet	9	50%	75%
Physical Activity	**14**	34%	62%
Diabetes	3	73%	81%
Hypertension	7	64%	81%
Depressive symptoms	**23**	12%	50%
Self-reported health	3	22%	65%
Health care	4	41%	46%
Pregnancy	2	46%	62%
Violence	3	13%	19%

SUMMARY OF HARMONIZED SURVEY DATA AVAILABLE BY DOMAIN

Self-reported health outcomes (diabetes diagnosis or access to health care) are the most **heterogenous**

Domain	Number of core harmonised variables*	% of surveys with all harmonised variables in the domain	% of NHS with at least one harmonised variable in the domain
		Adults	
Demographics	6	64%	100%
Socioeconomic status	15	67%	100%
Alcohol	6	33%	81%
Tobacco Use	8	26%	92%
Anthropometry	6	54%	96%
Diet	9	50%	75%
Physical Activity	14	34%	62%
Diabetes	3	73%	81%
Hypertension	7	64%	81%
Depressive symptoms	23	12%	50%
Self-reported health	3	22%	65%
Health care	4	41%	46%
Violence	3	13%	19%

NHS = National Health Surveys

SUMMARY OF HARMONIZED SURVEY DATA AVAILABLE BY DOMAIN

Questions related to **diagnosis** of non-communicable diseases, **self-reported health**, **access to health care**, and perception of **violence** are poorly developed in survey questionnaires

NHS = National Health Surveys

Domain	Number of core harmonised variables*	% of surveys with all harmonised variables in the domain	% of NHS with at least one harmonised variable in the domain
		Adults	
Demographics	6	64%	100%
Socioeconomic status	15	67%	100%
Alcohol	6	33%	81%
Tobacco Use	8	26%	92%
Anthropometry	6	54%	96%
Diet	9	50%	75%
Physical Activity	14	34%	62%
Diabetes	**3**	73%	81%
Hypertension	7	64%	81%
Depressive symptoms	23	12%	50%
Self-reported health	**3**	22%	65%
Health care	**4**	41%	46%
Violence	**3**	13%	19%

SUMMARY OF HARMONIZED SURVEY DATA AVAILABLE BY DOMAIN

Domain	Number of core harmonised variables	% of NHS with all harmonised variables in the domain	% of NHS with at least one harmonised variable in the domain
		Children	
Demographics	5	68%	100%
Socioeconomic status	16	85%	100%
Anthropometry	8	99%	100%

SUMMARY OF HARMONIZED SURVEY DATA AVAILABLE BY DOMAIN

Domain	Number of core harmonised variables	% of NHS with all harmonised variables in the domain	% of NHS with at least one harmonised variable in the domain
Children			
Demographics	5	68%	100%
Socioeconomic status	16	85%	100%
Anthropometry	8	99%	100%

- Number of domains available for comparison across countries was much lower than for adults.
- The only health-related domain that was possible to harmonize was that of anthropometric measures
  - for which the harmonization process mostly consisted of the conversion of different measurement units used for weight and height.

CHALLENGES FOR SURVEY HARMONIZATION

1. Disagreement in the definition of risk factors

Original Variable Name	Respondents universe	Original Question	English Translation	Original Responses	English Translation	Original Coding	Skip pattern go to
CIDI01	All	¿Alguna vez un doctor, una enfermera u otro profesional de la salud le dijo que tenía diabetes o azúcar alta en sangre?	Has a doctor, nurse, or other health professional ever tell you that you have diabetes or "high sugar" in your blood?	No	No	2	CIDI05
				Si	Yes	1	CIDI02 (if female) BIDI03 (if male)
				No sabe/no recuerda	Don't Know/Don't Remember	9	CIDI05
CIDI02	CIDI01 = yes and sex = female	¿Eso ocurrió cuando estaba embarazada?	Did this occur while you were pregnant?	Si	Yes	1	CIDI03
				No	No	2	
				No sabe/ no recuerda	Don't Know/Don't Remember	9	

National Survey on Health Risk Factors (ENFR) AR, 2005

Original Variable Name	Respondents universe	Original Question	English Translation	Original Responses	English Translation	Original Coding	Skip pattern go to	Notes
H7/403	401= 1	¿Le ha dicho algún profesional en salud que padece de diabetes o que tiene el azúcar alta en la sangre?	Has some health personnel ever told you that you have diabetes or high glucose levels?	si	Yes	1	404	
				no	No	0	407	
				NS/NR	don't know/ NA	99		
H8/404	403 = 1	¿Qué edad tenía usted cuando le dijeron por primera vez que era diabético(a)?	How old were you when you were diagnosed with diabetes?	edad	age	numeric	405	

Central America Diabetes Initiative (CAMDI) CR, 2005

CHALLENGES FOR SURVEY HARMONIZATION

2. Lack of consistency in categories or measurement units used for an indicator

Original Variable Name	Respondents Universe	Original Question	English Translation of Original Question	Original Responses	English Translation of Original Responses	Original Coding	Skip pattern go to
N001	V0025=1	Em geral, como o(a) Sr(a) avalia a sua saúde?	In general, how would you rate your health?	Muito boa	Very Good	1	Next question
				Boa	Good	2	
				Regular	Fair	3	
				Ruim	Poor	4	
				Muito ruim	Very Poor	5	
				Não aplicável	Not Applicable	mssing	

PNS, BR 2013

Original Variable Name	Respondents Universe	Original Question	English Translation of Original Question	Original Responses	English Translation of Original Responses	Original Coding	Skip pattern go to
QH40	All adults	Estado de salud	Health Status	Excelente	Excellent	1	Next question
				Muy Bueno	Very good	2	
				Bueno	Good	3	
				Regular	Regular	4	
				Malo	Poor	5	

ENSIN, CO 2010

CHALLENGES FOR SURVEY HARMONIZATION

3. Discrepancy in scales and questionnaires used for retrieving information about similar health behaviors or health outcomes that leads to a different assessment of the outcomes.

- *Physical activity* (International Physical Activity Questionnaire - **IPAQ vs** Global Physical Activity Questionnaire - **GPAQ**)
- *Fruit and vegetable consumption* (frequency questionnaires based on **servings vs** frequency questionnaires based on **portions**)

10 GOOD PRACTICES AND RECOMMENDATIONS IN HANDLING AND HARMONIZING HEALTH SURVEY DATA

1. Identify sources of **data representative of large populations** to assure great representation across countries/ cities and over time.

2. Use of harmonization **approaches that are rigorous but flexible** to accommodate differences across surveys.

3. **Identify and collate survey questions** and responses by selected domains under study (i.e., health risk factors)

4. Compare **wording questions** across survey with special attention to **answering patterns** to understand whether questions across surveys share same respondent universe (and therefore are comparable).

5. **Compare and contrast** quality of questions and **questionnaires** with surveys conducted by other Institutions

10 GOOD PRACTICES AND RECOMMENDATIONS IN HANDLING AND HARMONIZING HEALTH SURVEY DATA

6.Propose harmonized variable definitions and response categories with **attention to differences in wording across countries**.

7.Apply harmonization processes that can **balance specificity** of the information retrieved **versus the amount** of countries/ unit of analysis that can include the selected data.

i.Creation of ***multiple versions*** due to country differences that did not allow a single harmonized variable (e.g., diabetes, hypertension).

ii.***Unit conversion*** (e.g., height, weight).

iii.***Collapsing categories*** (e.g., education, self-rated health).

8. **Use of standardized questionnaires** and scales for the harmonization of questions that assess health behaviors through different set of questions (such as physical activity, or dietary habits)

10 GOOD PRACTICES AND RECOMMENDATIONS IN HANDLING AND HARMONIZING HEALTH SURVEY DATA

9.**Create harmonization protocols** that can document the decision taken during the harmonization process with special attention to:

i. The ***systematization and standardization*** of variable names across different years and versions of surveys

ii. The ***maintenance of consistency*** in the coding of answers and creation of answer categories across different years and versions

iii. The incorporation in the documents of the ***original questionnaires*** used for the harmonization.

10.Create **log documents** that can track the different iterations made during the harmonization process

ACKNOWLEDGEMENT – SALURBAL DMC

Kari Moore, MS

Goro Yamada, PhD

Jordan Hernandez, MPH

Jessica Uruchima, MPH

Steve Melly, PhD

Usama Bilal, PhD

Ana Diez- Roux

Ran Li, MS

THANK YOU


DREXEL UNIVERSITY
Urban Health
Collaborative
Dornsife School of Public Health


SALURBAL
Salud Urbana en América Latina


WorldFAIR