Modeling Populations in Latin America and the Caribbean

Instructions: Click on the link to access each author's presentation.

Organiser: Andrés Gutiérrez

Chair: Rolando Ocampo Alcántar

Discussant: Andrew Tatem

Participants:

Leesha Delatie-Budair: Administering censuses in Jamaica: challenges and solutions

Andrés Gutiérrez: ECLAC and UNFPA approach to model populations in Latin America and the Caribbean

Sabrina Juran:* UNFPA Efforts and Support to Censuses and Modeling of Populations in Latin America and the Caribbean

Christian Garces: Ecuadorian Experiences in the 2023 Household and Population Census

* Work presentation not available or non-existent

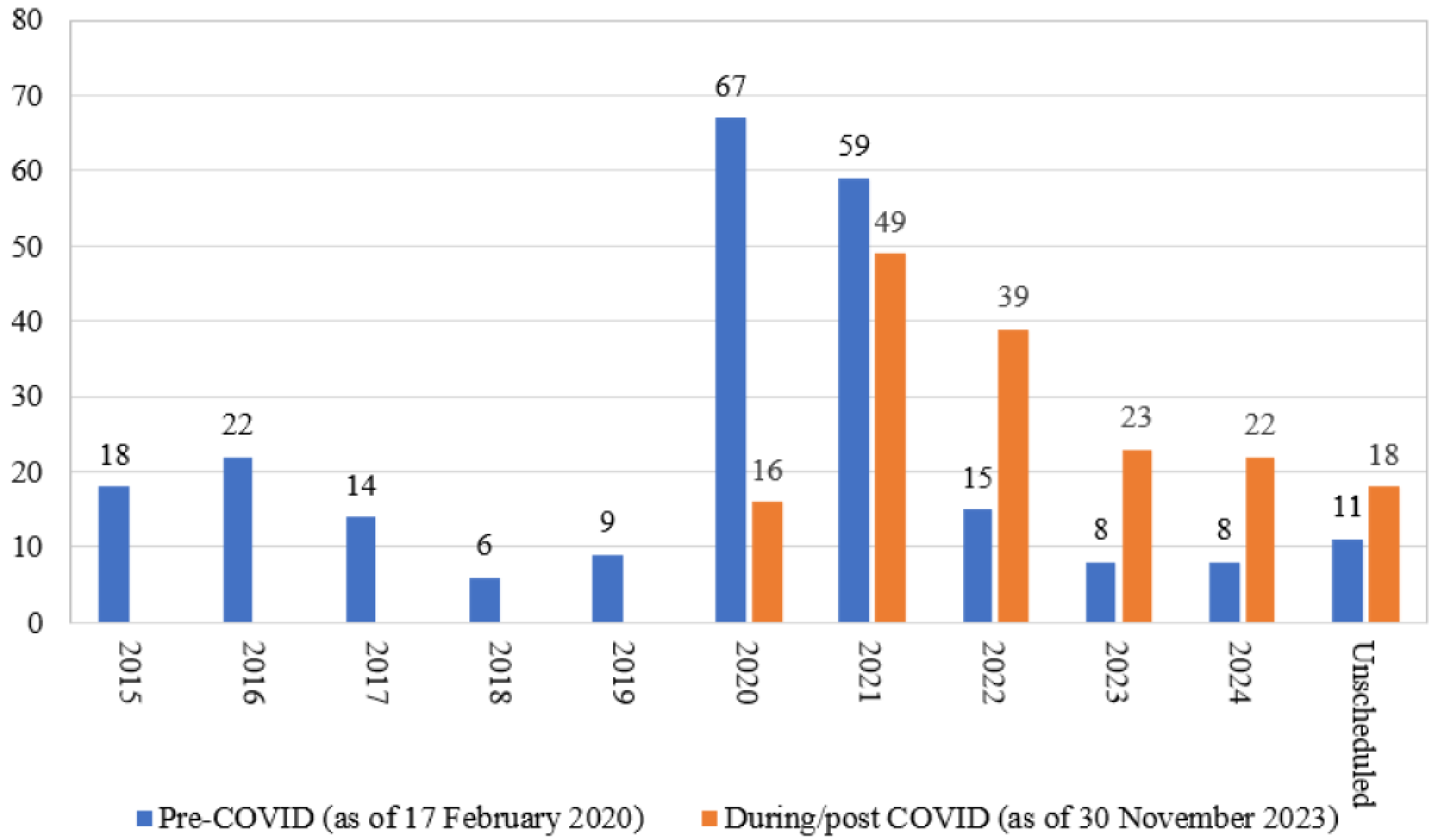
Administering censuses in Jamaica

Challenges and Solutions

Global Context

The 2022 Population and Housing Census is part of the **2020 World Programme on Population and Housing Censuses**

Number of countries or areas that have conducted, plan to conduct or have not scheduled a population and housing census in the 2020 round, by year



Source: Report of the Secretary-General on Population and housing censuses presented at the Fifty-fifth session of the UN Statistical Commission, 27 February–1 March 2024

“This census round has been particularly complex for the countries of our region, having to face not only technical challenges but also political, social, economic and communication challenges. Many of these challenges were magnified after the COVID-19 pandemic, which marked a before and after in the management of an operation of the magnitude of the population and housing censuses.”

Comment provided by: Dominican Republic on behalf of Latin American and Caribbean countries of ECLAC at the Fifty-fifth session of the UN Statistical Commission, 27 February–1 March 2024

> "Data quality is one of the major concerns of population and housing censuses conducted under the pressure of the COVID-19 pandemic. The high risk to the quality of census data emanates from adjustments to census processes and procedures motivated by the pandemic such as the extension of the duration of enumeration of the population and late changes to the design of field operations in order to reduce face-to-face interactions with respondents … Such impacts could reduce the comparability of census results from the current round with those from previous rounds."

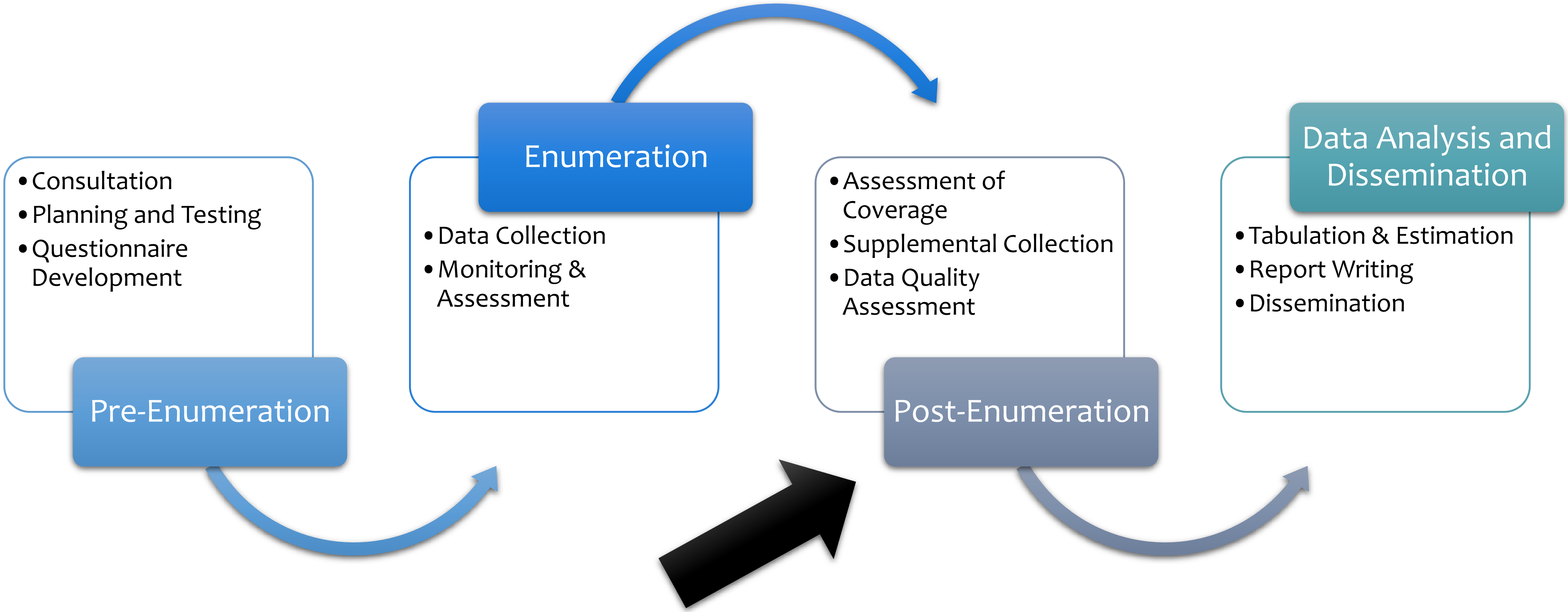
Para 2 Report on the results of UNSD surveys on the impact of the COVID-19 pandemic on 2020 round of population and housing censuses at the Fifty-fifth session of the UN Statistical Commission, 27 February–1 March 2024

Points of Note (UNSD)

- "… it is evident that the pandemic has exerted a significant adverse impact on the conduct of censuses … "
- "The circumstances of the pandemic that posed challenges to census-taking also created opportunities for innovation"
- "… the combined census methodology, which involves obtaining some of the census data from administrative sources and the remainder from field-based data collection. A combined census is often the first step towards a fully register-based census."
- "In parts of a country where enumeration is not possible, satellite imageries combined with existing data sources have enabled the estimation of population distributions for such areas."

Population censuses do not always manage to list all households and their populations throughout the country.

Census Process (simplified)



Pre-Enumeration

- Consultation
- Planning and Testing
- Questionnaire Development

Enumeration

- Data Collection
- Monitoring & Assessment

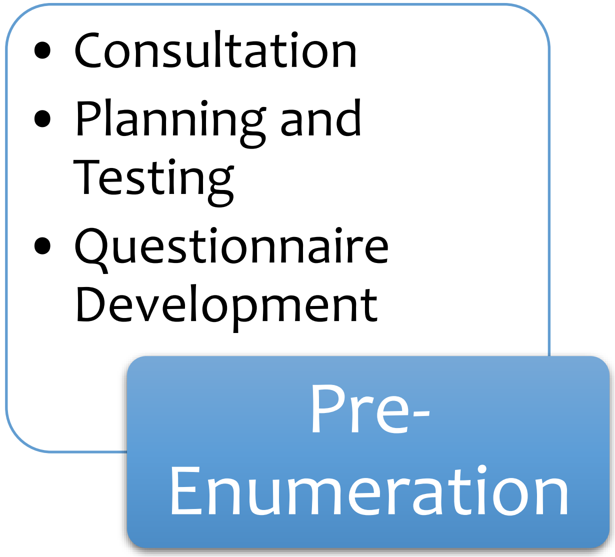
Post-Enumeration

- Assessment of Coverage
- Supplemental Collection
- Data Quality Assessment

Data Analysis and Dissemination

- Tabulation & Estimation
- Report Writing
- Dissemination

Census Process (simplified) – Key Challenges



- Consultation
- Planning and Testing
- Questionnaire Development

Pre-Enumeration

Consultation

- Wide-scale

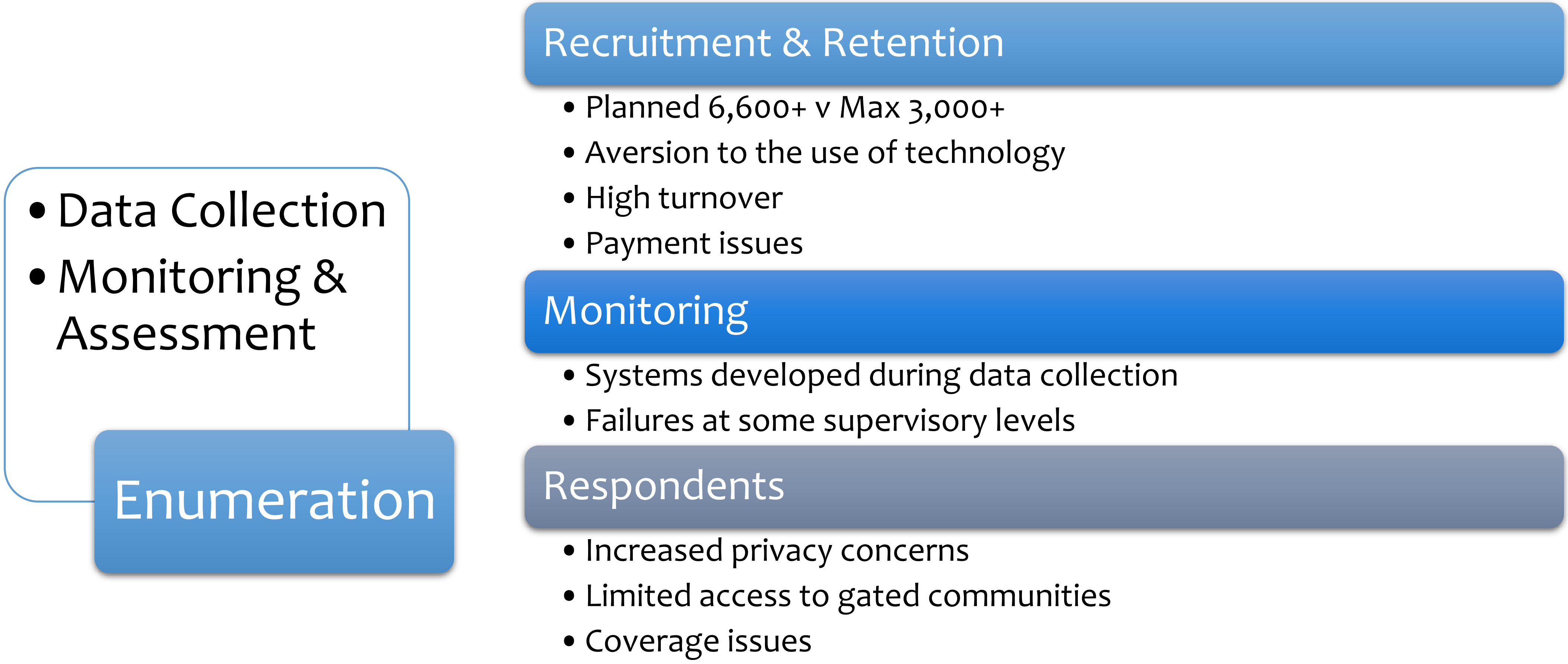
Planning and Testing

- Disrupted by global Pandemic
  - Pivoted to virtual training
  - Additional budget items (PPE)
  - Procurement delays
  - Turn-over of key personnel (HQ)
  - Not all systems were fully tested prior to the start of data collection

Questionnaire Development

- Completed as planned

Census Process (simplified) – Key Challenges



Enumeration

- Data Collection
- Monitoring & Assessment

Recruitment & Retention

- Planned 6,600+ v Max 3,000+
- Aversion to the use of technology
- High turnover
- Payment issues

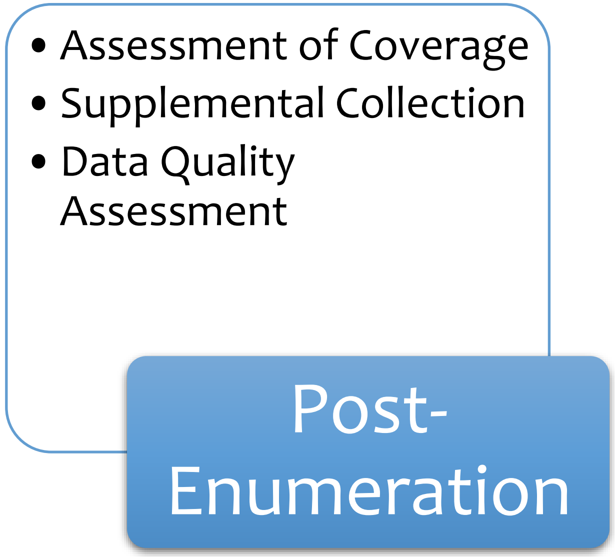
Monitoring

- Systems developed during data collection
- Failures at some supervisory levels

Respondents

- Increased privacy concerns
- Limited access to gated communities
- Coverage issues

Census Process (simplified) – Key Challenges



- Assessment of Coverage
- Supplemental Collection
- Data Quality Assessment

Post-Enumeration

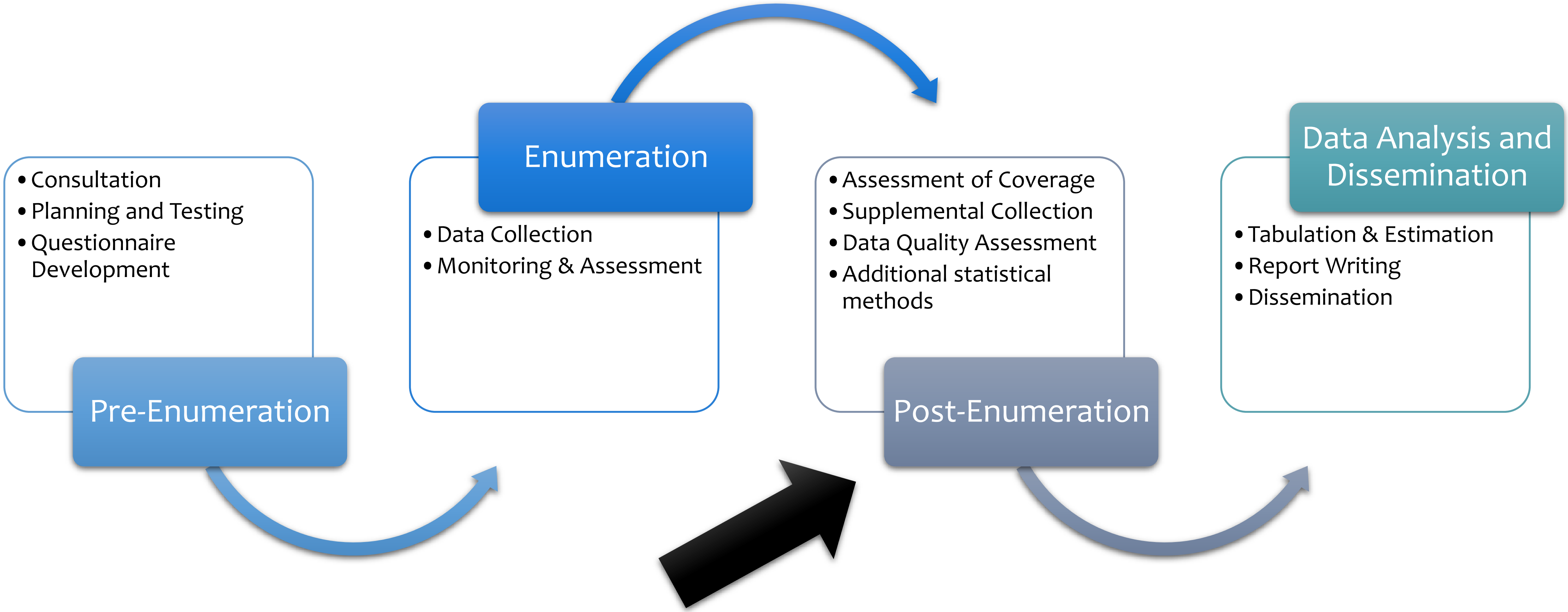
Politicizing of the Census

- Undermines the integrity of the process
- Garners unnecessary media attention

Solutions are Technically Complex

- Supported by UN ECLAC and CELADE
- Competent staff, but burnt-out

Modified Census Process (simplified)



Pre-Enumeration

- Consultation
- Planning and Testing
- Questionnaire Development

Enumeration

- Data Collection
- Monitoring & Assessment

Post-Enumeration

- Assessment of Coverage
- Supplemental Collection
- Data Quality Assessment
- Additional statistical methods

Data Analysis and Dissemination

- Tabulation & Estimation
- Report Writing
- Dissemination

Ongoing Work and Next Steps

- Partial VR in every ED
- Acquisition of satellite imagery, building footprints and administrative data
- Post-Census Web Survey to help assess the undercount
- Application of advanced statistical techniques
- Estimation of the count by age and sex
- Assessment of other indicators
- Publication of results

Advanced Statistical Techniques

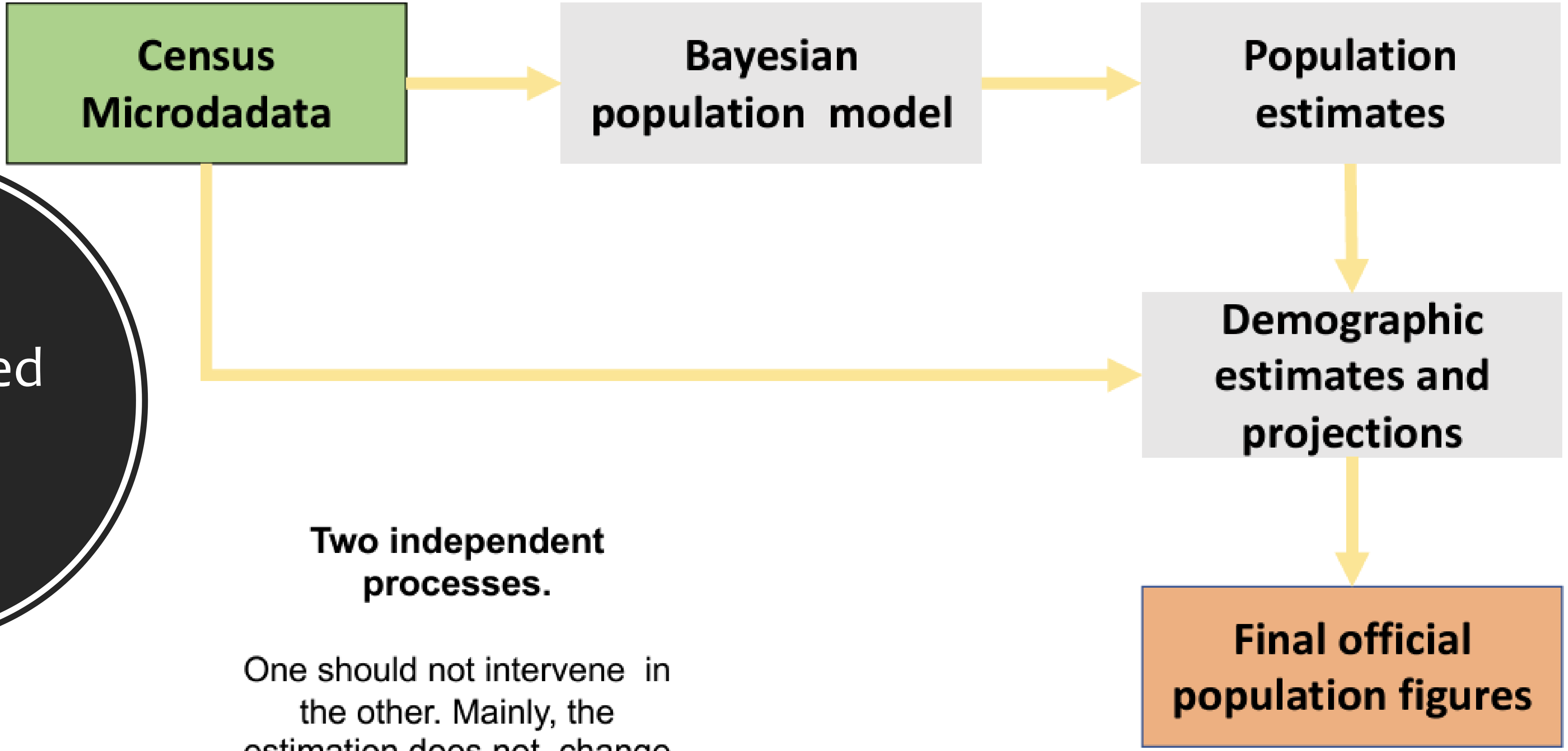
Population models

- Relate observed population data from the census or other surveys to other data sets to predict the population in areas where census information is incomplete.
- Designed specifically for each country based on available inputs and expected objectives.
- Models can be designed to make estimates various levels.

Statistical mixed models

- Bayesian
- Incorporate heterogeneity in unobserved areas.
- Uses covariates e.g. satellite (lights, building footprints), geospatial (roads, infrastructure), or cartographic variables.

The Planned Process



Census Microdadata

Bayesian population model

Population estimates

Demographic estimates and projections

Final official population figures

Two independent processes.

One should not intervene in the other. Mainly, the estimation does not change the data!

Thank You!

IAOS-ISI 2024
MEXICO CONFERENCE

ECLAC and UNFPA approach to model populations in Latin America and the Caribbean

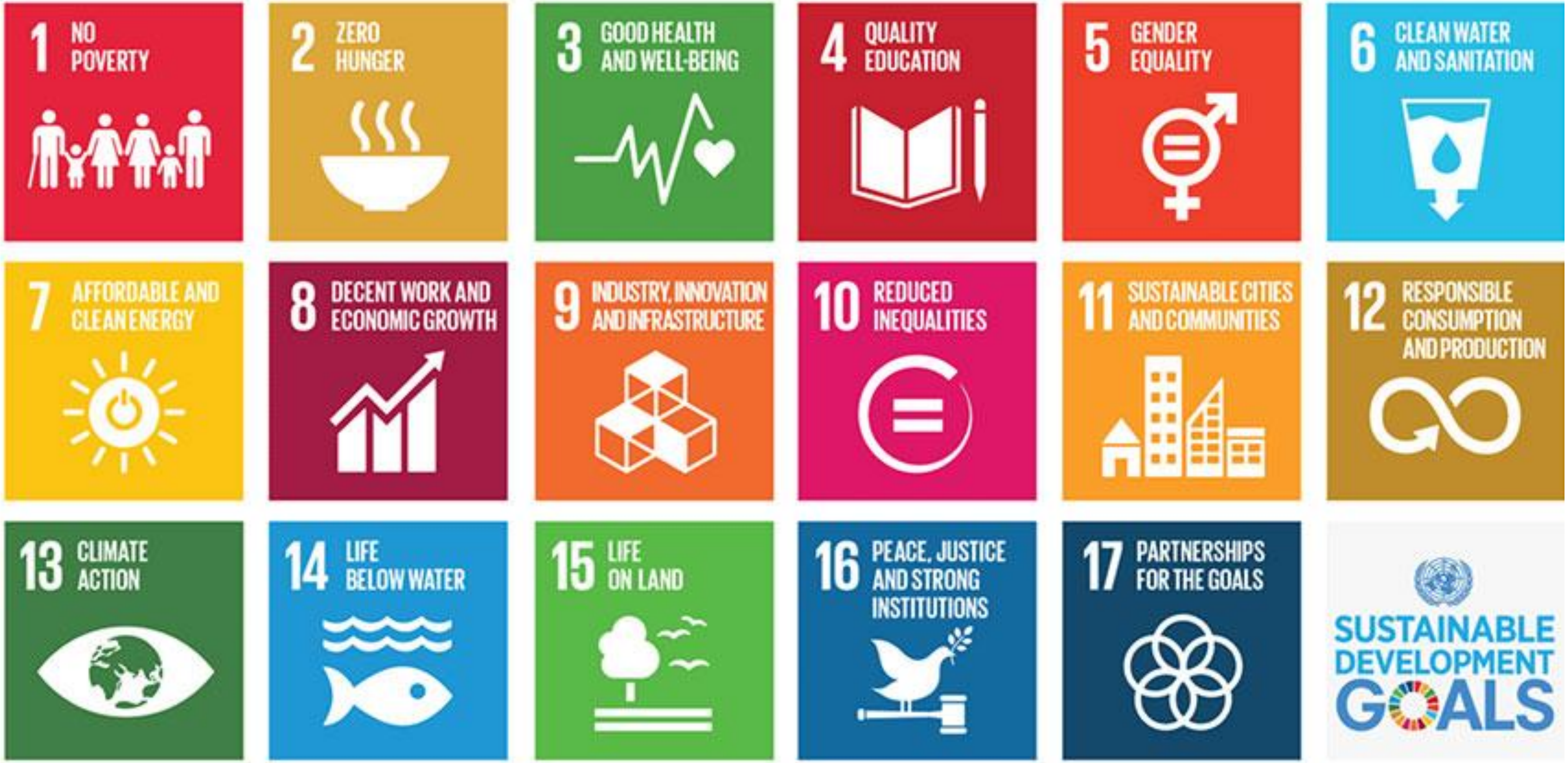
Andrés Gutiérrez, PhD.
Regional Adviser on Social Statistics
andres.gutierrez@un.org

INEGI
International Statistical Institute
IAOS
IMPROVING OFFICIAL STATISTICS

Why do we do the things we do?

An SDG perspective

SUSTAINABLE DEVELOPMENT GOALS



1 NO POVERTY

2 ZERO HUNGER

3 GOOD HEALTH AND WELL-BEING

4 QUALITY EDUCATION

5 GENDER EQUALITY

6 CLEAN WATER AND SANITATION

7 AFFORDABLE AND CLEAN ENERGY

8 DECENT WORK AND ECONOMIC GROWTH

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

10 REDUCED INEQUALITIES

11 SUSTAINABLE CITIES AND COMMUNITIES

12 RESPONSIBLE CONSUMPTION AND PRODUCTION

13 CLIMATE ACTION

14 LIFE BELOW WATER

15 LIFE ON LAND

16 PEACE, JUSTICE AND STRONG INSTITUTIONS

17 PARTNERSHIPS FOR THE GOALS

SUSTAINABLE DEVELOPMENT GOALS

Make cities and human settlements inclusive, safe, resilient and sustainable

SDG 11: Sustainable communities

- Target 11.1.: By 2030, ensure access for all to adequate, safe and affordable housing and basic services and upgrade slums.
  - Indicator 11.1.1: Proportion of urban population living in slums, informal settlements or inadequate housing.
- Target 11.1.: By 2030, enhance inclusive and sustainable urbanization and capacity for participatory, integrated and sustainable human settlement planning and management in all countries.
  - Indicator 11.3.1: Ratio of land consumption rate to population growth rate.

Censuses and recent experiences in Latin America and the Caribbean

The problem of coverage

- Censuses are massive statistical operations that try collecting data from all areas in the country in a certain period of time.
  - Some countries tried to expand the collection period to lower the under-coverage, implying a tremendous effort in resource mobilization.
  - This solution did not prove to be as effective as expected, and the lower coverage rates kept in some areas.
- The censuses should stop their collection stage after multiple extensions.
  - In several countries returning to collection in the areas of lower coverage was not an option due to limited budget.
  - Incomplete collection along the countries was a common issue.

Some challenges in population censuses

- Population censuses do not always manage to list all households and their populations throughout the country.
  - Complete omission of dwellings or misidentification of the occupancy status of the dwelling.
  - Complete or partial omission of people inside the dwellings.
  - Complete or partial omission of certain geographical areas due to problems of planning of field work, accessibility or security among others during the census enumeration.
- Most of the countries in LAC region are experiencing these kind of challenges in their censuses.

Some challenges in population censuses

- Some countries that have not made the census may face problems getting accurate and precise counts of people.
  - Obsolescence of figures based in old and outdated censuses.
  - Recent migration phenomena increased the need for up to date figures.
  - Need for prediction of counts in some districts and regions

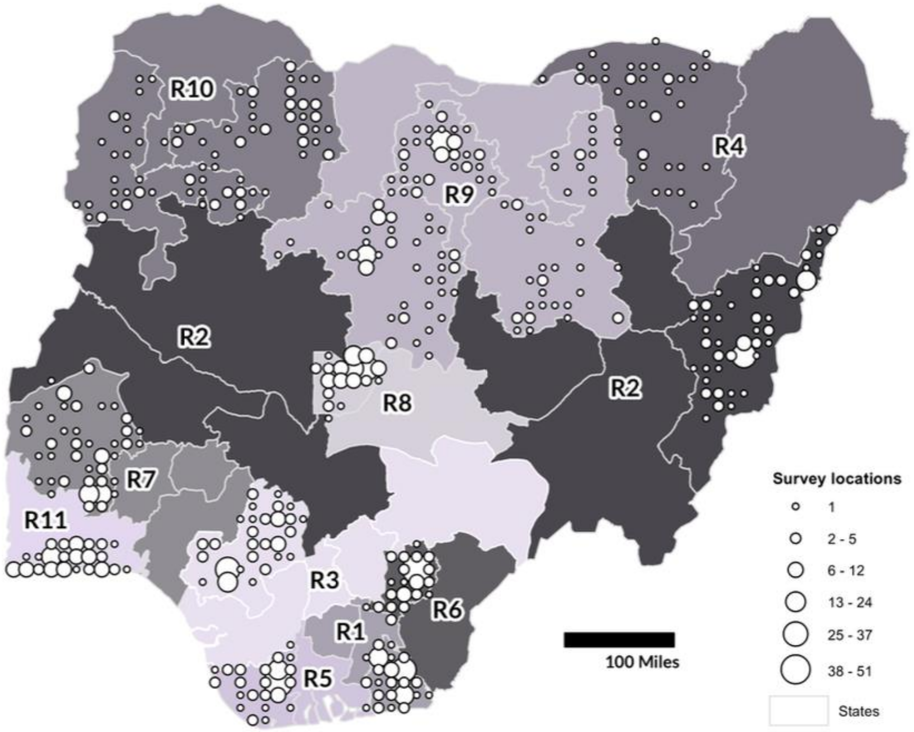
Parsimonious solution

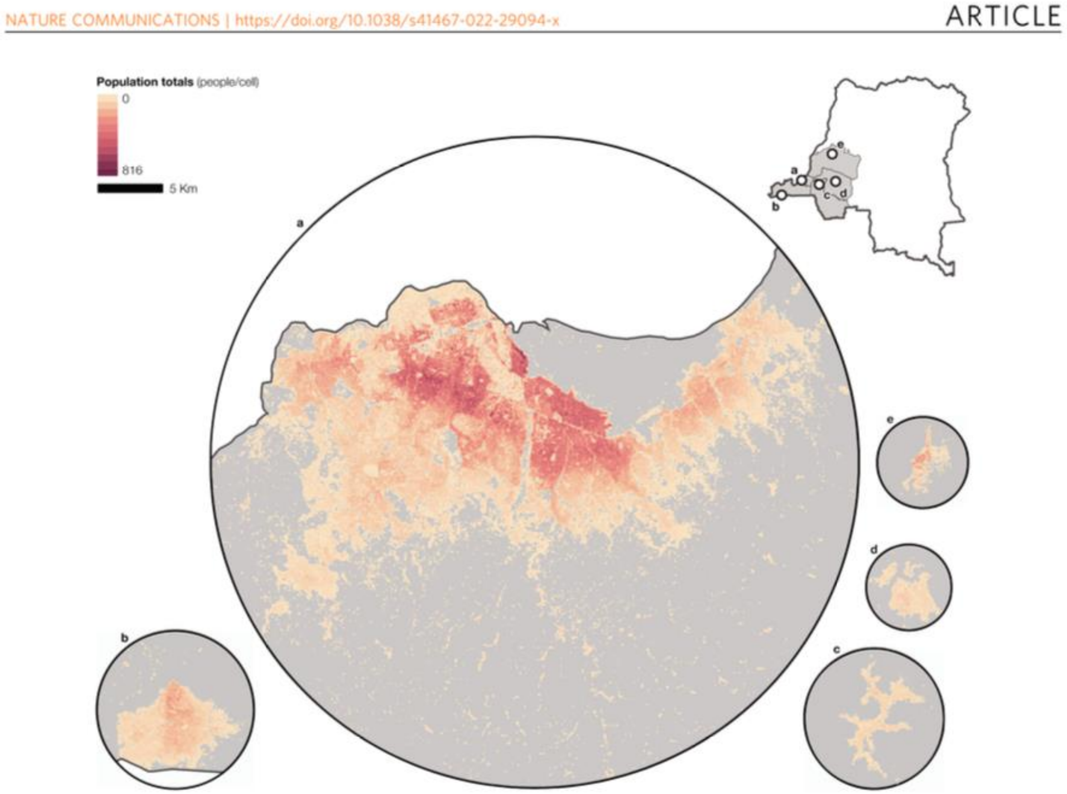
- When incomplete enumeration of areas is a problem in the census, we can rely on statistical models to predict counts of people (along with their demographic structure: age and sex).
    - Model-based estimates of counts represents a new approach to the problem of complete or partial omission.
    - The rationale behind these kind of models is borrowing strength from complete areas.
    - This approach uses remote sensing covariates that should be available for all of the areas in the country.
- In the literature we find a lot of experiences with similar models:
    - Boo, et. al (2022), Leasure, et. al (2020), Berg (2023)
    - ECLAC and UNFPA join venture in Latin America and the Caribbean

Population models

The approach of ECLAC and UNFPA

Models based on enumeration surveys



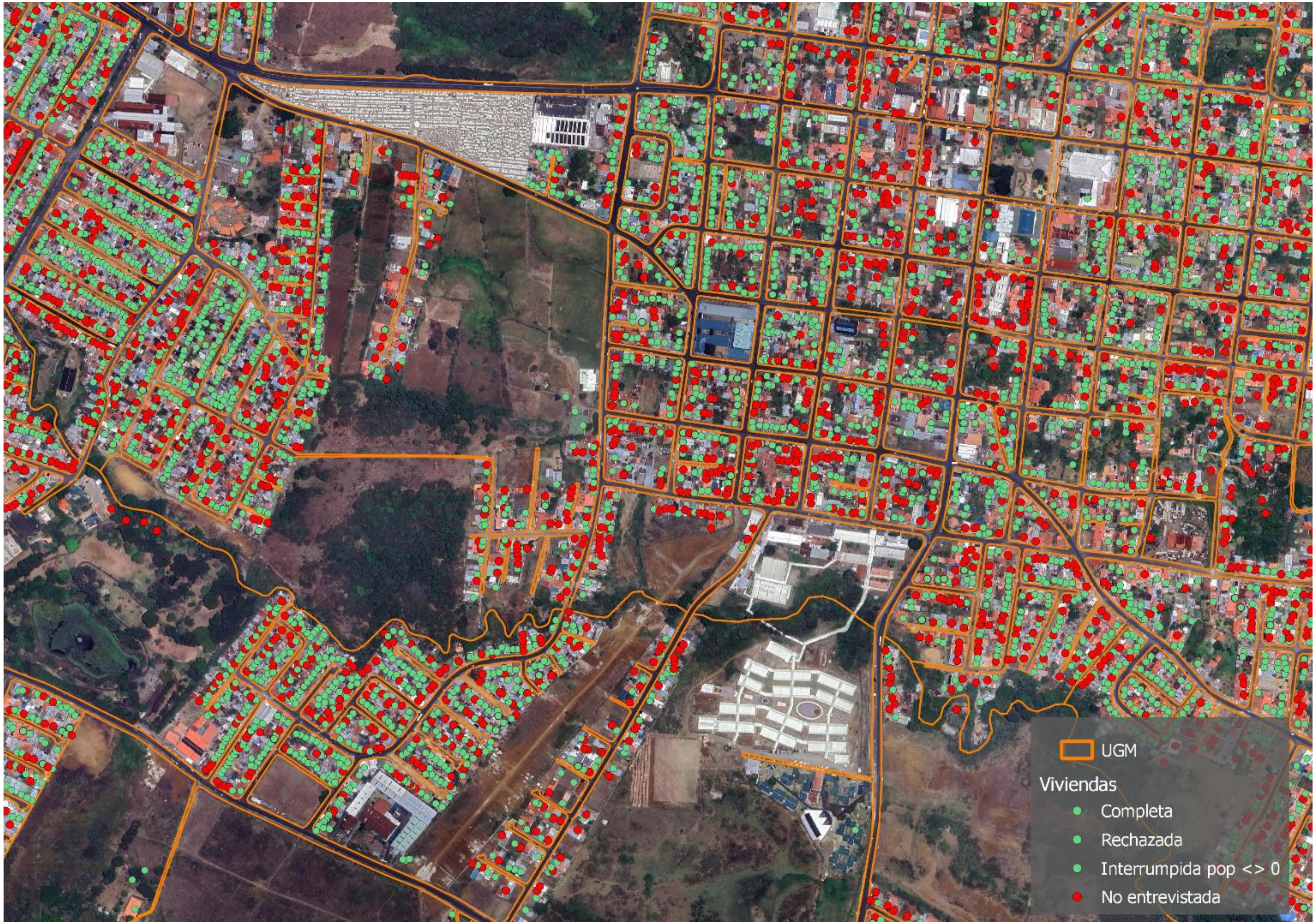

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-022-29094-x ARTICLE

ECLAC and UNFPA population models

- In our context, statistical models relate observed population data from the census to other data sets (available from administrative records or satellite imagery) in order to predict the population in areas where census information is incomplete.
- They are designed specifically for each country based on available inputs and expected objectives.
- Models can be designed to make estimates at grid level (1 km, 100 m, etc.), statistical sectors or other geographical or administrative levels, depending on the needs and the quality and quantity of information available.

Main characteristics

- Our population models have three characteristics:
  - They are Bayesian to be able to add previous information to the observed areas.
  - They are mixed to incorporate heterogeneity in unobserved areas.
  - Covariates always include satellite imagery (lights, building footprints), geospatial information (roads, infrastructure), or cartographic variables.


UGM
Viviendas
Completa
Rechazada
Interrumpida pop <> 0
No entrevistada

The Poisson GLMM for counts

We define the dwelling-level Poisson GLMM as in Berg (2022). Assume:

$$y_{ij} \mid \mu_{ij} \sim Poisson(\mu_{ij})$$
$$\mu_{ij} = N_j \, D_{ij}$$

Where y_{ij} represents the number of people in dwelling i and enumeration district j. N_j is the number of dwellings in enumeration district j. Also, D_{ij} is the average density in the dwelling and it related to the outcome through the following link function:

$$\log(D_{ij}) = \boldsymbol{x}_{ij}\boldsymbol{\beta} + u_j$$

Prior information and posterior distribution

The prior distributions for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are as follows:

$$\beta_p \sim \mathit{Normal}\,(0{,}10000)$$
$$u_j \sim \mathit{Normal}\,(0{,}\sigma_u^2)$$
$$\sigma_u^2 \sim \ Inverse - Gamma(0.0001{,}0.0001)$$

Therefore, the Bayesian estimator for the number of people in dwelling i from ED j is given as

$$\tilde{\theta}_{ij} = E(y_{ij} \mid \mu_{ij})$$

The parameter of interest

The aim of the research will always be estimating the number of people in the country

$$t_y = \sum_{All\ EDs} \sum_{All\ Dwellings} y_{ij}$$

However, this parameter can be decomposed as follows:

$$t_y = \sum_{Complete\ EDs} \sum_{Complete\ Dwellings} y_{ij} + \sum_{Incomplete\ EDs} \sum_{Incomplete\ Dwellings} y_{ij}$$

Predictive approach

This way, the proposed Bayesian predictor is given by the following expression:

$$\hat{t}_y = \sum_{Complete\ EDs} \sum_{Complete\ Dwellings} y_{ij} + \sum_{Incomplete\ EDs} \sum_{Incomplete\ Dwellings} \tilde{\theta}_{ij}$$

This expression is similar to Molina and Rao (2010) Empirical Best Predictor in the context of poverty maps and small area estimation models.

The Multinomial GLMM for age-sex counts

We also define a municipal-level Multinomial GLMM to predict the probability of people being in each of the 40 age-sex groups (20 x 2). This way:

$$\boldsymbol{N}_d \sim Multinomial(\boldsymbol{p}_d)$$
$$\boldsymbol{p}_d = (p_{d,1,1}, \dots, p_{d,2,20})$$

Where $\boldsymbol{N}_d = (N_{d\,1\,1}, \dots, N_{d\,2\,20})'$, and $N_{d,k,l}$ represents the number of people in municipality d belonging to the sex k and age group l. Also,

$$\log\left(\frac{p_{d\,i\,j}}{p_{d\,1\,1}}\right) = \boldsymbol{z}_{dij}\boldsymbol{\gamma} + e_{dij}$$

Technical assistance in the region

- ECLAC and UNFPA join efforts have benefited the following countries in the last two years:
  - Costa Rica
  - Ecuador
  - Dominican Republic
- We are currently working with the following countries:
  - Barbados
  - Guyana
  - Jamaica

The role of covariates

Satellite Imagery (ED-level)

- We access this information trough Google Earth Engine, which provides facilities to analyze and obtain this data through the Javascript and Python programming languages, and recently since 2021 in R with the rgee package.
- Among the main advantages of information based on remote sensing is the ease of access to data with deep geographic coverage that is impossible to obtain by traditional means such as surveys or administrative records.

- *Building footprints*
- *WorldPop projections*
- *Urban cover fraction*
- *Rural cover fraction*
- *Crops_cover fraction*
- *Altitude in meters above sea level*
- *Travel time to the nearest medical center*
- *Travel time to the nearest school*

Administrative data (municipal-level)

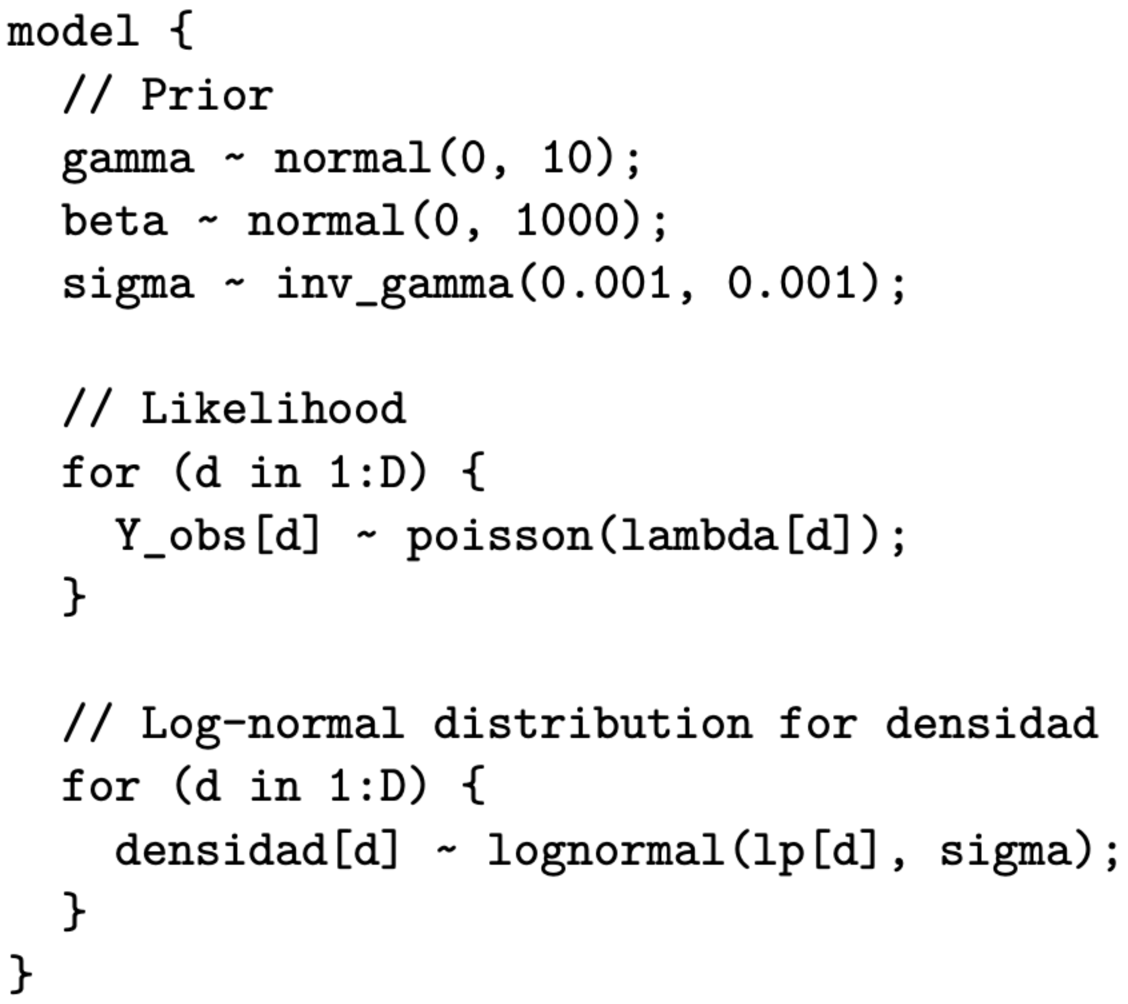
- In each country, valuable information can be found in administrative records.
- Also, we can find important covariates in the most recent census along with cartographic data available in the NSO.

- *Telecommunication access*
- *Access problems*
- *High crime rates*
- *Primary education enrollment*
- *informal settlements*
- *Indigenous area*
- *Protected area*

MCMC convergence and predictions

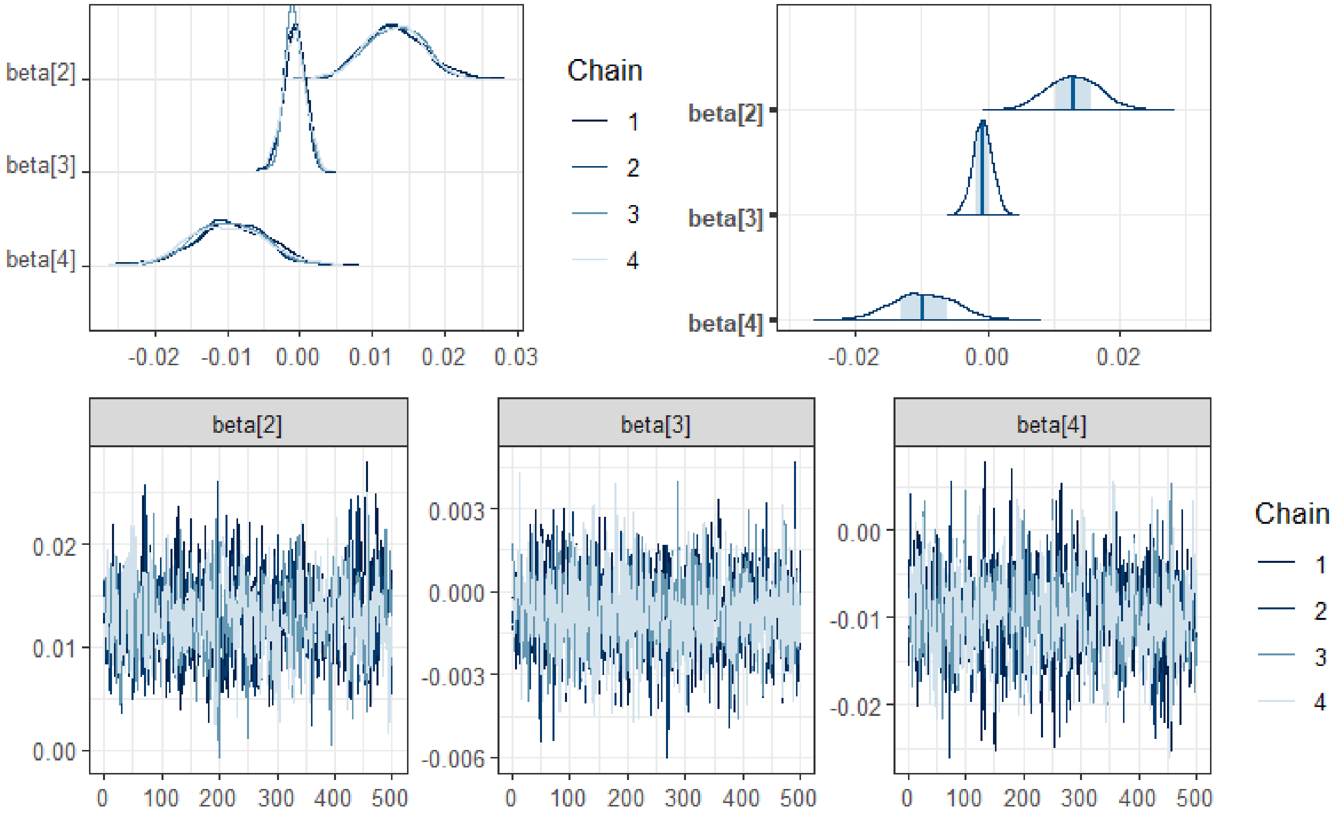
Software

- As these Bayesian computations are complex, we use our own coding in STAN.
    - STAN is an advanced Markov Chain Monte Carlo sampler that uses Hamiltonian algorithms.
    - It is easy to use and available in different platforms (Python, R, etc.)
    - It allows for computing parallelization making the process more efficient in the presence of this massive data sets.

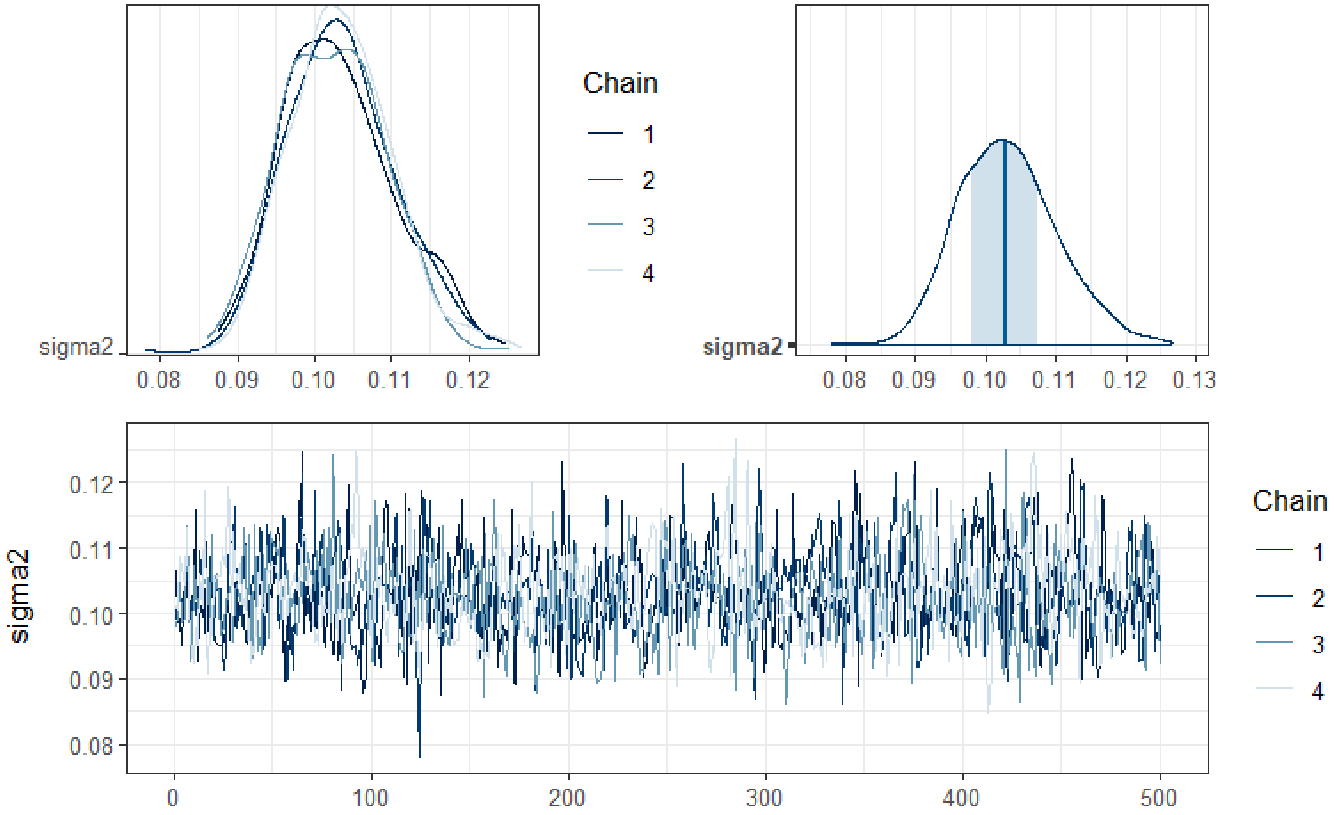


```
model {
  // Prior
  gamma ~ normal(0, 10);
  beta ~ normal(0, 1000);
  sigma ~ inv_gamma(0.001, 0.001);

  // Likelihood
  for (d in 1:D) {
    Y_obs[d] ~ poisson(lambda[d]);
  }

  // Log-normal distribution for densidad
  for (d in 1:D) {
    densidad[d] ~ lognormal(lp[d], sigma);
  }
}
```

Chaind for fixed effects coefficients

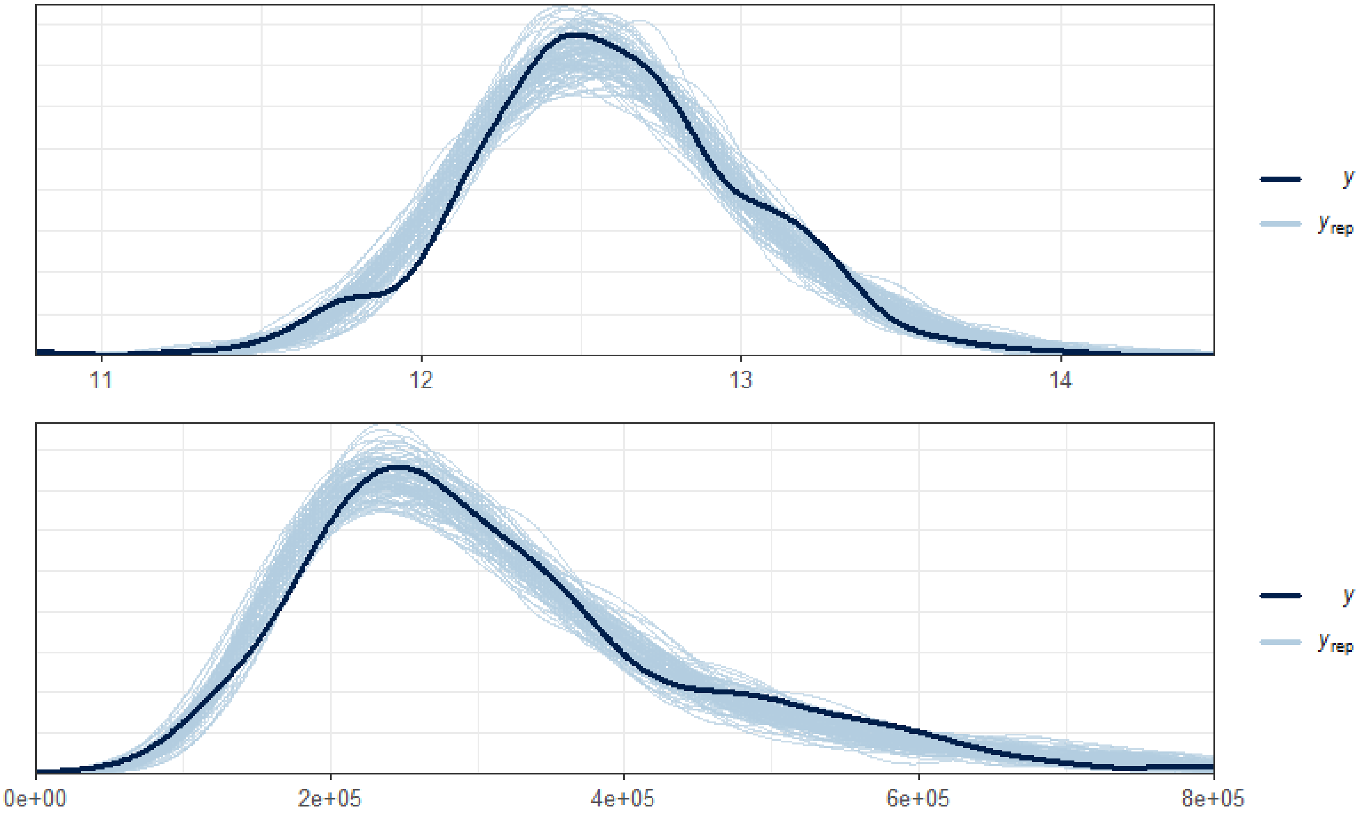


Chains for the variance of random effects

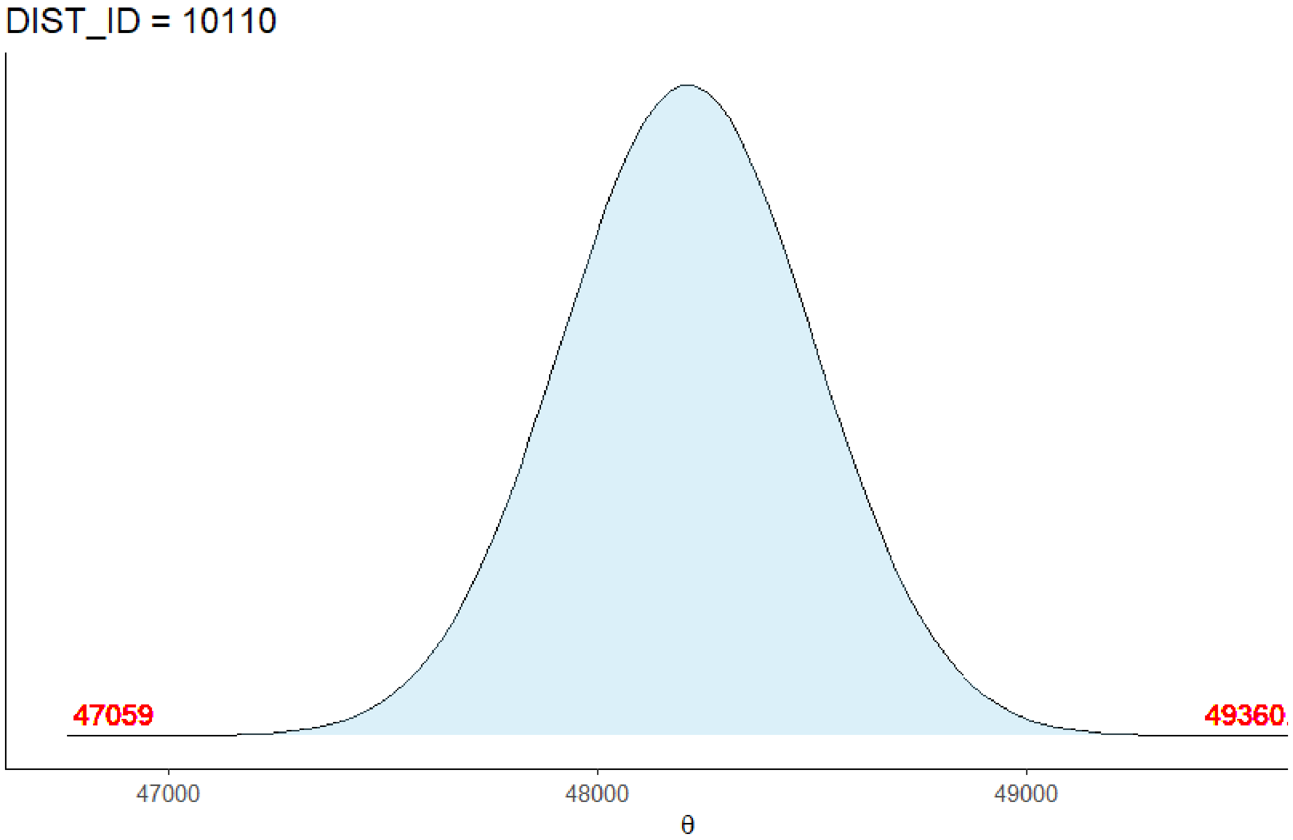


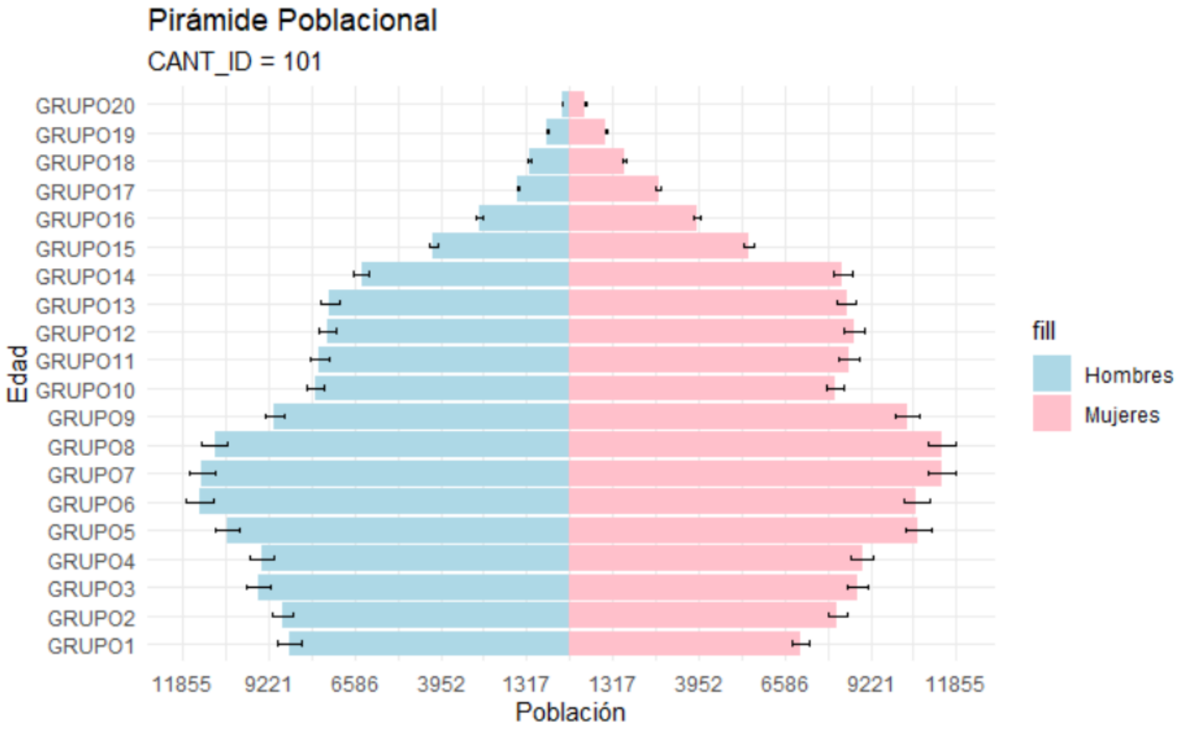
Posterior predictive checks

Log-scale and untransformed

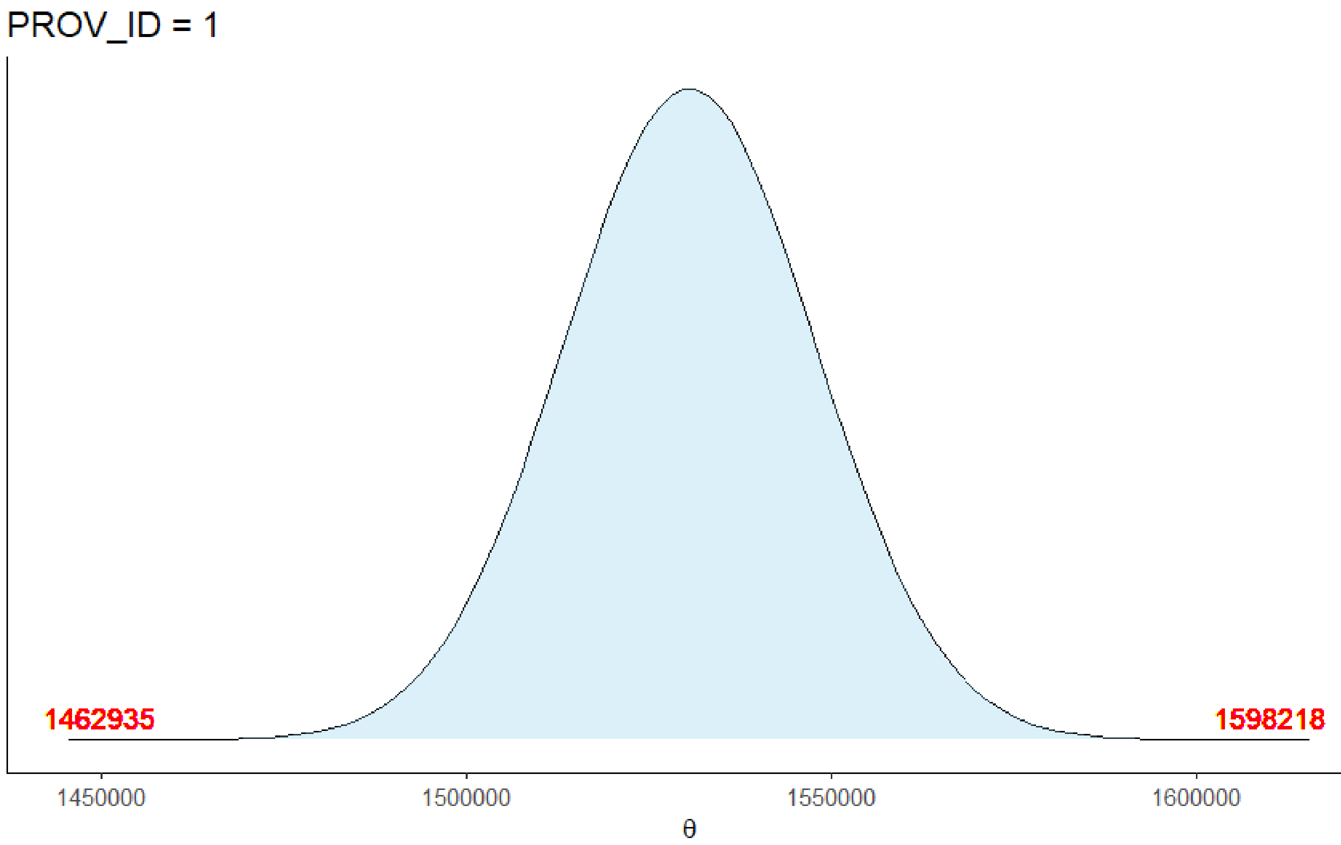


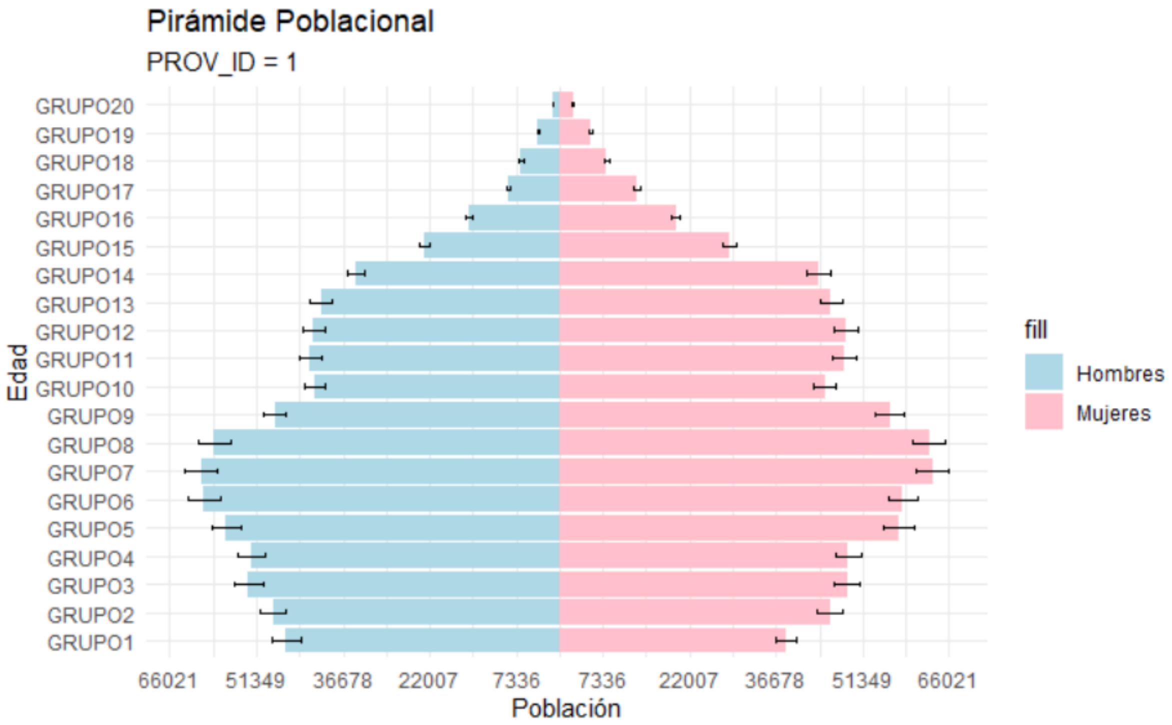
Municipal estimates



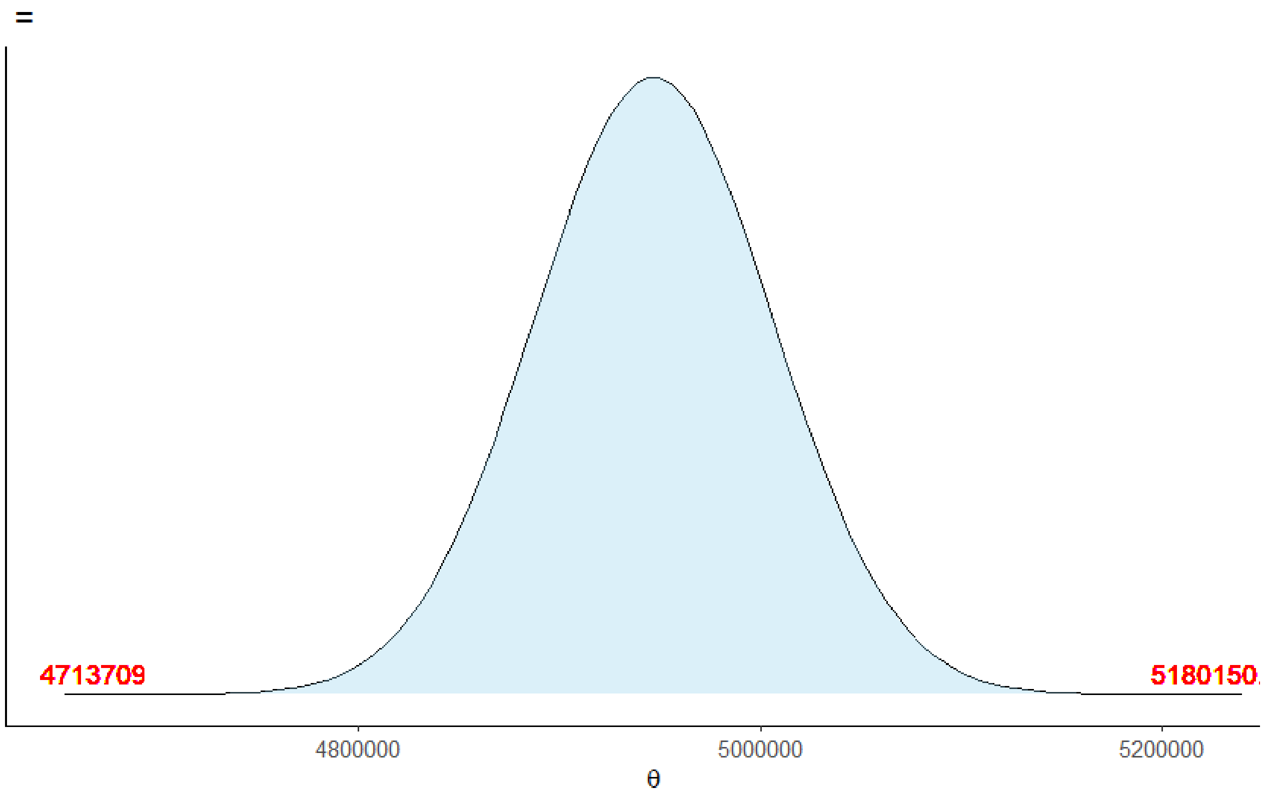


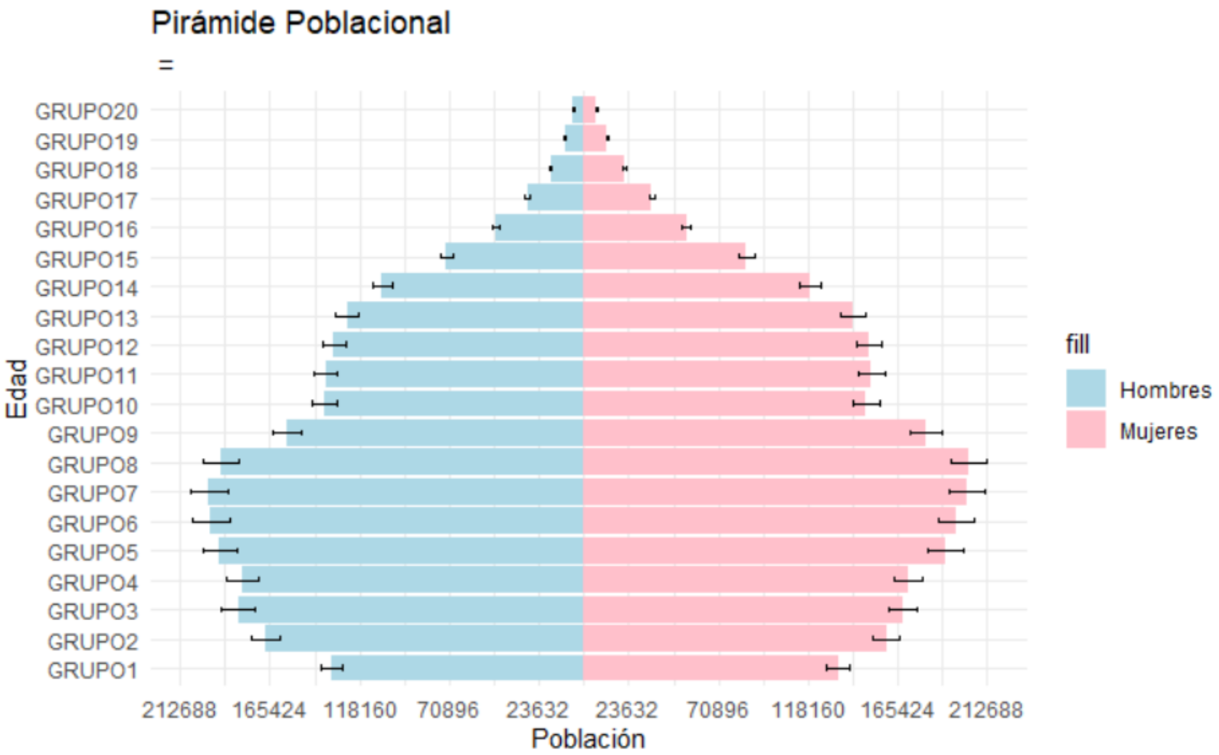
State estimates

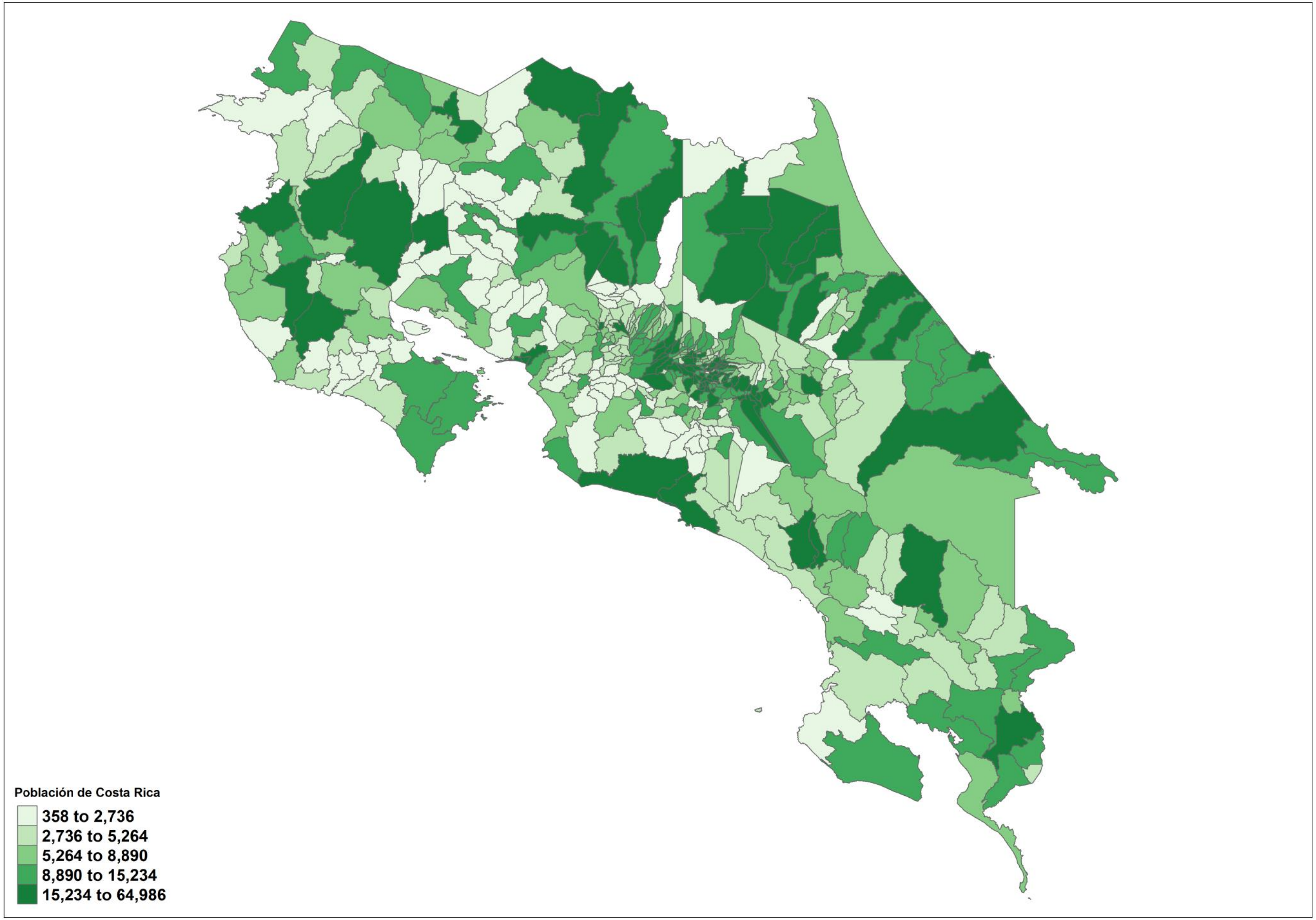




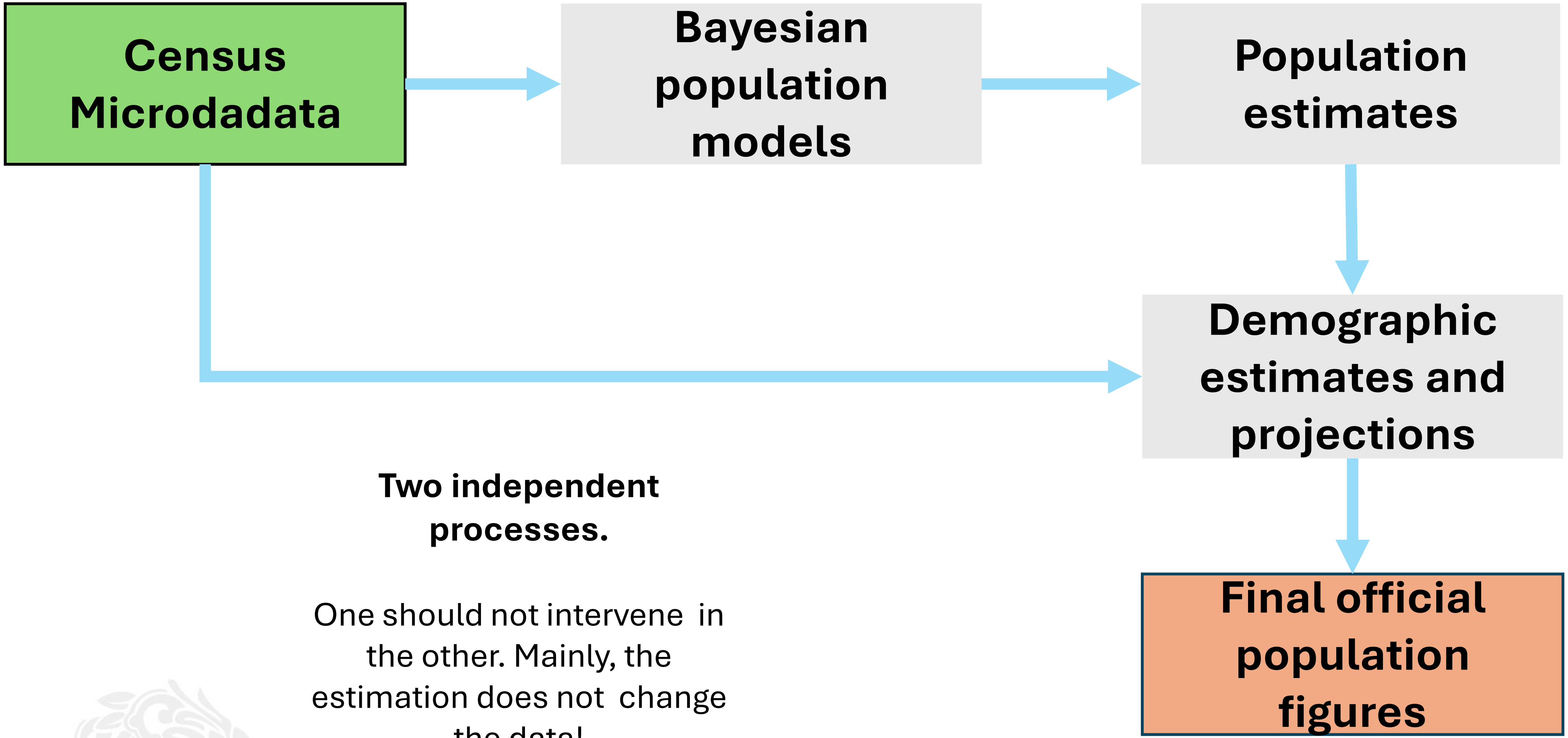
National estimates






Población de Costa Rica
358 to 2,736
2,736 to 5,264
5,264 to 8,890
8,890 to 15,234
15,234 to 64,986

One final word!



Census Microdadata

Bayesian population models

Population estimates

Demographic estimates and projections

Final official population figures

Two independent processes.

One should not intervene in the other. Mainly, the estimation does not change the data!

IAOS-ISI 2024
MEXICO CONFERENCE

Thank you

INEGI

International
Statistical
Institute

IAOS
IMPROVING OFFICIAL STATISTICS

INEC | Buenas cifras, mejores vidas

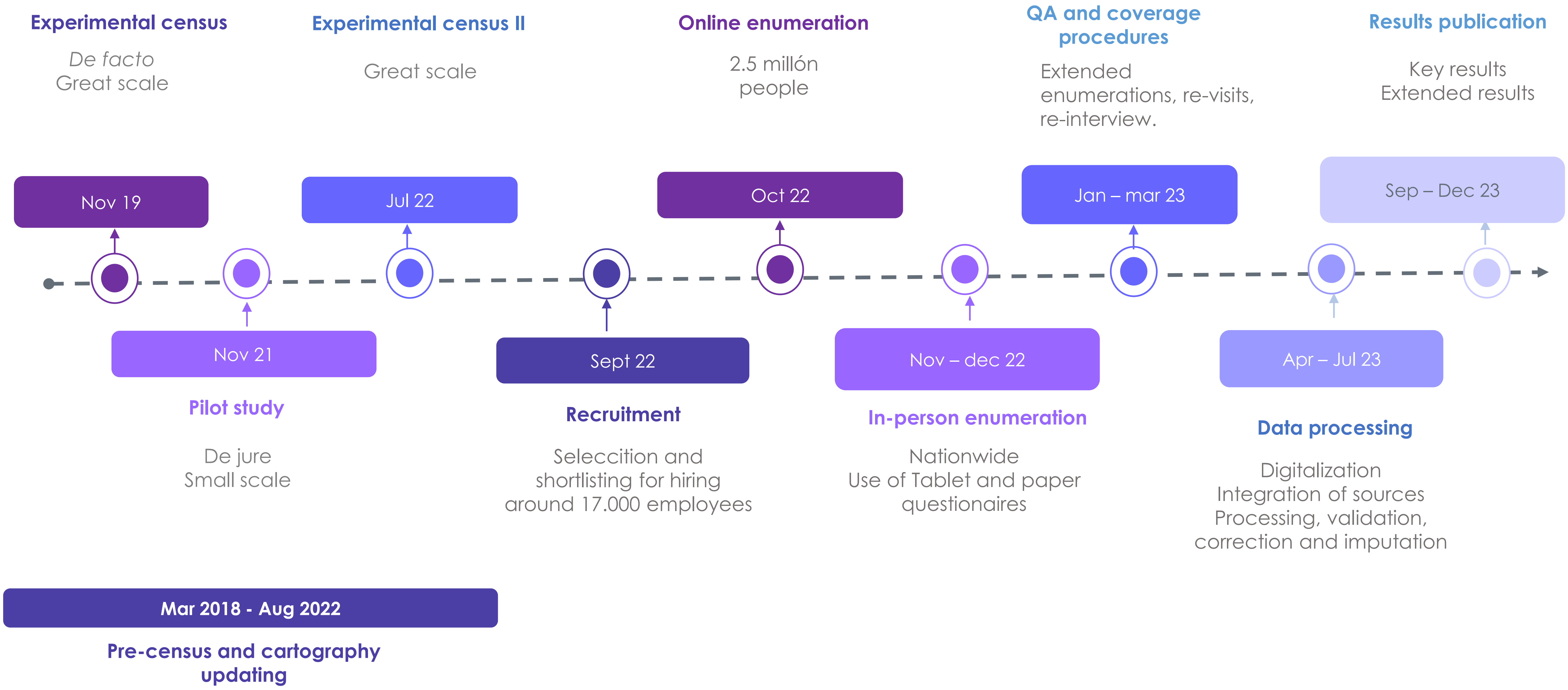
Challenges and opportunities in Census

Ecuador – 2020 round

May, 2024

www.ecuadorencifras.gob.ec

What it took to get us here?



Mar 2018 - Aug 2022

Pre-census and cartography updating

Nov 19

Experimental census

De facto
Great scale

Nov 21

Pilot study

De jure
Small scale

Jul 22

Experimental census II

Great scale

Sept 22

Recruitment

Seleccition and
shortlisting for hiring
around 17.000 employees

Oct 22

Online enumeration

2.5 millón
people

Nov – dec 22

In-person enumeration

Nationwide
Use of Tablet and paper
questionaires

Jan – mar 23

QA and coverage procedures

Extended
enumerations, re-visits,
re-interview.

Apr – Jul 23

Data processing

Digitalization
Integration of sources
Processing, validation,
correction and imputation

Sep – Dec 23

Results publication

Key results
Extended results

Challenges of counting Ecuador

A pandemic, insecurity and more

Covid-19

- Initial planning for **Q4 2020 with direct interviews using senior year students.**
- Sanitary emergency declared in **march 2020**, limited mobility, **increased concern over public health.**

Insecurity

- Days prior to field collection, several **incidents by organized crime** were perpetrated.
- **Loss of up to 20% of recruited personnel.**
- **Difficult to investigate areas** that require specialized logistics.

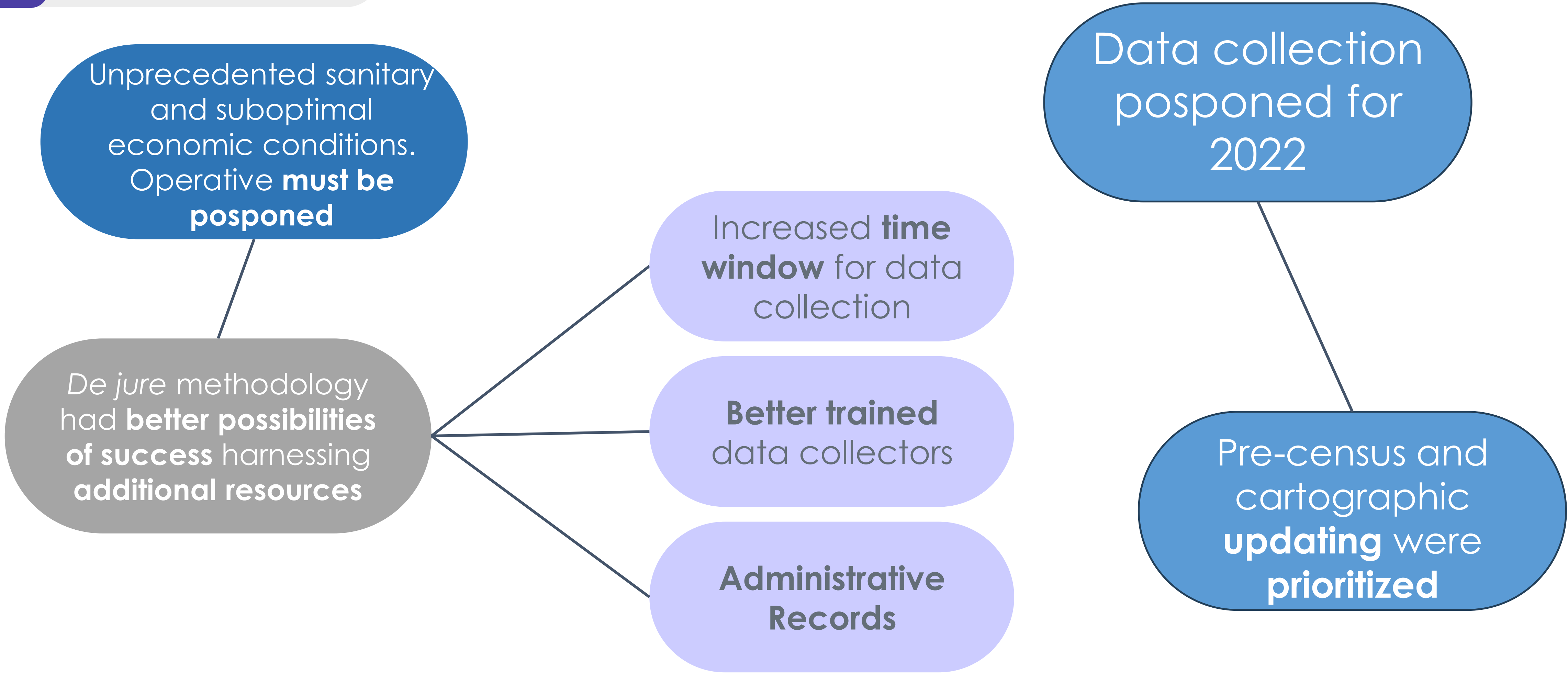
Structural complications

- **Increased rejection rate.** Bigger cities and insecurity. Labor market dynamics and household composition lead to increased difficult to find at home.
- **Tendency to omit certain populational groups.** Specially younger and Elder population

Facing challenges and creating opportunities

What to do with a pandemic?

Covid-19



Unprecedented sanitary and suboptimal economic conditions. Operative **must be posponed**

De jure methodology had **better possibilities of success** harnessing **additional resources**

Increased **time window** for data collection

Better trained data collectors

Administrative Records

Data collection posponed for 2022

Pre-census and cartographic **updating** were **prioritized**

How to deal with insecurity?

Insecurity

Articulation & planning:

- Articulation with all levels of territory
- Continous monitoring of red flagged areas
- Specialized operatives for specific territories

Administrative actions:

- Deconcentrated administrations managed hiring and re-recruitment at each territory
- Use of shortlisting for readily available replacements

Extension in the collection window:

- Re-visits
- Re-interviews
- Planning of suitable timing and strategy for data collection

How to deal with reality?

Structural complications

Rejections

Omited pop. groups

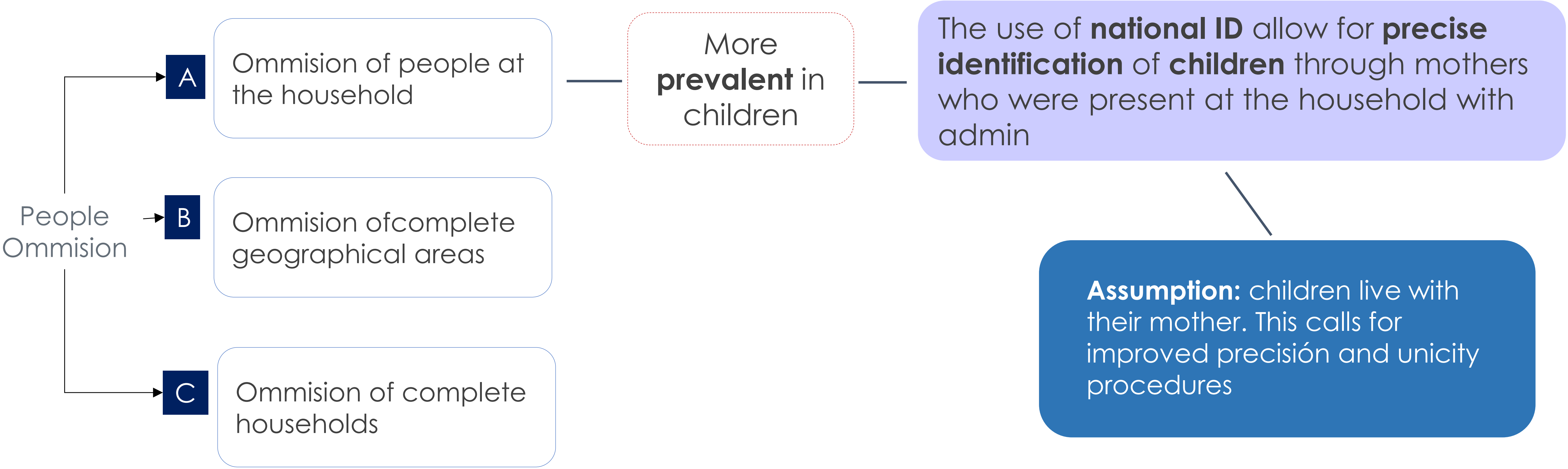
Nobody at home

Creating opportunities

Processing and analysis

Under 12 years old ommited

Ommision in censuses



People Ommision

A Ommision of people at the household

B Ommision ofcomplete geographical areas

C Ommision of complete households

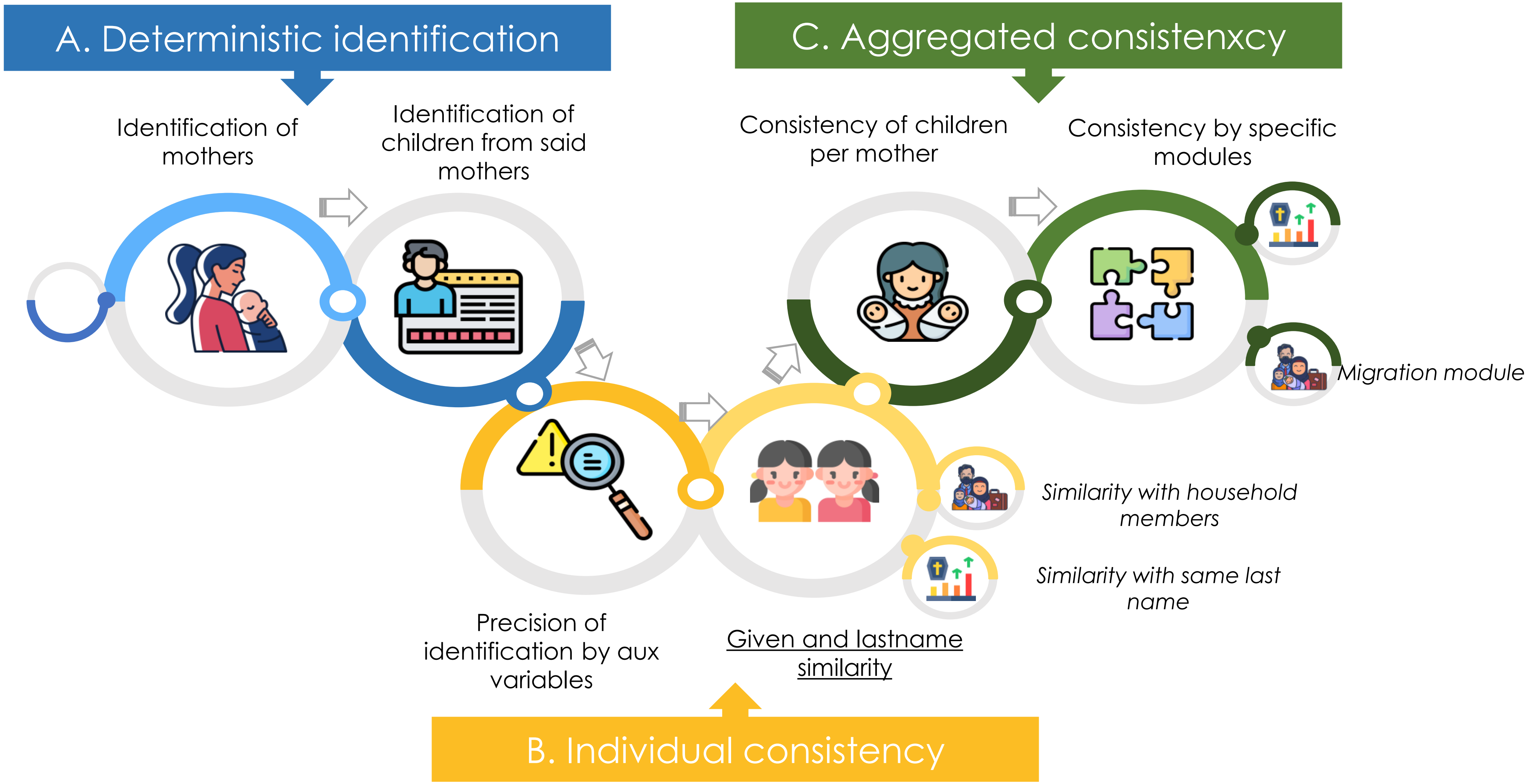
More **prevalent** in children

The use of **national ID** allow for **precise identification** of **children** through mothers who were present at the household with admin

Assumption: children live with their mother. This calls for improved precisión and unicity procedures

Source: CEPAL - UNFPA. 2014. «Los datos demográficos: alcances, limitaciones y métodos de evaluación.» *CEPAL - Serie Manuales N° 82 175.*

Omission in under 12 years old



A. Deterministic identification

Identification of mothers

Identification of children from said mothers

B. Individual consistency

Precision of identification by aux variables

Given and lastname similarity

Similarity with household members

Similarity with same last name

C. Aggregated consistenxcy

Consistency of children per mother

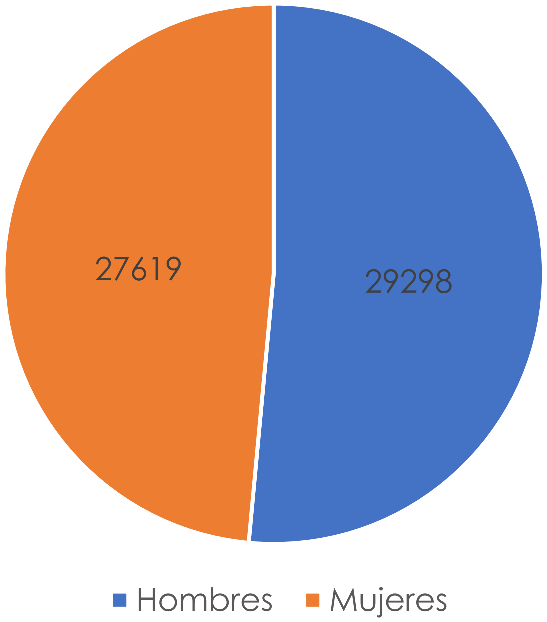
Consistency by specific modules

Migration module

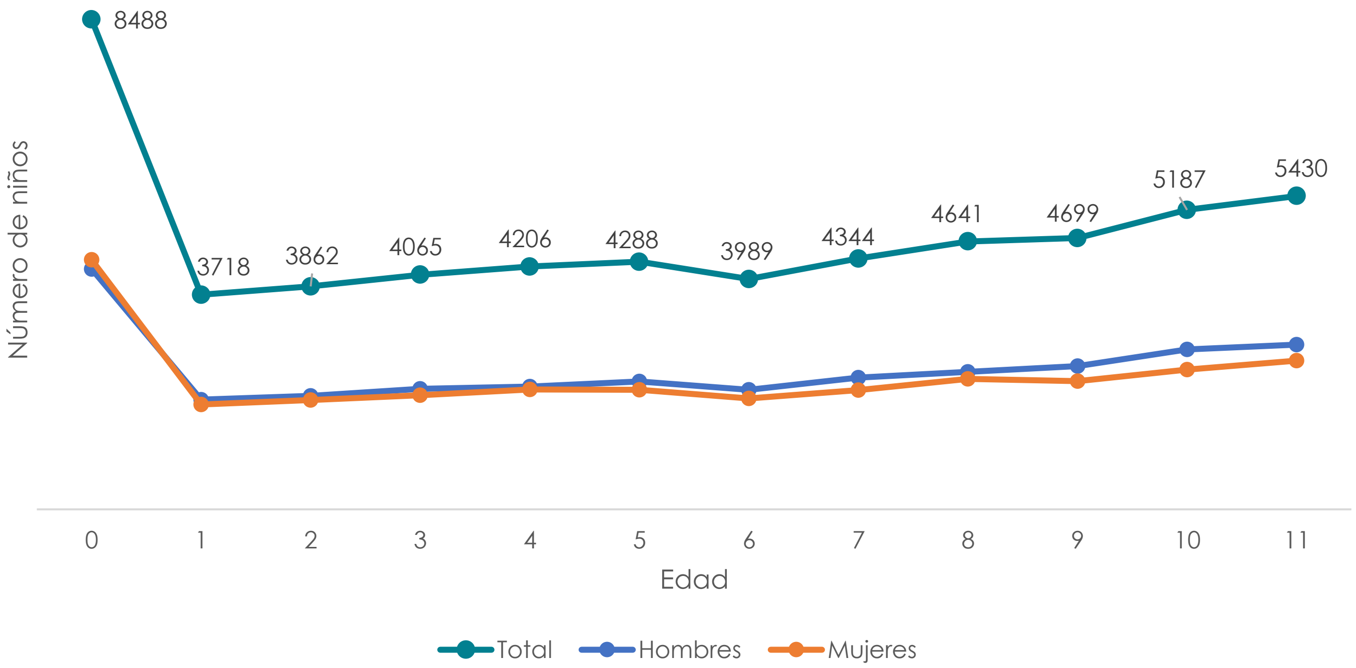
Under 12 years old identified and recuperated

- Over **56.000 children**
- Still children left out from **unidentified mothers** (No valid ID)

Children by sex



Children by age



Estimation of noninterviews

Housing units occupation

- De jure method enumerates usual place of residency; thus it commonly includes **estimation methods** for **noninterviews** of occupied housing units.
- Estimating population from noninterviews is crucial in assuring **comparability with *de facto* censuses** (INE Uruguay, 2011).

Housing units definitions, Ecuador 2022

Condition	Number	%
Occupied	4.821.690	72,9%
Vacant	765.205	11,6%
Seasonal or temporary	612.494	9,3%
Noninterviews	**240.528**	**3,6%**
Under construction	151.749	2,3%
Group quarters	19.869	0,3%
Total	**6.611.535**	**100%**

According to the census, over **240.000 noninterviews (3.64%)** were registered out of **6.6 million housing units**

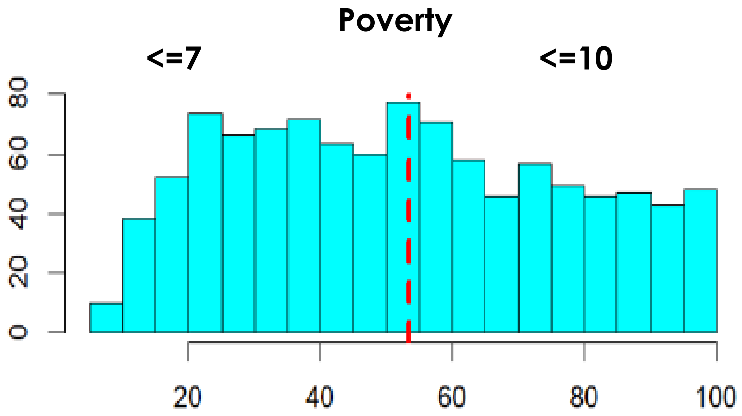
Hotdeck Dynamic imputation by class with single donor

This method assures the **highest likelikood of similarity** between **noninterview household** and **donor,** minimizing bias.

1. Strata construction

- Geographical and socioeconomic strata

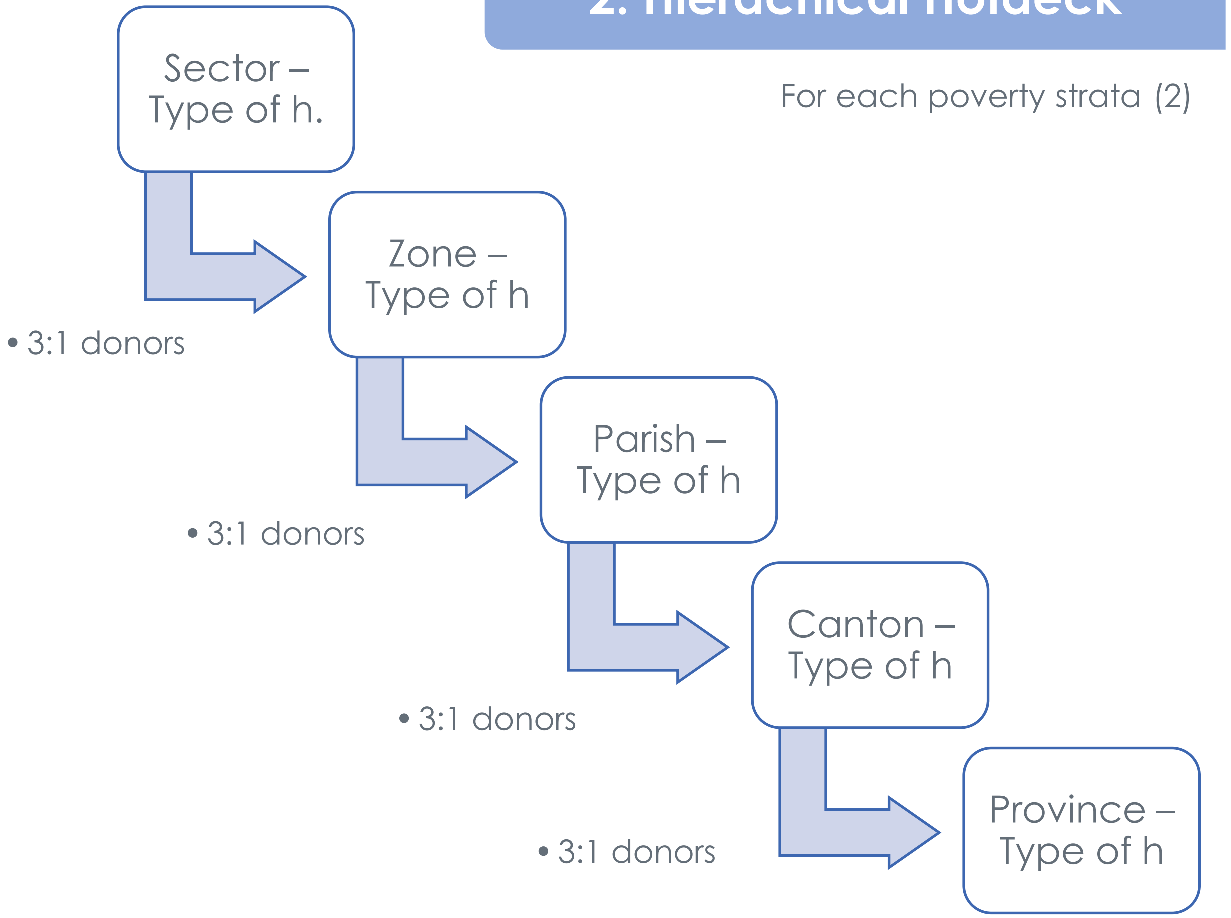
Parish level by poverty, housing unit type and strata



1: House 2: Apartment 3: Others	1: House or apartment 2: Others

2. Hierachical Hotdeck

For each poverty strata (2)



Sector – Type of h.

- 3:1 donors

Zone – Type of h

- 3:1 donors

Parish – Type of h

- 3:1 donors

Canton – Type of h

- 3:1 donors

Province – Type of h

Noninterviews estimated population

	Viviendas			Personas		
Provincia	Sin imputación	Con imputación	(%)	Sin imputación	Imputado	(%)
Azuay	339.524	13.308	3,8%	759.668	41.941	5,2%
Bolívar	86.557	951	1,1%	196.283	2.795	1,4%
Cañar	107.646	1.954	1,8%	221.377	6.201	2,7%
Carchi	63.268	1.312	2,0%	168.628	4.200	2,4%
Cotopaxi	191.521	1.574	0,8%	465.387	4.823	1,0%
Chimborazo	223.639	3.222	1,4%	462.963	8.970	1,9%
El Oro	251.379	16.439	6,1%	662.243	52.349	7,3%
Esmeraldas	202.772	9.317	4,4%	523.089	30.811	5,6%
Guayas	1.530.194	61.714	3,9%	4.192.240	199.683	4,5%
Imbabura	168.674	5.257	3,0%	452.882	16.997	3,6%
Loja	193.809	5.204	2,6%	468.770	16.651	3,4%
Los Ríos	329.316	8.141	2,4%	873.006	25.646	2,9%
Manabí	577.957	15.269	2,6%	1.542.855	49.985	3,1%
Morona Santiago	74.224	1.762	2,3%	186.440	6.068	3,2%
Napo	47.109	613	1,3%	129.613	2.062	1,6%
Pastaza	44.139	1.037	2,3%	108.551	3.364	3,0%
Pichincha	1.170.028	78.984	6,3%	2.848.914	240.559	7,8%
Tungurahua	234.954	1.811	0,8%	558.248	5.284	0,9%
Zamora Chinchipe	46.945	865	1,8%	108.179	2.794	2,5%
Galápagos	13.668	176	1,3%	28.086	497	1,7%
Sucumbíos	77.793	1.813	2,3%	193.145	5.869	2,9%
Orellana	68.341	1.072	1,5%	178.485	3.681	2,0%
Santo Domingo De Los Tsáchilas	184.889	6.259	3,3%	473.403	19.566	4,0%
Santa Elena	142.681	2.474	1,7%	377.428	8.307	2,2%
Total	**6.371.027**	**240.528**	**3,6%**	**16.179.883**	**759.103**	**4,5%**

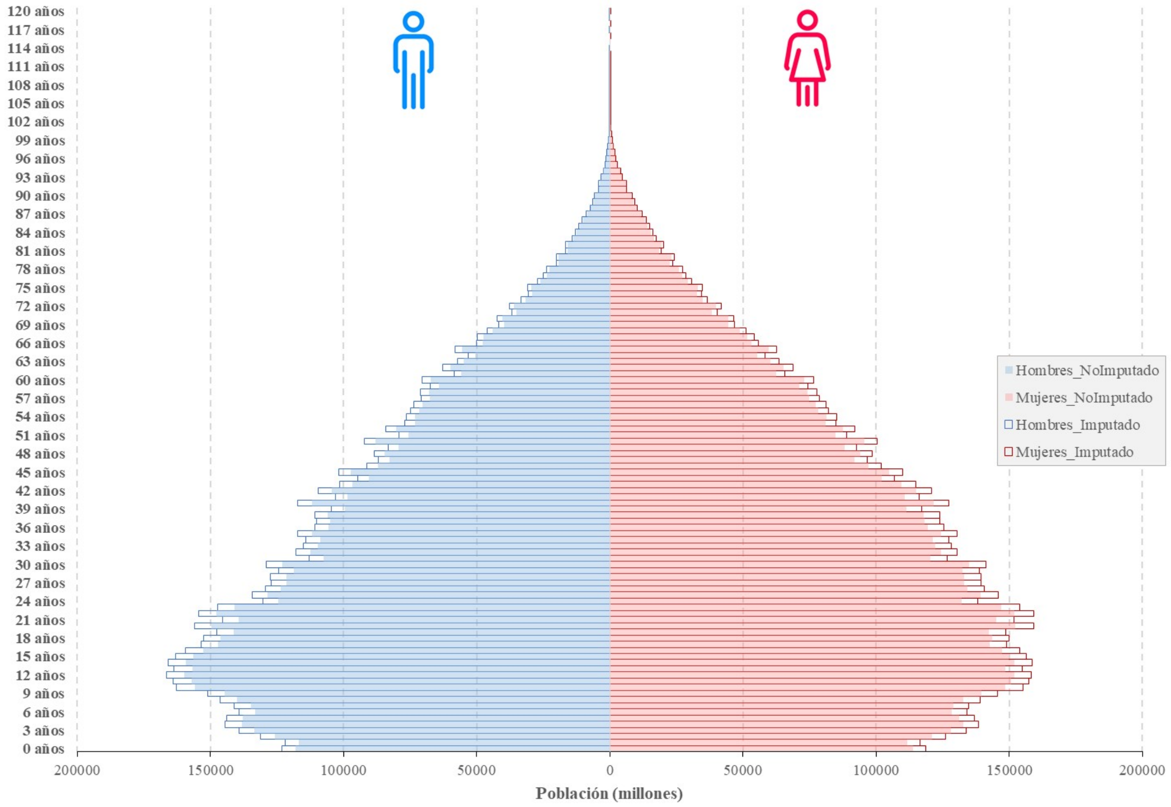
759.000 inhabitants for a total of 16.9 million count in 2022

Pichincha, Guayas and El Oro were provinces with most noninterviews. This derives from difficult to find at home due to labor market dynamics and rejections due to insecurity concerns.

Robustness check of noninterviews estimation

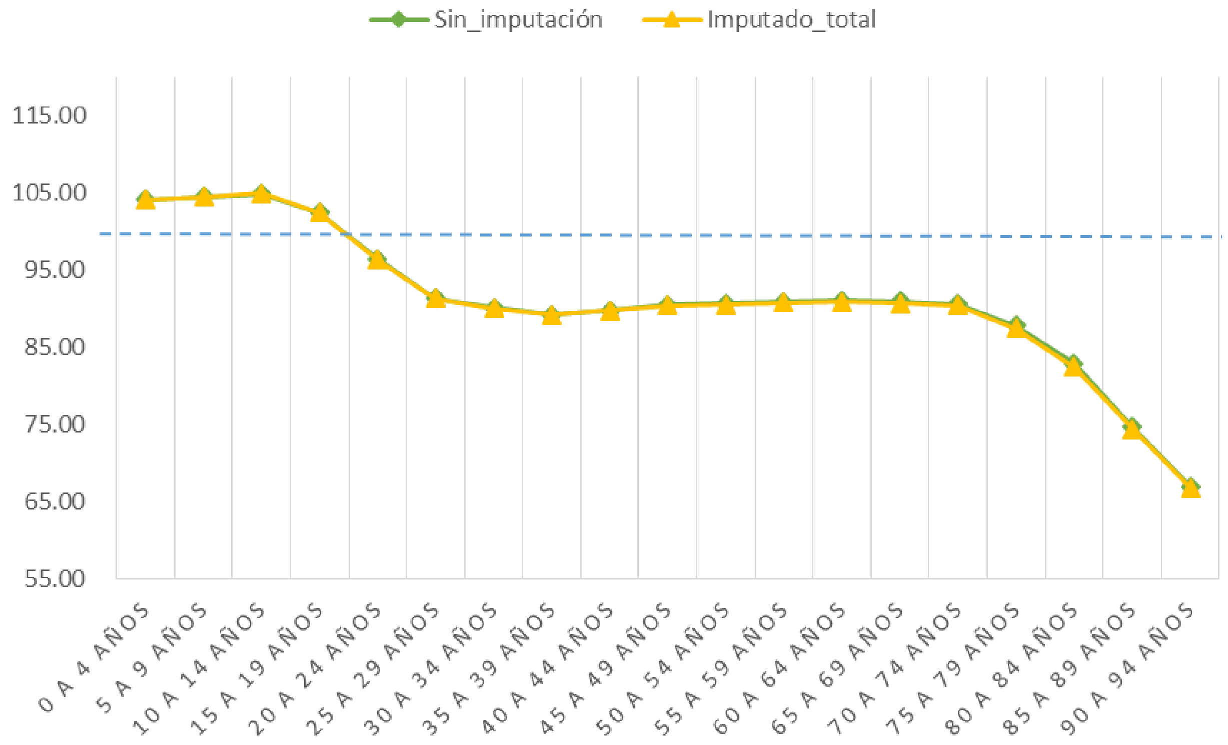
Appropiate distribution of estimated population **across age and sex** ensures **unbiasedness**

Population pyramid



Fuente: Elaboración propia con información del Censo de población y vivienda del INEC, 2022.

Masculinity index



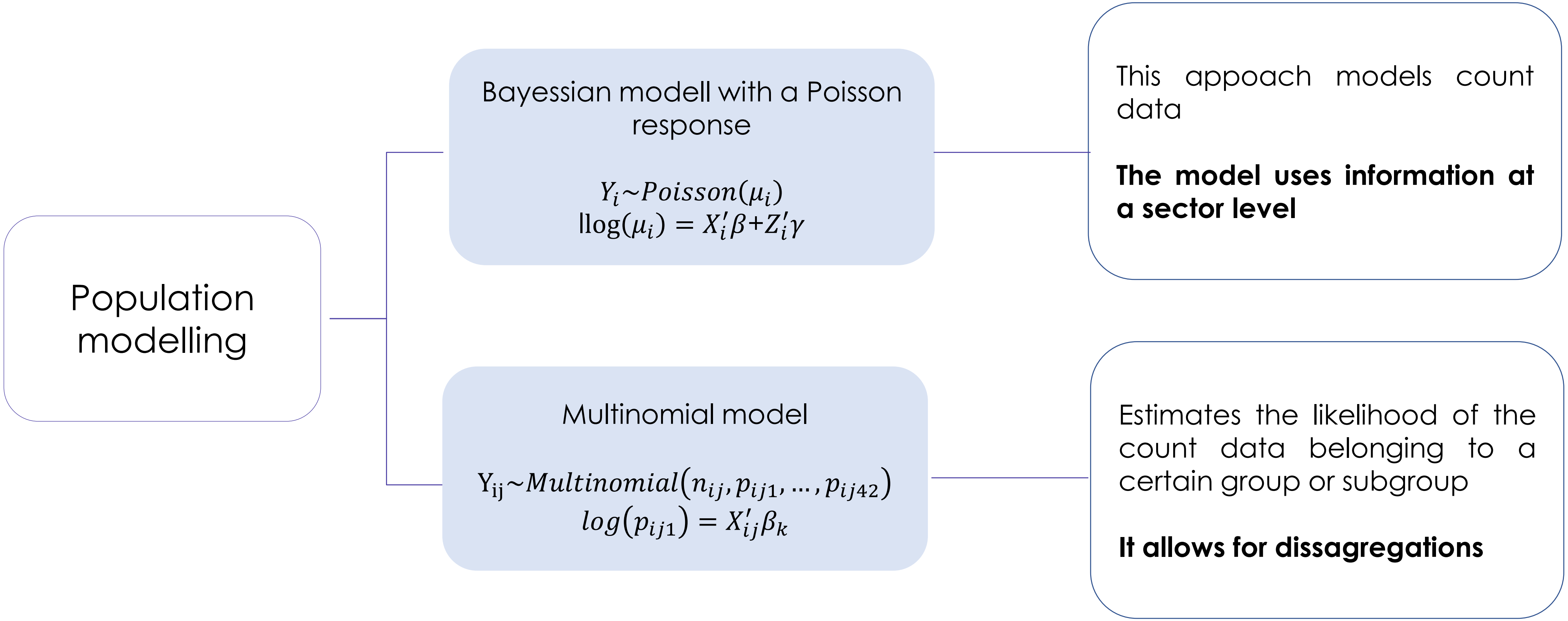
Robustness check of noninterviews estimation

Buenas cifras, mejores vidas

Several demographic indicators with and without imputation

Indicators	Original	Imputation	Total
Population			
Total	16.179.883	759.103	16.938.986
Men	7.886.776	365.747	8.252.523
Women	8.293.107	393.356	8.686.463
Average age	31,9	32,88	31,94
Ratios			
Masculinity index (x100)	95,1	92,98	95
Children/women (x100)	28,34	25,43	28,21
Dependency ratios			
Total	52,66	49,87	52,53
Young	38,98	35,74	38,84
Elder	13,67	14,13	13,69
Digital preference ratios			
Myers index	2,09	2,09	2,1
UN index	14,04	14,13	14,11
Wipple index	103,57	103,56	103,54

Population modelling



Population modelling

Bayessian modell with a Poisson response

$$Y_i \sim Poisson(\mu_i)$$
$$\text{llog}(\mu_i) = X_i'\beta + Z_i'\gamma$$

This appoach models count data

The model uses information at a sector level

Multinomial model

$$\mathrm{Y_{ij}} \sim Multinomial(n_{ij}, p_{ij1}, \dots, p_{ij42})$$
$$log(p_{ij1}) = X_{ij}'\beta_k$$

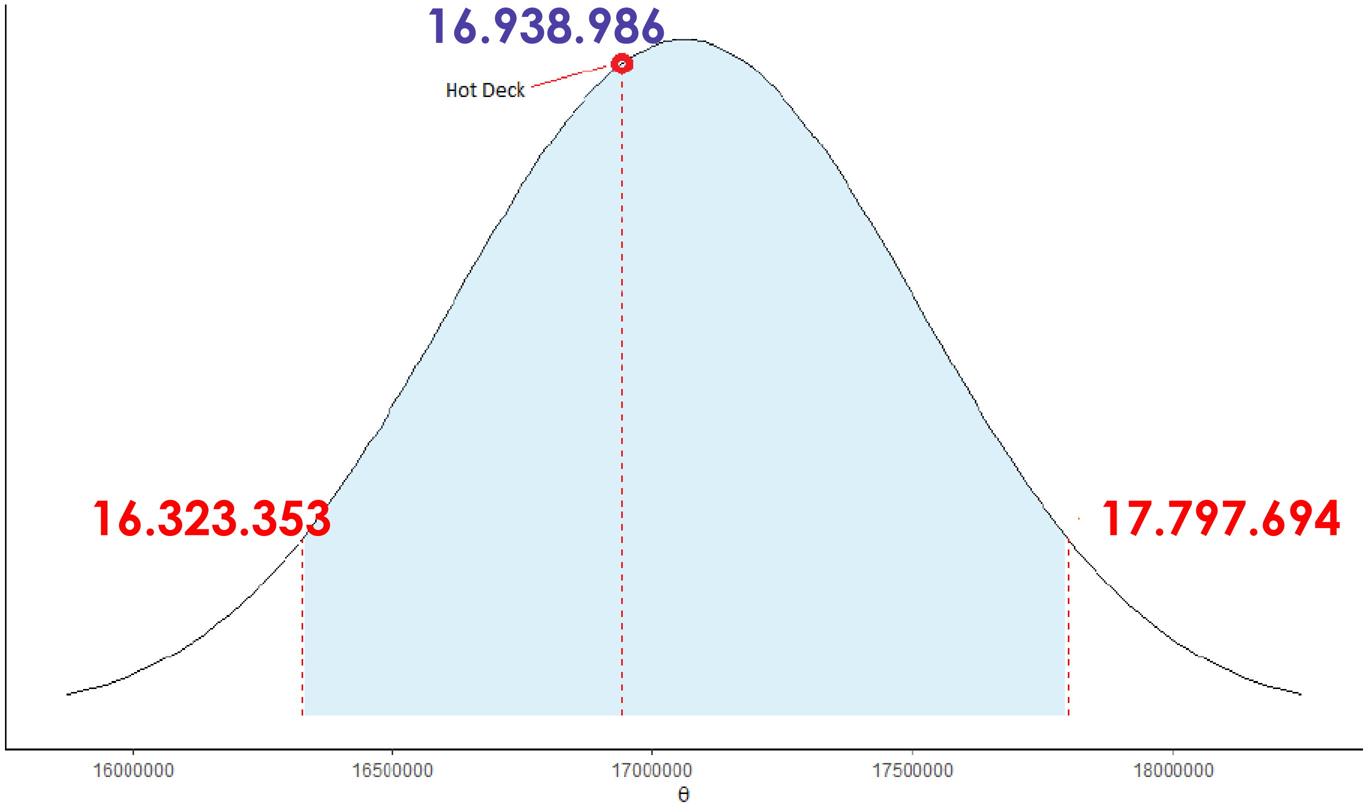
Estimates the likelihood of the count data belonging to a certain group or subgroup

It allows for dissagregations

Robustness check of noninterviews estimation

Bayessian model's predicted value is very **similar** to the **hotdeck** estimation, with a wide credibility interval

Posterior distribution of estimated population



| Fuente: CEPAL (2023).

Estimated population

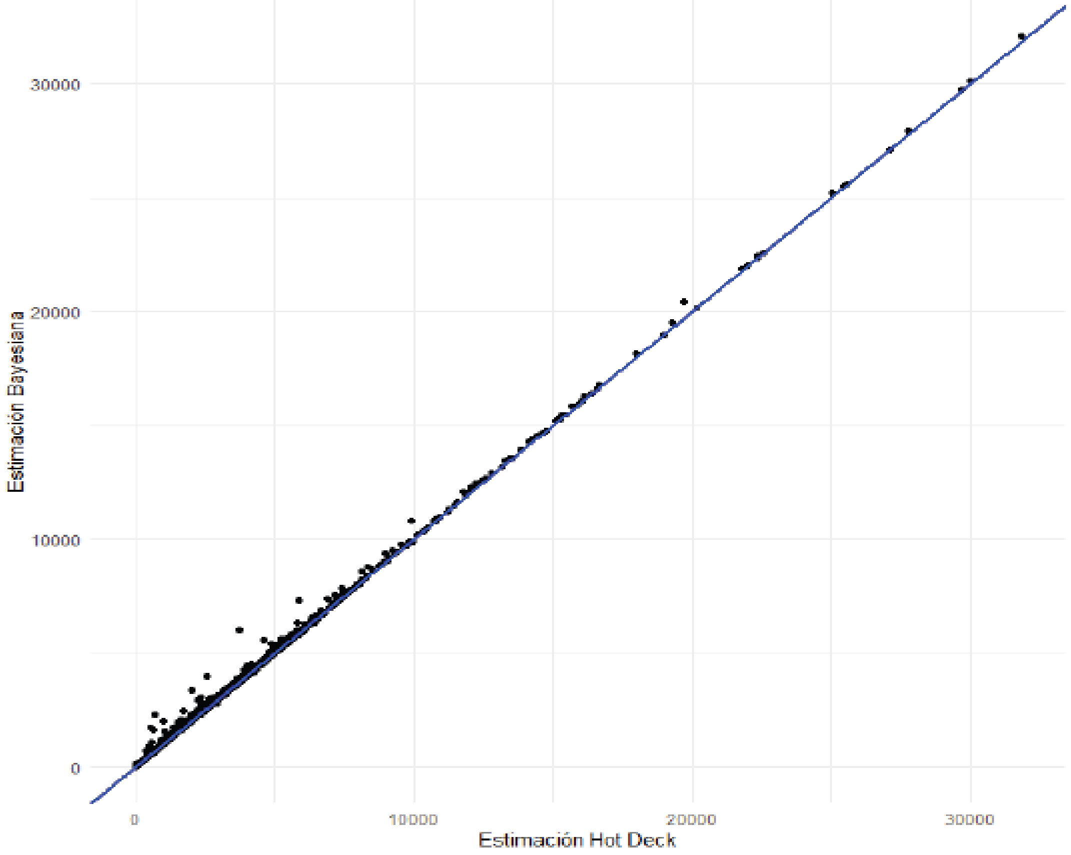
Total de personas	Error estándar	Límite inferior	Límite superior
17.060.523	449.494	16.323.353	17.797.694

Robustness check of noninterviews estimation

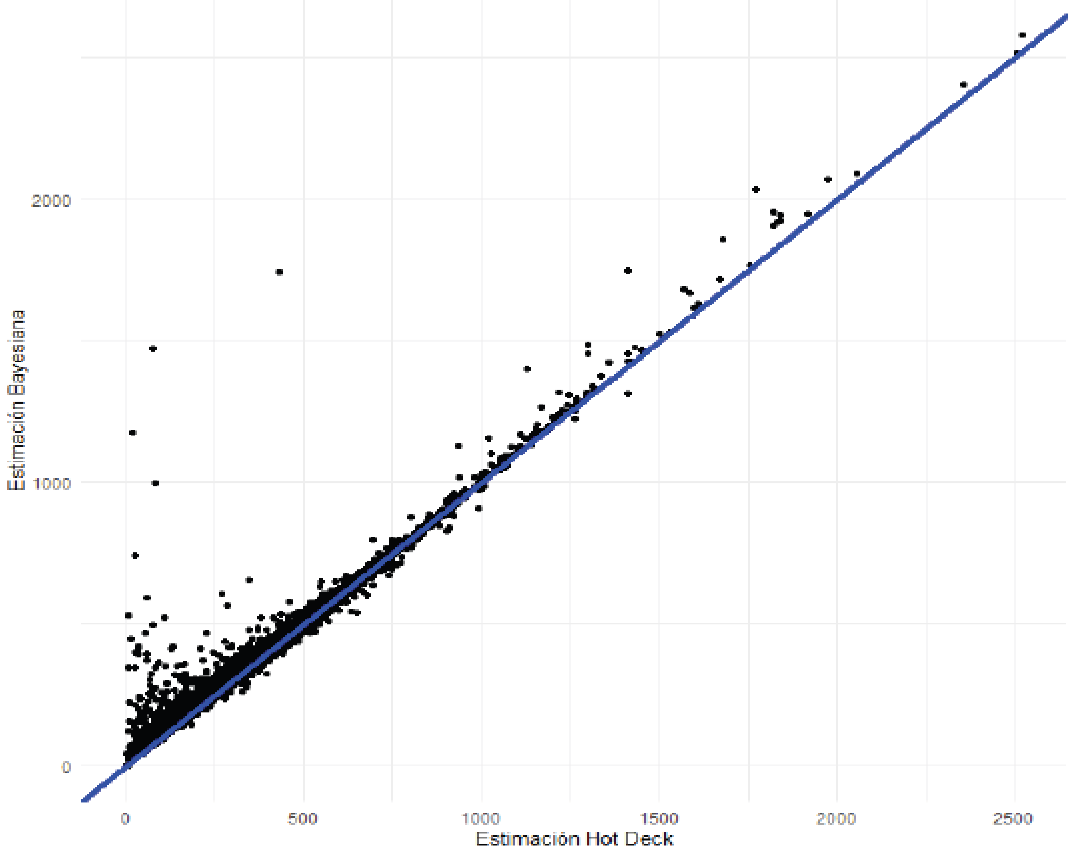
Buenas cifras, mejores vidas

At a sector level (52.932 sectors) both estimations are very close. We observe a slight tendency to overestimate through the bayessian method in very few sectors

Population by zone



Population by sector



Conclusions

- The pandemic and the insecurity crisis meant that careful and strategic planning had to be perform in order to preserve the integrity and quality of the census data. On the other hand, structural difficulties such as nonresponse, noninterviews and omission defects have to be treated after enumeration is done and relying on auxiliary data and analitical techniques.

- Precision procedures in administrative records allow for population recuperation, and proven statistical techniques such as hotdeck imputation used by several countries are a Good approach to tackle noninterviews.

- Population modelling had good results in Ecuador 2022 census data matching hatdeck imputation and helped to check robustness of counted population with noninterviews.

INEC Buenas cifras, **mejores vidas**

@InecEcuador

@ecuadorencifras

@ecuadorencifras

INECEcuador