



DANE

Administrative and survey data integration for work statistics:

Labor market indicators and poverty in Colombia

César López

**Director, Directorate of Methodology and Statistical Production
(DIMPE) National Administrative Department of Statistics (DANE), Colombia**

cmlopeza@dane.gov.co

Content



- 1. Great Integrated Household Survey (GEIH)**
- 2. Conflict-affected population and the Single Registry of Victims (RUV)**
- 3. Integration of the Victims' Registry with the Household Survey and the Production of Labor Market Indicators**
- 4. Integration of records for the measurement of monetary poverty**
- 5. Statistical Base Register Integration System (SIReB)* – DANE**

A map of the Republic of the Congo, colored in a solid light blue. The map shows the country's irregular coastline and its internal administrative divisions, which are outlined in thin black lines. To the left of the map, there are two small black dots, likely representing the country's position on a world map.

- Coverage. Nationally representative coverage comprising:**

- 32 departmental capital cities and metropolitan areas
- 5 secondary cities
- 23 out of 32 departments

Great Integrated Household Survey (GEIH)

Observation and analysis units:



Dwellings



Households



Individuals

Reference framework. For the labor market, the framework is based on the guidelines and recommendations of the International Labour Organization (ILO) and the International Conference of Labour Statisticians (ICLS).

- Data collection is carried out every week of the year
- Results are published on a monthly basis
- Indicators available on a monthly, quarterly, and annual basis, depending on the study domain

Context of the conflict-affected population

53 million – total national population

10 million – conflict-affected population

 **8.9 M** Forcibly displaced persons (main cause)

 **7.8 M** Individuals benefiting from State support

 **50.2%** Women

 **49.8%** Men

 **13.1%** Afro-Colombians

 **6.8%** Indigenous peoples

Challenges for field data collection

Despite being a subpopulation with high prevalence in the country, their identification for the calculation of labor force indicators presents several challenges:

- Fear of acknowledging their condition as conflict-affected persons
- Expectation of receiving monetary assistance from the State
- Risk of re-experiencing trauma / revictimization





Single Registry of Conflict-Affected Persons (RUV)

Integration of the Victims Registry
with household survey data



Single Registry of Conflict-Affected Persons (RUV)



The RUV serves to identify and maintain records of individuals affected by acts of violence in the context of the armed conflict, in accordance with the Victims and Land Restitution Law (Law 1448 of 2011)

Variables in the registry

- Identification variables (ID codes, type and number of document, name)
- Sex
- Geographic location
- Sociodemographic information
- Conflict-related events experienced

Registry issues

- Duplicates
- Errors in type and number of documents
- Inconsistencies in dates of birth
- Typographical errors



1- Initial loading and diagnostic

Databases are imported in their original formats, variable names are standardized, and a uniform encoding (UTF-8) is applied. This ensures consistency in data reading and enables the early detection of missing values, duplicates, or inconsistencies.

2- Harmonization and standardization of key variables

Key variables such as sex and identification number are harmonized by correcting errors, removing leading zeros, and eliminating fictitious records.

3- Integration validation with the household survey

Duplicates are detected using control keys, and phonetic techniques (e.g., Soundex) are applied to identify name variations. This process strengthens database integrity and reduces errors in matching with the household survey



4- Multi-year consolidation

Historical data are integrated with current data to track recurring cases and facilitate trend analysis. This enriches the overall perspective and strengthens the detection of anomalies across different years.

5- Integration with the population base registry

The population base registry enables validation of beneficiaries' identities against the official health database. Its integration prevents duplications and enhances traceability, ensuring more accurate identification in data linkages.

6- Exact linkage by identifier

A one-to-one match is carried out using document type and number. When an exact match is found, the record is flagged as linked, representing the most reliable form of matching.



7- Phonetic name linkage

For records not linked through exact identifiers, phonetic algorithms and name normalization are applied. This process recovers cases lost due to typographical errors, expanding the coverage of the integration.

8- Integration with the household survey

The household survey provides the population framework as well as information on income and living conditions. Integrating it with administrative records enriches the dataset and enables more accurate measurement of the impact of government transfers.

9- First direct match with the household survey

A many-to-one (m:1) match is carried out using primary identifiers (type_id, number_id). Matching cases are flagged as linked, forming the initial integration dataset.

10- Second match with an alternative identifier

Auxiliary identifiers created during the cleaning process are used to recover unmatched cases from the first round. This expands coverage while maintaining data quality.

Labor Market of the Conflict-Affected Population

Proposals for adjustment factors



Adjustment of the expansion factor

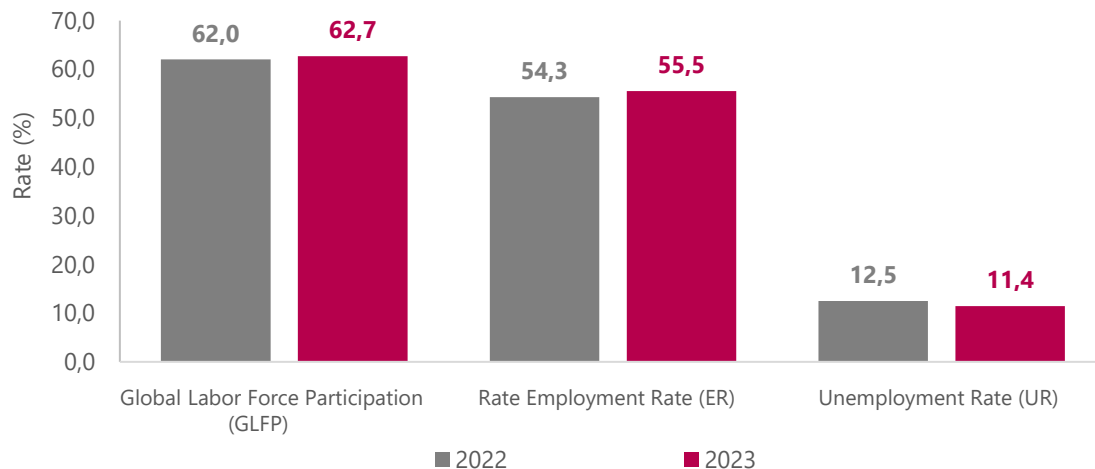
Since linking the information and calculating the expanded totals from the survey showed that the expanded total was approximately 10% lower than the registry, a correction of the expansion factor was applied to estimate the total RUV population.

Quasi random methods (Currently in research)

Quasi-randomization methods are being tested under the assumption that the sample information linked with the registry originates from a non-probability sample. The objective is to derive model-based weights that are doubly robust and capable of reducing potential biases in the estimates

Labor Market of the Conflict-Affected Population

Labor force participation rate (LFPR), employment rate (ER), and unemployment rate (UR) of the conflict-affected population 2022 – 2023



Mercado Laboral de la Población Víctima



¿Dónde está usted?

Conceptos básicos

Información 2022 - 2023

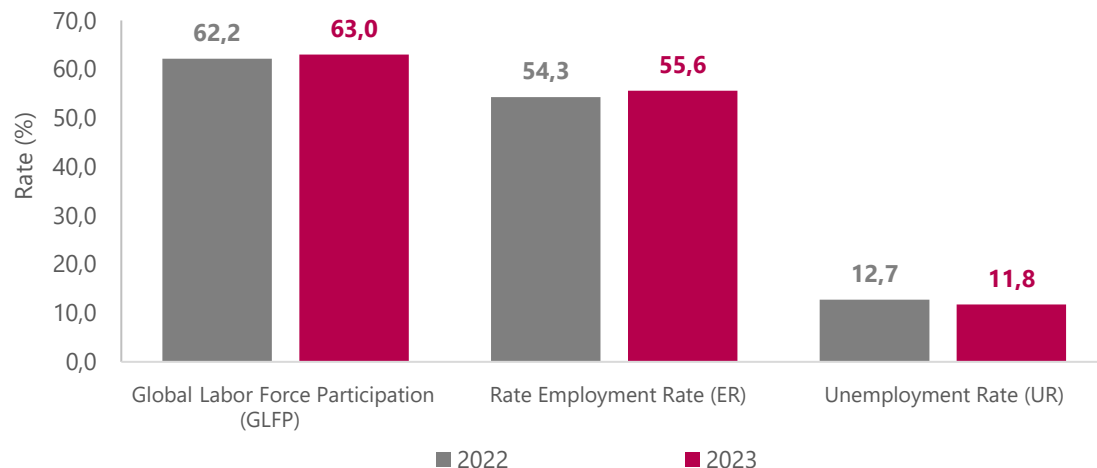
Para el año 2023, para la población víctima se tienen los siguientes indicadores de mercado laboral: la tasa global de participación fue 62,7%, la tasa de ocupación fue 55,5% y la tasa de desocupación fue 11,4%. Para el año 2022 estos indicadores fueron 62,0%, 54,3% y 12,5%, respectivamente.



Boletín técnico

Labor Market of the Conflict-Affected Population

Labor force participation rate (LFPR), employment rate (ER), and unemployment rate (UR) of the **population affected by forced displacement** 2022 – 2023



Source: DANE, Great Integrated Household Survey (GEIH), and Unit for Comprehensive Attention and Reparation to Victims (UARIV), Single Registry of Victims (RUV)

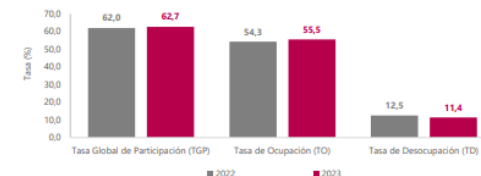


Boletín técnico

Bogotá D.C., 7 de febrero de 2025

Principales indicadores de mercado laboral de la población víctima 2022 - 2023

Gráfico 1. Tasa global de participación (TGP), Tasa de ocupación (TO) y Tasa de desocupación (TD) de la población víctima Total nacional 2022 - 2023



Fuente: DANE, Gran Encuesta Integrada de Hogares (GEIH) y Unidad para la Atención y Reparación Integral a las Víctimas (UARIV), Registro Único de Víctimas (RUV).

- Introducción
- Indicadores de mercado laboral de la población víctima.
- Indicadores de mercado laboral de la población víctima de desplazamiento forzado.
- Indicadores de mercado laboral de la población víctima indígena.
- Indicadores de mercado laboral de la población víctima negra, afrocolombiana, raizal y palenquera.
- Ficha metodológica
- Glosario

CDH-019-PDF-001-0-002-V18



Integration of records for the measurement of monetary poverty



Problem description



- High levels of monetary poverty and extreme poverty.
- Government policy of cash transfers to households as a tool to reduce poverty.
- Limited knowledge and mistrust regarding DANE's role, with some expecting subsidies or government programs, or associating it with tax collection.
- During the pandemic, field operations were adjusted, and administrative records were incorporated to complement survey data.

Integration of data from cash transfer programs in the GEIH



Source	Administrative record	Type of payment	2020	2021	2022	2023	2024
Department for Social Prosperity (DPS)	Familias en Acción (conditional cash transfer program)	Regular payment	X	X	X		
	Tránsito Renta Ciudadana (temporary cash transfer)	Additional payment	X	X		X	
	Renta ciudadana (permanent cash transfer)	Regular payment					X
	Jóvenes en Acción (youth cash transfer)	Regular payment	X	X	X	X	
		Additional payment	X	X			
	Renta Joven (youth basic income)	Regular payment					X
	Colombia Mayor (CM)(pension support for older adults)	Regular payment	X	X	X	X	X
		Additional payment	X	X			
Bogotá City Government	Compensación de IVA (VAT Compensation)	Regular payment	X	X	X	X	X
	Bogotá Solidaria (2020)	One-off payment	X				
	Ingreso Mínimo Garantizado (2021-2024) (basic income))	Regular payment		X	X	X	X
Medellín City Government	Economic Support for Older Adults and People with Disabilities (2020–2024)	Regular payment	X	X	X	X	X

Source: GEIH, DANE – Administrative Records. Institutional support (DPS, SDP, and Medellín Mayor's Office).

Integration and calculation process



- Analysis of the integrity and consistency of the annual registry delivered by the entities responsible for the transfer
- Quality verification
- Consultation with the source of general results and validation of the historical series
- Consultation with the independent poverty expert committee prior to publication
- Publication of results with and without income adjustment

Contribution of cash transfer programs to the reduction of monetary poverty (percentage)

National level, major domains and 23 capital cities
2023 - 2024



To ensure transparency in the dissemination process, results are published both with and without the adjustment resulting from the integration of administrative records.

Domain	With administrative records adjustment		Without administrative records adjustment	
	2023	2024	2023	2024
National	34,6	31,8	37,4	34,1
Urban Areas	31,8	28,6	33,7	30,2
Rural and dispersed population center	44,0	42,5	49,5	47,1
Quibdó	62,0	59,6	63,5	61,2
Riohacha	47,9	48,8	50,7	50,8
Valledupar	51,4	47,5	52,6	48,9
Sincelejo	46,4	46,0	49,0	47,7
Florencia	45,4	43,5	47,3	45,5
Cartagena	42,8	41,1	43,7	42,0
Popayán	41,1	38,8	42,8	40,0
Santa Marta	41,5	37,9	42,8	39,4
Neiva	43,8	36,7	44,9	38,1
Cúcuta	37,7	35,5	39,9	37,0
Montería	36,2	32,1	38,4	34,1
Ibagué	36,9	31,3	37,9	32,4
Pasto	32,1	30,6	33,6	32,1
Bucaramanga	31,5	30,3	32,1	31,0
Barranquilla	29,8	29,7	31,1	30,8
Tunja	31,5	28,7	32,4	29,3
Armenia	32,6	27,8	33,3	28,8
Villavicencio	30,4	25,8	31,6	27,0
Pereira	26,4	24,5	27,3	25,3
Cali	24,3	23,6	25,5	24,1
Medellín	25,3	22,1	25,8	22,6
Bogotá	24,1	19,6	26,5	21,5
Manizales	18,9	16,7	19,8	17,4



Statistical Base Register Integration System (SIReB)* – DANE

Statistical Base Registers and Statistical Business Register

*Temporary name

Transition towards a Register-Based Statistical System

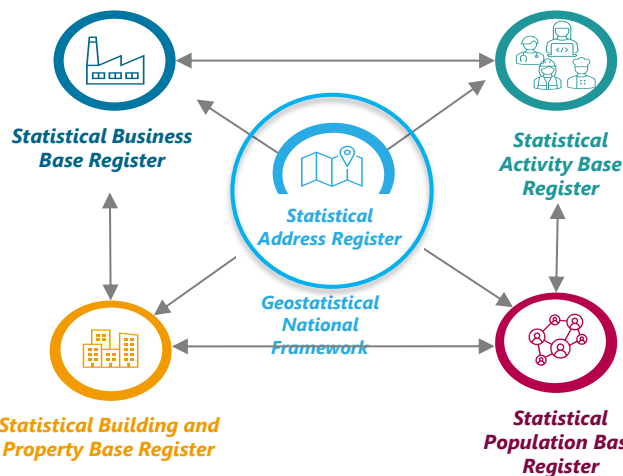


To advance towards a register-based statistical production system, under the principles of system coherence and consistency

Administrative Data Providers Management

- Enterprises
- Establishments and support units
- Enterprise groups

- Paid employment or labor activities
- Study activities and other activities related to the labor market



- Real estate property
- Land use
- Buildings and dwellings

- Persons and demographic events
- Households and dwellings

**Satellite
Statistical
Registers**

**Statistical Register of Labor
Relations – RELAB**

Migration Register – REMI

**Foreign Trade Statistical
Register – RECEX**

**Thematic Statistical
Directories (DESP,
DIREDU, DESA)**

**Ethnic Population
Statistical Register – REPE**

Censuses

Survey

**Alternative Data
Source**

Quality Process aligned with the Statistical Production Process (GSBPM)

Statistical Base Registers



Specify Needs

- Monitoring of the **needs matrix** to identify missing information attributes in the base registers required by internal and external user
- **Engagement of new data providers** to incorporate specialized thematic information

Design

- Review of information sources used in the base registers through the **Input-Oriented Output Quality framework**
- **Peer review** of administrative data sources used for the base registers

Build

- Semi-annual review of ETL processes and techniques to optimize the efficient preparation and structuring of administrative data sources

Collect

- Monitoring of information delivery under the principle of **shared responsibility**
- Tracking of **data delivery agreements** with external providers and issuing alerts on delays



Evaluate

- Internal: Application of **quality metrics** (timeliness, coverage, completeness, and coherence of information) and random audits validating results directly with statistical units
- External: **Feedback on inconsistencies** found in list frames, as part of the shared responsibility framework

Disseminate

- Review of **statistical products** prior to dissemination (documentation, annexes, presentations, and Geovisor functionality)
- Verification of compliance with the **dissemination schedule** for statistical directories and satellite registers

Analyse

- Historical assessment of the frequency of statistical units by thematic attribute in periodic cuts, enabling the identification of outliers and potential processing errors

Process

- Monitoring of inclusion and exclusion criteria for registers according to the data provider and possible changes in thematic precision and coverage
- Monitoring compliance with the integration schedule for sources in the base registers on a semi-annual basis

Conclusions and Further Advances in Data Integration with Household Surveys

- Despite the challenges in integrating the RUV and ensuring its quality, it has proven to be a useful alternative for estimating basic labor market indicators for a hard-to-measure population that requires special attention to guarantee their rights
- DANE already uses statistical registers (RREE) integrated with household surveys; however, their use is still very limited and specific.
- The institution is currently being strengthened through the creation of a new Statistical Registers Unit, which will provide the institutional capacity needed to move forward.
- Agreements are being established with data providers to obtain key registers, along with the analysis of various population and business surveys, with the goal of improving production and reducing the response burden on data sources.



DANE

Administrative and survey data integration for work statistics:

Labor market indicators and poverty in Colombia

César López

Director, Directorate of Methodology and Statistical Production

(DIMPE) National Administrative Department of Statistics (DANE), Colombia

cmlopeza@dane.gov.co