DANE
INFORMACIÓN PARA TODOS

Expert Meeting on Measuring Poverty and Inequality

Poverty Measurement During COVID: Administrative Records as Contrast and Data Source

Table of content

1. Introduction

Since 2012, the national Policy document "CONPES 150" established that the Colombian National Statistics Office (DANE, By its acronym in Spanish) is the institution in charge of estimating and publishing poverty incidence monitoring indicators based on two approaches: monetary poverty and multidimensional poverty. Regarding monetary poverty, CONPES 150 determines that the official source of information will be the Great Integrated Household Survey (GEIH), and for multidimensional poverty the source is the National Quality of Life Survey (ECV).

Due to the COVID-19 pandemic during 2020, DANE had to quickly implement a process of adapting the GEIH to a telephone operation, between March and July, for the 23 main cities and metropolitan areas. This in turn implied reducing the questionnaire that was traditionally used. Consequently, during this period only income sources corresponding to salaries and net earnings were collected, which represent about 65% of the aggregate income of Colombian households.

Considering the above, and that the measurement of income for the calculation of poverty incidence rates includes additional sources of income, between November and December 2020 DANE implemented a telephone recovery operation to re-contact households in the sample from March to July in the 23 cities and metropolitan areas.

Additionally, DANE, with the support of the Committee of Experts on Poverty, developed different analyses of the consistency of the information from the recovery operation that allowed contrasting the different sources of income using administrative records as sources of contrast. From this evaluation, it was found that in the case of the sources of income from government allowances (monetary transfers) and pensions, there was a decrease in the coverage of the income recovery operation in relation to the universe that the GEIH traditionally identifies. This could be associated with a possible memory bias of the respondents.

This led DANE and the Committee of Experts on Poverty to thoroughly evaluate the feasibility of integrating the administrative records of government allowances with the GEIH to improve the measurement of these sources of income.

2. Actions implemented by DANE to ensure the continuous production of the main labor market indicators

Due to the mandatory lockdown and isolation measures implemented during 2020 in the context of the early stages of the COVID-19 pandemic, DANE had to adapt the GEIH statistical operation to a telephone operation between March and April for the 23 main cities and metropolitan areas. In population centers and dispersed rural areas, and in the other head cities (urban domains other than the 23 cities), the GEIH was still carried out in person with a complete form.

a. Length of the questionnaire (March-April):

- The questionnaire was reduced from 200 to 39 questions to guarantee the regular production of labor market indicators.
- The questions associated with wages and net earnings, which represent approximately 65% of the aggregate income of Colombian households, were maintained in this form.
- As stated in DANE (2020) the reduced form included questions on age, sex, labor classification, and characteristics of the employed such as economic activity, occupation performed, income from the main activity, occupational position and hours worked (usual and effective). In addition, it retained the migration and identification module.

b. Sample design:

Likewise, DANE (2020) clarifies that the sample for the GEIH telephone and face-to-face operation is the same, and that the sample design did not change the selection procedure. That is, the blocks defined to collect the information in the original design were not affected by the modification of the collection method for the 23 cities and metropolitan areas between March and July 2020.

c. Income data recovery operation:

Given that between March and July 2020 the GEIH questionnaire was reduced, in order to guarantee the implementation of the telephone operation in the 23 main cities and metropolitan areas, between November and December 2020 DANE carried out a complementary telephone operation to recover the information of the variables associated with the following sources of income: First Activity Monetary Income (IMPA) other than wages and net earnings, Second Activity Income (ISA), In-kind income (IE), Monetary Income of the Unemployed and Inactive (IMDI), Interest and dividends (IOF1), pensions and retirements (IOF2), food and household assistance (IOF3H), government allowances (IOF3I) and leases.

d. Recovery operation questionnaire

- For the implementation of the income information recovery operation, the pollsters used a web application that included all the questions of the missing income sources for the period March-July 2020 for the 23 main cities and metropolitan areas.
- The income recovery questionnaire also included questions on housing and household characteristics that are necessary for the implementation of the income imputation procedure.

Table 1. Income Information Recovery Operation Form

23 cities and Metropolitan Areas - March to July 2020

Chapter	Question	Month March	April	May	June	July
Dwelling Data	What is the predominant material of the external walls of the house?	✔	✔	✔	✔	✔
	What is the predominant material of the floor of the house?	✔	✔	✔	✔	✔
	Which of the following services does the dwelling have?	✔	✔	✔	✔	✔
Household Data	Including living-dining room, how many rooms does this home have?	✔	✔	✔	✔	✔
	The sanitary service used by the household is:	✔	✔	✔	✔	✔
	How do you mainly dispose of garbage in this household?	✔	✔	✔	✔	✔
	Where this household mainly obtains water for...?	✔	✔	✔	✔	✔
	In which of the following places do the people in this household prepare their food?	✔	✔	✔	✔	✔
	What energy or fuel do you mainly cook with in this household?	✔	✔	✔	✔	✔
	The dwelling occupied by this household is:	✔	✔	✔	✔	✔
	How much is the monthly amortization payment?	✔	✔	✔	✔	✔
	If you were to pay rent for this home, how much would you estimate you would have to pay monthly?	✔	✔	✔	✔	✔
General Characteristics (for all household members)	Sex	⚠	⚠	⚠	⚠	⚠
	Which is the birthday for...?	⚠	⚠	⚠	⚠	⚠
	How old are you?	⚠	⚠	⚠	⚠	⚠
	What is the relationship of ... to the head of household?	✔	✔	✘	✘	✘
	Are you a member, contributor or beneficiary of any social security health institution?	✔	✔	✘	✘	✘
	To which of the following social security health regimes are you registered?	✔	✔	✘	✘	✘
Education (3 years of age or older)	Do you know how to read and write?	✔	✔	✘	✘	✘
	Are you currently enrolled to a preschool, school, or univertisy?	✔	✔	✘	✘	✘
	Which is the highest educational level achieved by ... and the last year or grade achieved in this level?	✔	✔	✘	✘	✘
Employed (12 years of age or older)	For how long have ... been continuosly working at this enterprise, business, industry, office, firm or farm?	✔	✔	✔	✔	✔
	Before discounts, how much did you earn last month in this job?	⚠	⚠	⚠	⚠	⚠
	Last month did you receive overtime income? how much did you receive? did you include this amount in last month's income that you previously reported?	✔	✔	✔	✔	✔
	In addition to your salary in cash, last month did you receive food as part payment for your work? how much do you estimate you received?	✔	✔	✔	✔	✔
	In addition to your salary in cash, last month did you receive housing as part payment for your work? how much do you estimate you received?	✔	✔	✔	✔	✔
	Does ... normally use company transportation to commute to work (bus or car)? how much do you estimate he/she received?	✔	✔	✔	✔	✔
	In addition to your salary in cash, last month did you receive other income in kind as part payment for your work (appliances, clothing, non-food products such as bonds?)? how much do	✔	✔	✔	✔	✔
	which of the following allowances did ... receive last month? how much? did you include this value in last month's income that you reported above (food, family transportation, or education)?	✔	✔	✔	✔	✔
	In addition to your salary in cash, last month did you receive bonuses as part payment for your work? how much? did you include this value in last month's income that you reported above?	✔	✔	✔	✔	✔
	in the last 12 months did you receive service bonus, christmas bonus, vacation bonus, permanent per diem, work bonuses, work accident payments? How much?	✔	✔	✔	✔	✔
	What was ...'s net profit or net fees from that activity, business, profession, or farm last month?	⚠	⚠	⚠	⚠	⚠
	How many months does the amount you received correspond to?	✔	✔	✔	✔	✔
	How many people does the company, business, industry, office, firm, farm or site where ... works have in total?	✔	✔	✘	✘	✘
	Is ... currently contributing to a pension fund?	✔	✔	✘	✘	✘
	In addition to ...'s main occupation, did he have another job last week?	✔	✔	✔	✔	✔
	How many hours does ... worked in that second job last week?	✔	✔	✔	✔	✔
	In that second job ... is:	✔	✔	✔	✔	✔
	How much did ... receive or earn on that second job?	✔	✔	✔	✔	✔
Unemployed (12 years of age or older)	Has ... been looking for a job for the first time or has ... worked before for at least two consecutive weeks?	✔	✔	✘	✘	✘
	Did ... recieved or earned any income for working last month? How much?	✔	✔	✘	✘	✘
Inactive (12 years of age or	Did ... recieved or earned any income for working last month? How much?	✔	✔	✘	✘	✘
Other income (12 years of age or older)	Last month did you receive pension and/or retirement income?	✔	✔	✘	✘	✘
	Last month did you receive income from rents, retirement, disability or personal substitution pensions, alimony? Last month's amount:	✔	✔	✘	✘	✘
	during the last 12 months, did you receive money from other households, persons or institutions; money for interest, dividends, profits or severance pay?	✔	✔	✘	✘	✘
	during the last 12 months, did you receive money from other households, persons or institutions; money for interest, dividends, profits or severance pay? Last month's amount:	✔	✔	✘	✘	✘

According to the literature (Ayhan & Işiksal, 2005) retrospective surveys may face phenomena of omission, under- or over-reporting in questions related to income due to possible biases that may derive from human memory. For this reason, the use of administrative records, as a source of contrast, provides a reference point to evaluate the accuracy and coverage of the information collected in the income recovery operation.

Administrative records can be used to evaluate these measurement biases either by comparing the microdata and aggregate amounts observed in the surveys and those reported in the administrative data (ECLAC, 2020; NATIONS U., 2012). For the 2020 exercise, DANE used administrative records as a source of contrast and imputation and used 2019 as a counterfactual to assess the relevance of using them. Therefore, for 2019 and 2020, a coverage analysis was performed, which consisted of contrasting the value observed in the survey with the administrative record value; and performing a reporting bias analysis.

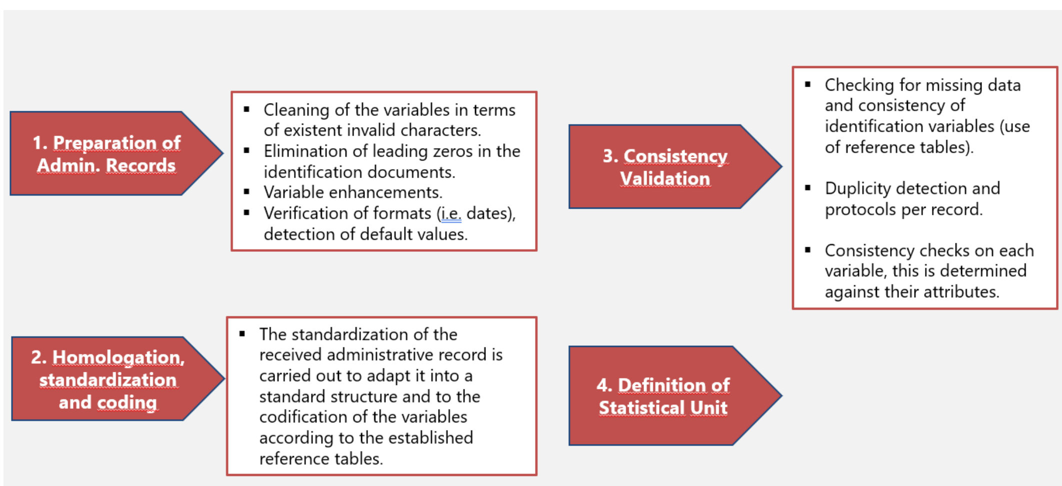
This procedure was performed for income sources IOF2 (pensions) and IOF3I (government allowances). In the first case, the amount of pension allowance reported in the Integrated Social Security Form (PILA) was used, while in the second case the monthly amount of total collections observed in the portability zone of each individual during the 2020 period was used.

3. Administrative Records Management

Data from administrative records per se do not serve a statistical purpose (United Nations 2007; Wallgren and Wallgren 2007; Zhang 2011), so before performing any direct imputation it is necessary to solve challenges such as quality in integration keys, homologation of administrative units to statistics, and measuring the accuracy of indicators after matching between datasets.

Figure 1. Management of Administrative Records

Input Protocol



Source: DANE, 2021

Figure 1 describes the process of preparing the administrative records, where step 1 consists of reviewing the integration keys. Within this analysis, the type and number of the identity document are evaluated, initial zeros and special characters are corrected, as well as detection of default values.

Finally, the Base Statistical Registry of Persons (REBP)[1] is used as a reference table to validate the suitability of the number and the consistency of the names associated with it[2].

In the following steps, we proceed to standardize, validate the consistency of the variables, determine the reference periods to be used in the registry and transform each record into a statistical record.

In the case of the IOF2, the administrative record PILA is the source used to provide information on pension payments in 2020. The following considerations are made:

- Payment transactions are included up to twelve months after the first day of contribution of the contribution periods other than the health subsystem or up to the payment transactions made in February 2021.
- The variable derived from the pension allowance per person was constructed, which results from the total sum of the pension allowance field that a contributor may have in a period according to the administrative record and did not exclude records based on the report of new entries. Thus, the pension payments corresponding to different types of pensions for the same contributor were aggregated in the total per period. The zero-pension type is not part of the variable domain according to Technical Annex 3 of Resolution 2388 of 2016 updated to December 21, 2020. Consequently, it was not considered in the statistics of pension payments.

On the other hand, for the records that make up the IOF3I variable related to government allowances, the total amounts received by the beneficiaries included in the administrative records of: Más Familias en Acción (FA)[3], Jóvenes en Acción (JA)[4], Colombia Mayor (CM)[5] and Otras Ayudas del Gobierno (Other Government Allowances) are used. It should be clarified that the last-mentioned item includes aid initiated in 2020 such as Ingreso Solidario[6], VAT compensation, and programs of the Bogotá, Medellín and Bucaramanga municipalities. In this case, the following aspects are taken into consideration:

- By using the administrative records for each institutional aid, it is possible to estimate the total amount received during the reference year and to identify the portability zone (i.e. permanence within the programs).

[1] Statistical Base Population Register (BSPR): It is created from the transformation of one or several administrative registers, such as the Civil Status National Registry, so that the objects and variables are adjusted to meet statistical needs (Wallgren & Wallgren, 2014). It is also possible to create a statistical record from processing administrative records together with other statistical records (Decision 780 CAN. 2013. Pg. 2).

[2] Another check that falls within the methodological framework is the identification of duplicates and inconsistencies.

[3] Familias en Acción is a program that supports the population living in poverty and extreme poverty, through the conditional delivery of health and education incentives.

[4] This is a national government program that supports young people in their technical, technological and/or professional training by providing conditional cash transfers to improve their capabilities and opportunities for social mobility and welfare conditions.

[5] The objective of "Colombia Mayor", is to increase protection for the elderly who are abandoned, do not have a pension, or live in extreme poverty or destitution, through the delivery of a monthly economic subsidy.

[6] The Ingreso Solidario Program is a cash transfer support provided by the National Government to households living in poverty, extreme poverty and economic vulnerability to mitigate the impact of the emergency caused by COVID-19.

- Given the existence of accumulated payments, retroactive payments and dynamics of entry and exit of new aid, the monthly amount of total collections observed in the portability zone of each individual during the 2020 period is used for each of the government's institutional aid administrative records.
- In 2020 within the framework of the Covid-19 pandemic, aids such as extraordinary payments of CM, FA, JA, VAT compensation, Ingreso Solidario, Bogotá Solidaria, Bono Vital de Bucaramanga, and Medellín monetary aids were initiated[7]. Most of them started in March 2020. Those aids referred to as ordinary payments are recurrent transfers systematically captured by the survey.
- In the case of the extraordinary payments of the CM, JA, FA programs, they are added to the programs' own payments, while the other aids that arise since 2020 are included in the category of other aids.

Following the standardization, and prior to the process of variable substitution, it is important to note that, although the GEIH does not ask in a mandatory way the fields of type and number of identity document since 2018, thus the completeness and consistency (evaluated with the REBP) of these fields allows linking the survey with different datasets. However, and due to the fact that there are still respondents who do not disclose their ID or that the number reported presents quality findings, these are included in the analysis set to avoid under-reporting of benefits, or in the case of pensions where although the PILA has a broad coverage of pensions it does not include some special regimes such as the military.

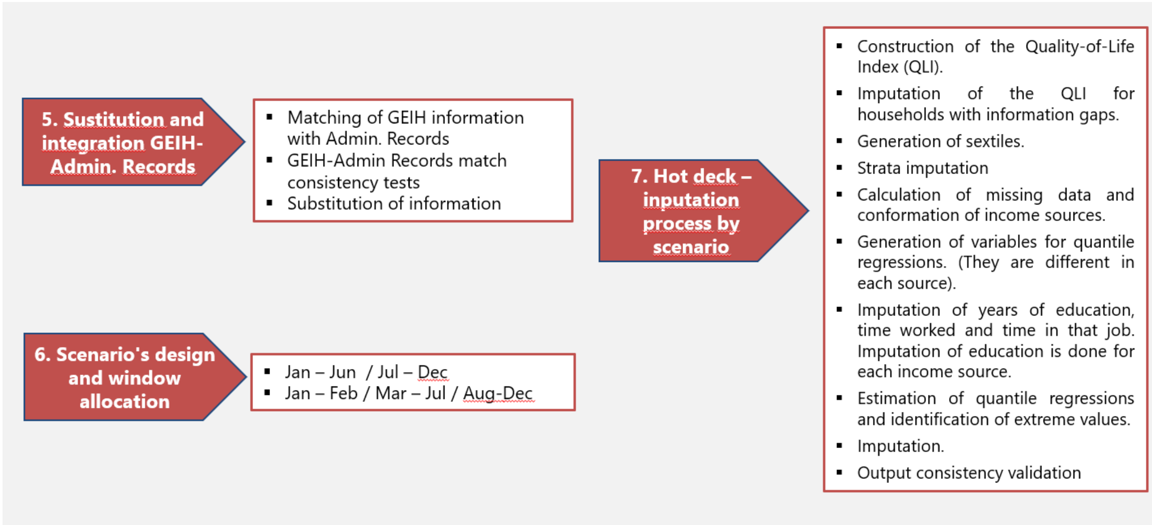
Based on this treatment, in step 5 included in Figure 2, and with the purpose of enriching the information collected by the GEIH in its questions P750S2A1 (Amount received for government allowances) and P7500s2a1 (Amount received last month for pensions or retirement due to old age, disability or substitution) is linked to the information collected by the different administrative records[8]. It is important to note that the item "Other government allowances " in the GEIH stores information on other programs in addition to those mentioned above, so the higher value observed between the data sets is preserved.

Figure 2. Management of Administrative Records

Output Protocol

[7] It is important to note that in the income measurement methodology defined by the Mission for the Splicing of the Employment, Poverty and Inequality Series (Mesep), in-kind assistance is not part of current household income.

[8] It is important to highlight that in the case of Familias en Acción, the program holders and beneficiaries were identified. This made it possible to improve the match between the survey and the Administrative Records.



Source: DANE, 2021

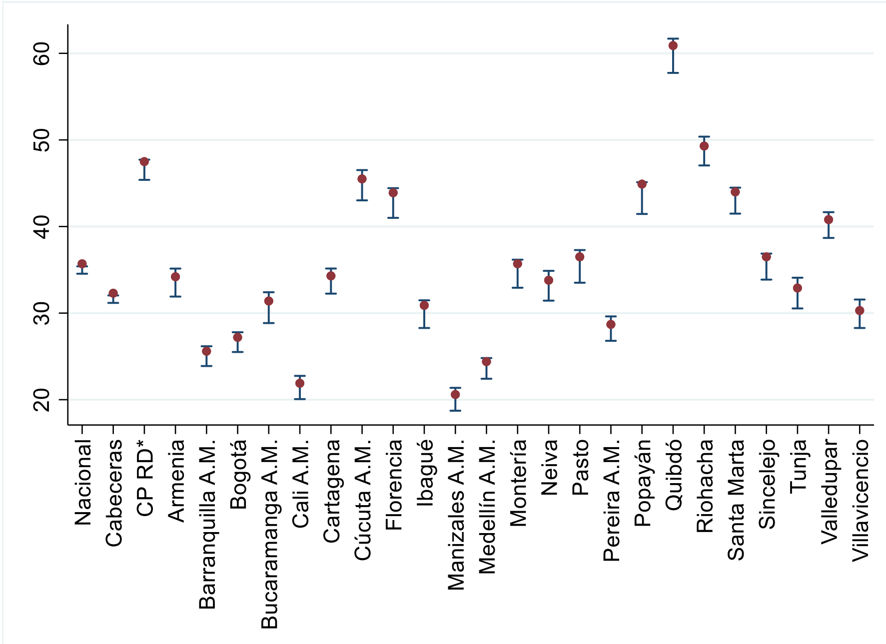
Additionally, DANE and the Committee of Experts on Poverty determined that the partition of the sample for the imputation of income for 2020 should be done respecting the nature of the data collection operations. In other words, as shown in point 6 of Figure 2, the imputation process was developed by organizing a partition of the sample into three blocks: January-February, March-July and August-December. Finally, an income imputation procedure is performed to correct for missing data. The income imputation methodology consists of a procedure for detecting missing and extreme values that are subsequently imputed following the process described in step 7 of figure 2.

4. Consistency Result Validation

Given that the measurement of income poverty in Colombia has always been done with information from household surveys and considering that for the construction of the aggregate household income variable in 2020 information from administrative records was also used, DANE estimated income poverty and extreme poverty for 2019 by integrating administrative records with the GEIH for that year. The results show that the published monetary poverty estimates are within the confidence intervals of the figures calculated with GEIH and administrative records for 2019. (See Figure 3)

Figure 3. Published incidence of monetary poverty

Confidence intervals estimated from GEIH-2019 and Administrative Records.

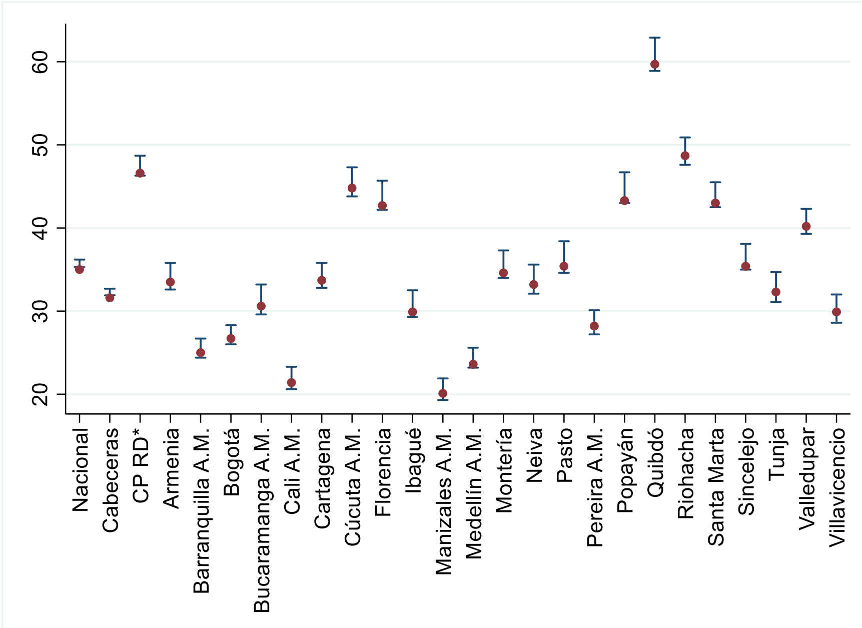


Source: DANE, match GEIH – Familias en Acción, Jóvenes en Acción y Colombia Mayor, 2019.

Likewise, Figure 4 shows that the 2019 monetary poverty incidences estimated from the GEIH and administrative data are located within the published confidence intervals for the 2019 figures (i.e., intervals resulting from the official 2019 estimate without the use of administrative records).

Figure 4. Incidence of income poverty 2019 GEIH-RRAA

Published confidence intervals.



Source: DANE, match GEIH – Familias en Acción, Jóvenes en Acción y Colombia Mayor, 2019.

This suggests that the comparability among the poverty series is not affected by the use of administrative records. However, it is important to highlight that the integration of the Great Integrated Household Survey (GEIH) with administrative records is a methodological procedure that allowed us to address reporting biases in some income sources, and consequently improved the measurement of the income variable for 2020. Additionally, the implementation of a complementary telephone operation to recover the missing income information for the March-July period, for the 23 cities and metropolitan areas, allowed to guarantee the permanence of the methodology for the construction of the aggregate income of Colombian households.

References

- Ayhan, H. Ö., & Işiksal, S. (2005). Memory recall errors in retrospective surveys: A reverse record check study. Quality and Quantity, 38(5), 475-493.
- CAN (2013), Decisión 780.
- DANE (2020), Comunicado de Prensa – Acciones para garantizar la continuidad de la Gran Encuesta Integrada de Hogares (GEIH), Bogotá D.C.
- DANE (2021), presentación rueda de prensa resultados de pobreza monetaria 2020
- DANE y DNP (2012), Misión para el Empalme de las Series de Empleo, Pobreza y Desigualdad (Mesep), Bogotá D.C.
- DNP (2012), documento CONPES Social, Metodologías oficiales y arreglos institucionales para la medición de la pobreza en Colombia
- Ministerio de Salud y Protección Social. Anexo técnico 3 de Resolución 2388 de 2016. Actualizado a 21 de diciembre de 2020
- NATIONS, U. (2012). Canberra Group Handbook on Household Income Statistics.
- CEPAL, N. (2020). Continuity of household surveys after the coronavirus disease (COVID-19) pandemic.
- United Nations. Commission économique pour l'Europe, & United Nations. Economic Commission for Europe. (2007). Register-based statistics in the Nordic countries: review of best practices with focus on population and social statistics.
- Wallgren, A., & Wallgren, B. (2007). Register-based statistics: administrative data for statistical purposes (Vol. 553). John Wiley & Sons.
- Wallgren, A., & Wallgren, B. (2014). Register-based statistics: statistical methods for administrative data. John Wiley & Sons.
- Zhang, L. C. (2011, August). Topics of statistical theory for register-based statistics. In ISI conference, Dublin (pp. 22-26).