Universidad del Rosario | Facultad de Economía

SERIE DOCUMENTOS DE TRABAJO

No. 285

Septiembre de 2022

Impact of the collection mode on labor income data. A study in the times of COVID19

Andrés García-Suaza

José Lobo

Sergio Montoya

Juan Ordóñez-Herrera

Juan D. Oviedo-Arango

Universidad del Rosario, Facultad de Economía
Cl. 12c #4-59. Bogotá, D. C., Colombia
PBX: (031) 2970200 ext. 4163

facultadeconomia@urosario.edu.co
https://www.urosario.edu.co/Facultad-de-Economia/Inicio/
https://repository.urosario.edu.co/handle/10336/20291

Impact of the collection mode on labor income data. A study in the times of COVID-19

ANDRÉS GARCIA-SUAZA[*], JOSÉ LOBO[†], SERGIO MONTOYA[‡], JUAN ORDÓÑEZ-HERRERA[§], JUAN D. OVIEDO-ARANGO[**]

ABSTRACT

The strict confinement implemented by the National Government of Colombia to contain the expansion of the pandemic caused by COVID-19 generated challenges in data collection operations through household surveys. As a result, the surveys with face-to-face collection methods migrated to a remote mode, through telephone surveys, which could have changed the possible reporting biases of variables, such as income. This paper studies the effect of the change in the information collection model in the Great Integrated Household Survey (*Gran Encuesta Integrada de Hogares*) of Colombia on the report of labor income. To do this, we exploit the geographical variation in implementing collection methods and an integration of the survey with a social security administrative record to quantify the variation on the report.

Keywords: Household surveys, measurement bias, labor income, administrative data, COVID-19, Colombia.

JEL codes: C83, C81; J31.

---

[*] Assistant Professor at Universidad del Rosario. Email: andres.garcia@urosario.edu.co.
[†] Economist at Superintendencia Nacional de Salud. Email: jlobo@unal.edu.co.
[‡] Research officer at OECD. Email: sergio.montoya@oecd.org
[§] Advisor, DANE. Email: jsordonezh@dane.gov.co.
[**] Ex-Director DANE and Professor at Universidad del Rosario. Email: jdoviedo@gmail.com.
The authors thank the National Administrative Department of Statistics (DANE) since this work arises from the need to understand the changes in the collection operations of the household survey in 2020. Authors thank comments and suggestion by Jhon Quinchua. The results and possible errors that may occur are the authors' responsibility, and they do not commit their institutions of affiliation. This work is part of the Alianza EFI-Colombia Científica research ecosystem projects under contract code 60185 and FP44842-220-2018.

1 Introduction

During the second quarter of 2020, governments around the world imposed confinement and social distancing measures to contain the spread of the pandemic caused by the SARS-CoV-2 coronavirus. This situation brought important challenges to the National Institutes of Statistics (NISs) to monitor the main economic and social variables, one of the main inputs for decision-making in public policies. In this context, continuous statistical operations usually collected monthly through face-to-face interviews, such as household surveys, were affected. They had to be redesigned to maintain the data collection process. The United Nations Statistical Division (UNSD, 2020) documented that 96% of NISs ceased collecting data in person, either partially or totally.

Initially, the restrictions on implementing face-to-face interviews made it necessary to conduct the surveys by telephone through Computer-Assisted Telephone Interviews (CATI). Also, to reduce the length of the questionnaires, in line with the United Nations (UN) and the International Labor Organization (ILO) recommendations,[1] whose purpose was to maintain the data quality. This change in the data collection process was temporary, as the severity of the confinement and social distancing policies progressively decreased. However, although the NISs normalized their operations, it is necessary to evaluate the impact of those changes, particularly on labor market indicators (UNSD, 2020).

The income variable is probably one of the most relevant data collected through household surveys since it provides information on the socioeconomic situation and the standard of living of

[1] For example, the ILO (2020) produced a series of working papers called COVID-19: Guidance for labor statistics data collection, which discusses key elements for conducting rapid surveys; shorter questionnaires containing essential information are defined, and tips are provided for capturing relevant data during the COVID-19 crisis, such as the prevalence of remote work and telecommuting, inactivity, joblessness, and job loss, among others.

households and serves as an input for the measurement of poverty (Glewwe, 2007; Lohmann, 2011; Burton et al., 2020). Some studies recognize that measurement errors associated with income data collection affect measures of unemployment, inequality and poverty (van Praag et al., 1983; Ravallion, 1988; Chesher & Schluter, 2002; Figari et al., 2012; Angel et al., 2018; Fessler et al., 2018; Ward & Edwards, 2021). For this reason, it is crucial to understand the impact of the data collection process on revenue measurement.

The National Administrative Department of Statistics of Colombia (DANE) implemented mixed collection methods according to its operating regions. For example, during strict lockdowns (March to July 2020), they kept doing face-to-face interviews in less urbanized and rural areas, while data in major cities and metropolitan areas was collected remotely through CATI. The nature of this distinction constitutes a natural experiment to understand the impact generated by the change in the collection mode. An element that will serve as a fundamental input for this work is the Statistical Registry of Labor Relations developed by DANE to monitor the effects on the labor market. This source was built by integrating social security administrative records that were not altered by the changes in the collection method of survey collection and therefore served as a counterfactual for the measurement of possible biases because of the change in the collection method. Motivated by the above, this paper aims to study the impact of the temporary change in collecting data from the Colombian household survey, Great Integrated Household Survey (GEIH, for its Spanish acronym), in the measurement of income in the corresponding period from May to July 2020.

Similarly, it is pertinent to remember the study of biases in the measurement of income is a crucial instrument to validate the comparability of the relevant variables collected in a household, as well as to determine methodological designs that combine different interview methods to reduce operating costs. In this sense, this work proposes exploiting the temporal and spatial variation in the data collection process and integrating administrative data into the survey to quantify the bias in self-reported income. Specifically, we will study whether income measurement is sensitive to implementing CATI-type collection methods instead of the usual face-to-face mode, also known as Computer-Assisted Personal Interviewing (CAPI). To do this, we will use a quasi-experimental design, particularly event-study-type specifications. They allow calculating whether there is a difference in declared income because of the application of CATI from May to July 2020 in Colombia. The selection is based on the hypothesis that using information from administrative records makes it possible to establish a reference value for labor income and control for possible unobserved heterogeneities associated with the COVID-19 situation. The study also explores possible heterogeneous effects by population groups, such as gender, education, and age groups, and checks the robustness of the identification strategy.

There is extensive literature examining the sources of income measurement errors in household surveys. Angel et al. (2019) present a general framework of the primary sources of error in income measurement, in particular: socially desirable behaviors, socioeconomic characteristics, survey design, and learning effects, which refers to the strategies of the respondents to remember the information. All sources can produce both over-reporting and under-reporting of income. For example, biases have been observed regarding socially desirable behaviors, possibly related to characteristics of households and individuals: people with lower incomes over-report their income to avoid revealing their current economic conditions. In contrast, those with high incomes are inclined to declare lower values. These patterns are known as mean reversion errors (see, e.g., Bound et al., 1994; Moore et al., 2000; Kim & Tamborini, 2014). Other factors associated with the income variable, e.g., the target measurement variable (gross income or disposable income), are the amounts, the wording of the questions, the reference period, and the survey mode, which also affect its measurement (Moore et al., 2000; Canberra Group, 2011; Kuhn, 2019;). For further discussion of measurement error in survey data, see Horowitz and Manski (1995) and Bound et al. (2001).[2]

The existing literature comparing data collection processes based on CAPI and CATI states that these tend to cause a bias in coverage and representativeness. Ellis and Krosnick (1999) and Bowling (2005) show that a CATI-based survey might have a higher representation of higher-income, more educated, and younger individuals (see also Jordan et al., 1980; de Leeuw, 1992). This is related to the lack of telephone services coverage in other population groups. Studies in this line have documented other differences from the operational point of view. CAPI offers some advantages because of the physical and visual contact between the respondent and the interviewer, allowing additional communication tools to guide and create trust and legitimacy for the respondents (see Statistics Canada, 2010).

The literature points out that CAPI is a more efficient method for long questionnaires while reducing the non-response rate (Roberts et al., 2010; Klausch et al., 2013; Economic and Social Research Council [ESRC], 2019), which also could improve the data quality on income. Besides, Canberra Group (2011) provides a general guide to the main considerations in income data collection and argues that face-to-face interviews can produce higher quality data. While CAPI

[2] In practice, the rapid response during the period of strict confinement to enable the implementation of household surveys using CATI required overcoming challenges both in data collection and at the methodological level. Regarding the difficulties in the collection process, the ILO (2020a) maintains that the change from CAPI to CATI requires a comprehensive administrative registry of telephone numbers that allows for preserving the sampling structure, designing protocols to replace non-response, quickly training interviewers, and adjusting the battery of questions in the survey questionnaire. Regarding methodological issues, protocols to correct incomplete data (e.g., correction for expansion factors) and a greater understanding of post-data collection measurement bias are required to preserve comparability over time. Consequently, NISs have applied strategies (e.g., reusing old samples or previously contacting households using introductory letters to encourage participation) to maintain the sampling design (ILO, 2020b).

could over-report socially desirable behaviors, such as lower rates of alcohol consumption or a higher proportion of voters (Holbrook et al., 2003). Jackle et al. (2010), Schräpler et al. (2010), Caeyers et al. (2010), Statistics Canada (2010), and Lynn and Kaminska (2013) conduct extensive studies comparing the impact of using CAPI, CATI, and other methods of interviewing.

Therefore, this paper provides evidence on the measurement biases associated with the mode of collection. For this purpose, we analyze a large-scale household survey and exploit the redesign of the survey in Colombia as an experimental scenario to understand the impact of these two measurement errors. The results obtained suggest that switching to the CATI operation was associated with a decrease in reported income by approximately 4 percentage points (p.p). This result is consistent across different specifications, even using only the household survey. These findings are more pronounced for men than for women, where there is no statistically significant result. It also highlights the significant results among the most educated and individual over 25 years old.

It is important to note that integrating survey data and administrative data allows us to understand the sources of measurement error better and contrast the patterns of mean reversion, as previously analyzed by Bollinger (1998), Bound and Krueger (1991), Kreiner et al. (2013), and Meyer and Mittag (2021), among others. In this way, the results derived from this work are also relevant to designing methodological frameworks that allow for improving survey data based on the use of administrative records (see c.f., Abowd & Stinson, 2013; van der Klaauw, 2014), as well as to expanding the knowledge about the comparability and complementarity of the observation unit of these two information sources (see Kapteyn & Ypma, 2007; Meyer et al., 2015, for a discussion).

We organized this document into six sections. In the second, we describe the changes caused in the collection methods of the household survey during 2020, as well as the use and integration of administrative records and surveys. The third describes the data and the empirical strategy. The fourth presents the main results, and the fifth the robustness checks. Finally, we provide some concluding remarks in the last section.

2 Changes in collection methods in the household survey

Because of the mobility restrictions caused by the COVID-19 coronavirus pandemic in 2020, mainly in urban centers, DANE implemented different measures that made it possible to guarantee the continuity of the production of the primary labor market indicators derived from the GEIH. This adaptation process occurred between March and July 2020. It comprised the implementation of the GEIH through a telephone operation and a questionnaire shorter than usual[3] for the 23 main cities and Metropolitan Areas (MA). A face-to-face process was maintained for the domains known as the Other cities (urban areas of municipal capitals other than the 23 main cities and MAs) and other populated centers and dispersed rural areas (rural areas) with the complete GEIH questionnaire.

The change in the collection method could cause a sample selection[4] associated with the absence of telephone contact information (Economic Commission for Latin America and the Caribbean [ECLAC], 2020). However, DANE made a significant logistical effort that made it possible not to restrict the selection to households that appeared in a telephone directory. They maintained the sampling in a stratified manner and by geographic conglomerates. After the

[3] The GEIH form was cut from 200 to 39 questions between March and April 2020 (Departamento Nacional de Estadísticas [DANE], 2020) and increased to 60 questions from May to July 2020.

[4] DANE (2020) clarifies that the sample selection procedure between March and July 2020 for the 23 main cities and metropolitan areas was not modified.

geographical units were selected, different communication channels were used to get the telephone numbers. If it was impossible to get telephone numbers, they deployed officials in the field to get them. If the household could not be contacted this way, another household from the same geographic cluster was selected. Therefore, it is possible to assume that the sample's representativeness has not changed significantly during the period of application of the CATI operation.

To validate this assumption, Table 1 presents the proportion of women calculated from the expanded samples (i.e., expansion factors) for different periods and geographic disaggregations. Non-significant variations are observed in these statistics, evidencing that there were no problems associated with sampling, which is essential to reduce possible effects that confound the impact on the collection method. In particular, the proportion of women remains stable at around 50.7% at the national level. Across all three survey domains (metropolitan areas, other cities, and rural areas), proportions also remained stable. In this analysis, we also observed that there was no significant variation. People under 25 years old consistently correspond to 42% of the population during the two years analyzed, those between 25 and 54 correspond to 40%, and people over 55 years old correspond to approximately 18% (see Table 2).

Table 1. The proportion of women in the GEIH sample 2019-2021.

Period	Metropolitan areas	Other Cities	Rural	Total
Jan-Mar 2019	51.85	51.34	47.15	50.66
Apr-Jun 2019	51.85	51.34	47.14	50.66
Jul-Sep 2019	51.85	51.34	47.14	50.66
Oct-Dec 2019	51.85	51.34	47.14	50.66
Jan-Feb 2020	51.85	51.34	47.14	50.66
May-Jul 2020	51.84	51.34	47.14	50.66
Aug-Dec 2020	51.84	51.34	47.14	50.66
Jan-Mar 2021	51.83	51.35	47.14	50.66
Apr-Jun 2021	51.83	51.36	47.14	50.66

Source: Authors' calculations. Metropolitan areas refer to the 23 main cities; Other cities are the rest of the urban population.

Table 2. Proportions by age groups in the GEIH 2019-2020 sample

Period	Under 25	25 - 54	Over 54
Jan-Mar 2019	42.49	40.16	17.35
Apr-Jun 2019	42.37	40.17	17.45
Jul-Sep 2019	42.26	40.18	17.56
Oct-Dec 2019	42.14	40.20	17.66
Jan-Feb 2020	42.05	40.21	17.75
Mar-Jul 2020	41.87	40.22	17.91
Aug-Dec 2020	41.72	40.24	18.04
Jan-Mar 2021	41.57	40.25	18.18
Apr-Jun 2021	41.45	40.26	18.29

Source: Authors' calculations.

3 Data and empirical strategy

3.1 Data

Two sources of information are used to estimate the causal impact of the change in the collection method on labor income records: the household survey and a statistical record built from social security payment data. The first source of information comprises the GEIH, which collects the main variables that characterize the Colombian labor market. In 2020, this survey collected average information from 19,152 households per month. It generated representative information for geographic domains of the main 23 cities, their metropolitan areas, and the rest of the country's geographic coverage. The GEIH has a modular structure, in which sociodemographic information is included, as well as variables of employment characteristics, such as the income of the employed. For the proposed analysis, we used the data corresponding to the period from January 2019 to June 2021.

To have a reference measurement of income that the change in the collection method has not affected, the second source of information is from the Statistical Registry of Labor Relations (RELAB, for its Spanish acronym) produced by DANE. In this source, it is possible to observe the employee-employer relationship of the dependents[5] through the payments made to the social security system. The RELAB includes information on a subset of labor relations, which, when integrated with other DANE statistical operations, allows labor relations to be characterized by demographic characteristics, such as sex or age.

DANE uses RELAB as a source of information to analyze the Colombian labor market situation and the consistent analysis of different statistical operations of the labor market and economic activities. The main input of the RELAB is the administrative record[6] of the Social Security Integrated Form (PILA, for its Spanish acronym) of the Ministry of Health and Social Protection of Colombia. The income variable is constructed from the information of the payment of social security subsystems, such as health or pension, reported by employers or contributors. In addition, quality rules apply to raw data, addressing duplicate or redundant information for administrative or legal purposes only.

Integrating the information at the individual level from the GEIH household survey and the RELAB statistical register is done through a deterministic match with the identification keys available in the two sources of information. This integration makes it possible to control the possible heterogeneities found between geographic domains and evaluate the potential biases of the change in the GEIH collection mode when it goes from face-to-face to non-face-to-face operations (by telephone) from March to July 2020. The sample from this integration corresponds to a group of workers with formal employment relationships. Although it is not a representative

[5] In the RELAB, the dependents correspond to the GEIH occupational positions of workers or employees of a private company or the government and day laborers or peons. These occupational positions are also known as salaried, dependent, or employed workers.

[6] Administrative data have become a valuable source of information for generating statistics and developing technical analyzes even when their design does not initially respond to the technical specifications for statistical purposes (Wallgren and Wallgren, 2007; Zhang 2011; Kuhn, 2019;).

sample of the total Colombian labor market, because of its high rates of informality, the consistency in the data allows evaluation of the impacts of using different collection methods

3.2 Methodology

Considering the temporal variation in the implementation of the collection models, CATI and CAPI, and the geographic heterogeneity, the proposed identification strategy exploits the cross-sectional and temporal variation. The cross-sectional variation is because the telephone data collection strategy was used only in the 23 main metropolitan areas. In contrast, the temporal variation comes from applying this procedure only between March and July 2020. Finally, the methodological design is used because one of the two data sources had the previously described changes. At the same time, the other source did not suffer modifications since the process by which the RELAB data is generated did not change during 2020. In this way, the effect of implementing CATI is derived from the change in the discrepancy in the income report between the survey and the administrative record.

It is natural to find discrepancies between self-reported labor income and the RELAB record. Abowd and Stinson (2013) argue that differences between sources should not be confused with measurement errors since both sources are sensitive to measurement errors. In the literature, Nordberg et al. (2001) found that income estimates derived from administrative records are reliable and higher than income derived from surveys, except for low-income households. By comparing surveys and administrative records, we attempt to characterize survey income reporting biases. Along these lines, a predominant effect has been found at the ends of income distributions. For example, over-reporting is observed in the low-income population and under-reporting in high-income households (see Bound et al., 2001; Gottschalk and Huynh, 2010; Kim and Tamborini, 2012; Consolini and Donatiello, 2013; Kim & Tamborini, 2014).

If these differences are systematic, a change in the GEIH collection process configures an exogenous shock that allows identifying its impact. This means that if all the other variables remain constant, the income reported in the sources will maintain a constant difference that the collection through CATI could affect because of the characteristics of this method. The variation in the discrepancy constitutes an estimate of the causal effect.

In this scenario, an event study strategy is adequate for inferring the impact of interest. In particular, the following specification is used:

$$\log(y_{itz}^{GEIH}) - \log\left(y_{itz}^{RELAB}\right) = \alpha_0 + X_i\beta + \sum_t \theta_t D_i T_t + \gamma_t + \sum_z \phi_z \, \mathrm{zone}_{iz} + \varepsilon_i \tag{1}$$

Where y_i^{GEIH} is the monthly salary or income from the main activity of the individual i reported in the household survey, y_i^{RELAB} is the income calculated from what is reported in the RELAB. Thus, the left part of the equation measures the discrepancy in logarithmic units between the sources of information. D_i is an indicative variable activated for people living in the 23 metropolitan areas (i.e., the treatment group); T_t is an indicative variable activated for the period corresponding to implementing mixed collection methods. The effect of interest will be measured by the parameter θ_t corresponding to the period March-July 2020. That is, θ_t measures the changes in the discrepancy in the income variable regarding a reference period established at the beginning of the shift in collection mode.

Additionally, it is controlled by a comprehensive battery of location (zone, denoted by z) variables built from the interaction between the department and geographical area (metropolitan area, other cities, or rural areas), whose effect is measured by ϕ_z. γ_t denotes a set of time fixed effects. Finally, X_i is a vector of controls that includes sex, age group (<25, 25-54, >55), educational level achieved, number of hours worked during the last week (at the time of the survey), rate of unemployment,[7] sector of activity of the worker (ISIC revision 4 at the sector level), classification of the worker's job (ISCO to one digit).

This specification is estimated for the period between January 2019 and June 2021 where t corresponds to quarters and taking January and February 2020 as a base reference period. We exclude March since it was a survey adaptation period, the collection process was interrupted this month. The CATI application period (from April to July 2020) was defined as a single period and measured the impact of the change in the collection method.

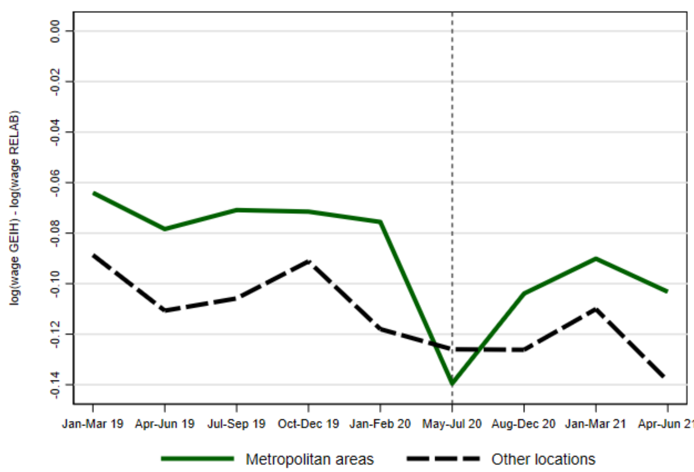
We should note that the integration between GEIH and RELAB generates a sample of workers that is not representative of the total labor market, given that it corresponds to salaried workers linked to the social security system. However, this sample has an important economic interpretation, as it is associated with the formal component of employment. The sample represents the total RELAB since the socioeconomic characteristics of this sub-sample correspond to the entire registry, as evidenced in Table A.1 and Table A.2 in the Annex. In short, the observations obtained from the matching between GEIH and RELAB follow similar socio-demographic characteristics to the universe of all formal workers identified with GEIH.

The validity of the proposed identification strategy rests on the assumption of parallel trends. Here, it establishes that between the treatment and control group, i.e., between metropolitan areas and the rest, the discrepancy between the employee's response in the survey and in RELAB is constant before the intervention. To validate this assumption, it is required that the evolution of the discrepancy in income be similar, regardless of the possible biases in each of the information

[7] For metropolitan areas, we use the unemployment rate for each area; for other cities, we use the national unemployment rate for all other cities. Similarly, for rural areas, we take the rural unemployment rate at the national level. We did not take these last indicators at the city or municipality level since the survey does not have enough observations in these domains to make a good measurement of unemployment by the municipality.

sources and regional patterns. The macroeconomic impact caused by Covid-19 and associated measures may be more significant in metropolitan areas. This is not a problem for the proposed methodology if both sources of information reflect this macroeconomic impact.[8] **Figure 1** presents the evolution of the average discrepancy in the reports for two geographical domains.

Figure 1. Average of the difference of labor income between GEIH and RELAB by geographical areas. 2019-2021.



Source: Authors' Calculations. RELAB-GEIH matching.

It is notable that the difference between the geographic domains is relatively constant over time, which is in favor of the assumption of parallel trends. This result is also corroborated through the estimation of the event study specification. It is important to note that the implemented identification strategy does not depend on the existence of parallel wage trends between metropolitan areas and the rest of the country, which could be a more demanding assumption. Even if the trends before treatment are parallel, the trends after that may not be, given that large cities were more exposed to Covid-19 infections and the respective preventive measures. To give

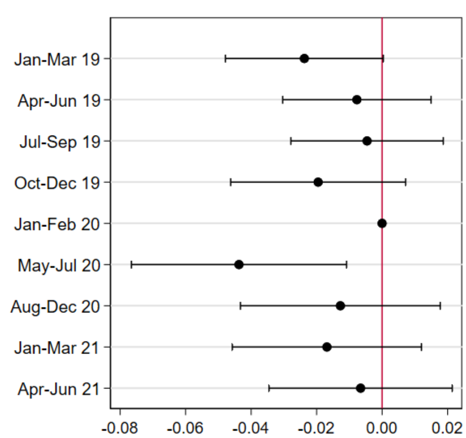
[8] Alternatively, if one of the sources of information reflects a greater impact of the pandemic than the other, but this discrepancy is similar between the metropolitan areas and in the other cities and rural areas, the assumption continues to be corroborated.

evidence of the robustness of the results, estimates using this alternative assumption of parallel trends are presented. Lastly, to my first glance, it appears that there is a drop in the discrepancy between GEIH and RELAB, which is consistent with a reduction in GEIH self-reporting, i.e., apparently the change in collection method from CAPI to CATI would be generating a downward reporting bias in labor income.

4 Results

Figure 2 and **Table 3** present the results of the estimation of the event study specification. **Table 3** shows the result of sequentially including controls and additional fixed effects so that in the complete specification (column 4), it is observed that a change in the discrepancy's pattern in the labor income is equivalent to 4.4 p.p with respect the base reference period. For the other periods, the effect is not significant at a significance level of 5%, which also implies that, once the collection mode was restored to CAPI, the discrepancy between the sources returned to the levels before the temporary changes of GEIH data collection. This implies that implementing the CATI collection method caused reductions in self-reported income.

Figure 2. Estimated impact of data collection mode on reported labor income



Source: Authors' calculations. RELAB-GEIH matching. The lines represent a 95% confidence interval.

Table 3. Estimated impact of data collection method on reported labor income

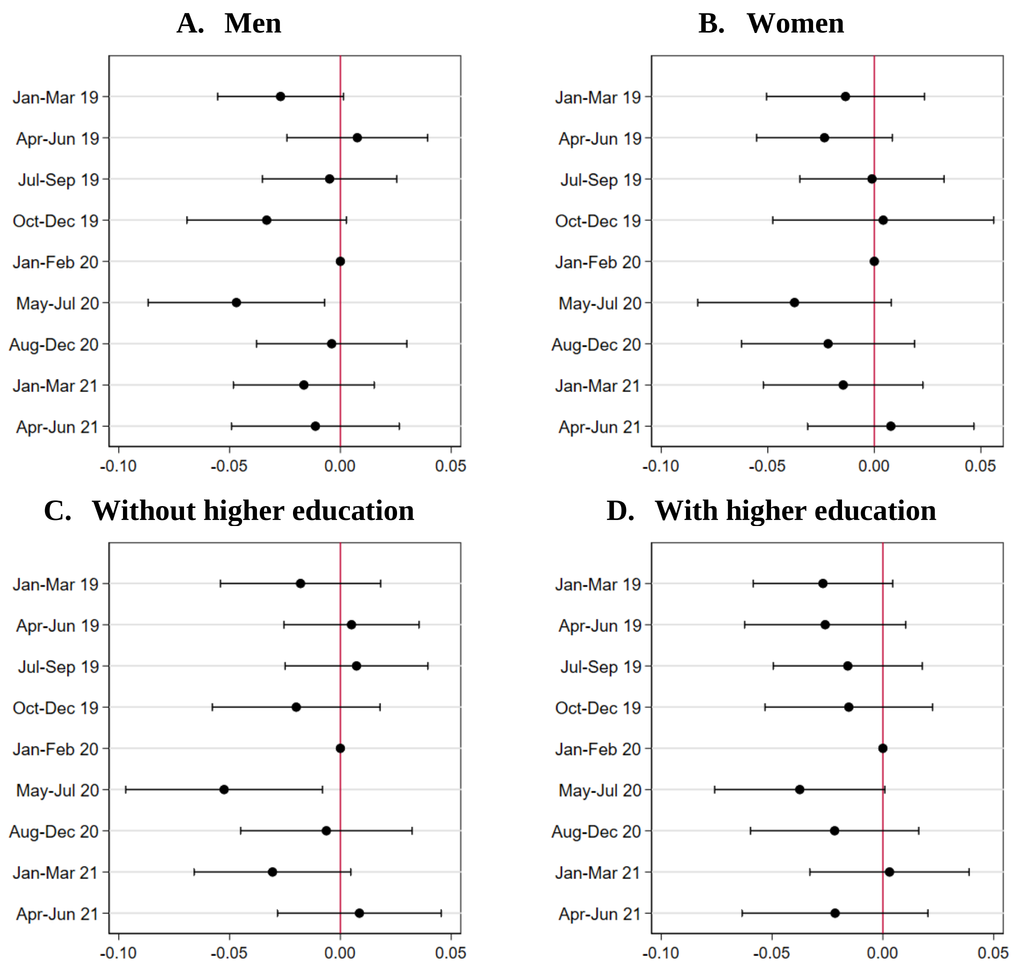
	(1)	(2)	(3)	(4)
Jan-Mar 19 × MA	-0.021	-0.024*	-0.024*	-0.024*
	(0.013)	(0.012)	(0.012)	(0.012)
Apr-Jun/19 × MA	-0.012	-0.011	-0.009	-0.008
	(0.012)	(0.012)	(0.011)	(0.011)
Jul-Sep/19 × MA	-0.009	-0.009	-0.006	-0.005
	(0.013)	(0.013)	(0.012)	(0.012)
Oct-Dec/19 × MA	-0.026*	-0.025*	-0.022	-0.020
	(0.014)	(0.013)	(0.014)	(0.013)
May-Jul/2020 × MA	-0.055***	-0.054***	-0.052***	-0.044***
	(0.015)	(0.015)	(0.015)	(0.016)
Aug-Dec/2020 × MA	-0.020	-0.020	-0.017	-0.013
	(0.014)	(0.014)	(0.015)	(0.015)
Jan-Mar/2021 × MA	-0.022	-0.019	-0.017	-0.017
	(0.015)	(0.014)	(0.014)	(0.014)
Apr-Jun/2021 × MA	-0.007	-0.007	-0.006	-0.007
	(0.014)	(0.014)	(0.014)	(0.014)
Observations	178,660	178,660	178,660	178,653
R-squared	0.004	0.020	0.024	0.044
Individual controls	NO	YES	YES	YES
Month FE	YES	YES	YES	YES
Activity FE	NO	NO	NO	YES
Occupation FE	NO	NO	NO	YES
Dpt-class FE	NO	NO	YES	YES

Source: Authors' work using RELAB-GEIH match.

These results indicate the aggregate impact of the collection method. However, this change in the report may vary according to the interviewee's profile. It has been documented that income reporting errors can vary regarding sociodemographic characteristics. This literature found that more educated individuals, women, and older individuals tend to report income more accurately. This would imply that in the same way that different patterns are generated in the income report when the CAPI method is used, changing the collection mode could stress said patterns. Therefore, changes in reporting patterns by sex, age, and educational level are explored using the same identification strategy.

Figure 3 presents the results, which show that men education tend to have higher levels of income under-reporting after implementing the CATI method. With women, the impact is of lesser magnitude and not significant. In the analysis by educational level, it is obtained that the under-reporting is of a similar magnitude between the groups analyzed and is significant in both cases.

Figure 3. Estimated impact by sex and education level



Source: Authors' work using RELAB-GEIH match. The lines represent a 95% confidence interval.

These results imply that faced with the change in the collection method, all the population groups vary in their response patterns with a marked age pattern. An additional analysis by age group shows that individuals between 25 and 54 years of age have a significant impact of similar magnitude to the change in the collection method, while for the other two age groups this effect is not significant (see Figure A.1). Likewise, by household typology (according to the presence of children in the household), this effect seems to be explained by the group without children in the household (see Figure A.2). This implies that the observed impact seems to be explained by men over 25 years of age without children in the household.

5 Robustness checks

To provide additional evidence that validates the results found, four additional exercises are presented that serves as a robustness check: i) a placebo test in which populations whose survey method for the GEIH did not change are compared; ii) an estimate in which the set of control observations is restricted, iii) an estimate in which only the information from the household survey and not the administrative record was used; and, finally, iv) an estimation that uses only RELAB records in which no effect is expected, given the stability in this source of information.

In the placebo test, all workers living in the 23 cities and MAs were removed from the sample, workers from the Other cities area were taken as the treated population, and workers in rural areas as the control. As expected, no significant effect was found (see **Table 4**, column 1). This exercise serves as a verification of the assumption of parallel trends in the GEIH and RELAB reports. The challenge regarding this assumption is the coincidence in the telephone's time operation and the mobility restriction measures, given that the course of the pandemic could modify the reporting trends in the most affected areas,[9] which could be confused with the effect of interest in this study. However, the estimate of the placebo effect compares areas most affected by the pandemic (other

[9] It must be considered that for this aspect to make the estimation of this study difficult, it is not enough that the pandemic has changed salary trends; It would be necessary that the trends in the discrepancy of reports between the different sources of information have changed. In other words, reporting incentives have changed differently for surveyed workers compared to employers who report to the administrative registry. Furthermore, this change in incentives is different in metropolitan areas.

cities compared to rural areas) and does not find significant effects. This is consistent with the absence of changes in the trend in the report.

Table 4. Results of robustness exercises

	(1)	(2)	(3)	(4)
Jan-Mar 19 × MA	-0.001	-0.025*	0.013	0.031*
	(0.027)	(0.013)	(0.013)	(0.016)
Apr-Jun/19 × MA	0.014	-0.014	-0.005	0.016
	(0.025)	(0.014)	(0.018)	(0.017)
Jul-Sep/19 × MA	0.004	-0.008	-0.002	-0.005
	(0.027)	(0.014)	(0.016)	(0.019)
Oct-Dec/19 × MA	-0.024	-0.016	0.019	0.025
	(0.030)	(0.015)	(0.015)	(0.016)
May-Jul/2020 × MA	0.005	-0.044**	-0.037*	-0.000
	(0.025)	(0.019)	(0.020)	(0.017)
Aug-Dec/2020 × MA	-0.010	-0.011	-0.010	0.008
	(0.026)	(0.019)	(0.015)	(0.014)
Jan-Mar/2021 × MA	-0.030	-0.007	0.013	0.032*
	(0.026)	(0.018)	(0.012)	(0.017)
Apr-Jun/2021 × MA	0.034	-0.017	-0.005	0.001
	(0.025)	(0.015)	(0.018)	(0.019)
Observations	21,998	171,763	327,932	178,641
R-squared	0.056	0.043	0.571	0.458
Individual controls	YES	YES	YES	YES
Month FE	YES	YES	YES	YES
Activity FE	YES	YES	YES	YES
Occupation FE	YES	YES	YES	YES
Dpt-class FE	YES	YES	YES	YES

Source: Authors' work using RELAB-GEIH matching.

Another criticism of the implemented methodological design might be that the 23 main cities has a different labor market than the control set, particularly in rural areas. The comprehensive set of controls captures much of these differences, but as additional check we present an estimation in which we took only workers in other cities as a control, which should be similar to workers in metropolitan areas. **Table 4**, column 2, indicates the estimation results. The magnitude of the effect estimated in this way is close to that found in the original estimate of this study (4.5 p.p. vs. 4.3 p.p.), which allows us to conclude that including rural areas in the initial estimate does not generate significant distortions. There is a significant, albeit small, effect in the first quarter of 2019.

In turn, column 3 of **Table 4** presents the estimate of the event study specification using only data from the GEIH for salaried workers, as follows:

$$\log(y_{itz}^{GEIH}) = \alpha_0 + X_i\beta + \sum_t \theta_t D_i T_t + \gamma_t + \sum_z \phi_z \, \text{zone}_{iz} + \varepsilon_i \tag{2}$$

Here, the counterfactual of the change in the collection method is built on the level of labor income instead of the discrepancy between data sources. The comparison based exclusively on GEIH may include confounder effects of CATI implementation with other differential effects that the pandemic itself may have generated across regions. The latter justifies using RELAB as a reference to compare changes un GEIH income report. However, the estimated parameters are similar to the initial estimate (3.7 p.p. vs. 4.4 p.p.). Despite requiring stronger assumptions, this estimate allows exploring another possible objection to the assumptions of the original hypothesis of this study: whether the employers that report to the administrative registry have incentives to change their reporting strategy between March and July 2020. Also, whether these changes present differentiated patterns in the metropolitan areas, the main estimate would be the combination of the effect described and the impact of the change in the collection operation. The results suggest that the employer reporting effect does not affect the estimates.

Finally, the fourth robustness check that only uses information from the RELAB (**Table 4**, column 4) shows a non-significant effect of 5% significance, which implies that the impacts found in the GEIH comparison can be attributed to the change in the collection method. This occurs if a source without variation in this process does not present evidence consistent with the results found, so the possibility of effects that generate additional biases can be ruled out. The sum of the robustness exercises validates the impact found and presents solid evidence on the impact of changes in the collection method on the income variable, essential for labor market analysis and poverty estimates.

6 Concluding remarks

Household surveys are instruments with a great capacity for collecting detailed information on household income, which also provides input on relevant indicators for public policies. For this reason, estimating income in a precise and comparable way over time is particularly important for the National Institutes of Statistics and policymakers. The information collection method is one of the possible sources of under-reporting and over-reporting. In the context of COVID-19, it has been of particular interest because of changes in the collection methods of household surveys because of the measures implemented to contain the pandemic.

The case of Colombia presents characteristics that are particular for studying the impact of changing the collection method. During the pandemic, differential collection methods were implemented by geographic domain, adding the availability of administrative data that served as a reference since this source of information was not affected by operational changes in the collection. Using a quasi-experimental design, it is estimated that the change in the collection method generated an average reduction of 4.4 percentage points in the labor income report. This level of under-reporting is higher for men and individuals between 25 and 54 years of age.

This exercise allows quantifying the impact of the collection method in a large-scale operation in a non-experimental context, as well as taking advantage of integrating administrative data with surveys to generate learning that leads to effective designs based on mixed collection methods that guarantee the quality of sensitive variables such as income.

References

Abowd, J. M., & Stinson, M. H. (2013). Estimating measurement error in annual job earnings: A comparison of survey and administrative data. *Review of Economics and Statistics*, 95(5), 1451–1467.

Angel, S., Disslbacher, F., Humer, S., & Schnetzer, M. (2019). What did you really earn last year?: explaining measurement error in survey income data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(4), 1411–1437.

Angel, S., Heuberger, R., & Lamei, N. (2018). Differences between household income from surveys and registers and how these affect the poverty headcount: Evidence from the Austrian SILC. *Social Indicators Research*, 138(2), 575–603.

Bollinger, C. R. (1998). Measurement error in the current population survey: A nonparametric look. *Journal of Labor Economics*, 16(3), 576–594.

Bound, J., Brown, C., Duncan, G. J., & Rodgers, W. L. (1994). Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics*, 12, 345–368.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In: *Handbook of econometrics* (Vol. 5, pp. 3705–3843). Elsevier.

Bound, J., & Krueger, A. B. (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?. *Journal of labor economics*, 9(1), 1-24.

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of public health*, 27(3), 281-291.

Burton, J., Crossley, T. F., Fisher, P., Gaia, A., & Jäkle, A. (2020). *Understanding and reducing income reporting error in household survey* (No. 2020-01). Understanding Society at the Institute for Social and Economic Research.

Caeyers, B., Chalmers, N., & Weerdt, J. (2010). A Comparison of CAPI and PAPI through a Randomized Field Experiment. *Available at SSRN 1756224*.

Canberra Group. (2011). Canberra Group Handbook on Household Income Statistics. United Nations.

Chesher, A., & Schluter, C. (2002). Welfare measurement and measurement error. *The Review of Economic Studies*, 69(2), 357-378.

Consolini, P., & Donatiello, G. (2013). Improvements of data quality through the combined use of survey and administrative sources and micro simulation model. *M. Jäntti*, 1000, 125-139.

de Leeuw, E. D. (1992). *Data quality in mail, telephone and face to face surveys*. TT Publikaties, Plantage Daklaan 40, 1018CN Amsterdam.

Departamento Nacional de Estadísticas [DANE]. (2020). Comunicado de Prensa – Acciones para garantizar la continuidad de la Gran Encuesta Integrada de Hogares (GEIH).

Economic and Social Research Council [ESRC]. (2019). ESRC Question Bank FACTSHEET 6. https://www.ukdataservice.ac.uk/use-data/guides/methods-and-software-guides

Economic Commission for Latin America and the Caribbean [ECLAC]. (2020). Recommendations for eliminating selection bias in household surveys during the coronavirus disease (COVID-19 pandemic).

Ellis, C. H., & Krosnick, J. A. (1999). Comparing telephone and face to face surveys in terms of sample representativeness: a Meta-Analysis of Demographics Characteristics. Universidad de Michigan, NES.

Fessler, P., Kasy, M., & Lindner, P. (2018). Survey mode effects on measured income inequality. *The Journal of Economic Inequality*, 16(4), 487–505.

Figari, F., Iacovou, M., Skew, A. J., & Sutherland, H. (2012). Approximations to the truth: Comparing survey and microsimulation approaches to measuring income for social indicators. *Social Indicators Research*, 105(3), 387–407.

Glewwe, P. (2007). Measurement error bias in estimates of income and income growth among the poor: analytical results and a correction formula. *Economic Development and Cultural Change*, 56(1), 163–189.

Gottschalk, P., & Huynh, M. (2010). Are earnings inequality and mobility overstated? The impact of nonclassical measurement error. *The Review of Economics and Statistics*, 92(2), 302-315.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, 67(1), 79–125.

Horowitz, J., & Manski, C. (1995). Identification and Robustness with Contaminated and Corrupted Data. *Econometrica*, 63, 281–302.

International Labour Organisation. (2020a). Capturing impacts on employment and unpaid work using Rapid Surveys Guidance to data producers to maintain labour force survey data collection COVID-19: Guidance for labour statistics data collection.

International Labour Organisation. (2020b). Guidance to data producers to maintain labour force survey data collection.

Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the effect of data collection mode on measurement. *International Statistical Review*, 78(1), 3-20.

Jordan, L. A., Marcus, A. C., & Reeder, L. G. (1980). Response styles in telephone and household interviewing: A field experiment. *Public Opinion Quarterly*, 44(2), 210–222.

Kapteyn, A., & Ypma, J. Y. (2007). Measurement error and misclassification: A comparison of survey and administrative data. *Journal of Labor Economic*s, 25(3), 513–551.

Kim, C., & Tamborini, C. R. (2012). Do survey data estimate earnings inequality correctly? Measurement errors among black and white male workers. *Social Forces*, 90(4), 1157-1181.

Kim, C., & Tamborini, C. R. (2014). Response error in earnings: An analysis of the survey of income and program participation matched with administrative data. *Sociological Methods & Research*, 43(1), 39-72.

Klausch, T., Hox, J. J., & Schouten, B. (2013). Assessing the mode-dependency of sample selectivity across the survey response process. Statistics Netherlands: The Hague.

Kreiner, C. T., Lassen, D. D., & Leth-Petersen, S. (2013). Measuring the accuracy of survey responses using administrative register data. Evidence from Denmark. NBER Working Paper, 19539.

Kuhn, U. (2019). Measurement of income in surveys. Swiss Centre of Expertise in the Social Sciences FORS, 02, Version 1.0.

Lohmann, H. (2011). Comparability of EU-SILC survey and register data: The relationship among employment, earnings and poverty. *Journal of European Social Policy*, 21(1), 37–54.

Lynn, P., & Kaminska, O. (2013). The impact of mobile phones on survey measurement error. *Public Opinion Quarterly*, 77(2), 586–605.

Meyer, B. D., & Mittag, N. (2021). Combining administrative and survey data to improve income measurement. *Administrative Records for Survey Methodology*, 297–322.

Meyer, B. D., Mok, W. K., & Sullivan, J. X. (2015). Household surveys in crisis. *Journal of Economic Perspectives*, 29(4), 199–226.

Moore, J. C., Stinson, L. L., & Welniak, E. J. (2000). Income measurement error in surveys: A review. *Journal of Official Statistics*, 16(4), 331–362.

Nordberg, L., Penttilä, I., & Sandström, S. (2001). A study on the effects of using interview versus register data in income distribution analysis with an application to the Finnish ECHP-survey in 1996. Statistics Finland.

Ravallion, M. (1988). Expected Poverty under Risk-Induced Welfare Vulnerability. *Economic Journal*, 98(3), 1171–1182.

Roberts, C., Eva, G., Allum, N., & Lynn, P. (2010). Data quality in telephone surveys and the effect of questionnaire length: A cross-national experiment. ISER Working Paper Series.

Schräpler, J. P., Schupp, J., & Wagner, G. G. (2010). Changing from PAPI to CAPI: Introducing CAPI in a longitudinal study. *Journal of Official Statistics*, 26(2), 239–269.

Statistics Canada. (2010). Survey Methods and Practices. Minister of Industry

United Nations Statistics Division [UNSD]. (2020). Planning and Implementing Household Surveys Under COVID-19.

Van der Klaauw, B. (2014). From micro data to causality: Forty years of empirical labor economics. *Labour Economics*, 30, 88-97.

van Praag, B., Hagenaars, A., & van Eck, W. (1983). The Influence of Classification and Observation Errors on the Measurement of Income Inequality. *Econometrica*, 51(4:1093–1108).

Ward, J. M., & Edwards, K. A. (2021). CPS Nonresponse During the COVID-19 Pandemic: Explanations, Extent, and Effects. *Labour Economics*, 72, 102060.

Wallgren, A. y Wallgren, B. (2012). Estadísticas basadas en registros: Aprovechamiento estadístico de datos administrativos, INEGI, México.

Annex

Table A 1. Comparison of the proportion of workers by sex of the integrated sample GEIH-RELAB and the sample of formal employees GEIH

Period	Formal GEIH		GEIH -RELAB		Differences in percentage points	
	Men	Women	Men	Women	Men	Women
Jan-Mar19	54.6%	45.4%	53.7%	46.3%	0.9	-0.9
Apr-Jun19	54.9%	45.1%	53.8%	46.2%	1.1	-1.1
Jul-Sep19	54.9%	45.1%	54.0%	46.0%	0.9	-0.9
Oct-Dec19	54.8%	45.2%	54.1%	45.9%	0.8	-0.8
Jan-Feb20	54.6%	45.4%	54.0%	46.0%	0.6	-0.6
May-Jul20	53.6%	46.4%	53.0%	47.0%	0.6	-0.6
Aug-Dec20	54.8%	45.2%	54.1%	45.9%	0.8	-0.8
Jan-Mar21	55.1%	44.9%	54.7%	45.3%	0.4	-0.4
Apr-Jun21	54.8%	45.2%	54.0%	46.0%	0.8	-0.8

Source: Authors' calculations. RELAB-GEIH matching. Formal GEIH corresponds to employees with social security contributions.

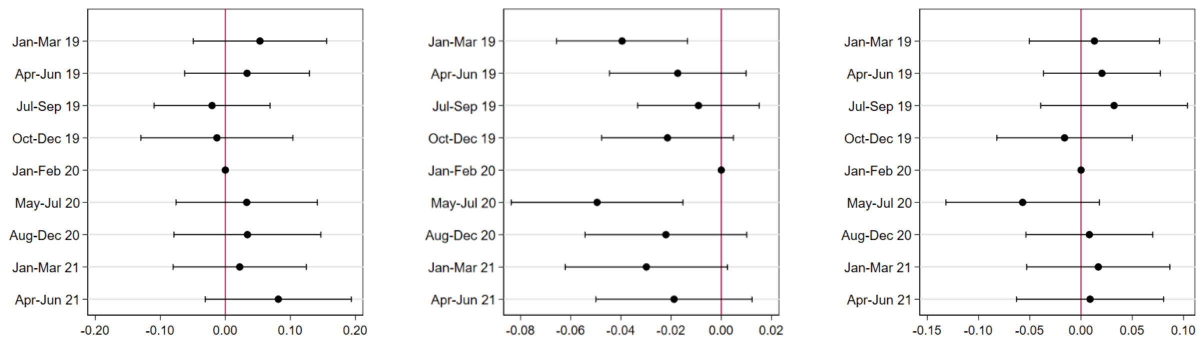
Table A 2. Comparison in the composition of the age group of the integrated sample GEIH-RELAB and the sample of formal employees GEIH

Period	Formal GEIH			GEIH -RELAB			Differences in percentage points		
	< 25	25 a 54	55+	< 25	25 a 54	55+	< 25	25 a 54	55+
Jan-Mar19	12.1%	76.7%	11.2%	10.9%	77.4%	11.7%	1.2	-0.6	-0.6
Apr-Jun19	11.8%	76.7%	11.5%	10.8%	77.2%	12.0%	1.0	-0.5	-0.5
Jul-Sep19	12.0%	76.6%	11.5%	11.1%	77.0%	11.9%	0.9	-0.5	-0.5
Oct-Dec19	12.2%	76.2%	11.6%	11.1%	77.0%	11.9%	1.1	-0.8	-0.3
Jan-Feb20	11.5%	76.2%	12.2%	10.5%	76.9%	12.6%	1.1	-0.7	-0.4
May-Jul20	9.6%	77.3%	13.1%	8.9%	77.7%	13.4%	0.7	-0.4	-0.3
Aug-Dec20	10.4%	76.9%	12.7%	9.5%	77.3%	13.2%	0.9	-0.4	-0.5
Jan-Mar21	10.2%	77.3%	12.5%	9.8%	77.4%	12.8%	0.4	-0.1	-0.4
Apr-Jun21	10.1%	77.3%	12.6%	9.1%	77.6%	13.3%	1.0	-0.3	-0.6

Source: Authors' calculations. RELAB-GEIH matching. Formal GEIH corresponds to employees with social security contributions.

Figure A 1. Estimated impact by age

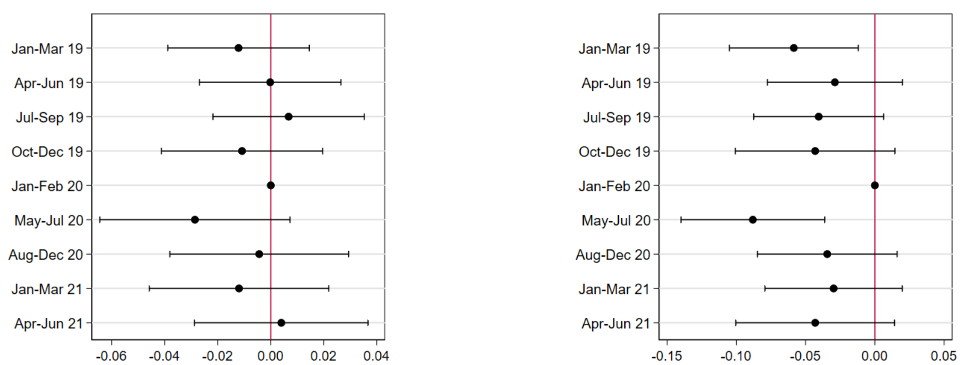
Less than 25 **25-54** **More than 55**



Source: Authors' calculations. RELAB-GEIH matching. The lines represent a 95% confidence interval.

Figure A 2. Estimated impact by household type

With children **Without children**



Source: Authors' calculations. RELAB-GEIH matching. The lines represent a 95% confidence interval.