MONITORING DATA COLLECTION FOR HOUSEHOLD BUDGET SURVEY AND DIGITAL TRANSFORMATION: AN INTEGRATED APPROACH TO IMPROVE SURVEY TECHNIQUES

Loredana De Gaetano, Laura Capparucci, Elena Cezza,
Claudia Fabi, Edoardo Raimondi

Abstract. The Household Budget Survey is a key pillar for understanding consumption dynamics and informing national political and economic decisions. The accuracy and reliability of the data collected depend, to a large extent, on the effectiveness of the survey process. This study aims to explore data monitoring for the Household Budget Survey in detail, analysing current challenges and possible solutions to optimize survey techniques and adapt them to the new needs of the socio-economic context. It seems necessary to exploit the potential offered by technological and methodological innovation and reduce the use of paper-based survey questionnaires in favour of data acquisition systems more suitable for certain respondent profiles, such as web questionnaires and applications for mobile devices. Particular emphasis is placed on training personnel involved in data collection, who are asked to acquire new interdisciplinary skills, and on the need to implement meticulous and continuous quality control strategies and tools.

1. Introduction

The Household Budget Survey is a fundamental source of information for understanding consumer behaviours and socio-economic dynamics (Istat, 2023).

Monitoring data collection plays a crucial role in ensuring the accuracy and efficiency of the process, as well as the quality of the collected data. It enables the identification of critical issues and facilitates timely intervention during the data collection phase, resulting in more reliable outcomes and an in-depth understanding of household spending behaviours (Barcherini *et al.*, 2002).

At the European level, ensuring high data quality, interoperability, and integrity has become a strategic priority, driven by the growing demand for publicly available data that can be easily reused for various purposes (Publications Office of the European Union, 2021).

Over the past years, as part of the European Strategy for Data, the European Commission has promoted the adoption of harmonized standards, protocols, and reference frameworks for data gathering and processing, to improve data consistency and accessibility across sectors and member states (European Commission, 2020).

Key initiatives have been launched, including the Rolling Plan for ICT Standardisation, the European Interoperability Framework, and the ISA² Programme[1], to promote semantic operability. These initiatives encourage the adoption of semantic assets, such as shared data models and Core Vocabularies, and machine-readable formats, to facilitate data exchange and reuse across different applications and systems (European Commission, 2019, 2021).

In the age of technological innovations, digital transformation enables the automation of data collection through online tools and mobile applications. Additionally, it offers significant opportunities to enhance the efficiency of data collection, conduct advanced analyses, and improve overall information management. The transition toward digital and automated data collection methods aligns with European guidelines to modernize statistical processes and leverage technology for better data governance. These efforts are particularly relevant for large-scale surveys like the Household Budget Survey, as they foster comparability, accuracy, and consistency in socio-economic data collection.

Following an outline of the current monitoring activities for the survey, this paper presents the results of an analysis conducted on the notes recorded by the interviewers in the initial and final questionnaires of the 2023 survey edition. The study specifically focused on the challenges encountered, along with improvement suggestions provided by both respondents and interviewers. Based on these insights, and with the perspective of introducing a mobile application for expenditure recording, new monitoring reports are proposed.

Lastly, the document highlights the benefits of advanced survey tools on interviewers' fieldwork, as well as their role in the modernization of the data collection process.

2. Monitoring the quality of the survey: methods and tools

The current monitoring activities of the survey's quality make use of traditional processing techniques based on SAS datasets of microdata relating to contacts and outcomes recorded by the interviewers during the survey period. Since the Household Budget Survey provides for at least three occasions of contact between

---

[1] Among the initiatives funded by the ISA² programme are the "SEMIC Study on Data Quality Management", which explores how semantic methodologies—such as controlled vocabularies, shared data models, and standard metadata schemas—can enhance data quality by linking data governance principles with interoperability frameworks, and the "Data Quality Guidelines for the Publication of Datasets in the EU Open Data Portal" project, which provides EU data providers with recommendations to prepare and publish high-quality data sets, in alignment with the FAIR (Findable, Accessible, Interoperable, Reusable) principles.

the sample family and the interviewer to whom it is assigned (initial interview, intermediate visit, and final interview), the indicators that are developed also refer to all three occasions in which the interaction between respondents and interviewers could generate critical issues, interruptions, and abandonments.

The focus of the analysis is placed on the completion of the interviews, both within the sextuplets of families extracted to be part of the sample and on the individual families, but also the performance of the interviewers, in a specific reference period (quarter/month/fortnightly).

As regards the analysis of the performance of the interviewers, tables are drawn up which, starting from the contact attempts recorded for each reference month, highlight the most important characteristic elements of the work carried out. For example: attempts made (at home or by telephone), families contacted, families interviewed, minimum, median, average and maximum duration of contacts made and of the final interview, average number of members within the families interviewed. Below is an exemplary extract (Table 1), relating to the reference month "December" of the 2023 survey year.

Table 1 – *Performance indicators of the interviewers - reference month 12, year 2023.*

Interviewer code	Attempts	By phone	At home	Completed interviews	Duration of final interview (hh:mm)				Ncomp
					average	median	Min	max	
R_1	99	31	66	13	0:40	0:38	0:28	1:04	2.2
R_2	29	0	29	8	0:28	0:29	0:14	0:35	1.8
R_3	28	7	21	6	1:31	1:29	1:11	1:52	3.0
R_4	70	0	70	15	0:25	0:24	0:15	0:33	1.7
R_5	67	7	60	12	0:34	0:34	0:31	0:41	2.8
R_6	52	1	51	10	0:19	0:20	0:13	0:25	1.8
…	…	…	…	…					
TOTAL	12,202	1,748	10,347	2,339	0:36	0:31	0:09	7:08	2.1

Source: calculations based on Istat data.

The monitoring of performance indicators allows for observing the presence of any critical issues that deserve additional training intervention within the survey network, for example in correspondence with a too high number of families contacted compared to families with complete interviews - symptomatic of excessive use frequent replacement procedures that allow moving on to a substitute family if the first one does not intend to collaborate - or the duration of the final interview is too short, on average, compared to the overall performance of the entire network - a phenomenon that could hide a hasty behaviour on the part of the interviewer and therefore less accurate in gathering information.

Furthermore, such a widespread survey across the territory cannot lack the support of synthetic indicators that allow the monitoring of the survey by territorial

aggregations. At a regional level, summary tables are drawn up which represent the main interest rates: loyalty rate, drop rate, rejection rate, and interruption rate.

Tables 2 and 3 show the regional indicators referring to month 12, year 2023, both taking the sextuplets as a reference base, i.e. the aggregates of 6 families extracted to form a homogeneous group – the first family and five potential substitutes – and all families contacted, regardless of their position in the sextuplet.

Through these monitoring tables, it is possible to analyse the different performances of the survey network across the territory and to identify in which phase of the complex process of completing a family interview the greatest critical issues occur (initial interview, intermediate visit, and final interview).

In fact, between the initial and final interviews, the survey provides that time is left for families to fill out a paper diary of the expenses incurred and it is precisely this commitment that can lead to a lack of willingness on the part of families to continue with the interview. This can occur especially if the network is unable to support the respondents in their daily recording of the information requested by the diary.

Table 2 – *Sextuplets indicators and completeness rate by Region, month 12, year 2023.*

Region	Assigned	Completed initial interview	Completed intermediate visit	Completed final interview	Completeness rates		
					Initial	Intermediate	Total
Piemonte	162	151	151	145	93.21	100.00	89.51
Valle d'Aosta	44	42	42	42	95.45	100.00	95.45
Lombardia	378	327	325	321	86.51	99.39	84.92
Trentino-Alto Adige	122	103	101	99	84.43	98.06	81.15
Veneto	178	168	167	161	94.38	99.40	90.45
Friuli-Venezia Giulia	62	62	62	61	100.00	100.00	98.39
Liguria	76	72	72	69	94.74	100.00	90.79
Emilia-Romagna	198	141	141	131	71.21	100.00	66.16
Toscana	204	193	193	190	94.61	100.00	93.14
Umbria	78	75	75	73	96.15	100.00	93.59
Marche	88	73	73	73	82.95	100.00	82.95
Lazio	250	213	212	200	85.20	99.53	80.00
Abruzzo	84	77	76	73	91.67	98.70	86.90
Molise	62	59	59	58	95.16	100.00	93.55
Campania	198	182	181	171	91.92	99.45	86.36
Puglia	130	119	119	117	91.54	100.00	90.00
Basilicata	56	54	54	52	96.43	100.00	92.86
Calabria	92	81	81	75	88.04	100.00	81.52
Sicilia	186	176	176	172	94.62	100.00	92.47
Sardegna	62	57	57	56	91.94	100.00	90.32
TOTAL	2,710	2,425	2,417	2,339	89.48	99.67	86.31

Source: calculations based on Istat data.

Finally, Table 3 reports the main indicators per family, therefore considering all the families contacted as a whole, regardless of their position in the sextuplet.

This data allows us to plan targeted interventions on the interviewers' network, intervening in those areas where the dropout rate is too high and where there is a tendency to replace the family by hastening the process that leads to substitutions, to allow for the higher rate of possible fidelity to the extracted base sample.

There are then other indicators to complete the general picture represented so far. The main ones monitor the number of interviews carried out for each day of the week, making it possible to observe, in a survey in which the time factor is so relevant, whether there is a concentration of interviews on the survey calendar on specific days of the week. Overall, the entire survey process is monitored at its nodal points, to be able to ascertain its fluidity and continuity over time and intervene where critical issues arise at a local or even individual level.

Table 3 – *Indicators on assigned families, loyalty rates, and drop by Region, month 12, year 2023.*

Region	Assigned	Contacts	Completed	Rates[2]		
				Fidelity	Dropout	Refusal
Piemonte	972	381	145	40.69	61.68	13.19
Valle d'Aosta	264	85	42	38,10	50,59	0.00
Lombardia	2.268	990	321	36,76	67,37	14,84
Trentino-Alto Adige	732	232	99	45,45	56,90	10,61
Veneto	1.068	432	161	32,92	62,50	42,22
Friuli-Venezia Giulia	372	158	61	29,51	61,39	9,28
Liguria	456	189	69	26,09	63,49	12,50
Emilia-Romagna	1.188	515	131	31,30	72,23	29,84
Toscana	1.224	490	190	44,21	58,98	33,22
Umbria	468	176	73	36,99	56,82	22,00
Marche	528	266	73	23,29	72,56	5,70
Lazio	1.500	734	200	33,50	70,44	11,03
Abruzzo	504	179	73	43,84	59,22	30,19
Molise	372	143	58	44,83	59,44	20,00
Campania	1.188	442	171	49,71	61,09	17,78
Puglia	780	368	117	37,61	67,93	55,60
Basilicata	336	123	52	42,31	57,72	22,54
Calabria	552	299	75	24,00	73,91	43,44
Sicilia	1.116	421	172	52,91	57,96	41,39
Sardegna	372	155	56	51,79	63,87	18,18
TOTAL	16,260	6,778	2,339	38,91	64,64	23,88

Source: calculations based on Istat data.

---

[2] The loyalty, dropout, and refusal rates are calculated as follows. The fidelity rate is given by the ratio of interviews completed on the base family compared to the total number of interviews completed. The dropout rate is given by the ratio between the total of definitive negative outcomes (i.e. refusal, prolonged absence of the family, etc.) and the total of families with at least one contact attempt. Finally, the refusal rate is calculated as the ratio between the number of refusals (recorded before the start of the initial interview) and the total of definitive negative outcomes.

3. Monitoring the survey: analysis of the interview notes field

In the perspective of introducing a digital tool for expense annotation, an analysis was conducted on the notes recorded in the initial interview and final interview questionnaires of the 2023 survey edition. The feedbacks from both interviewers and respondents provide a comprehensive view of the challenges encountered during the data collection phase. Therefore, this analysis can inform adjustments to certain aspects of data collection aimed at reducing the statistical burden on respondents, adopting a bottom-up approach (see Table 4).

During the reference year, a total of 2,808 comments were recorded. From these, notes deemed most pertinent to the intended purposes were selected and analysed using an interpretative approach (Guidicini and Castrignano, 1997). As a result, four distinct macro-topics were identified:

- **difficulties in making contact**: this group includes cases where the interviewers are unable to reach the family or the family is absent during the data collection period.
- **Difficulties in completing the questionnaire**: this category encompasses notes about the complexity and specificity of the questions, language comprehension issues, and instances where injury or illness hindered the ability to make expenditures. In many cases, to overcome these challenges, respondents opted to provide receipts for their expenses, which were then transcribed by the interviewer or a trusted person.
- **Willingness to participate**: this group includes issues such as distrust towards the interviewers or scepticism about the survey's purpose, the perception of the commitment as burdensome and time-consuming, concerns about privacy, and apprehensions regarding the collection, use, and protection of the provided data.
- **Optimization of response methods**: this category encompasses requests for modernizing the survey technique and for advance notification regarding the data collection period and the types of expenses to be recorded.

Each identified issue contributes to a significant statistical disturbance, often resulting in interruptions or refusals to participate in the survey. The results of the analysis suggest that implementing a mobile application for expense recording could effectively mitigate these problems and offer valuable insights for enhancing the data collection process. For instance, incorporating a predefined list of products to be recorded, sending in-app notifications to increase engagement, adding smart features such as receipt scanning, including a FAQ section and tooltips for better question comprehension, providing multilingual support, integrating internal consistency

checks would likely reduce the response burden and improve the accuracy and completeness of the collected data (De Vitiis *et al.*, 2023). An additional feature that could be implemented in the App is the integration of summary reports. This functionality would enable families to have a comprehensive overview of their expenses categorized by type, potentially promoting expenditure recording.

Table 4 – *Issues identified from the analysis of interview notes and proposed features for the mobile application.*

Critical issues encountered	Possible solutions to implement in the App
• Difficulty in contact • Family absent during the survey period • Complexity and high specificity of the questions • Language difficulties • Health problems • Lack of trust or interest • Survey too long and burdensome • Privacy concerns • Questions perceived as excessively intrusive • Mode of completion considered outdated • Short notice on the period and mode of completion	• Pre-loaded product list • In-app notifications • Receipt scanner • FAQ and tooltips • Multilingual support • Summary reports • Respondents' engagement programs

Source: authors' elaborations on Istat data.

4. Proposals for new monitoring reports

Introducing a mobile application for recording household expenses necessitates focused attention on implementing advanced new reporting. The newly proposed summary reports are intended to fulfil a dual purpose: first, to act as a survey supervision tool for the Statistical Institute, and second, to serve as a means of monitoring personal expenses for families, as proposed in the preceding paragraph. A primary objective is to supervise the use of the mobile application and the input data, providing the Statistical Institute with a system to enhance the realization of the intended sample and ensure consistency in expenditures recording. In this context, it is crucial to also monitor the performance of interviewers, whose roles require redefinition with the introduction of the App.

Table 5 presents some proposals for monitoring reports to be integrated into the infrastructure supporting the new application. Each category of the report –

household spending trends, data quality, interviewers' fieldwork efficiency, and application usage – is accompanied by a set of indicators to be produced (Istat, 2022).

Table 5 – *Proposed reports and related monitoring indicators.*

Type of report	Indicators
Report on household spending trends	• Analysis of weekly expenses (line graph processing, identification of sudden peaks or dips in expenses, and analysis of their causes). • Detailed monthly analysis of major spending categories such as groceries, transportation, housing, healthcare, entertainment, etc. (processing pie and/or bar charts to visualize the percentage of spending allocated to each category, identifying categories with the highest expenses, and comparing the results with the established budget).
Report on data quality	• Verification of correct categorization of expenses, completeness of input data, and adherence to defined procedures. • Identification of missing or incorrect data. • Identification of deviations from guidelines and necessary corrective actions, such as additional training or clarifications on procedures.
Report on interviewers' fieldwork efficiency* **(initial interviews, final interviews, and diary completion via the App)*	• Evaluation of the average time taken by interviewers to complete household expense registrations. • Analysis of response times to assigned tasks and any discrepancies compared to expected timelines. • Detailed quality control of interviewers' expense recording (accurate categorization of expense items, consistency with actual family consumption expenditure through cross-checking with receipts) and identification of causes for discrepancies (such as typing errors or missing data). • Analysis of mutual feedback and interactions among interviewers during the survey period. Identification of collaboration issues and promotion of initiatives to improve knowledge sharing among interviewers. • Continuous monitoring of how frequently interviewers update family expense data in the mobile application (identifying delays in data recording and underlying causes such as technical issues or difficulties in obtaining information, implementing measures to ensure timely data updates and improved responsiveness to monitoring needs). • In-depth examination of interviewers' utilization of specific features of the mobile application.
Report on application usage	• Monitoring of respondents' interactions with the application, including access times, frequency of data updates, and usage of specific features, such as receipt scanning. • Analysis of usage trends, including peaks or lows in the utilization of specific functionalities.

This initial set of indicators will be subject to continuous assessments by the competent Istat Directorates, data collectors, and the respondents themselves, to ensure they remain relevant and effectively meet the evolving needs of their users.

5. The role of the interviewer in the digital transition

The digital transition aims to simplify the data collection process, reducing the burden on respondents and the survey network, while enhancing the accuracy and timeliness of data collection.

In addition to the potential benefits mentioned in paragraph 3, this section addresses aspects related to the role of interviewers. Firstly, digitally collected data are immediately available and do not require manual transcription, thereby relieving interviewers of this task and significantly reducing the likelihood of human error. Moreover, the entire survey network can benefit from automatic reporting provided by the mobile application. The collected data can be promptly analysed and converted into preliminary reports, facilitating the identification of common spending patterns among users and enabling early detection of anomalies. Additionally, advanced reporting features offer enhanced customization and flexibility, according to the users' specific needs.

Nevertheless, alongside the opportunities, this transformation is not without its challenges. In this regard, it is important to emphasize that interviewers will continue to play a key role in ensuring the accuracy and completeness of survey data. The adoption of a mobile application entails a redefinition of their role and skill set, potentially including the capability to receive periodic updates on diary compilation status and provide support to respondents. An important consideration is the security of the collected personal data, which must be managed appropriately to preserve users' privacy (De Vitiis *et al.*, 2023). Interviewers will need comprehensive training on the use of digital tools, data interpretation in a digital context, and associated risks.

6. Conclusions

Optimizing data collection is essential for improving the accuracy and efficiency of the Household Budget Survey and addressing identified issues through monitoring activities. Digital tools, particularly mobile applications, offer significant advantages in this regard.

For respondents, an application would simplify the data collection process, allowing them to enter information quickly and accurately from their mobile devices.

This reduces the risk of human errors and enhances the reliability of the data collected. Furthermore, respondents gain quick and intuitive access to their personal expenditure data, enhancing transparency and the effectiveness of communication with interviewers. This capability proves particularly valuable in situations where the timeliness of information is crucial (i.e., managing survey completion difficulties).

For interviewers, the adoption of mobile applications for expenditure recording relieves them of the burden of manual data entry and enables them to focus on more significant monitoring activities. Moreover, these tools provide detailed real-time reporting of collected data, allowing the detect of trends, anomalies, and areas of interest promptly. This capability facilitates prompt responses to critical situations and supports informed adjustments to the fieldwork plan.

Implementation of a mobile application also entails addressing new and specific requirements. Interviewers play a pivotal role in ensuring the success and sustainability of this change. Comprehensive training and ongoing support are essential to facilitate the transition from paper-based methods to the adoption of new technologies. Active involvement of interviewers in the testing phases is required to identify potential issues early on and pinpoint opportunities for improvement.

With appropriate measures to address weaknesses and threats related to the project, the modernization of the Household Budget Survey has the potential to pave the way for further innovations in other socioeconomic surveys.

References

BARCHERINI S., CALIA P., FILIPPUCCI C., GRASSI D. 2002. Qualità del processo di produzione nell'indagine sui consumi dell'Istat. In FILIPPUCCI C. (a cura di) *Strategie e Modelli per il Controllo della Qualità dei Dati*, Franco Angeli, Milano, pp.135-162.

DE VITIIS C., DE FAUSTI F., INGLESE F., PEREZ M. 2023. Smart Surveys: Methodological Issues and Challenges for Official Statistics. In *Second workshop on methodologies for official statistics* | Roma, 6-7 dicembre 2023.

EUROPEAN COMMISSION. 2019. Data quality management. Drafted under the *2016.07 SEMIC: Promoting semantic interoperability amongst the EU Member States*. Available at: https://interoperable-europe.ec.europa.eu/sites/default/files/document/2019-09/SEMIC%20Study%20on%20data%20quality%20management.pdf

EUROPEAN COMMISSION. 2020. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee

and the Committee of the Regions: COM(2020) 66 final "A European strategy for data". Brussels, 19 February 2020.

EUROPEAN COMMISSION. 2021. Rolling plan for ICT standardisation – Shaping Europe's digital future. Available at: https://interoperable-europe.ec.europa.eu/collection/rolling-plan-ict-standardisation/rolling-plan-2021

GUIDICINI P., CASTRIGNANO M. 1997. *L'utilizzo del dato qualitativo nella ricerca sociologica*. Milano: Franco Angeli.

ISTAT. 2022. Indagine sulle spese delle famiglie - Scheda standard di qualità Anno 2022. Available at: https://www.istat.it/scheda-qualita/indagine-sulle-spese-delle-famiglie-3/

ISTAT. 2023. Indagine sulle Spese delle Famiglie. Available at: https://www.istat.it/it/archivio/71980.

PUBLICATIONS OFFICE OF THE EUROPEAN UNION. 2021. *Data.europa.eu: Data Quality Guidelines*. Luxembourg: Publications Office of the European Union. Available at: https://data.europa.eu/doi/10.2830/79367

_____

Loredana DE GAETANO, Istat, degaetan@istat.it
Laura CAPPARUCCI, Istat, capparuc@istat.it
Elena CEZZA, Istat, elena.cezza@istat.it
Claudia FABI, Istat, claudia.fabi@istat.it
Edoardo RAIMONDI, Istat, edoardo.raimondi@istat.it