INTEGRATING ADMINISTRATIVE DATA WITH ISTAT SURVEYS FOR MUNICIPAL HOMELESSNESS RISK ASSESSMENT

Angela Chieppa, Marija Mamolo

Abstract. This study investigates the insights and analytical potential of integrating ISTAT data sources concerning the homeless population. The main goal is to derive new municipal classifications associated with different levels of expected homelessness rates, crucial for designing tailored sample surveys. Comprehensive direct enumeration of the homeless population is cost-intensive and often impractical, while accurate territorial indicators could guide surveys only in specific areas.
The surveys sources considered are the National Survey on Homelessness in Italy (2011 and 2014), annual surveys on residential socio-assistance facilities, the permanent census of non-profit institutions, and annual surveys on administrative population registers. Furthermore, the Integrated Data Base of Usual Residents (AIDA) has been integrated, since the AIDA individual occurrences, known as Signs of Life (SoL), combine information from many administrative sources to enhance the analysis of specific population profiles.
This paper begins with an introduction to the challenges of data collection on homelessness and the importance of integrating various data sources. It then reviews existing ISTAT surveys and administrative databases, detailing their possible role in estimating the homeless presence at municipal level. The methodology section describes the municipal dataset built to analyse integrated sources and the computation of 'risk' indicators that are supposed to be associated with homeless presence.
Results highlight opportunities for improvement of data quality and implications for data collection and estimation. Future research goals include updated *ad hoc* surveys and the design of an information system for continuous monitoring of homelessness in Italy.

1. Introduction

The accurate and timely collection of data on the homeless population is crucial for effective policy development and targeted service provision. However, this task is challenging due to the transient nature of homelessness and the substantial costs and complexities associated with conducting ad hoc surveys. In 2010, with the closure of the European Year for Combating Poverty and Social Exclusion, the European Parliament called for an ambitious European strategy on homelessness. Since 2008, the Parliament has increasingly emphasized the need for standardized definitions and measurements of homelessness across Member States (Edgar et al., 2007). Establishing precise and shared definitions and methods is essential to

accurately measure homelessness and capture its multifaceted nature (Leterme et al., 2023).

FEANTSA, the European Federation of National Organizations Working with the Homeless, has developed in 2005 a classification that serve as a crucial framework for understanding and measuring homelessness across Europe. The classification is called ETHOS, European Typology of Homelessness and Housing Exclusion (Amore et al., 2011). ETHOS encompasses various forms of homelessness, including rooflessness, and insecure or inadequate housing. ETHOS Light, a simplified version, offers a harmonized definition for statistical purposes. This framework not only standardizes data collection and supports comparability across countries but also constitutes a reference point for planning new data acquisition methods, ensuring target identification for potential *ad hoc* surveys, and providing an essential basis for evaluating the integration potential of different data sources.

Recent comparative reports on data acquisition related to homelessness across various countries highlight that there is still no comprehensive monitoring system at the EU level. Countries employ diverse measurement methods and intervention policies (Busch-Geertsema V. et al., 2010) but recent years have seen a growing trend towards harmonization and improved data acquisition (Busch-Geertsema V. et al., 2014). These reports also indicate that traditional sources for data acquisition exhibit significant gaps: while the population census serves as a powerful mechanism for data collection, it is primarily designed for a general target population and must be suitably adapted to effectively capture information on homelessness. *Ad hoc* surveys can yield valuable snapshots of the homeless population; however, their high costs necessitate improvements to ensure longitudinal analysis and maximize their utility (O'Sullivan et al., 2020). Administrative data and registers present substantial limitations due to under-coverage of the homeless population, which, by its nature, evades registration. Nonetheless, they remain potential indirect sources of information, complementing other data collection efforts (Zindato et al., 2023).

Since 2018, Italy's Permanent Population Census strategy has integrated administrative sources with a set of surveys (ISTAT, 2022) specifically designed to enhance the quality of ISTAT's statistical population register data. Homelessness figures, as part of the broader census outputs, are a key focus of this methodological effort. The integration of these sources and the development of targeted surveys for homelessness present new opportunities that require further investigation. To address these challenges, a dedicated task force comprising thematic experts, methodologists, and census specialists has been formed. The goal of this work is to transition from sporadic to continuous statistical production for homelessness statistics, similar to the approach used for other census population statistics.

This paper investigates the potential utility of existing ISTAT surveys in estimating the homeless population and explores how administrative data can enhance insights into municipal classifications of homelessness. It begins with an examination of available ISTAT sources, details the creation of an integrated dataset for this study, and presents initial findings from multidimensional analyses, including clustering. Finally, it discusses key findings and outlines future steps for advancing these methodologies.

2. Available Data Sources

In planning *ad hoc* surveys targeting the homeless population, ISTAT relies on the ETHOS classification as a foundational framework. For example, a Point-In-Time enumeration survey is scheduled for 2025, targeting roofless individuals and shelter users, key categories within ETHOS Light. When investigating innovative methodologies and multiple data sources, however, the framework is expanded to encompass broader ETHOS categories and data from sources with wider target populations, thereby supporting the development of a multi-source information system and related indicators, which forms the core of this paper. This flexible approach aids in constructing monitoring indicators and covariates for estimation models, focusing on correlations across target groups rather than exact alignment.

Thus, a further distinction can be drawn among the various data sources considered, based on their potential to directly measure the target population—where homelessness is a primary focus—or to provide indirect insights through related populations. In this latter category, administrative registers, for instance, yield indirect information via statistical data referred to as "Signs of Life" (SoL) (Bernardini *et al.*, 2024a)., which potentially correlate with aspects of homelessness without explicitly targeting it.

2.1. National Surveys on homeless population (2011, 2014).

The most recent *ad hoc* surveys on the homeless population in Italy, carried out by ISTAT, date back to 2011 and 2014 (ISTAT, 2014; Inglese et al., 2021). These surveys provided relevant insights, including some socio-demographic characteristics, a list of services for homeless people by municipality, and estimates of the homeless population at the national level and for some larger municipalities. These surveys represented a significant methodological and innovative milestone in Italian official statistics, as they conducted a nationwide assessment of the homeless population using indirect sampling, yielding promising results. Nonetheless, critical organizational and methodological issues arose, underscoring the importance of conducting supplementary control investigations using alternative survey

techniques. The data estimated from these surveys still represent the most accurate and comprehensive source of information about the homeless population in Italy.

2.2. Surveys on administrative municipal registers: people registered in special addresses to support Census dissemination.

The Permanent Population Census, conducted by ISTAT, integrates administrative data with sample surveys designed specifically to improve the quality of individual data in the statistical registers (Bernardini *et al.*, 2024). This approach allows for detailed census results with specific classifications and territorial levels typical of census dissemination. For households and dwellings, two annual surveys are conducted on samples of addresses/households present in the registers. For people not belonging to these two main population groups, additional surveys and quality improvements are necessary. To address this, ISTAT collaborates annually with municipalities to conduct specific surveys (ISTAT, 2022) and validation using municipal administrative registers (*anagrafi*) related to population segments not previously included in the sample surveys of the Permanent Census. These segments are: 1) People living in institutional households. 2) People residing in authorized camps or in tolerated and spontaneous settlements. 3) People registered at dedicated addresses, real or fictitious, specifically designed to register homeless individuals and people without a stable dwelling/address.

These surveys aim to fill gaps in the statistical registers and ensure that the Permanent Census reflects the entire population in all its components. However, analysis of these data and municipal metadata confirmed that these special addresses, although intended for the homeless, are now used to administratively manage many other irregular situations, such as women and minors who are victims of violence or people living in dwellings that do not meet legal requirements.

2.3. Surveys on residential social and care facilities.

ISTAT conducts an annual survey on the availability of socio-assistance and socio-health residential facilities and the types of users they assist, allowing for precise documentation of both the users and the resources dedicated to this form of territorial assistance (ISTAT, 2023). The survey is conducted online via an electronic questionnaire. It covers all public or private facilities providing residential services (assisted accommodation with overnight stays) of a socio-assistance and/or socio-health nature. These facilities accommodate individuals in need for various reasons, such as elderly individuals living alone or with health issues, persons with disabilities, minors without guardianship, young women in distress, victims of gender-based violence, but also foreigners or Italian citizens facing economic difficulties and social distress. This latter category could be directly associated with

the homeless condition. The survey takes place between October and February each year.

2.4. Surveys supporting the Permanent Census of Non-profit Institutions.

The Permanent Census of Non-profit Institutions in Italy offers a comprehensive statistical overview of the sector through a Statistical Register, which integrates administrative and statistical sources to annually update data on the main characteristics of institutions (Della Queva *et al.*, 2023). Additionally, a periodic sample survey is conducted every three years to gather additional information on specific topics. This survey serves to verify and supplement the data in the register. This non-profit system facilitates the extraction of data related to homelessness by identifying organizations dedicated to homeless services, such as soup kitchens.

2.5. Administrative Signs of Life to derive "indirect" information on homeless presence.

The AIDA thematic register (Integrated Data Base of Usual Residents) utilizes administrative sources to compile population statistics. It incorporates a new statistical integrated individual data concept called "Sign of Life" (SoL) for individuals recorded in administrative databases like employment, education, tax returns, earnings, retirement benefits, and residency permits. Some Signs of Life signify specific individual activities that clearly indicate a sustained period and location, such as employment or holding a rental contract. Others pertain to individual conditions or statuses, such as dependents listed in tax returns.

Information on SoL of the individuals registered as residents in each municipality can be used to compute indicators and rates associated with homeless presence. For example, the percentage of people located in a certain municipality who do not record any SoL of employment activity or are not associated with any housing sign (no rental, no property) can be considered. Alternatively, specific indicators can be calculated by working with a subset of "at-risk" individuals. For instance, rates based solely on the SoLs of individuals registered at special addresses designed for homeless in the municipal registries can be calculated.

3. Setting up the Municipal Dataset for Experimental Purposes

To analyse the available information on the homeless population in Italy using ISTAT official statistics, a municipal dataset integrating the following sources was built: national surveys on homeless population (2011, 2014), the annual survey on population registers (2021), the annual survey on residential social and care facilities (2020), the Permanent Census of Non-Profit Institutions (2021), and AIDA (2021).

The dataset is at a municipal level and includes, among other variables, the demographic size of municipalities, the presence and number of services for homeless people derived from the survey sources, and the number of people recorded in administrative population registers at special addresses. Additionally, the following three indicators are computed based on the SoL related to the people of each municipality:

- % RU - Municipal Population Register Under-coverage: the proportion of people not recorded in the Municipal Population Register but added during Census validation because recognized as residents, on the total number of individual records in the municipal population register.
- % RH - Registered Homeless: the percentage of individuals recorded at special addresses designed for homeless people on the total number of individual records in the municipal population register.
- % DP - Discontinuous Presence: the percentage of individual records with irregular SoL on the total number of individual records in the municipal register. The irregular SoL includes presence in administrative sources with discontinuous or sporadic duration patterns.

The choice of these rates depends on specific hypotheses:

1. The total under-coverage in the municipal population registers correlates with the under-coverage of specific population targets, such as the homeless, implying that higher total under-coverage in the register indicates higher under-estimation of the homeless population.
2. The proportion of registered homeless individuals in each municipality is associated with the proportion of people registered as homeless at special addresses.
3. Discontinuous administrative patterns are likely linked to homeless individuals or those at risk of homelessness.

Some rates considering only the SoL of people registered at special addresses are also computed. These specific rates represent the type/profile of individuals recorded as homeless and can help differentiate among municipalities through an indirect evaluation of the effective homeless population. Specific rates for each SoL class are computed, focusing on those relevant to differentiate individual profiles of people recorded in the population register. Among others, the following specific rates are computed:

- % RH_stable_signs: (total number of individual records registered at special addresses with stable/continuous administrative signs)/(total number of individual records registered at special addresses)
- % RH_no_signs: (total number of individual records registered at special addresses with no administrative signs)/(total number of individual records registered at special addresses)

- % RH_pensions: (total number of individual records registered at special addresses with pensions or support income signs)/(total number of registered at special addresses)

The hypothesis is that different levels of these rates are correlated with different levels of homeless presence. For example, the %RH_no_signs (percentage of registered homeless individuals with no signs of life) are supposed to be positively correlated with the number of homeless individuals in the municipality. Conversely, the %RH_stable_signs (percentage of registered homeless individuals with stable signs) are supposed to be negatively correlated with the number of homeless individuals in the municipality.

4. Classification of Municipalities Based on Homeless Presence Evidence

Four main groups of municipalities were identified based on direct evidence of homeless presence, as shown in Table 1.

Table 1 - *Municipalities according to direct evidence of homeless presence.*

Presence of services for homeless people	No homeless in the pop. register	People registered in special homeless addresses	*Total*
No services	5,656	1,985	**7,641**
With services	50	213	**263**
Total	**5,706**	**2,198**	**7,904**

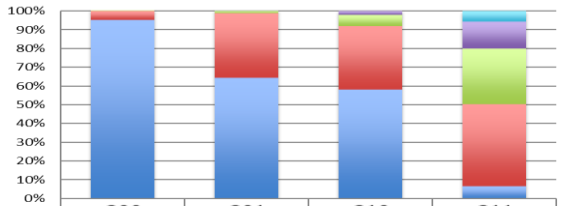
This initial classification of Italian municipalities defines four groups, which provide a framework to tailor data collection strategies for investigating the homeless phenomenon:

- *Group G00 – No Services, No Registered Homeless*. This group includes 5,656 municipalities. It can serve as a layer from which a few municipalities can be periodically sampled to verify the actual absence of homeless individuals.
- *Group G01 – No Services, Presence of Registered Homeless*. Comprising 1,985 municipalities, where there is a need to improve the design and monitoring of current surveys on services. Additionally, more detailed information is required on individuals registered at special addresses to confirm if they are indeed homeless, given the lack of available services.
- *Group G10 – Only Services*. 50 municipalities. There is a need to verify the accuracy of registry data, as it appears to be unreliable. An ad hoc survey on the homeless population is necessary in these municipalities.

- *Group G10 – Services and Registered homeless.* This group has to be sampled in homeless survey, since it encompasses the 213 towns with both services and people registered at special addresses.

Using the other variables and rates present in the municipal dataset, the characteristics of each group were analysed in more detail. The analysis started with demographic size and then studied the differences according to the risk indicators. Significant differences among groups were revealed.

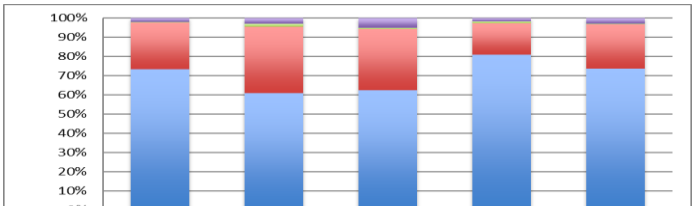
Figure 1- *Municipal population size of emerging groups.*



	G00	G01	G10	G11
>250th	0	0	0	12
100to250th	0	0	1	31
50to100th	4	26	3	63
10to50th	275	679	17	93
<10th	5377	1280	29	14

As shown in Figure 1, Group G00 primarily consists of small municipalities. Group G01 predominantly comprises municipalities with fewer than 50,000 inhabitants. Group G10 includes medium-sized municipalities and larger cities like Latina. Group G11 mainly encompasses larger municipalities but also includes small municipalities with fewer than 10,000 inhabitants.

Figure 2 - *Distribution of groups across Italian regions.*



	North-west	North-east	Center	South	Main Islands
G11	2,07%	3,17%	5,27%	1,91%	2,86%
G10	0,23%	1,29%	0,52%	0,95%	0,39%
G01	24,31%	34,68%	31,71%	16,32%	23,05%
G00	73,39%	60,86%	62,50%	80,82%	73,70%

Figure 2 illustrates the percentage composition of groups across Italian regions. Group G00 dominates in all regions, especially in the South (80.82%) and Main Islands (73.70%), representing municipalities without services nor registered homeless. Group G01 is prominent in the Northeast (34.68%) and Center (31.71%), highlighting areas with registered homeless individuals despite lacking services.

For each group, further evidence of the risk of the presence of homeless people was evaluated using administrative signs presented in Table 1. For instance, the percentage of people registered at special homeless addresses in municipalities with no services but with registered homeless individuals (G01) is similar to that in municipalities with services and registered homeless individuals (G11). This result calls for further socio-demographic analysis of the individuals registered at special addresses to better understand the homeless phenomenon and to question the absence of services. In the group of municipalities with services but without individuals registered at special addresses (G10), indicators related to population register under-coverage and the presence of unstable populations suggest the presence of homeless people who might not be registered at special addresses or at all.

Table 2. *Risk rates computed for each group of Municipalities (average value).*

Group	% RH Registered Homeless	% RU Register Undercoverage	% DP Discontinuous Presence
G00 – No Services,No Registered Homel.	-	0.1%	8‰
G01 – No Services, Presence Registered H.	1.6	0.18%	7‰
G10 – Only Services	-	0.17%	11‰
G11 –Services and Registered homeless	1.7	0.23%	8‰
All Municipalities	*0.4*	*0.12%*	*8‰*

5. Cluster Analysis Results: Municipalities According to Risk Indicators

The groups discussed in precedent paragraph can also be segmented internally. The administrative risk rates and available variables in the municipal dataset were used to create additional subgroups within the four main groups already identified. Cluster analysis was performed, one for each group, using the k-means method. This resulted in a final classification, shown in Table 3, in which the clusters resulting are labeled according to the evaluation of their statistical profiles. In table 4, clusters profiles are shown by reporting average values and counts of the main analysis dimensions. Some anomalous profile/cluster also emerge, where anomaly is detected when values of average cluster indicators are outliers of pertaining groups.

Table 3 - *Municipal classification derived from cluster analysis.*

Group	Subgroup/ Cluster	Label of final municipal class	Frequency
G00:	G00-1	*Very small Municipalities, no risk*	3,860
No Services,	G00-2	*Small Municipalities, no risk*	1,640
No Registered Homeless	G00-3	*Risk of homeless presence*	108
	G00-4	*Anomalous/Risk*	48
G01	G01-1	*Anomalous/Risk*	13
No Services,	G01-2	*Very low risk of homeless presence*	1,559
Registered Homeless	G01-3	*High Risk of homeless presence*	413
G10	G10-1	*With night shelters*	28
Services,	G10-2	*Other services*	22
No Registered Homeless			
G11	G11-1	*Considerable presence of homeless*	205
Services,	G11-2	*Bigger Municipalities*	6
Registered Homeless	G11-3	*Milano*	1
	G11-4	*Rome*	1
Total nr. Municipalities			*7,904*

Table 4 - *Clusters results by average group rates and counts.*

Subgroup /cluster	Nr. Municipalities	Municipal Residents	Registered Homeless	%RH	%RH no_signs	%DP	%RU
	(frequency)	*(average of municipal counts)*			*(average of municipal rates)*		
G00-1	3,860	2,137		-	-	5.20	0.03
G00-2	1,640	4,753		-	-	7.50	0.22
G00-3	108	4,153		-	-	30.01	0.72
G00-4	48	1,843		-	-	202.33	0.26
G01-1	13	1,304	71	43.58	0.05	12.91	0.13
G01-2	1,559	9,930	12	1.37	0.08	5.47	0.12
G01-3	413	12,845	12	1.03	0.09	12.21	0.38
G10-1	28	10,600		-	-	14.20	0.12
G10-2	22	25,345		-	-	7.59	0.23
G11-1	205	63,312	124	1.60	0.07	7.49	0.23
G11-2	6	619,332	2,685	3.82	0.08	10.55	0.38
G11-3	1	1,349,930	8,541	6.33	0.08	15.94	0.63
G11-4	1	2,749,031	22,182	8.07	0.16	8.92	0.42

6. Concluding Remarks and Current Work

This study underscores the critical need for updated survey data and the incorporation of additional variables on the homeless population to improve the accuracy and depth of findings. Despite these limitations, the integrated approach demonstrates that significant improvements in data quality in predicting the presence of homeless population in each Municipality can be achieved at minimal cost. By combining administrative data with traditional survey methods, valuable insights are gained that enhance the investigations on homelessness.

The results of this study suggest several practical benefits. Firstly, integrating various data sources can significantly improve the quality and reliability of information on the homeless population. Secondly, minor adjustments to existing surveys can lead to more accurate and comprehensive data collection, which is crucial for effective policy-making and resource allocation. Finally, classifying Italian municipalities based on available data sources can aid in developing targeted sampling strategies and predictive models, thereby providing more accurate estimates of the homeless population.

Looking ahead, current activities and future goals focus on several key areas. The analysis will be enhanced by incorporating updated sources and additional variables, which will provide a more detailed and nuanced understanding of homelessness. ISTAT is set to conduct ad hoc surveys on the homeless population, beginning with a Point-in-Time survey in 2025. This survey will provide a snapshot of the homeless population at a specific moment, offering valuable data for the statistical validation of the proposed risk indicators and for suggesting additional ones.

Moreover, the development of an information system that could continuously monitor and update data on the homeless population is currently under study. This system will leverage the added value of integrating various sources and components.

In conclusion, while there are challenges and limitations, the integration of administrative data with traditional survey methods offers a promising step forward. By leveraging these combined data sources, a more comprehensive and accurate understanding of the homeless population can be gained, ultimately leading to more effective interventions and support systems.

References

AMORE K., BACKER M., HOWDEN-CHAPMAN P. 2011. The ETHOS Definition and Classification of Homelessness: An Analysis, *European Journal of Homelessness*, Vol. 5, No. 2, pp. 19-37.

BERNARDINI A., CHIEPPA A., TAMBURRANO T. 2024. Discovering individual profiles from administrative signs of life useful for the estimation of Census results,

RIEDS - Rivista Italiana di Economia, Demografia e Statistica, Vol. 78, No. 1, pp. 15-24.

BUSCH-GEERTSEMA V. et al. 2010. Homelessness and homeless policies in Europe: Lessons from research. In *Conference on Homelessness*, Vol. 9.

BUSCH-GEERTSEMA V., et al. 2014. *Extent and Profile of Homelessness in European Member States: A Statistical Update.* FEANTSA, Brussels.

DELLA QUEVA S., LEPORANICO V., NICOSIA M., STOPPIELLO S. 2023. The nonprofit sector in Italian Inner Areas, *Rivista di Statistica Ufficiale*, Vol. 77-4.

EDGAR B., HARRISON M., et al. 2007. *Measurement of Homelessness at European Union Level.* Brussels: European Commission.

GOSME L. 2014. Key steps towards a European Union homelessness policy, *Journal of European Social Policy*, Vol. 24, No. 3, pp. 289-299.

INGLESE F., MASI A. 2021. Una metodologia per la stima della homelessness attraverso le Unità di strada: il caso della città di Torino. *ISTAT WORKING PAPERS* N. 5/2019, Istat-Istituto Nazionale di Statistica, Roma.

ISTAT. 2014. *La ricerca nazionale sulla condizione delle persone senza dimora in Italia*. Roma: ISTAT.

ISTAT. 2022. *Piano Generale del Censimento Permanente della Popolazione e delle Abitazioni 2022-2031.*Roma: ISTAT.

ISTAT. 2023. *Le strutture residenziali socio-assistenziali e socio-sanitarie, 1° gennaio 2022.* Roma: ISTAT.

LETERME Y., DEVELTERE P. 2023. The European Platform on Combating Homelessness (EPOCH): From Policy to Practice and from Practice to Policy, *European Journal of Homelessness*, Vol. 17, No. 2.

O'SULLIVAN E., PLEACE N., et al. 2020. Distorting Tendencies in Understanding Homelessness in Europe, *European Journal of Homelessness*, Vol. 14, No. 3, pp. 109-135.

ZINDATO D., TRUSZCZYNSKI M. 2023. Hard-to-reach population groups in administrative sources: Main challenges and future work, *Statistical Journal of the IAOS*, Vol. 39, pp. 849-863.

Angela CHIEPPA, Istat, chieppa@istat.it
Marija MAMOLO, Istat, mamolo@istat.it