29 July 2026

Small area estimates of persons with severe functional limitations: an application by province, sex and age class[i]

1. Objectives of the work

This work describes a methodological experimentation aimed at producing highly disaggregated territorial estimates for the indicator of Prevalence of Persons with Severe Functional Limitations (GLF), collected through the latest available edition of the European Health Interview Survey (EHIS). The objective is to reliably estimate the prevalence of the phenomenon at provincial level, with disaggregation by sex and by two macro age classes (15–64 years and 65 years and over), by resorting to indirect estimation techniques based on small area models (Small Area Estimation – SAE). Such methods make it possible to overcome the limits of direct estimates alone, often not sufficiently precise in small domains, offering more solid and detailed information for the identification of health needs and the planning of care services in the territories.

The growing need for granular data to support public decisions, combined with the availability of new auxiliary sources at sub-regional level, has led Istat to intensify the development of advanced methodological solutions, aimed at improving the descriptive capacity of the statistical system and at strengthening the production of estimates at more detailed territorial levels. Sample surveys, such as those on health, constitute the basis for building health indicators, but their capacity to provide reliable estimates is limited to the regional level, due to the need to balance costs and sample size. However, in the health field, more detailed data are often necessary to monitor equity within regions and to plan the supply of services, especially in a context of demographic ageing. In Italy, the percentage of people over 75 (12.6%) exceeds the share of all EU countries, whose average is equal to 10.2% (provisional figure for 2024), making it even more relevant to have granular statistical information on the health and autonomy of the elderly population, in order to foresee the impact on welfare, traditionally supported by the family.

Sample surveys, such as EHIS, collect indicators in line with the ICF classification (International Classification of Functioning, Disability and Health) and adopt a bio-psycho-social approach, which also takes into account the environmental factors influencing social participation. However, sampling limits prevent reliable estimates from being obtained at sub-regional level, particularly in the smallest areas with scarce sample size.

The estimation method adopted uses data aggregated at provincial level, making it possible to produce reliable estimates in full compliance with the regulations in force on the protection of personal data and on statistical confidentiality, as provided for by Art. 9 of Legislative Decree No. 322/1989. Italian and European legislation also limits the use of data belonging to special categories of personal data, including health information, which may be processed only if provided for by specific provisions of law or regulation, with adequate measures to protect the rights of the data subjects (Art. 2-sexies, Legislative Decree No. 196/2003; Art. 9 of EU Regulation 2016/679).

In order to comply with these constraints, the information collected with the EHIS survey is used exclusively in aggregate form. The integration of aggregate data coming from the Integrated System of Registers (SIR) makes it possible to apply indirect estimation methods (Small Area Estimation, SAE), in line with the European recommendations (IESS Regulation No. 2019/1700, Art. 9), producing more detailed and precise estimates without linking personal data at elementary level.

This approach makes it possible to obtain interpretable and comparable values, avoiding the processing of sensitive data and guaranteeing that the disability condition, in particular, is not directly identifiable at individual level. The experimentation of SAE methods represents an opportunity to improve the quality of the estimates of the EHIS GLF indicator, while remaining fully within the regulatory constraints governing the access to and use of personal data.

The EHIS survey (European Health Interview Survey), carried out in a harmonized way in all the countries of the European Union, collects a wide range of internationally comparable indicators concerning health status, use of health services, prevention and the main determinants of health. These include the *functional limitations* — relating to sight, hearing, mobility and cognitive abilities — which constitute the objective of this first experimentation, as well as the *reduction of autonomy in the activities of daily living (ADL and IADL)*, for which SAE estimation methodologies are being tested and could subsequently be integrated into this experimental statistic. Such indicators, integrated with estimation techniques capable of producing trustworthy results also at provincial level, provide a relevant contribution to the strengthening of the territorial information offer. In particular, they make it possible to identify with greater precision the areas showing the greatest critical issues, favoring a more effective allocation of resources and a better support to social policies, based on evidence referring to territorial contexts more disaggregated than the traditional regional or national analyses.

Within this framework, the information derived from the EHIS survey could cyclically contribute to significantly widening also the information offer on the well-being of the

territories (BesT), the system developed by Istat to monitor equitable and sustainable well-being at sub-regional level (Istat, 2024).
The indicator of unmet care needs among the frail elderly could represent a further area of interest, also for monitoring aspects related to the new legislation on care for the elderly (Legislative Decree No. 29 of 15 March 2024). In summary, this experimentation, which currently focuses on the indicator *Severe Functional Limitations,* aims to provide an example of how it is possible to widen the information potential of health surveys, providing quality data at a more detailed territorial level, useful to better support health and social planning policies in an evolving demographic context and in order to calibrate such interventions on the basis of efficient and granular statistical information.

The final outcome of this experimentation makes it possible to feed the Istat Information System "Disabilità in Cifre" (Disability in Figures) with the indirect estimates of the indicator *Severe Functional Limitations* at provincial level, by gender and by macro age classes, in order to make them available to decision makers, researchers and interested users. The estimates presented highlight an infra-regional heterogeneity, offering a deeper insight into the informative picture beyond the average regional level.

1.1. Sampling design of the EHIS survey

The sampling design presents a general structure typically used in most of the social surveys on households carried out by the Italian National Institute of Statistics. This design is in fact based on a two-stage municipalities-households scheme, with a stratification of the municipalities according to their demographic size. The design of the 2019 survey was also integrated with that of the Permanent Population and Housing Census survey. The selected municipalities constitute a sub-sample of the 2850 municipalities present in the list sample selected for the round of the Permanent Census relating to 2018. The second-stage units are the households, drawn with a random selection criterion from the population registers for the sample municipalities with fewer than 1000 inhabitants and from the list of households selected for the 2018 Permanent Census for the sample municipalities with 1000 inhabitants and over. The final sample is composed of about 22,800 households, living in 835 municipalities of different sizes and distributed over the whole national territory.
The planned domains include the five Italian geographical macro-areas (North-West, North-East, Centre, South and Islands) according to the NUTS 1 classification, the Italian regions (NUTS 2) and the two autonomous provinces of Bolzano and Trento. Alongside these administrative domains, the so-called "wide areas" (aree vaste), were also included, sub-regional territorial aggregations made up of several contiguous municipalities sharing economic, social or service-related characteristics. These areas, often corresponding to aggregations of Local Health Authorities (ASL), show a higher functional homogeneity and are particularly useful for planning and monitoring purposes in the social and health care field. This concept, in fact, is often used in the health and socio-economic field to optimize

the management of public services and for statistical monitoring and planning purposes (ISTAT, 2018, Atlante statistico dei comuni, *https://aster.istat.it/*). The sampling design therefore stratified the population by broad area within each region, for a total of 67 areas over the whole national territory (with the exception of some regions which, over time, have concentrated all the health districts of the single ASLs into a single ASL: Marche, Molise and Sardinia), so as to be able to carry out possible sub-regional comparisons with the 2012 health survey, in which the wide area was a planned domain. For further details on the sampling strategy adopted, reference can be made to the methodological note (Istat, 2021).
The estimates produced by the survey are essentially estimates of absolute and relative frequencies, referring to households and individuals. The estimates are obtained through a calibrated estimator (Devaud D., Tillé Y., 2019; Deville J. C., Särndal C. E., 1992) with constraints on:

- distribution of the population in the 21 Italian regions (19 regions plus the autonomous provinces of Trento and Bolzano) by sex and seven age classes (0-14, 15-24, 25-44, 45-54, 55-64, 65-74, 75+);
- distribution of the population in the 5 territorial divisions by sex and nine age classes (0-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+);
- distribution of the population by citizenship (household of Italian citizens, household of foreign citizens, mixed household).

1.2. Some disseminated results

The regional-level results for the selected indicator show a rather articulated geography. The prevalences of *Severe Functional Limitations* (sensory, motor or cognitive) among people aged 15-64 are relatively low, equal to **3.5%**, corresponding to about **1.345 million** individuals, while among the over-65s the individuals with such severe difficulties in basic functions are about **3.86 million** (28.4% of the over-65 population). Regional differences are also relevant: the highest prevalences are found in Calabria, Sardinia and Sicily (35.6%, 35.2% and 34.6%), against the minimum value of Friuli Venezia Giulia (21.7%).
Among the approximately 3.9 million individuals over 65 with *Severe Functional Limitations*:

- 2.833 million (20.9%) have severe difficulties in walking or in climbing up and down stairs without the help of a person or aids;
- 1.874 million (13.8%) report severe difficulties in hearing or sight, even with aids;
- 1.113 million (8.2%) show severe difficulties in memory or concentration.

2. Methodology used

The provincial-level estimates by sex and for the two macro age classes (15-64 and 65 years and over) of the indicator relating to *Severe Functional Limitations* were obtained through the application of a well-established small area estimation methodology based on a mixed-effects model defined at area level (Fay-Herriot, 1979). This method makes it possible to combine the information coming from sample surveys with auxiliary data available at aggregate level, significantly improving the precision of the estimates where direct information alone is not sufficiently reliable.

Suppose we want to estimate an indicator for *D* subpopulations or domains. For each domain *d* (*d* = 1, 2, ..., D), we have a direct estimate of the indicator based on the sample survey, denoted by $\hat{\theta}_d$ and a vector of auxiliary variables X_d available from external sources (e.g. administrative or census data). The Fay-Herriot model is based on two components:

1. A **sampling model**, which describes the direct estimate as a function of the true value of the indicator and of a random sampling error:

   $$\hat{\theta}_d = \theta_d + e_d$$

   where e_d represents the sampling error, with zero mean and variance σ_e^2 known or estimated.

2. An **area-level regression model**, which expresses the true value of the indicator as a linear function of the auxiliary variables, plus a random area term:

   $$\theta_d = \boldsymbol{X}_d \beta + u_d$$

   where β is the vector of coefficients to be estimated and u_d is the random effect of area *d* with zero mean, variance σ_u^2 and independent of the sampling error.

By combining the two models, the overall mixed-effects model is obtained:

$$\hat{\theta}_d = \boldsymbol{X}_d \beta + u_d + e_d. \qquad (1)$$

The **resulting SAE estimator** is a weighted combination of the direct and the synthetic estimate,

$$\hat{\theta}_d^{sae} = \hat{\gamma}_d \, \hat{\theta}_d + (1 - \hat{\gamma}_d) \, \boldsymbol{X}_{\boldsymbol{d}}^{\boldsymbol{T}} \, \hat{\beta}, \qquad (2)$$

where $\hat{\beta}$ is the maximum likelihood estimate of β obtained under the normality assumption for the distribution of the sampling error and of the random area effect, and $\hat{\gamma}_d$ is the weight attributed to the direct estimator, given by:

$$\hat{\gamma}_d = \hat{\sigma}_u^2/(\hat{\sigma}_e^2 + \hat{\sigma}_u^2), \tag{3}$$

where:

- $\hat{\sigma}_e^2$ is the estimated variance of the direct estimates;
- $\hat{\sigma}_u^2$ is the variance of the random area effect estimated by the maximum likelihood method.

Equation (2) represents the empirical version of the best linear unbiased predictor of θ_d – EBLUP – under model (1), obtained by substituting the maximum likelihood estimates $\hat{\beta}$ e $\hat{\sigma}_u^2$ in the formula of the best linear unbiased predictor of θ_d. The latter is derived by minimising the mean squared error of prediction (Rao and Molina, 2015, chs. 5 and 6). In summary, the model adopted makes it possible to significantly improve the quality of small area estimates, thanks to an effective use of the auxiliary information available in the sub-domains of interest, integrated with the direct survey data.
In the application of the method, auxiliary variables were selected according to criteria of thematic coherence with the phenomenon of severe functional limitations, availability at provincial level with disaggregation by sex and by the two age classes considered, quality and stability of the source, territorial variability and statistical significance. The selection was carried out by combining theoretical evaluations, correlation analyses, checks on the predictive capacity of the covariates and stepwise procedures implemented with the R package embi, in order to strengthen the explanatory capacity of the model and to refine the predictions in the different segments of the population:

• the number of persons receiving a pension linked to a disability condition (invalidity pensions, work-injury benefits, civil invalidity, attendance allowance, war pensions), related to the population, used as a proxy of the territorial presence of population with recognised disability, since at the time an integrated national register on disability was not available;

• the hospital attractiveness index, derived from the Hospital Discharge Records (SDO) data processed by the Ministry of Health, as a synthetic indicator of the supply of and demand for complex health services. This index measures the capacity of an area to attract patients from other provinces or regions, reflecting hospital mobility flows and therefore represents an indicator of accessibility and quality of the health care supply.

In addition to the context variables at the level of the sub-domain of interest, the integration between demographic data and income information coming from administrative sources made it possible to enrich the characterisation of the reference population, providing further elements useful for the specification of the model. In particular, covariates referring to the sub-domains were considered, derived from the following information:

• distribution of the population by 7 age classes;
• distribution of the population by 3 educational levels (primary, secondary, university degree);
• rate of population at risk of poverty;
• quintiles of the equivalised income computed at national, regional and provincial level;
• distribution of the population by income source (work, pension, capital), grouped into 5 classes of increasing size;
• distribution of the population by 4 classes of the average number of weeks worked, obtained by dividing the year into quarters.

The integration between demographic data and income information coming from administrative sources made it possible to further enrich the characterisation of the reference population and to improve the specification of the model. The set of selected variables provides a coherent picture of the socio-demographic and territorial characteristics potentially associated with the distribution of severe functional limitations.

3. Implementation of the method and analysis of the results for the GLF indicator

The analysis presented in this section describes the estimation process of the indicator of *Severe Functional Limitations (GLF)* for unplanned domains, starting from the data collected within the latest available edition of the European Health Interview Survey (EHIS), referring to the year 2019. The aim is to produce reliable estimates for territorial and socio-demographic domains not directly represented by the survey, with a level of detail sufficient to meet the growing information needs linked to the planning of local health and social policies.

More disaggregated estimates are essential for several purposes: analysing socioeconomic and health inequalities across areas more accurately, particularly in the most vulnerable ones; assessing the care needs of frail older people and the burden of care, in relation to their level of functional independence; and supporting evidence-based planning at local level, especially with regard to the Essential Levels of Care (LEA) and the measures set out in the National Recovery and Resilience Plan
However, traditional sample surveys such as EHIS, while offering a solid information basis, present limits linked to the small sample size, which prevents the production of reliable estimates for territorial domains of small size, such as the provinces, especially when disaggregations by sex and age classes are required. This limit proved particularly critical in the 2019 edition of EHIS, where the reliability of the direct estimates remains

acceptable only at regional level, while at provincial level the uncertainty increases considerably.

To overcome these critical issues, the application of SAE estimation methods to the GLF indicator was tested, in order to produce estimates at provincial level, cross-classified by sex and by two macro age classes (15–64 years and 65 years and over), for a total of 428 domains. The approach adopted is characterised by some methodological innovations: the production of estimates for unplanned domains, i.e. in the absence of direct representativeness; the integration of auxiliary variables coming from administrative and demographic sources (INPS, censuses, registers); the use of functional transformations to improve the adherence of the model to the normality assumptions; and the use of benchmarking to guarantee the coherence of the indirect estimates with the official regional ones.

The direct estimates obtained show an acceptable error only at regional level, while at provincial level — especially when the disaggregation by sex and by the two macro age classes is considered — the uncertainty becomes too high. For this reason, the application of SAE methods allows remarkable improvements in terms of efficiency and precision. To assess the reliability of the direct estimates at such disaggregated levels, the following criterion based on the coefficient of variation (CV %) was adopted:

- estimates publishable without restrictions: CV % ≤ 16.6;
- estimates publishable with caution: 16.6 < CV % ≤ 33.3;
- estimates not recommended: CV % > 33.3.

The analysis of the direct estimates obtained from the 2019 EHIS survey for the GLF indicator highlights important critical issues in terms of precision. In particular, many estimates show high coefficients of variation, such as to compromise their reliability and the possibility of official dissemination. The classification of the direct estimates according to the CV reported in Table 1 shows that:

- only 83 estimates fall within the range considered publishable (CV ≤ 16.5%);
- 154 estimates are publishable with caution (16.5% < CV ≤ 33.3%);
- as many as 166 estimates turn out to be not publishable (CV > 33.3%);
- for 25 domains, no direct estimate could be computed due to the absence of sample.

Table 1 Direct estimates for the GLF variable at provincial level, Year 2019, classified according to the degree of precision

CV % of the estimates	Assessment	Estimates
[0; 16.6]	Publishable	83
(16.6; 33.3]	Publishable with caution	154
(33.3 and over]	Not publishable	166
Not available		25

These results highlight the inability of direct estimation methods to provide trustworthy and reliable estimates in the disaggregated domains, making the recourse to small area estimation methods indispensable. In this context, the application of SAE methods allows a substantial improvement of the coverage and of the quality of the information available for health monitoring in the territories, offering concrete support to the construction of evidence-based health and social policies. The results obtained within this experimentation are potentially replicable also for the next edition of the health survey (EHIS 2025), thus contributing to consolidating the systematic use of SAE methods in official production and to strengthening the national information system.

After having outlined the motivations and benefits linked to the use of SAE methods, we now turn to the statistical-methodological analysis of the results obtained. From the analysis of the estimates (Figure 1), it emerges that the values in the elderly population (65+) are generally higher than those relating to the 15-64 age group, with a partial overlap between the tails of the distributions. The distribution of the estimates thus appears as a combination of two distinct distributions for the two age classes. Figure 2 shows the distribution of the variable after a logarithmic transformation, which, while maintaining an asymmetry linked to the age classes, is on the whole more adequate than the distribution on the original scale.

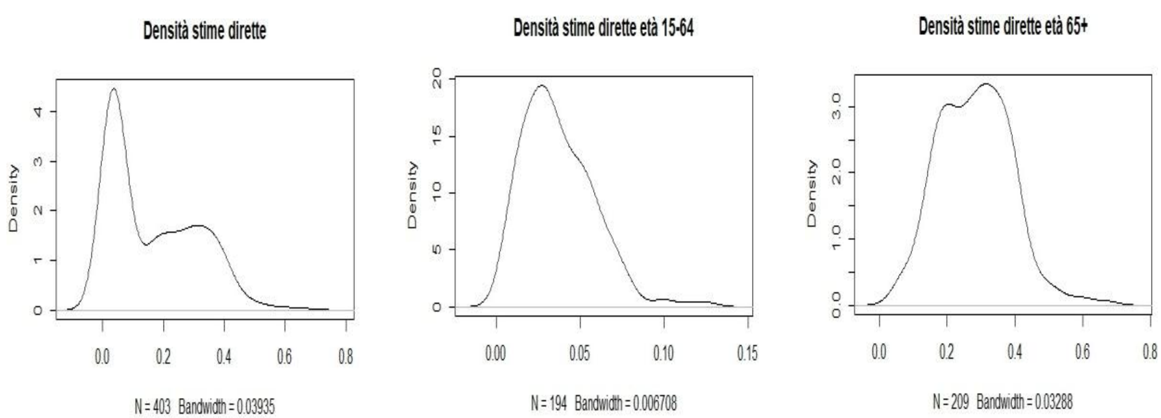

Figure 1: density of the direct estimates in the domains of interest

The Fay-Herriot estimator is based on the specification of the mixed-effects model described in Section 2, in which the sampling errors and the random effects are assumed to have zero mean and to be independent and normally distributed. However, the diagnostic analyses carried out on the standardized residuals and on the random effects show significant deviations from these assumptions. In particular, marked asymmetries are observed (skewness = –0.45 for the residuals and –1.56 for the random area effects), heavy tails (kurtosis equal to 5.24 and 8.64 respectively) and very low p-values in the Shapiro-Wilk tests (respectively < 1e-7 and < 1e-13), which indicate a clear violation of the normality assumption.

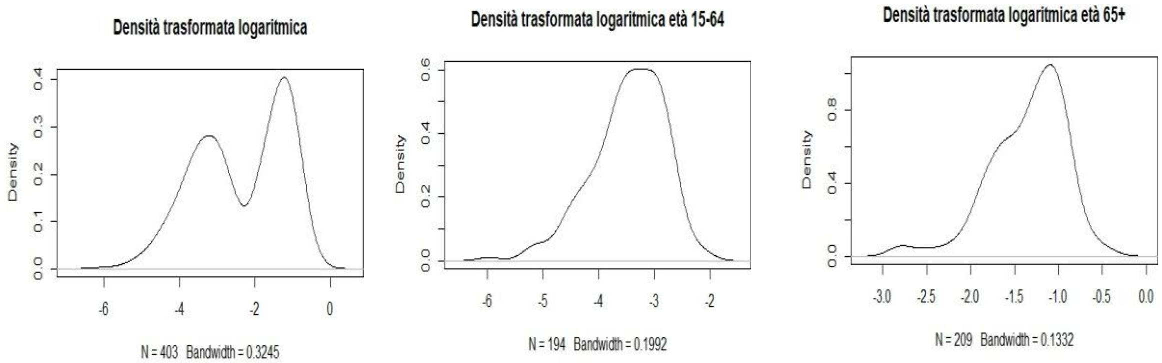

Figure 2: density of the log-transformed direct estimates in the domains of interest

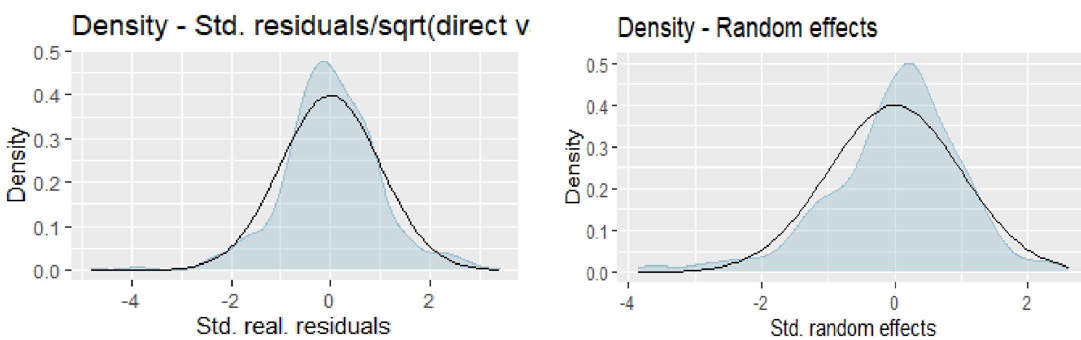

Figure 3: Distribution of the standardised residuals and random effects of the standard model.

A logarithmic transformation was applied to the direct estimates and their sampling variances to improve normality and stabilise the variance. Returning the model-based estimates to the original scale requires an inverse transformation, which is non-linear and therefore introduces bias. A bias-corrected back-transformation was consequently adopted, namely the crude correction based on the properties of the log-normal distribution proposed by Rao and Molina (2015, ch. 6) and implemented in the R package **emdi** through the `bc_crude` option.. Such a transformation contributed to obtaining a more symmetric distribution of the errors and of the random effects, reducing the impact of the outliers and improving the quality of the fit of the model. The results confirm the effectiveness of the intervention: both the residuals and the random effects show skewness values close to zero and more contained kurtosis. Moreover, the Shapiro-Wilk normality tests highlight a substantial improvement: the p-value for the standardised residuals rises to 0.066 (not significant), while for the random effects, although the p-value remains significant (0.002), a very relevant progress is observed with respect to the original model. These results are also illustrated in Figure 4.

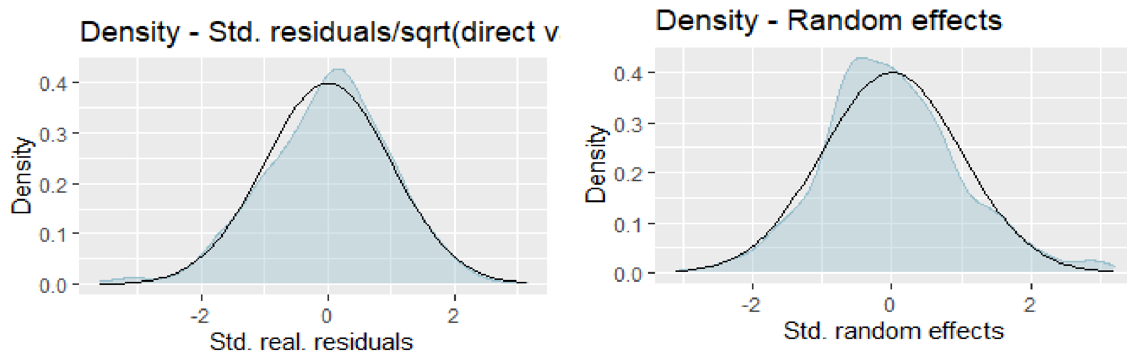

Figure 4: Distribution of the standardised residuals and random effects of the model with logarithmic transformation

The goodness-of-fit measures also confirm the improvement obtained with the application of the logarithmic transformation. In particular, the adjusted R^2, which takes into account the number of covariates included in the model and provides a more robust assessment of the predictive capacity, shows a significant increase: from 0.847 in the untransformed model to 0.920 in the transformed model. This increase indicates that a larger share of the variability observed in the direct estimates is explained by the model, confirming a better coherence between the data and the assumptions of the model itself. The increase in the adjusted R^2 reflects not only an improvement in the adherence of the model to the data, but also a greater efficiency in the estimation of the parameters, thanks to the greater stability of the variance and to the reduction of the influence of anomalous values.

In the domains, particularly when the sample size is small, the estimate of the variance associated with the direct estimates tends to be unstable and not very reliable. This instability is due to the fact that, with few data available, the sample variances may turn out to be randomly overestimated or underestimated, generating strongly fluctuating values from one area to another. This condition has significant effects in the Fay-Herriot model, where the γweights (defined in equation 3) regulate the balance between the direct estimate and the modelled component of the composite estimator (equation 2). In the presence of a lower estimated variance, the γ weight tends to increase the influence of the direct estimate; conversely, a higher variance attributes greater weight to the model component. However, if the variances are estimated in an unstable way, the resulting weights may be extreme or incoherent, reducing the overall robustness and precision of the estimator. For this reason, it is fundamental to adopt corrective strategies to stabilise the weights and to improve the reliability of the estimates produced, through a smoothing model of the variances associated with the direct estimates. After a selection procedure, the model employed was the following,

$$log[VAR\ (\hat{\theta}_d\ )^2] = \beta_0 + \beta_1 \log(n) + \beta_2 \log\ (\hat{\theta}_d\ ) \quad (4)$$

in which the sample variance is related to the sample size of the area and to the intensity of the observed phenomenon, making it possible to predict a more stable variance, less influenced by sampling noise. In other words, instead of directly using the observed variances, often too unstable, a “smoothed” estimate is obtained, which reduces the random variability and better reflects the underlying structure of the data.

As shown in Figure 5, the use of the smoothed variances contributes to stabilising the distribution of the γ weights between the two macro age classes considered. This leads to a composition of the weights less susceptible to random fluctuations, thus improving the adaptability of the model and increasing the reliability and precision of the estimates produced in the small areas.

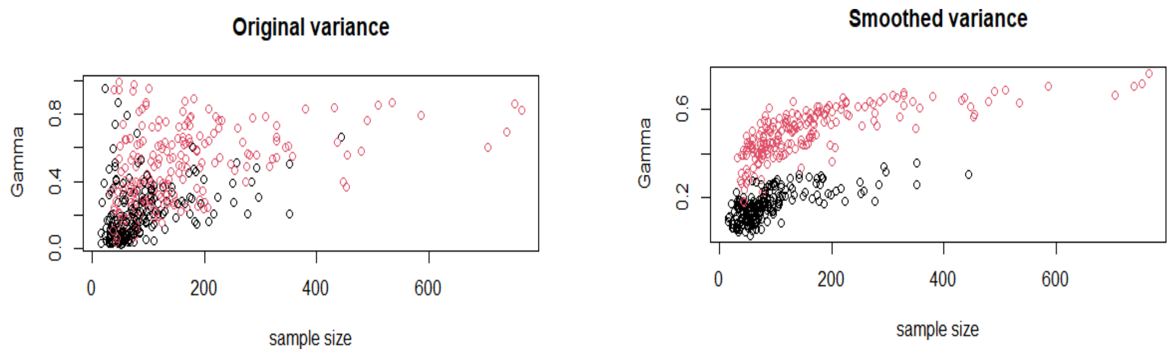

Figure 5: Values of the gamma coefficients (black: 15-64 class, red: over-65 class) with respect to the sample size of the domain for the FH model based on logarithmic transformation: original variance (left) and smoothed variances (right).

The comparison between the standard Fay-Herriot model and the one with logarithmic transformation (Figure 6) shows that the logarithmic transformation improves the coherence of the estimates, particularly in the areas with small sample size. The estimates obtained with the log-transformed model, moreover, allow further efficiency gains with respect to the application of the method on the original scale, guaranteeing greater stability and reliability of the estimates (Figure 7).

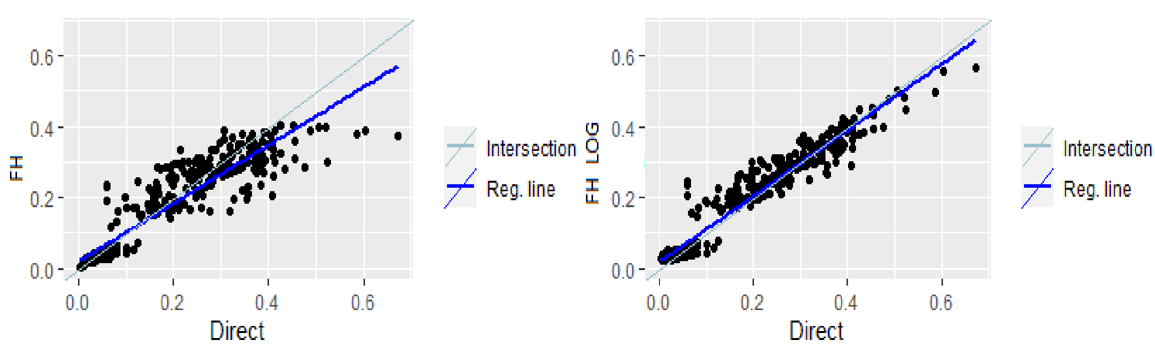

Figure 6: Behaviour of the FH and FH LOG estimates with respect to the direct estimates

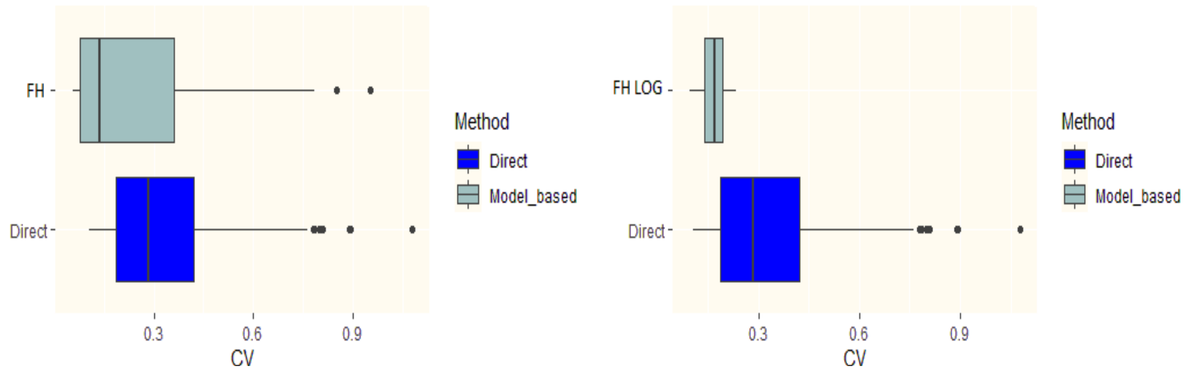

Figure 7: Distribution of the coefficients of variation of the FH and FH LOG estimates with respect to the direct estimates.

An essential aspect in the production of small area estimates is the guarantee of coherence among the results obtained at different levels of territorial disaggregation. To achieve this objective, the indirect model-based estimates were subjected to a benchmarking process, i.e. a realignment with respect to the corresponding direct estimates already published at regional level. This intervention not only ensures the alignment with the official estimates disseminated by the Institute in the planned domains, but also contributes to strengthening the robustness of the estimates produced with SAE methods in the unplanned domains. Moreover, a further element to be considered is the impact of benchmarking on the precision of the estimates. Table 2 reports the distribution of the coefficients of variation for three estimators — direct, SAE, and SAE with benchmarking — divided into the usual three precision classes:

- CV ≤ 16.6% (good precision),
- 16.6% < CV ≤ 33.3% (acceptable),
- CV > 33.3% (low precision).

The table clearly highlights the effectiveness of the SAE model, in particular of the one based on a logarithmic transformation of the parameter of interest. Indeed, all the estimates computed on the basis of this model (both SAE and SAE with benchmarking) fall within the acceptability threshold of the CV (< 33.3%), unlike the direct estimates, which present 166 cases with CVs above this threshold.

Table 2: Distribution of the CV% of the estimates classified according to three precision levels

Estimator	0-16.6%	16.6%-33.3%	>33.3%
Direct	83	154	166
SAE	197	231	0
SAE_BENCH	152	276	0

This result confirms that the logarithmic transformation significantly improves the precision of the small area estimates, making the distribution of the CVs more concentrated in the classes of higher reliability. It is important to underline that the

coherence process introduced very limited variations between the initial estimates (SAE) and the final ones (SAE_BENCH), proving the capacity of the adopted model to provide robust and efficient estimates, while maintaining a good coherence with the direct estimates.
The slight increase in the CVs observed after benchmarking is entirely coherent: it derives from the addition, in the computation of the MSE, of an extra-variability component introduced to take into account the uncertainty associated with the alignment of the SAE estimates with the direct ones. This represents a methodologically acceptable compromise, aimed at guaranteeing the coherence among the various estimation domains without excessively compromising the precision of the estimates.

In summary, the methodological process was articulated as follows:

• **Fay-Herriot model**: a linear mixed-effects model was adopted, defined on a specific transformation of the parameter of interest, in order to improve the compliance with the normality assumptions of the residuals and to stabilise the variance. This device made it possible to increase the robustness and reliability of the estimates produced.

• **Variance smoothing**: to address the instability of the sample variances, particularly marked in the small areas with small sample size, a smoothing model of the variances associated with the direct estimates was implemented. This model links the estimated variance to the sample size and to the intensity of the observed phenomenon, making it possible to obtain more stable variance estimates and to define more coherent weights of the SAE estimator, improving its fit and precision.

• **Integration and use of auxiliary sources**: auxiliary variables available at provincial level and classified by sex and by the two age classes were incorporated, selected for their explanatory significance with respect to the analysed phenomenon. These include demographic, socio-economic and health indicators coming from official sources.

• **Benchmarking of the estimates**: to guarantee coherence and compatibility among the different territorial levels, the indirect SAE estimates were subjected to a calibration process (benchmarking) with respect to the corresponding direct estimates available at regional level. This step ensured the preservation of the additivity between the provincial and the regional estimates, favouring a harmonious integration of the results with the official information system.

To complete the informative picture, marginal estimates can be obtained by aggregating those produced for more detailed domains defined by the cross-classification of province, sex and age class. This procedure makes it possible to provide reliable estimates also for unplanned domains, obtained through aggregation along one or more dimensions (for example by province only, sex only or age class only), guaranteeing the coherence with the estimation model adopted.

A particularly relevant case concerns the REG × Sex × Age class (15-64; 65+) domains, for which the direct estimates show relatively high coefficients of variation (CV for the 15-64 age class between 11% and 39%; CV for the 65-and-over age class between 5% and 26%). To improve the reliability also at this marginal level, the final small area provincial estimates can be aggregated, thus obtaining regional estimates classified by sex and macro age class, more efficient and coherent with the method and the estimation model used for the corresponding provincial estimates (Figure 8).

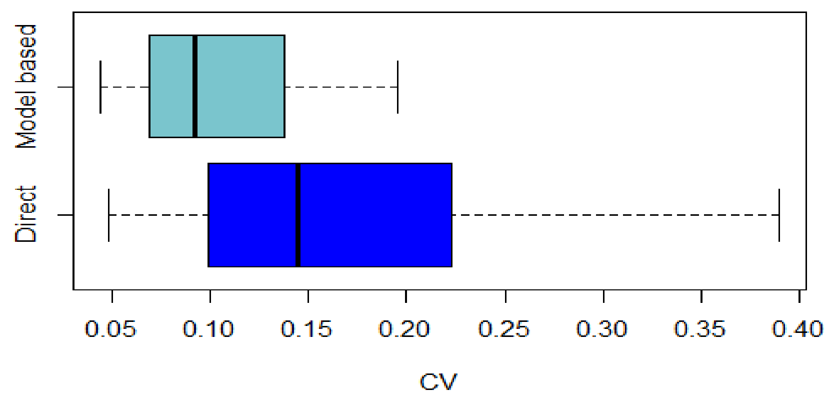

Figure 8: distribution of the CV% for the SAE – model-based – and direct estimates for the domains obtained as REGION x SEX x macro age class.

This approach makes it possible to maintain the coherence among the SAE estimates at different levels of disaggregation, avoiding the use of different models. The final estimates were obtained as weighted averages of the provincial estimates for each region, and the computation of the mean squared error (MSE) took into account the covariance among the provincial estimates within the same region. As shown in Figure 8, the estimates for the marginal domains thus obtained are characterised by CVs between 4% and 20%, showing a clear improvement with respect to the direct estimates.

4. Description of the results, assessment and validation of the SAE estimates

The results of the analyses carried out to assess and validate the final estimates produced for small areas are presented below. To verify the possible presence of bias attributable to the assumptions of the model, the SAE estimates were compared with the corresponding direct estimates, which — although affected by high variability — represent an unbiased estimate under the sampling design.

The linear regression analysis between the SAE and the direct estimates shows a strong and significant relationship: the slope coefficient is equal to 0.922, the intercept is close to zero and the coefficient of determination R^2 is equal to 0.956, confirming a high coherence of the results obtained with the two estimation approaches. As shown in Figure 9, the two subgroups relating to the two macro age classes considered (15–64 and

65+) are clearly distinguishable. The SAE estimates tend to concentrate around the bisector, with a slight compression of the extreme values, a typical feature of the shrinkage effect introduced by the model. This behaviour is coherent with what is expected: small area estimates may present a slight attenuation in the very high or very low values with respect to the direct estimates, without, however, evidence of systematic bias.

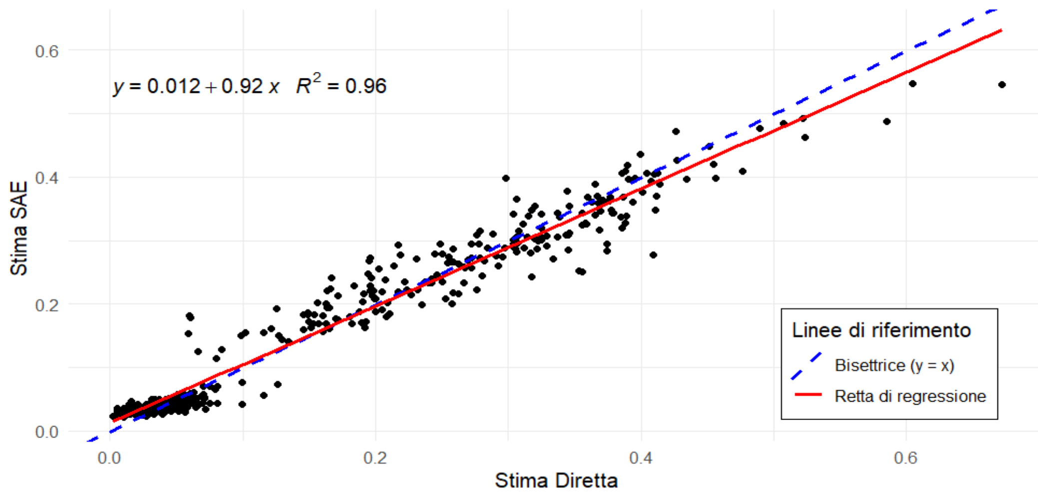

Figure 9 - Comparison between SAE and direct estimates.

Analysing the two age classes separately (Figure 10), it can be observed that in the 15-64 age group the relationship between the SAE and the direct estimates is less strong: the slope coefficient is equal to 0.367 and the R^2 drops to 0.55, indicating a higher variability with respect to the overall model. On the contrary, in the 65-and-over age group, the coherence between the two estimates turns out to be decidedly higher, with a slope coefficient of 0.763 and an R^2 equal to 0.872. This difference in performance can be partly explained by the different degree of precision of the direct estimates in the two age classes. In the 15-64 age group, in fact, the direct estimates tend to be more unstable not only because of the smaller sample size, but also because in this subpopulation the analysed phenomenon is far less frequent than in the 65+ subpopulation. The combination of these two factors – reduced sample size and low prevalence of the phenomenon – increases the statistical uncertainty, making the comparison with the model-based estimates more complex. On the contrary, in the elderly population, where the phenomenon is more widespread and the statistical signal stronger, the estimates are generally more stable and the correspondence between the two estimation approaches is higher.

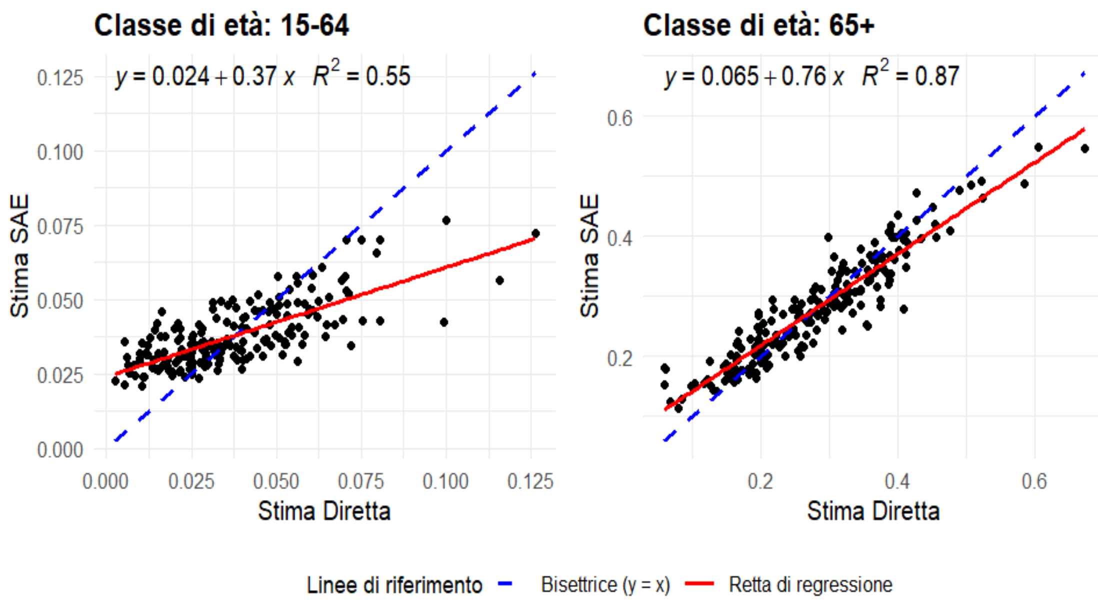

Figure 10: Comparison between SAE and direct estimates in the two macro age classes.

The analysis, further disaggregated by sex, is reported in Figure 11. When the models are separated by sex, for males in the 15-64 age group the regression coefficient is 0.309 with an R^2 of 0.516, while in the 65+ group the coefficient rises to 0.684 with a higher R^2, equal to 0.806. Females show a similar tendency: for the 15-64 age group the coefficient is 0.414 with an R^2 of 0.579, while for the 65+ group the coefficient increases to 0.718 and the R^2 reaches 0.864.

In all the comparisons, the p-value of the slope coefficient of the model is extremely significant, confirming that the relationship between the SAE and the direct estimates is solid in all the subgroups. This result suggests that the SAE estimates can be a good substitute for the direct estimates, with slight differences depending on the age class and on sex. The variations in the variability (R^2) among the subgroups suggest that the quality of the estimates might be slightly different, but the general model is confirmed to be very valid, especially for the 65+ group, where the fit of the model turns out to be particularly strong.

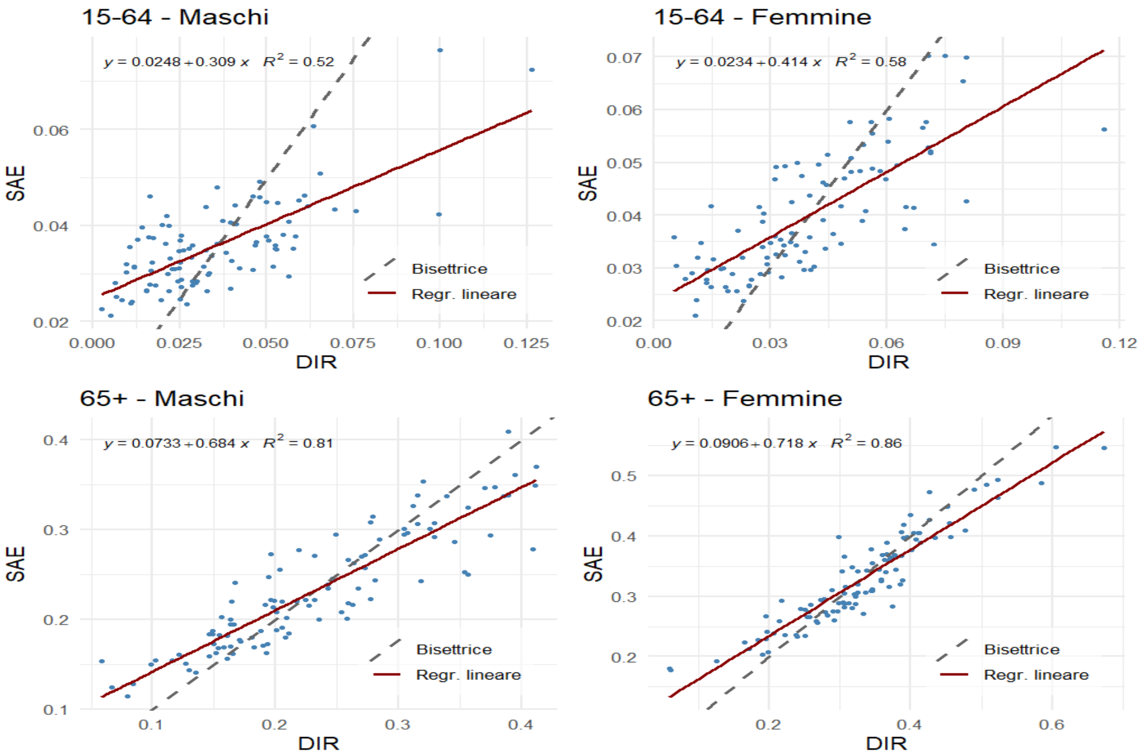

Figure 11 Comparison between SAE and direct estimates in the two age classes broken down by sex.

When small area estimation methods are applied, it is normal to observe differences between direct estimates and model-based estimates, especially in the domains characterised by high sampling instability. Such instability may derive both from the smaller sample size and from the lower diffusion of the phenomenon in the population considered, conditions which tend to produce direct estimates with high coefficients of variation.

In these cases it is therefore expected that the SAE estimates may deviate more from the direct estimates; this behaviour does not represent in itself a signal of model bias, but rather an intrinsic feature of small area estimation methods. If the analysis is based exclusively on the graphical comparison between direct and SAE estimates, possible signals of bias can be interpreted in a more informative way above all in the domains where the direct estimates are relatively more precise (i.e. with more contained coefficients of variation). For this reason, the previous graphical evaluation was complemented by a formal statistical verification, applying the Wald test of *goodness-of-fit* (Brown et al., 2001). This test makes it possible to formally verify the possible presence of a systematic bias component attributable to the model used, as it allows one to assess whether the differences between direct and SAE estimates are compatible with the statistical uncertainty associated with the two estimates. In particular, for each area the following contribution was computed:

$$GOF = (SAE - DIR)^2 / \left(Var(DIR) + MSE(SAE)\right)$$

where Var(DIR) represents the variance of the direct estimate and MSE(SAE) the mean squared error of the SAE estimate. The sum of such contributions approximately follows a χ^2 distribution with a number of degrees of freedom equal to the number of areas considered. High values of the test statistic would indicate the presence of discrepancies between the two types of estimates which cannot be explained by their statistical variability and therefore potentially attributable to bias introduced by the model

Applying the test to the whole set of territorial areas, no evidence of systematic bias emerges in the SAE estimates with respect to the direct estimates. The value of the Wald statistic turns out to be equal to 273.7 with 403 degrees of freedom (p-value ≈ 1), indicating that the differences observed between the two types of estimates are fully coherent with the level of statistical uncertainty associated with the estimates.

The analysis was then replicated separately for the two macro age classes considered. Also in this case no evidence of systematic bias emerges. For the 15–64 age group the Wald statistic is equal to 207.9 with 194 degrees of freedom (p-value = 0.235), while for the population aged 65 and over the statistic is equal to 65.8 with 209 degrees of freedom (p-value ≈ 1).

The analysis disaggregated by age class and sex confirms this result in almost all the subgroups. In particular, for females aged 15–64 the Wald statistic is equal to 84.1 with 97 degrees of freedom (p-value = 0.822), while for the 65+ population values equal to 31.3 (df = 105) for females and 34.6 (df = 104) for males are obtained, with p-values close to one.

A single deviation from the hypothesis of perfect coherence emerges for males in the 15–64 age group, for whom the Wald statistic is equal to 124 with 97 degrees of freedom (p-value = 0.035). This is, however, a borderline result, not unusual in applications of small area estimation methods, especially in the presence of direct estimates characterised by high sampling instability. A more in-depth diagnostic analysis of the individual contributions to the test statistic shows that this result is almost entirely attributable to a single territorial area, the province of Pordenone, characterised by a particularly unstable direct estimate and by a very high coefficient of variation. Excluding this observation, the test statistic for the males 15–64 group is reduced to 113.5 with 96 degrees of freedom (p-value = 0.108), thus showing no indication of systematic bias of the model also for this subgroup.

On the whole, the results of the test confirm that no evidence of systematic bias of the SAE estimates with respect to the direct estimates emerges. The observed differences are mainly attributable to the greater instability of the direct estimates in some domains characterised by reduced samples and by a lower diffusion of the phenomenon, conditions under which it is physiological that the model-based estimates deviate more from the direct ones.

In addition to the comparison between direct and SAE estimates, the validation of the SAE estimates was also based on the comparison of the territorial ranks deriving from the direct estimates and from the SAE estimates. The comparison of the ranks makes it possible to assess not only the values of the indicators, but also their relative position within the national ranking. Discrepancies in the ranks represent a useful indicator to identify potential inefficiencies in the direct estimates, often characterized by high variability in the small domains due to reduced samples. In parallel, such differences may highlight limits or assumptions not fully satisfied in the adopted SAE models, thus providing useful diagnostic elements for the evaluation and optimization of small area estimation methodologies.

The analysis of the rank correlations between the direct and SAE estimates for the GLF indicator showed in table 3 , provides important indications on the coherence between the two sets of estimates in the different population subgroups. The overall correlation between the ranks of the two types of estimates is equal to 0.806, indicating a good concordance. This suggests that the SAE methodology is able to preserve the territorial distribution observed in the direct estimates.

Table 3 Overall and subgroup rank correlation.

Sex	Age Class	Level	Rank Correlation (Spearman)
F	15-64	Subgroup	0.775089
F	65+	Subgroup	0.923917
M	15-64	Subgroup	0.587892
M	65+	Subgroup	0.906252
F	All	Marginal by sex	0.853387
M	All	Marginal by sex	0.757781
All	15-64	Marginal by age	0.681853
All	65+	Marginal by age	0.915176
All	All	Overall	0.805626

Analysing the data by sex, a higher coherence emerges among females ($\rho = 0.853$) with respect to males ($\rho = 0.758$). Such a difference might reflect a higher reliability of the SAE estimates for the female population or, on the contrary, a higher variability of the male data, which the model manages to capture with less precision.

The most interesting picture emerges from the disaggregation by age class. The 65+ group shows a very high correlation ($\rho = 0.915$), confirming the robustness of the SAE model in a population segment in which the GLF indicator is more widespread and the direct estimates turn out to be relatively more reliable. On the contrary, in the 15–64 age group the correlation drops to 0.682, indicating a lower overlap between the rankings produced by the two estimators. This discrepancy can be partly attributed to a greater instability and imprecision of the direct estimates, due to the lower prevalence of the phenomenon at younger ages.

The analysis by combined subgroups of sex and age provides further elements of reflection. The highest correlations are observed among the elderly: 0.924 for women and 0.906 for men, confirming the solidity of the SAE estimates in this group. On the contrary, among subjects aged 15–64, the correlation is still good for women (0.775), while for men it drops to 0.588. What has been observed also emerges clearly in Figure 12, which shows how the rank differences between direct and SAE estimates are more marked in the first age class, with a more pronounced effect for male individuals within the 15-64 group.

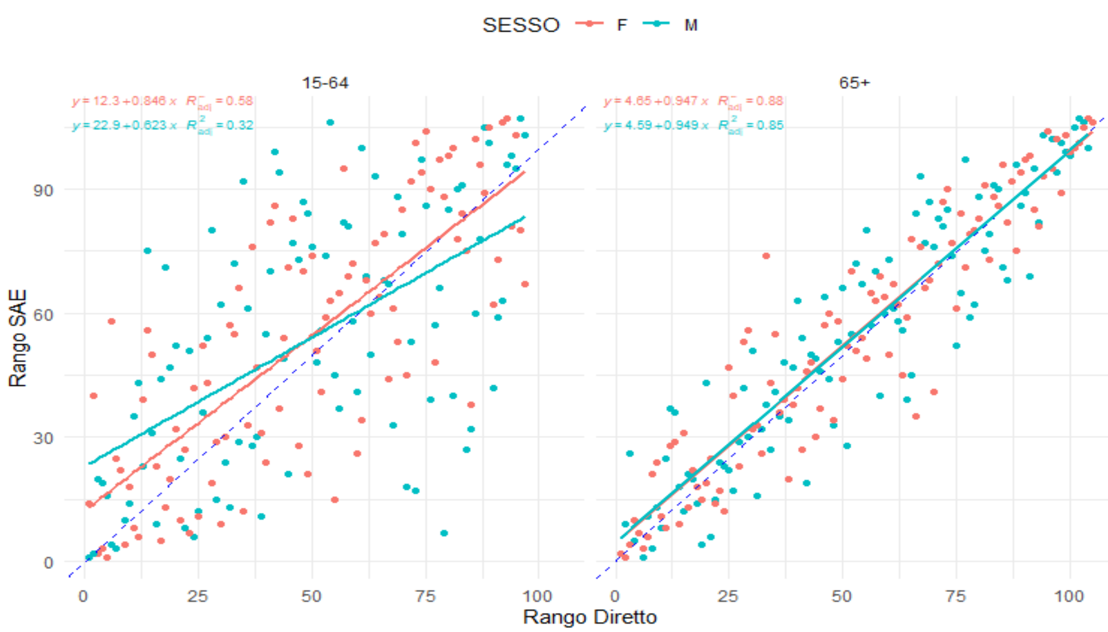

Figure 12 Comparison of the ranks of direct vs SAE estimates by Age class and Sex.

To assess whether the absolute rank differences between direct and SAE estimates depend on the precision of the estimates, measured through the coefficient of variation (CV), a non-parametric test and a regression model were applied. The non-parametric Kruskal-Wallis test makes it possible to compare the distribution of the rank differences among three CV classes, defined by dividing the CV into the intervals [0, 0.16], (0.16, 0.33], (over 0.33) and named respectively “Low”, “Medium” and “High”, verifying whether at least one class differs significantly from the others. In parallel, the linear regression model with the rank differences as response variable and predictors such as the estimated variance of the direct estimates, the sample size and the reference population makes it possible to quantify and isolate the effect of each factor on the rank change. The analysis of the results suggests that the rank differences between direct and SAE estimates vary significantly among the different CV classes; the Kruskal-Wallis test has a very low p-value (4.019e-07), indicating that at least one of the CV classes has rank differences significantly different from the others. The regression model, instead, deepens this relationship by considering also other variables such as the sample variance of the direct estimates, the number of sample units, the value of the direct estimate and the surveyed population. From the results can be observed that:

- The sample variance has a positive and significant effect on the rank difference (coefficient 530, $p = 0.0168$), confirming that a higher uncertainty in the direct estimates is associated with wider shifts in the ranks.
- The direct estimate has a negative and highly significant effect (coefficient −36.25, $p < 0.001$): areas with higher direct estimates are associated with smaller changes in rank than areas with lower ones.
- The number of sample units and the surveyed population do not turn out to be significantly associated with the rank difference.

In summary, both the non-parametric test and the regression model indicate that the precision of the direct estimates (represented by the CV and/or by the variance) is a key factor in determining how much the relative position of the domains changes when moving from the direct estimates to the SAE estimates. Less precise direct estimates are more subject to rank variations, while the value of the direct estimate plays an inverse role in modifying these differences.
Figure 13 illustrates the variation of the positions in the ranking between the direct and the SAE estimates, divided by the three classes of the coefficient of variation (CV). It can be clearly observed that the most marked changes in the ranks occur for the direct estimates with higher CVs, i.e. those characterised by a greater instability and a lower precision.

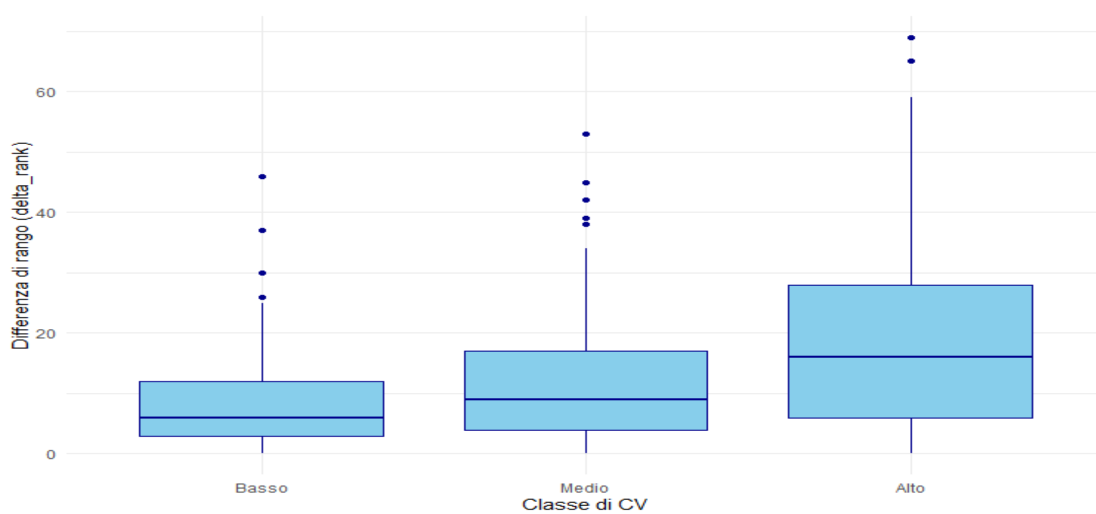

Figure 13 Rank difference by CV classes of the direct estimates.

In summary, the application of the SAE methodology makes it possible to improve the reliability of the estimates in the small domains while maintaining a high coherence with the direct estimates and preserving, on the whole, the territorial ranking. The estimates turn out to be particularly solid for the elderly population, while in some subgroups characterized by greater sampling instability — such as men under 65 — the comparison between direct and model-based estimates may present a relatively higher variability, without evidence of systematic bias.

5 Conclusions

The methodological experimentation presented constitutes a relevant example of how it is possible to increase the granularity of official statistics production for equitable and sustainable well-being indicators, indicating a possible evolution of official statistics, in this case in the health dimension. The application of SAE methods to the Severe Functional Limitations indicator made it possible to fill an important information gap, offering reliable territorial estimates at provincial level and with disaggregations by sex and age class.

In a demographic context characterized by profound changes and by a structural ageing, the availability of trustworthy data on the health and autonomy of the elderly population represents a fundamental requirement to guide health and care policies. Direct survey estimates alone, due to the sampling limits in the small domains, would not be able to meet this growing need. The experience carried out here demonstrates, instead, that the adoption of indirect estimation techniques and the integration with auxiliary sources — above all of an administrative nature — can strengthen and enhance the information assets already collected, improving at the same time the territorial coverage and the quality of the estimates. All this in full compliance with privacy regulations, thanks to the exclusive use of aggregate data.

The innovation introduced is twofold. On the product side, new detailed statistical evidence is made available, capable of returning infra-regional differences which are crucial for the planning of services. On the process side, the systematic use of advanced estimation methodologies is affirmed, which increases the adaptive capacity of the statistical system with respect to emerging information needs.

From the institutional point of view, the experimentation strengthens the role of Istat in responding to new knowledge challenges and paves the way for future developments: SAE methods may be extended not only to other health indicators, such as the loss of autonomy in daily activities (ADL/IADL), but also to other domains of well-being and society, using data coming from the main Istat surveys, such as the Labour Force Survey, EU-SILC and AVQ, thus contributing to enriching the information offer of the "Benessere dei Territori" (BesT) system.

Looking ahead, the regular use of SAE methods may constitute a stable pillar of official statistical production. Limiting ourselves to health indicators, such techniques will be able to concretely support territorial planning, the monitoring of the Essential Levels of Care (LEA) and the evaluation of the planned interventions. This experimentation also takes on a strategic value in the European context, being fully consistent with Regulation 2019/1700 (IESS), which encourages the use of innovative methods to improve the coverage of the harmonised indicators at Union level. The experience gained by Istat thus contributes to strengthening the role of Italy in the context of European social statistics. Finally, the results obtained are potentially replicable already starting from the next edition of the health survey (EHIS 2025), consolidating the use of SAE methods in official production and durably strengthening the capacity of the national information system to respond to increasingly complex demographic and social challenges.

References

Brown G., Chambers R., Heady P., Heasman D. (2001) Evaluation of small area estimation methods: an application to unemployment estimates from the UK LFS. In: *Proceedings of statistics Canada symposium*. Statistics Canada, vol 2001, pp 1–10

Deville J. C., Särndal C. E. (1992), Calibration estimators in survey sampling, *Journal of the American Statistical Association*, 87(418), 376-382.

Devaud D., Tillé Y. (2019), Deville and Särndal's calibration: revisiting a 25-years-old successful optimization problem, *TEST*, 28(4), 1033-1065.

Fay R. E., Herriot R. A. (1979), Estimates of income for small places: an application of James-Stein procedures to census data, *Journal of the American Statistical Association*, 74(366), 269-277.

Eurostat (2018), European Health Interview Survey: Methodological manual, URL: https://ec.europa.eu/eurostat/documents/3859598/8762193/KS-02-18-240-EN-N.pdf/5fa53ed4-4367-41c4-b3f5-260ced9ff2f6?t=1521718236000

Istat (2021), Condizioni di salute e ricorso ai servizi sanitari in Italia e nell'Unione Europea - Indagine EHIS 2019 - Nota metodologica, URL: https://www.istat.it/ws/fascicoloSidi/553/Nota%20metodologica.pdf

Istat (2024), Il Benessere equo e sostenibile dei territori. Report regionali – Anno 2024, URL: https://www.istat.it/comunicato-territoriale/il-benessere-equo-e-sostenibile-dei-territori-report-regionali-anno-2024/

Ministero della Salute (2024), *Monitoraggio dei Livelli Essenziali di Assistenza (LEA)*, URL: https://www.salute.gov.it/new/it/tema/livelli-essenziali-di-assistenza/il-sistema-di-monitoraggio/

Rao, J. N., Molina, I. (2015). *Small area estimation*. John Wiley & Sons.

[i] This methodological note was edited by Michele D'Alò. Section 1 was written by Lidia Gargiulo and Laura Iannucci; Sections 2, 3, 4 and 5 by Michele D'Alò. The SAE estimates of the indicator under examination are the result of a team effort coordinated by Michele D'Alò, to which Maria Giovanna Ranalli (University of Perugia), Andrea Fasulo and Francesco Isidori (DIRM/DCME/MEB) contributed, with the thematic support of Isabella Corazziari (DIRM/DCME/MEB), Lidia Gargiulo e Laura Iannucci (DISD/DCSW/SWC).