Vienna Yearbook of Population Research 2025 (Vol. 23), pp. 69–82
https://doi.org/10.1553/p-cp2h-ne2d

DEBATE

Assessing national vs subnational population heterogeneities in a global context

Jesús García-Gómez[1,2], Juan Galeano[1,2] and Albert Esteve[1,2]

ABSTRACT Quantitative demographic research is often conducted at the national level, largely due to data limitations and a lack of hypotheses about within-country variation. Yet national-level data can obscure important internal diversity, leading to potentially misleading conclusions about demographic patterns. Using the CORESIDENCE Database, we analyse subnational variation in household arrangements across 142 countries divided into 1788 subnational units. We study four key household variables: household size, number of children, number of spouses or partners and number of other members. Results reveal substantial internal heterogeneity, particularly in Africa and Asia, where subnational variation often rivals or exceeds national-level variation. In contrast, European countries show greater homogeneity, making national data more representative. These findings emphasise the importance of integrating subnational data into demographic research. Relying solely on national-level indicators risks masking meaningful demographic patterns and misrepresenting population dynamics. As traditional census data become increasingly scarce, fine-grained data are crucial for capturing the complexity of demographic variation within countries.

KEYWORDS Subnational data • Household arrangements • Household structures • Population heterogeneities

Introduction

Quantitative research in social sciences and demography, often based on censuses and surveys, and on nationally representative data, is typically conducted at the national level. This is partly due to data limitations that prevent fine-grained disaggregation, and partly because clear hypotheses about within-country differences are often lacking. However, extensive evidence demonstrates the importance of adopting a subnational perspective in demographic studies (Canudas-Romo et al., 2024; Kashnitsky et al., 2017), moving beyond the traditional urban-rural dichotomy (Kulu and Boyle, 2009; Riederer and Buber-Ennser, 2019). A pivotal example is the influential study on fertility decline in Europe, which leveraged subnational data (Coale and Watkins, 1986) to uncover substantial regional variations, providing a more detailed understanding of the demographic transition that both confirmed and challenged elements of the original demographic transition theory.

✉ Jesús García-Gómez, jgarciagomez@ced.uab.es
1 Universitat Autònoma de Barcelona (UAB), Bellaterra, Spain
2 Center for Demographic Studies (CED), Bellaterra, Spain

More recently, research examining the relationship between development and fertility has similarly relied on subnational data to refine theoretical models (Fox et al., 2019). However, outside Europe, comparative studies at the subnational level remain scarce, with a few notable exceptions (Crombach and Smits, 2022; Gershman and Rivera, 2018; Lesthaeghe and Esteve, 2016; Pezzulo et al., 2017).

Examining the factors underlying these subnational differences is crucial for understanding the role of contextual and macro-social influences. For example, why do some regions within a country exhibit different fertility levels or variations in life expectancy? Are these differences driven by economic factors, long-standing cultural legacies or structural aspects such as urbanisation, age composition or educational attainment? When subnational disparities persist even after controlling for individual characteristics, uncovering their origins becomes particularly valuable for theoretical advancement. For instance, Eurostat data for the year 2023 (Eurostat, 2024) show that fertility rates in Spain vary between regions, ranging from 0.84 in the Canary Islands to 1.36 in the Region of Murcia. Similarly, in Germany, fertility rates differ significantly between regions, ranging from 1.14 in Leipzig to 1.53 in Schwaben. Life expectancy also shows notable regional variation. In France, there is a life expectancy difference of 3.7 years between Nord-Pas-de-Calais and Île-de-France, while in Italy, life expectancy ranges from 81.6 years in Campania to 85.1 years in the Autonomous Province of Trento. In some cases, regional differences within a single country may even exceed those between neighbouring countries. For instance, in 83 federal regions of Russia in 2017, life expectancy for men ranged from 60.3 years in the Chukotka autonomous area to 78.6 years in the Republic of Ingushetia (Scherbov et al., 2019); while in Turkey in 2023, the total fertility rate varied from a low of 1.13 in Bartın to a high of 3.27 in the province of Şanlıurfa (Turkish Statistical Institute, 2024). In other cases, subnational differences may be much subtler. For example, the total fertility rate in Slovenia ranges from 1.43 in Zahodna to 1.61 in Vzhodna, while the total fertility rate in Denmark varies from 1.42 in Hovedstaden to 1.57 in Sjælland (Eurostat, 2024).

Building on this argument in favour of taking the subnational scale into account, this article presents a focused study on household structure across the globe. The aim is to identify which regions exhibit the greatest heterogeneity, and to assess whether heterogeneity at the national scale is greater or lesser than heterogeneity at the subnational scale.

Goal of the paper

Which countries and regions exhibit the highest subnational variation in household arrangements? To what extent does this variation call into question the representativeness of national-level data? And, do the available household data allow us to adequately address these questions? These are critical considerations for demographers and social science researchers conducting cross-national research. A key challenge lies in defining the study's geographic scope (i.e., the number of countries or regions) and selecting the appropriate unit of analysis (countries, regions or subnational units). Although these decisions would ideally be driven by scientific rationale or theory, most cross-national studies default to using the country as the unit of analysis, often leaving subnational heterogeneity unexplored.

https://doi.org/10.1553/p-cp2h-ne2d

To tackle these issues and identify the most appropriate scale for large-scale comparative studies, this article presents a straightforward analysis assessing global heterogeneity, and comparing national-level and subnational-level variation. Our analysis begins by examining internal variation within countries. To contextualise subnational differences, we then measure national-level variation across broad subcontinental regions and compare the two. We argue that the subnational level becomes increasingly relevant and necessary when internal variation significantly exceeds variation at the national level. To capture potential geographic differences, we group countries into continental regions and subregions for our analysis.

We draw on data from the subnational dataset of the CORESIDENCE Database (SN-CoDB) (Galeano et al. (2024) for a detailed description). The CORESIDENCE Database is an original resource that enables comparative analysis of household arrangements at the subnational level from a global perspective. We focus on household arrangements not only due to the extensive available data on these arrangements, but also because they are important for understanding family dynamics. Our analysis includes various characteristics of households, such as size and internal composition, providing a comprehensive overview of household structures across different societies. Households, which are typically composed of kin, play a central role in economic activities and are vital for individual socialisation. Thus, households are essential for both social and economic reproduction, reflecting and shaping cultural and economic dynamics.

Data and methods

The CORESIDENCE Database

The SN-CoDB provides 176 family and household indicators for a total of 719 country-year samples, covering 149 countries that are subdivided into 3950 subnational units. For our study sample (Figure 1), we selected the most recent sample dated after the year 2000 (with the only exception being Iraq, for which the most recent available sample dates back to 1997), for a total of 142 countries across the world. These samples come from different data sources, with most coming from the Demographic and Health Surveys (DHS) (ICF, 2023) program (61 samples), the International Integrated Public Use Microdata Series (I-IPUMS) (Ruggles et al., 2024) (34 samples) or the Multiple Indicator Cluster Surveys (MICS) (UNICEF, 2023) program (27 samples). Finally, for European countries only, we use 10 samples from the European Union (EU) Labour Force Survey (EU-LFS) (Eurostat, 2023a) and nine samples from the EU Statistics on Income and Living Conditions (EU-SILC) survey (Eurostat, 2023b), while for Australia, we use data from the Household, Income and Labour Dynamics in Australia (HILDA) (Wilkins, 2021).

We use four levels of analysis by aggregation. The highest level is the continental region. We consider five continental regions (Africa, America, Asia, Europe and Oceania), which are divided into 17 subcontinental regions (see Figure 1 for the list of subcontinental subregions). Continental groupings follow the classifications developed by the United Nations, the DHS program and the European Union Statistical Office. The third level is the country,

https://doi.org/10.1553/p-cp2h-ne2d

Figure 1 Distribution of country samples used in the analysis by continental and subcontinental region, data source and year

Source: Subnational dataset of the CORESIDENCE Database, drawing on data from DHS, I-IPUMS, MICS, EU-LFS, EU-SILC and HILDA.

https://doi.org/10.1553/p-cp2h-ne2d

and the fourth level is the subnational unit. In this study, we use the first administrative level of internal division. The number of subnational units per country varies widely. The number of subnational units is positively associated with the country's population (Pearson correlation coefficient equals 0.62; coefficient of determination equals 0.38), but is moderately associated with the country's area (Pearson correlation coefficient equals 0.36; coefficient of determination equals 0.13). Altogether, our study is based on 142 countries and 1788 subnational units. The average number of subnational units per country is 16.6, the median number is 10, and the interquartile range goes from six to 15.

Variables of interest

We study national and subnational variation according to four household characteristics. The *first* variable is household size, which simply measures the number of people living in the household. As we rely on the definition of household used in each data source, the difference between the *de jure* and the *de facto* definition could affect cross-country comparability. However, the comparison of subnational units is not affected. The *second* variable is the number of children of the household head (reference person or householder, depending on the terminology used in different sources), or of the household head's spouse or partner, living in the household. The *third* variable is the number of marital spouses or partners of the head living in the household. Since polygamy is not practised in most countries, this variable could be included as a dichotomous variable (presence or absence of the household head's spouse or partner in the household). However, we decided to include it as a numerical variable to reflect the fact that in 46 subnational units (mostly in Western Africa and Southern Asia) and two countries (Burkina Faso and Niger), the average number of marital spouses or partners of the head living in the household is greater than one. Thus, if this variable were dichotomised, it would obscure the existing diversity. Furthermore, the calculation process explained later is not affected by this decision, meaning that the results for most of the sample, in which polygamy is not practised, would be identical regardless of whether this variable was treated as numerical or dichotomous. The *fourth* variable is the number of other household members (both relatives and non-relatives). This category includes individuals who are not classified as the head, a spouse or partner of the head or a child of the head.

Analytical strategy

To measure national and subnational variation, we use the coefficient of variation (CV). The CV is a standardised measure of dispersion that is defined as the ratio of the standard deviation to the mean. It fits our purposes because it can be interpreted as the mean deviation from the mean, expressed as a percentage of the mean. For instance, if the distribution of the average household size in four subnational units is four, five, five and six; the average household size is five, the mean deviation to the mean is 0.5 and the CV is 10%. The CV is a dimensionless number, so it is appropriate when comparing variation across

https://doi.org/10.1553/p-cp2h-ne2d

different variables, or samples, with different means. More importantly, it is also useful when comparing variation at different levels of spatial aggregation (see Figure 4): first, it reports the variation at each level of analysis in an easily interpretable way; second, it allows for straightforward comparisons between the variation existing at each level. We replicated the analyses using other measures of variation such as ANOVA, focusing on the inter- and intra-group variance, or multilevel modelling, looking at the intraclass correlation (not shown). The results obtained using these alternative approaches were consistent with the CV approach, and we opted for the latter because of its reported advantages.

We first examine the CV for each country across the four analysed household variables (results shown in Figure 3). The national value of each variable is treated as the mean value for each country. Subsequently, the population-weighted mean of the subnational units' relative variation to the national value is calculated. We consider weighting by population appropriate to control for the possible effect of small countries or subnational units masking general trends observed in a country or continental or subcontinental region. Unweighted analyses yield similar results leading to the same conclusions. In the second step, we compare variation at both the subnational and the national level, with Figure 4 illustrating each continental and subcontinental region. The x-axis represents the CV by subnational unit. To obtain the CV by subnational unit, we calculate the population-weighted mean of the national CV (calculated for Figure 3) of each continental and subcontinental region. The y-axis represents the CV by country. To obtain the CV by country, the continental and subcontinental region values are treated as the mean values for each group of countries. Subsequently, the population-weighted means of the countries' relative variation to the continental and subcontinental region values are calculated. Table 1 summarises the information of Figure 4 to aid in the interpretation of the results. Specifically, Table 1 presents the ratio between the variation at the national and the subnational level for each continental and subcontinental region. This ratio is provided for both the four household arrangement variables and their combined total.

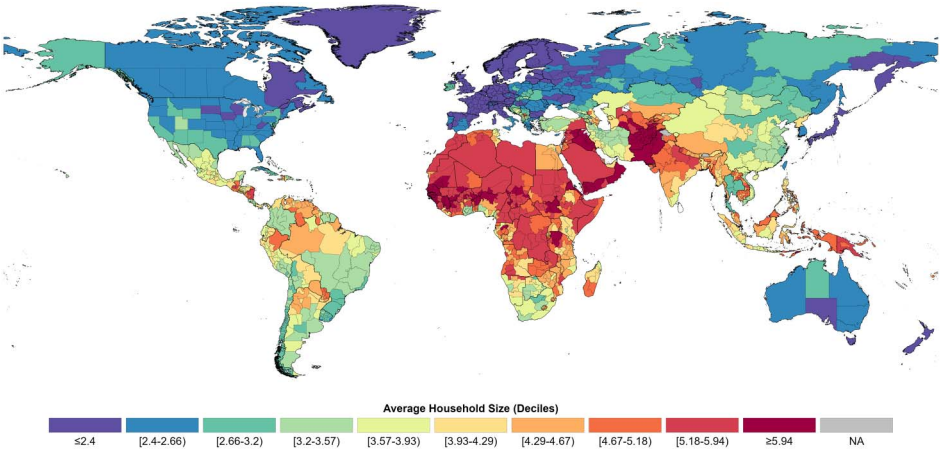
Results

To show the geographic scope of the CORESIDENCE Database and the regional distribution of our main variable of interest, we present a world map at the subnational level of the average household size in Figure 2.[1] While the national-scale version of the same map has already been published (Esteve et al., 2024), the higher level of detail in this version enables us to test a whole new set of hypotheses. For context, the range of household size amplitude varies from 1.9 (Finland) to 8.4 (Senegal) at the national scale, and from 1.7 (Berlin, Germany) to 13.2 (Basse, Gambia) at the subnational scale.

1 Note that while the analyses are conducted using the subnational dataset of the CORESIDENCE Database, Figure 2 uses the subnational harmonised dataset. Although both datasets are identical for most countries, the harmonised dataset includes a reduced number of subnational units for certain countries for geographical representation reasons (see Galeano et al. (2024) for a detailed description of both datasets). To provide contextual information, we use data from the United Nations (2022) database for countries for which we do not have CoDB data at the subnational level.

https://doi.org/10.1553/p-cp2h-ne2d

Figure 2 Average household size by subnational unit or country. Most recent observation since the year 2000 in the CORESIDENCE Database



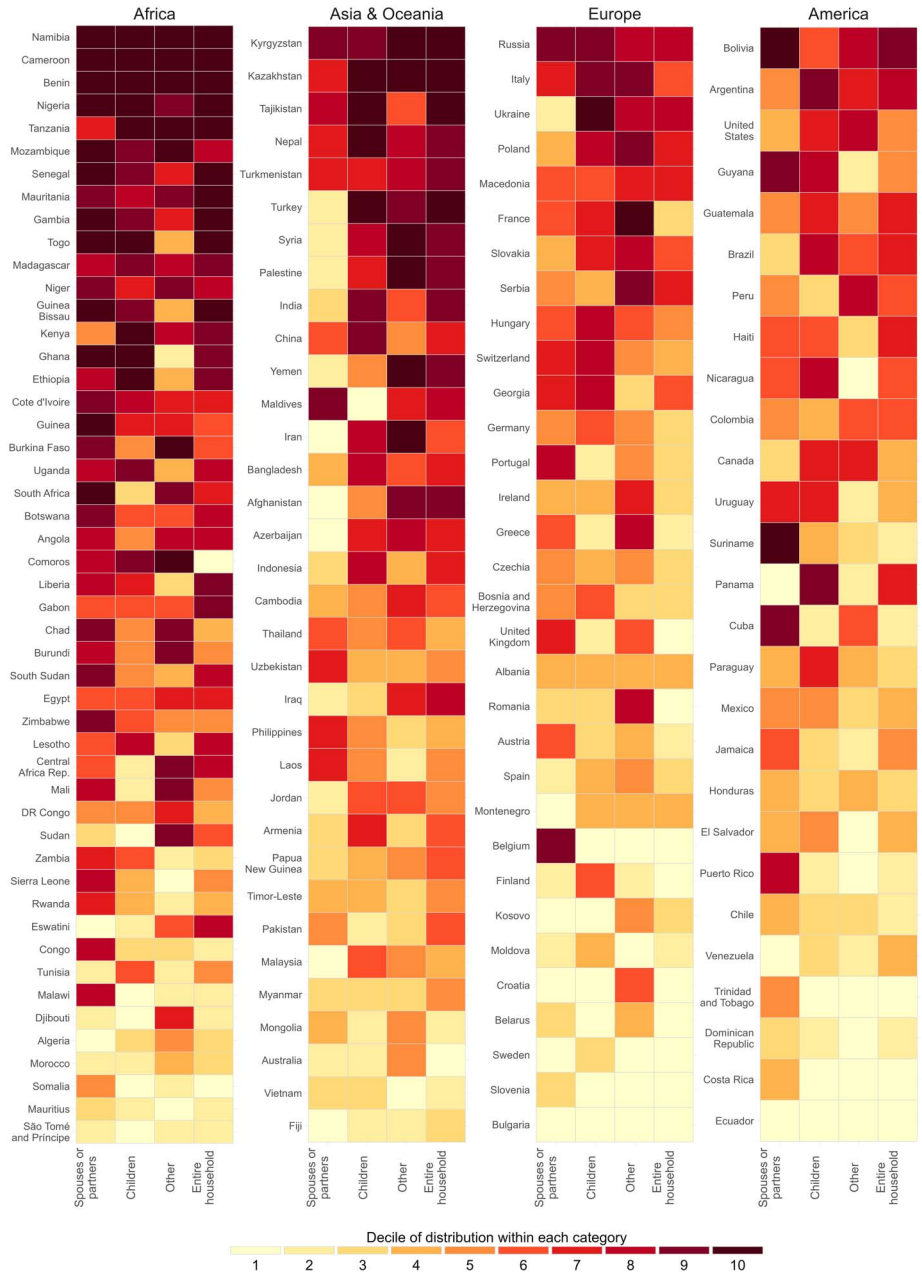
Source: Subnational harmonised dataset of the CORESIDENCE Database for subnational units, drawing on data from DHS, I-IPUMS, MICS, EU-LFS, EU-SILC and HILDA; United Nations Database on Household Size and Composition for countries.

Figure 3 shows the results of the internal variation of countries for the four household variables. Specifically, it displays the decile in which the CV for each country is classified for each variable. Countries are grouped by continental region and sorted by CV level. The countries that occupy the highest positions, which belong to the top deciles, are those with the greatest internal differences, i.e., the most heterogeneous ones. Figure 3 reveals that in every continental region, there are countries with high and with low internal variation. However, most heterogeneous countries are in Africa (Namibia, Cameroon, Benin, Nigeria, Tanzania, Senegal and Mozambique) and in Asia (Kyrgyzstan and Kazakhstan). European and American countries show the lowest levels of internal heterogeneity, with only a few cases (Russia, Italy and Ukraine in Europe; Bolivia and Argentina in America) showing relatively high levels of internal variation. Although it is not visible in Figure 3, the four household variables studied do not have the same levels of internal variation. The household variable that displays the highest levels of CV is the average number of other members living in the household, followed, by a considerable margin, by the average number of children and the number of spouses or partners.

Once the internal variation has been analysed country by country, Figure 4 and Table 1 allow us to assess the relevance of subnational variation versus national variation within continental and subcontinental regions. Figure 4 displays the CV by subnational unit and by country, with the vertical axis indicating variation by country and the horizontal axis indicating variation by subnational unit. To improve the graphic visualisation of the results, the axis scales are adjusted to the data of each of the four panels. Since this scale adjustment affects comparability between panels and between axes, the graphs include lines representing the equations $y = 4x$, $y = 2x$, $y = x$, $y = 0.66x$, $y = 0.5x$ and $y = 0.25x$.

https://doi.org/10.1553/p-cp2h-ne2d

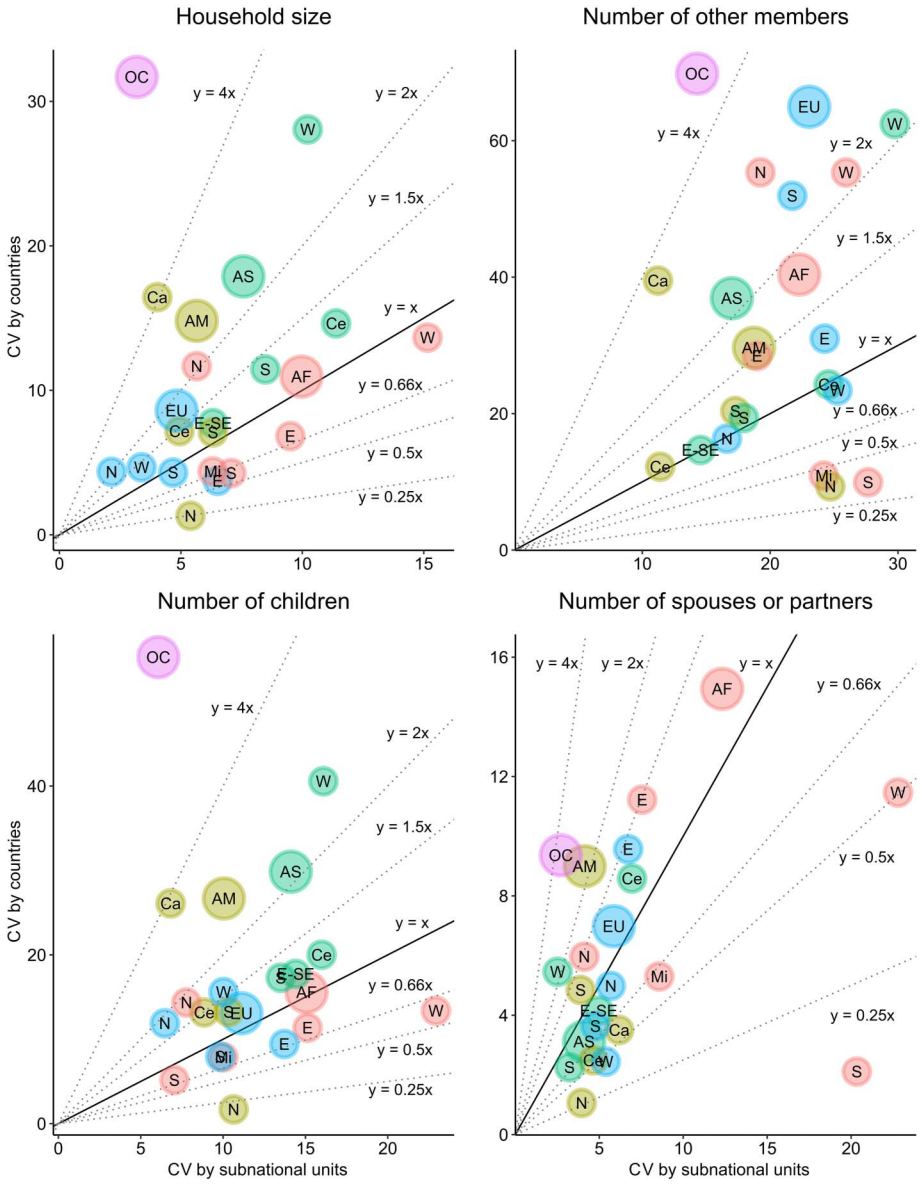
Figure 3 Coefficients of variation by country for four household characteristics



Source: Subnational dataset of the CORESIDENCE Database, drawing on data from DHS, I-IPUMS, MICS, EU-LFS, EU-SILC and HILDA.

https://doi.org/10.1553/p-cp2h-ne2d

Figure 4 National and subnational variation across four household characteristics



Source: Subnational dataset of the CORESIDENCE Database, drawing on data from DHS, I-IPUMS, MICS, EU-LFS, EU-SILC and HILDA. Note: Circle size depends on the level of country aggregation: large circles represent continental regions and small circles represent subcontinental regions. Designated colours group subcontinental regions into their continental regions. Continental regions abbreviations: AF: Africa; AM: America; AS: Asia; EU: Europe; OC: Oceania. Subcontinental regions abbreviations: Ca: Caribbean; Ce: Central; E: East; E-SE: East-Southeast Mi: Middle; N: North; S: South; W: West.

https://doi.org/10.1553/p-cp2h-ne2d

Table 1 Ratio between subnational and national variation by continental and subcontinental region across four household characteristics

Continental/subcontinental region	Total CV	Entire household	Spouses or partners	Children	Other
Africa	0.73	0.91	0.82	0.97	0.55
Eastern Africa	0.88	1.39	0.67	1.33	0.67
Middle Africa	1.73	1.44	1.62	1.27	2.22
Northern Africa	0.42	0.48	0.69	0.54	0.35
Southern Africa	2.89	1.66	9.63	1.36	2.79
Western Africa	0.93	1.11	1.99	1.71	0.47
America	0.48	0.38	0.46	0.38	0.63
Caribbean	0.33	0.24	1.76	0.26	0.28
Central America	0.85	0.69	1.87	0.67	0.93
North America	3.34	4.11	3.72	6.32	2.65
South America	0.83	0.89	0.80	0.78	0.85
Asia	0.49	0.42	1.32	0.47	0.46
Central Asia	0.87	0.78	0.81	0.80	1.01
Eastern South-Eastern Asia	0.91	0.82	1.20	0.81	0.99
Southern Asia	0.86	0.74	1.43	0.78	0.93
Western Asia	0.43	0.36	0.46	0.40	0.48
Europe	0.48	0.56	0.84	0.85	0.35
Eastern Europe	0.95	1.75	0.70	1.45	0.78
Northern Europe	0.82	0.49	1.14	0.54	1.02
Southern Europe	0.60	1.09	1.31	1.24	0.42
Western Europe	0.96	0.72	2.20	0.64	1.08
Oceania	0.16	0.10	0.29	0.11	0.20

Source: Subnational dataset of the CORESIDENCE Database, drawing on data from DHS, I-IPUMS, MICS, EU-LFS, EU-SILC and HILDA.

These lines facilitate the visualisation of the relationship between variation by country and variation by subnational unit in each continental and subcontinental region, eliminating the need to account for the different magnitudes of the horizontal and vertical axes of the panels. The continental and subcontinental regions that exhibit very similar national and subnational variation are situated close to the line $y = x$. Regions with national variation that is double the level of subnational variation are positioned near the line $y = 2x$, whereas regions with national variation that is half the level of subnational variation are located near the line $y = 0.5x$. Moreover, Table 1 shows the ratio between the subnational and the national variation. Regions with very similar national and subnational variation have a ratio close to one. Regions with national variation that is double the level of subnational variation

https://doi.org/10.1553/p-cp2h-ne2d

have a ratio close to 0.5, while regions with national variation that is half the level of subnational variation have a ratio close to two. An example will help illustrate the insights we can derive from Figure 4 and Table 1. Regarding the average number of children in Asia, Figure 4 shows that the CV by country is almost 30 and the CV by subnational unit is around 14. Therefore, the variation by country is equal to almost twice the variation by subnational unit. This result is reflected in the fact that the circle for Asia is just to the left of the line $y = 2x$ in the "Number of children" panel of Figure 4, and the ratio for Asia in the "Children" column of Table 1 is 0.47.

The overall results for the four key variables show that Africa tends to have CV of similar magnitude to that of the other continental regions, and that the variation by subnational unit is similar to the variation by country, with the two exceptions that in the number of spouses or partners variable the CV is much higher than that of the other regions, and in the number of other members variable the national variation is almost twice as high as the subnational variation. In both America and Asia, the CV are similar to those of other regions, and the national variation is higher than the subnational variation (with ratios close to 0.5). The only exception is the average number of spouses or partners in Asia, where the low CV at the country level and the higher variation across subnational units (ratio of 1.32) likely reflect differences in household composition (particularly the presence or absence of a partner) rather than polygamous arrangements, which are rare in the region. In Europe, both the national and the subnational variation tend to be low (except in the number of other members variable), and the national variation in the four variables exceeds the subnational variation (although the difference is small for the number of children and the number of spouses or partners). Oceania stands out as the region with the highest national variation in three of the four variables, and with the lowest subnational variation in the four variables. The combined total level of the CV of America, Asia and Europe is very close to 0.5, meaning that the variation by country is two times higher than the level of subnational variation. In other words, one-third of the total variation is at the subnational level. This ratio goes up to 0.73 in Africa (which means that 42% of the total variation is at the subnational level) and goes down to 0.16 in Oceania.

To summarise the results for the 17 subcontinental regions, we focus on the combined total CV. In three subcontinental regions (Middle Africa, Southern Africa and North America), the subnational variation is greater than the national variation. In another 10 subcontinental regions, the ratio is between 0.82 and 0.99, meaning that the subnational variation accounts for at least 45% of the total variation in most subcontinental regions. In the remaining four subcontinental regions, the ratio is between 0.33 and 0.6, indicating that the subnational variation accounts for between 25 and 38% of the total variation.

Conclusions and discussion

Subnational data prove to be important when analysing household arrangements, demonstrating that substantial variation exists within countries that national-level data may obscure. By utilising data from the CORESIDENCE Database, we have highlighted the value of considering both the national and the subnational level of analysis for improving

https://doi.org/10.1553/p-cp2h-ne2d

our understanding of the diversity of household structures worldwide. Our findings reveal that countries, particularly those in Africa and Asia, exhibit considerable internal variation in household characteristics. Specifically, the average household size and the number of other household members show the greatest variation, suggesting that economic and cultural factors influencing household structures are complex and context-specific. Furthermore, our analysis reveals that the ratio of subnational to national variation varies widely across different subcontinental regions. In Africa, subnational variation often exceeds national variation, indicating that relying solely on national data may overlook crucial aspects of household arrangements. In contrast, European countries tend to show greater uniformity, implying that national-level data may be sufficient in these contexts.

The results emphasise the need for researchers to carefully consider the level of analysis when studying demographic phenomena. While national data have traditionally dominated demographic research, our findings support the argument for greater integration of subnational data to better capture the nuances of household arrangements and, more broadly, of demographic indicators and population structures. Our results contribute to the debate about the appropriateness of using the country as the primary unit of analysis in comparative international studies. By highlighting the significance of subnational heterogeneity, we advocate for a more comprehensive approach that accounts for the variability within countries. This approach could lead to more accurate and nuanced interpretations of social change and inequalities that extend beyond the realm of household analysis. The example of households has served to illustrate, based on indicators available at the subnational level across many countries, the differences that exist both between and within countries. Similar studies could be conducted for fertility, which has been declining in most countries worldwide, as well as for mortality and migration.

However, conducting subnational studies relies on the availability of data. Historically, censuses have been the most appropriate sources for such analyses. Their international availability, particularly through initiatives like the I-IPUMS project, makes this kind of analysis feasible. However, in an increasing number of countries, traditional censuses are disappearing, and with them the availability of microdata samples. Although surveys provide richer conceptual detail, they cannot compete with the breadth of information available through censuses. Administrative records could provide an alternative, but their implementation and use in research are rare and are generally confined to a few Western countries, such as Nordic nations. As a final reflection, it is likely that in countries facing challenges such as low fertility, ageing populations and increased migration, demographic landscapes will become increasingly diverse across different areas of the same country. In this new reality, analysis at the subnational scale will be even more crucial.

Funding

CORESIDENCE is supported by an Advanced Grant funded by the European Research Council (ERC). Grant agreement ID: 101052787. Jesús García-Gómez also acknowledges funding from the Programa Investigo (European Union Next Generation).

https://doi.org/10.1553/p-cp2h-ne2d

ORCID iDs

Jesús García-Gómez https://orcid.org/0000-0003-3288-355X

Juan Galeano https://orcid.org/0000-0002-3682-1797

Albert Esteve https://orcid.org/0000-0001-9916-386X

References

Canudas-Romo, V., Shen, T., and Payne, C. F. (2024). National population growth rate, its components, and subnational contributions: A research note. *Demography*, *61*(3), 615–626. https://doi.org/10.1215/00703370-11372369

Coale, A. J., and Watkins, S. C. (1986). *The decline of fertility in Europe*. Princeton University Press. https://doi.org/10.1515/9781400886692

Crombach, L., and Smits, J. (2022). The demographic window of opportunity and economic growth at sub-national level in 91 developing countries. *Social Indicators Research*, *161*(1), 171–189. https://doi.org/10.1007/s11205-021-02802-8

Esteve, A., Pohl, M., Becca, F., Fang, H., Galeano, J., García-Román, J., Reher, D., Trias-Prats, R., and Turu, A. (2024). A global perspective on household size and composition, 1970–2020. *Genus*, *80*(1), 2. https://doi.org/10.1186/s41118-024-00211-6

Eurostat (2023a). European Union labour force survey. European Union. https://ec.europa.eu/eurostat/web/microdata/european-union-labour-force-survey

Eurostat (2023b). EU statistics on income and living conditions (EU-SILC) survey. European Union. https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions

Eurostat. (2024). *Total fertility rate by NUTS 2 region*. https://doi.org/10.2908/TGS00100

Fox, J., Klüsener, S., and Myrskylä, M. (2019). Is a positive relationship between fertility and economic development emerging at the sub-national regional level? Theoretical considerations and evidence from Europe. *European Journal of Population*, *35*(3), 487–518. https://doi.org/10.1007/s10680-018-9485-1

Galeano, J., Esteve, A., Turu, A., García-Roman, J., Becca, F., Fang, H., Pohl, M., and Trias-Prats, R. (2024). CORESIDENCE: National and subnational data on household size and composition around the world, 1964–2021. *Scientific Data*, *11*(1), 145. https://doi.org/10.1038/s41597-024-02964-3

Gershman, B., and Rivera, D. (2018). Subnational diversity in Sub-Saharan Africa: Insights from a new dataset. *Journal of Development Economics*, *133*, 231–263. https://doi.org/10.1016/j.jdeveco.2018.01.003

ICF (2023). Demographic and health surveys. Funded by USAID. https://dhsprogram.com/data/.

Lesthaeghe, R. J., and Esteve, A. (2016). *Cohabitation and marriage in the Americas: Geo-historical legacies and new trends*. Springer Nature.

Kashnitsky, I., de Beer, J., and van Wissen, L. (2017). Decomposition of regional convergence in population aging across Europe. *Genus*, *73*(1). https://doi.org/10.1186/s41118-017-0018-2

Kulu, H., and Boyle, P. J. (2009). High fertility in city suburbs: Compositional or contextual effects? *European Journal of Population*, *25*(2), 157–174. https://doi.org/10.1007/s10680-008-9163-9

Pezzulo, C., Hornby, G. M., Sorichetta, A., Gaughan, A. E., Linard, C., Bird, T. J., Kerr, D., Lloyd, C. T., and Tatem, A. J. (2017). Sub-national mapping of population pyramids and dependency ratios in Africa and Asia. *Scientific Data*, *4*. https://doi.org/10.1038/sdata.2017.89

Scherbov, S., Shulgin, S., Andruchowitz, S., Arkhangelsky, V., Ediev, D., Efremov, I., Nikitina, S., and Sanderson, W. (2019). *Russian Demographic Data Sheet 2019*. Russian Presidential Academy of National Economy and Public Administration (RANEPA), Russian Federal State Statistics Service (Rosstat), and International Institute for Applied Systems Analysis (IIASA).

Riederer, B., and Buber-Ennser, I. (2019). Regional context and realization of fertility intentions: The role of the urban context. *Regional Studies*, *53*(12), 1669–1679. https://doi.org/10.1080/00343404.2019.1599843

Ruggles, S., Cleveland, L., Lovaton, R., Sarkar, S., Sobek, M., Burk, D., Ehrlich, D., Heimann, Q., and Lee, J. (2024). *Integrated Public use Microdata Series International: Version 7.5*. IPUMS. https://doi.org/10.18128/D020.V7.5

Turkish Statistical Institute. (2024). Birth statistics, 2023. Republic of Türkiye. https://data.tuik.gov.tr/Bulten/Index?p=Birth-Statistics-2023-53708&dil=2

UNICEF (2023). Multiple indicator cluster surveys (MICS). UNICEF. https://mics.unicef.org/

United Nations, Department of Economic and Social Afairs, Population Division (2022). Database on household size and composition. https://www.un.org/development/desa/pd/data/household-size-and-composition

Wilkins, R (2021). HILDA Survey. HILDA Project, Melbourne Institute of Applied Economic and Social Research, University of Melbourne. https://melbourneinstitute.unimelb.edu.au/hilda

https://doi.org/10.1553/p-cp2h-ne2d