

Publication status: This preprint has not been published elsewhere.

Health Vulnerability Index and territorial clustering for Primary Health Care planning: a census-data and spatial analysis approach in Brazil

Misael de Souza Vieira, Gabriel Marcos de Moraes Nascimento, Lygia Bueno Fragoso, Elaine Leandro Machado, João Flávio de Freitas Almeida

<https://doi.org/10.1590/SciELOPreprints.18395>

Submitted on: 2026-10-13

Posted on: 2026-10-14 (version 1)

(YYYY-MM-DD)

Health Vulnerability Index and territorial clustering for Primary Health Care planning: a census-data and spatial analysis approach in Brazil

Misael de Souza Vieira Pinto¹ 
<https://orcid.org/0009-0001-2920-9209>

Gabriel Marcos de Moraes Nascimento¹ 
<https://orcid.org/0009-0006-8045-605X>

Lygia Bueno Fragoso² 
<https://orcid.org/0009-0008-8786-5483>

Elaine Leandro Machado³ 
<https://orcid.org/0000-0002-3226-3476>

João Flávio de Freitas Almeida^{1,*} 
<https://orcid.org/0000-0002-3884-217X>

¹*Department of Industrial Engineering, Universidade Federal de Minas Gerais (UFMG).
 Av. Antônio Carlos, 6627, Belo Horizonte, 31270-901, MG, Brazil*

²*Basic Health Unit, Primary Health Care, Dom Cabral, Belo Horizonte Government (PBH)*

³*Department of Preventive and Social Medicine, School of Medicine/UFMG*

*Corresponding author: joao.flavio@dep.ufmg.br

October 13, 2026

Abstract

Primary Health Care (PHC) in Brazil's Unified Health System (SUS) requires territorial planning tools capable of addressing social inequalities and supporting the Family Health Strategy (ESF). This study proposes a spatial decision-support framework that integrates a multidimensional Health Vulnerability Index (HVI) with spatially constrained clustering at the census-sector level. The HVI is constructed from variables in the 2022 Brazilian Demographic Census across the domains of Urban Infrastructure, Human Capital, Income and Employment, and Demographic Vulnerability. The resulting HVI was then used in a spatially constrained hierarchical clustering algorithm incorporating queen contiguity and ESF-aligned population thresholds to generate homogeneous and contiguous PHC planning regions. Results showed a 77.8% reduction in the number of planning units nationally while preserving spatial coherence and internal homogeneity. The proposed framework transforms fine-grained vulnerability data into operationally meaningful planning geographies, providing a scalable and reproducible basis for PHC needs-based territorial planning and future allocation models within SUS.

Keywords: Health vulnerability, Primary health care, Spatial clustering, Census data.

1 Introduction

Primary Health Care (PHC) is the foundational element of Brazil’s Unified Health System (SUS), serving as the main entry point to healthcare and the primary mechanism for addressing population healthcare needs through a territorial and community-oriented approach [1]. The Family Health Strategy (ESF), organized around multidisciplinary teams assigned to defined geographic catchment areas, operationalizes this territorial logic. Effective PHC planning therefore depends on accurately identifying local healthcare needs and defining coherent planning units that align service provision with local health demands and vulnerabilities [2–5].

In a context marked by profound territorial inequalities, rapid urbanization, and heterogeneous living conditions, supporting *need*-based PHC resource allocation remains a major challenge. This challenge stems primarily from the persistent mismatch between population *need*, defined as the expected volume of patient visits derived from demographic structure, clinical protocols, and social determinants of health, and observed *demand*, which corresponds to actual service utilization recorded in health facilities [6–8].

This mismatch is exacerbated by the limited ability of conventional spatial and analytical frameworks to represent intra-urban heterogeneity. Traditional administrative boundaries and aggregated vulnerability indices often obscure localized concentrations of deprivation, leading to ecological fallacies in which highly heterogeneous populations are treated as homogeneous units. As a result, the spatial “territory of governance” frequently diverges from the actual “territory of health needs” [9–11]. Many spatial planning approaches rely on arbitrary or non-contiguous territorial aggregations and are susceptible to the Modifiable Areal Unit Problem (MAUP) [12], particularly within SUS planning, where municipalities are commonly adopted as analytical units. Despite advances in vulnerability mapping, existing indices remain insufficiently granular for micro-level PHC planning, and spatial methods rarely integrate fine-scale vulnerability measurement with operationally coherent territorial design.

To address these limitations, we propose a territorial planning framework that integrates vulnerability assessment with spatially constrained regionalization to support operational public-sector planning. The framework first constructs a Health Vulnerability Index (HVI) from nationwide IBGE census data [13], combining indicators of *urban infrastructure*, *human capital*, *income and employment*, and *demographic vulnerability* following the methodological principles of the IPEA Social Vulnerability Index (IVS) [14]. It then applies spatially constrained clustering based on max- P -regions principles to generate contiguous, homogeneous, and population-balanced territorial units aligned with ESF operational requirements. Unlike previous studies, which typically address either vulnerability assessment or spatial regionalization in isolation, the proposed framework integrates both into a unified planning methodology. The resulting territorial units constitute an intermediate planning layer that can be directly incorporated into facility-location, workforce allocation, referral optimization, and health network design models [15–17]. By transforming fine-grained census information into operational planning geographies, the framework provides a scalable and reproducible decision-support tool for evidence-based PHC planning.

The remainder of this article is organized as follows. Section 2 reviews the literature on health vulnerability indices and spatial clustering in PHC planning. Section 3 presents the data and methodological framework, including the construction of the HVI and clustering procedures. Section 4 reports the empirical results and characterizes the territorial units obtained by the proposed framework. Section 5 discusses implications for PHC planning, limitations, and future research directions. Section 6 concludes the study.

2 Literature review

Vulnerability-based approaches have become central to territorial health planning by providing quantitative tools to represent spatial heterogeneity in population risk and need. The Social Vulnerability Index (SoVI), originally developed by Cutter et al. [18, 19], established a foundational framework for measuring place-based susceptibility using census-derived socioeconomic and demographic indicators. This formulation has been widely extended as a general approach for translating structural inequalities into spatially explicit metrics for planning and risk assessment [20].

In health systems, the SoVI framework has evolved into various Health Vulnerability Indices, which integrate socioeconomic, demographic, environmental, and health-related dimensions to support priority-setting in public health. In Brazil, SoVI-inspired approaches have been adapted for health planning at different spatial scales. Luiz et al. [9] proposed a municipal-level composite index combining demographic, economic, environmental, educational, and health variables, while Peixoto et al. [21] developed a census-tract-based HVI for intra-municipal analysis. Empirical studies have further linked vulnerability measures to PHC-related outcomes, including service coverage, disease prevalence, screening access, and oral health performance [10, 22].

Spatial analytical methods have been increasingly combined with vulnerability indices to improve their operational value in health planning. Geographic information systems, spatial scan statistics, and cluster detection techniques have been used to identify localized concentrations of risk and misalignments between service provision and population needs [23, 24]. Recent studies have further advanced spatial approaches to healthcare demand–supply analysis and primary care network redesign [15, 25], as well as patient clustering for resource allocation [26]. These approaches underscore the importance of spatial structure in interpreting vulnerability patterns. However, results remain sensitive to spatial aggregation choices, as highlighted by the Modifiable Areal Unit Problem (MAUP) [12].

To address these limitations, constrained spatial clustering methods have emerged to generate contiguous and internally homogeneous territorial units suitable for planning. Among these, the Max-P-Regions problem formalized by Duque et al. [27] partitions geographic space into the maximum number of regions that satisfy population and contiguity constraints while optimizing internal homogeneity. Such methods are particularly relevant for PHC, as they enable the construction of operational planning zones aligned with service allocation requirements [27].

A broader perspective on vulnerability measurement in Brazil is provided by [28], who conducted a scoping review of area-based deprivation measures used in health research. The review classified indices developed at different spatial scales, including municipalities, districts, neighborhoods, and census tracts, and synthesized their methodological characteristics, applications, and data sources. More importantly, it identified the absence of a nationwide, small-area, reproducible deprivation index suitable for supporting health research and public policy, highlighting an important methodological gap in the Brazilian literature.

Building upon this research agenda, GeoSES [29] represented an important advance by proposing a nationwide socioeconomic index for health research in Brazil. By combining multiple socioeconomic dimensions through principal component analysis, it provides a transparent and standardized measure for epidemiological studies at the municipal level. Nevertheless, because it was designed primarily to characterize socioeconomic context rather than support territorial health planning, it neither captures intra-urban heterogeneity at the census-tract scale nor generates operational planning regions for PHC.

Despite these advances, existing vulnerability-based approaches remain limited. We propose a reproducible nationwide Health Vulnerability Index at the census-tract level integrated with spatially constrained clustering. The resulting framework generates contiguous and operational

planning territories, overcoming the limited geographic scope and lack of planning-oriented regionalization found in existing vulnerability-based approaches.

3 Materials and Methods

This study employs two main methodological components: (i) the construction of a multidimensional Health Vulnerability Index and (ii) spatially constrained clustering, both using data from the 2022 Brazilian Demographic Census [13].

3.1 Health Vulnerability Index (HVI)

The HVI was developed as a composite index to support PHC planning, relying exclusively on variables from the 2022 Brazilian Census [13]. It is computed at the census-tract level across all 458,772 tracts in Brazil, providing a transparent and reproducible measure of health vulnerability.

Data preprocessing comprised the selection and harmonization of census-tract variables, as illustrated in Figure 1. Variables were selected based on their theoretical relevance to the social determinants of health and PHC needs. The resulting dataset was organized into four dimensions, whose indicators are presented in Tables 2, 4, 6, 8, while the corresponding Census variables are listed in Tables 3, 5, 7, 9.

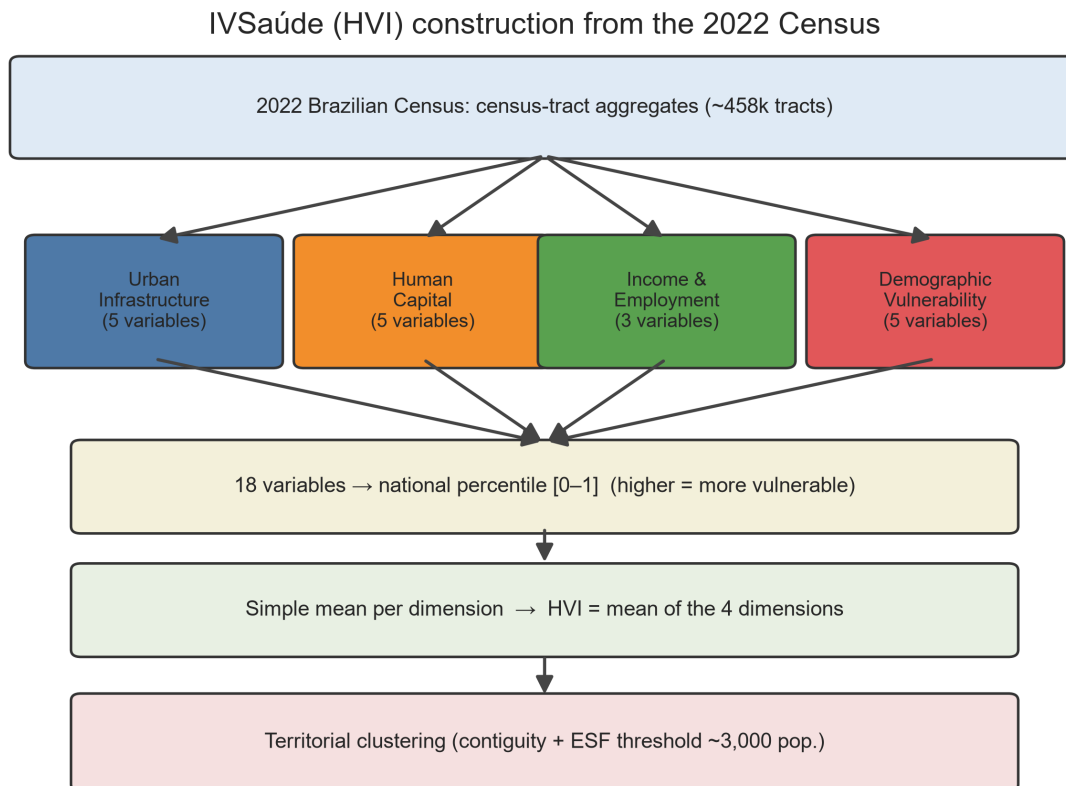


Figure 1: Methodological workflow for HVI construction and subsequent territorial clustering.

Urban Infrastructure (UI) comprises five indicators derived from IBGE Census variables related to inadequate water supply, sanitation, garbage collection, bathroom availability, and household overcrowding (more than two residents per bedroom). These indicators capture environmental conditions associated with higher health vulnerability, including increased risks of *Aedes aegypti* transmission, waterborne diseases, and other adverse urban health outcomes [30–32].

Human Capital (HC) includes five indicators reflecting educational attainment and family structure: illiteracy rates (among those aged 15+ and 15–29), female-headed households without a spouse, and illiteracy among household heads (overall and female). Low literacy levels limit health literacy and access to services, while female-headed single-parent households often face greater socioeconomic barriers [33–35].

Income and Employment (IE) is represented by three indicators: low household-head income (below half the minimum wage), adolescent household heads, and substandard housing. These capture socioeconomic deprivation linked to reduced healthcare access and inter-generational vulnerability [36–38].

Demographic Vulnerability (DV) constitutes an additional dimension relative to the IPEA Social Vulnerability Index (IVS), which has only three dimensions [14]. This dimension comprises five indicators: proportion of population aged 65+, aged 0–4, households headed by older adults, the demographic dependency ratio, and premature mortality (30–69 years). These reflect age-related healthcare needs and underlying health inequalities [39, 40].

To avoid arbitrary weighting, the following standardization procedure was adopted:

1. Each variable was transformed into a national percentile rank ($P_j \in [0, 100]$), with higher values indicating greater vulnerability.
2. The average percentile was calculated for each dimension.
3. The final HVI was obtained as the simple average of the four dimension scores:

$$\text{HVI} = \frac{\text{UI} + \text{HC} + \text{IE} + \text{DV}}{4}$$

3.2 Spatially Constrained Clustering

The most distinctive component of this methodology is the transformation of the HVI into operational planning geographies through spatially constrained clustering. A regionalization algorithm inspired by the Max- p Regions problem [27] was applied. The procedure partitions census tracts into the maximum number of regions that are (i) internally homogeneous with respect to the HVI and its sub-indices, (ii) spatially contiguous (using queen contiguity), and (iii) respect minimum population thresholds aligned with ESF operational requirements ($\approx 3,000$ inhabitants).

Contiguity was enforced via a spatial weights matrix based on queen contiguity. The algorithm balances internal homogeneity, measured by within-region HVI variance, with the creation of operationally viable planning units.

3.3 Validation and Experimental Design

The proposed methodology was validated through internal consistency of the HVI (Cronbach's alpha [41] and correlation with IPEA's IVS); spatial autocorrelation analysis (Moran's I [42, 43]); sensitivity tests on normalization thresholds and clustering parameters; and comparative performance across algorithms and geographic scales. Following, the notation, the statistics used in each procedure, and the resulting territories evaluation criteria are described.

Notation. Let $S = \{1, \dots, n\}$ be the set of census tracts, v_i the HVI of tract i , and p_i its resident population. Let $G = (S, E)$ be the queen contiguity graph, in which $(i, i') \in E$ whenever the polygons of tracts i and i' share at least one boundary point, and let $w_{ii'}$ denote the corresponding spatial weights. A territorial partition returned by the clustering procedure is written $\mathcal{R} = \{R_1, \dots, R_K\}$ and satisfies

$$\bigcup_{k=1}^K R_k = S, \quad R_k \cap R_l = \emptyset \quad \forall k \neq l, \quad (1)$$

$$G[R_k] \text{ is connected}, \quad k = 1, \dots, K, \quad (2)$$

$$P_k = \sum_{i \in R_k} p_i \leq \tau \quad \text{or} \quad |R_k| = 1, \quad k = 1, \dots, K, \quad (3)$$

where $G[R_k]$ is the subgraph of G induced by R_k and $\tau = 3,000$ inhabitants is the reference catchment of one ESF team. Condition (2) states that every planning unit is a single connected territory. Condition (3) states that no unit exceeds the caseload of a single team, with the unavoidable exception of census tracts whose own population already exceeds τ and which an aggregation procedure cannot split. Stating the population requirement as an upper bound, rather than as the lower bound of the canonical max- p formulation, follows the regionalization literature that admits either direction of the constraint [44, 45]. Constrained regionalizations of this family have been used to delineate health enumeration areas from census tracts under an explicit population requirement [46]. Among the partitions satisfying (1)–(3), the procedure favors those that are internally homogeneous with respect to the HVI and its sub-indices, in the sense made precise by (8) below.

Internal consistency. The coherence of the index was assessed with Cronbach's alpha,

$$\alpha = \frac{q}{q-1} \left(1 - \frac{\sum_{j=1}^q \sigma_j^2}{\sigma_T^2} \right), \quad T_i = \sum_{j=1}^q Y_{ij}, \quad (4)$$

where Y_{ij} is the value of component j in tract i , q the number of components, σ_j^2 the variance of component j across tracts and σ_T^2 the variance of their sum [41]. The statistic was computed within each dimension, over its indicators, and over the four dimension scores. Values above 0.70 were read as evidence that the components of a dimension measure a common construct, while values approaching unity were read as redundancy rather than merit, since they indicate that indicators duplicate one another instead of contributing distinct information [47]. The same reliability check is standard practice in the construction of census-based vulnerability indices [48, 49].

Convergent validity. The HVI was compared with the IPEA Social Vulnerability Index [14], which is published at the municipal scale. Tract scores were therefore aggregated to that scale by population weighting,

$$\text{HVI}_m = \frac{\sum_{i \in S_m} p_i v_i}{\sum_{i \in S_m} p_i}, \quad (5)$$

where S_m is the set of tracts of municipality m , and the two orderings were compared through the Spearman rank correlation [50], that is, the Pearson coefficient computed on the ranks of the M municipalities. In the absence of ties this is equivalent to

$$\rho = 1 - \frac{6 \sum_{m=1}^M d_m^2}{M(M^2 - 1)}, \quad (6)$$

with d_m the difference between the ranks assigned to municipality m by the two indices. Ranks were used rather than raw values because the two indices are expressed on different scales, so only their ordering of territories is comparable; the same rank-based comparison is used to assess agreement among established indices of social determinants of health [51]. Agreement supports convergent validity; divergence is expected and informative, since the HVI incorporates a demographic dimension that the IVS does not contain.

Spatial autocorrelation. Whether the index describes organized geographic variation rather than noise was examined with the global Moran's I statistic,

$$I = \frac{n}{\sum_{i \in S} \sum_{i' \in S} w_{ii'}} \cdot \frac{\sum_{i \in S} \sum_{i' \in S} w_{ii'} (v_i - \bar{v})(v_{i'} - \bar{v})}{\sum_{i \in S} (v_i - \bar{v})^2}, \quad (7)$$

computed on the same contiguity weights that constrain the clustering and compared with its expectation under spatial randomness, $\mathbb{E}[I] = -1/(n - 1)$. The statistic originates in Moran [42] for regular lattices and is used here in the general weighted form of Cliff and Ord [43]; significance was assessed by conditional permutation of the observed values across tracts. The same statistic has been applied to health vulnerability at the census-tract scale in Belo Horizonte [52]. A positive and significant I is a precondition for the approach as a whole: were vulnerability spatially random, aggregating contiguous tracts could only destroy information, and administrative boundaries would be as defensible as any alternative.

Criteria for the resulting territories. Three quantities were used to evaluate a partition. Internal homogeneity was measured by the within-region sum of squares of the HVI relative to its total dispersion,

$$\eta^2(\mathcal{R}) = 1 - \frac{\sum_{k=1}^K \sum_{i \in R_k} (v_i - \bar{v}_{R_k})^2}{\sum_{i \in S} (v_i - \bar{v})^2}, \quad \bar{v}_{R_k} = \frac{1}{|R_k|} \sum_{i \in R_k} v_i, \quad (8)$$

which equals one when the aggregation preserves all information on vulnerability and zero when the units are internally as heterogeneous as the country as a whole. The numerator is the within-cluster sum of squares minimized by Ward-type agglomeration [53], and the ratio is the proportion of explained inertia used to assess partitions under spatial constraints [54], and measures of within-region variance of this kind also support the choice of the number of regions in regionalization [55]. Operational adequacy was measured by the proportion of regions satisfying (3) and by the dispersion of their populations,

$$\text{CV}(\mathcal{R}) = \frac{1}{\bar{P}} \sqrt{\frac{1}{K-1} \sum_{k=1}^K (P_k - \bar{P})^2}, \quad \bar{P} = \frac{1}{K} \sum_{k=1}^K P_k, \quad (9)$$

balance being one of the three classical criteria of territory design, alongside contiguity and compactness [56]. The gain in managerial parsimony was measured by the reduction in the number of

planning units,

$$\Delta(\mathcal{R}) = 1 - \frac{K}{n}. \quad (10)$$

These criteria are in tension by construction and are therefore reported jointly: any partition can be made more homogeneous by producing more units, and more parsimonious by producing fewer.

Sensitivity and experimental design. The sensitivity tests varied the normalization thresholds adopted in the construction of the index and the parameters of the clustering procedure, in particular the population threshold τ , recomputing (8), (9) and (10) in each scenario and comparing the resulting orderings through (6). Experiments were run at two scales: nationally, over the 458,772 census tracts of the 2022 Census, to assess scalability and to quantify the reduction in planning units; and at the municipal scale, taking Belo Horizonte and its 5,113 census tracts as a case in which the resulting units can be confronted with the territorial structure in operation [57, 58]. The scenario analyses on τ characterize the planning configurations available under alternative ESF caseload assumptions.

Results of the validation. Three findings from these procedures shaped the specification reported above. First, the population requirement in (3) is rarely binding at the tract level: of the 450,088 tracts with valid population records, covering 202,561,627 residents, the median tract holds 428 inhabitants and only 21 tracts, or 0.005% of the total, exceed $\tau = 3,000$ on their own, the largest reaching 4,804 residents. Those tracts account for 0.04% of the national population, so the singleton exception in (3) is a formal safeguard rather than a practical limitation. Second, (4) was decisive in discarding an earlier specification of the index, which carried a health dimension proved to be constant across every tract of each of the 5,570 municipalities, carrying no intra-municipal resolution, and was virtually uncorrelated with the remaining dimensions (Pearson -0.08 and -0.04), driving the composite to a negative alpha that rose to 0.56 once the dimension was removed and substituted by the census-only demographic block described in Section 3. Third, vulnerability is spatially structured at the tract scale: in Belo Horizonte (7) yields $I = 0.15$ under queen contiguity, against $\mathbb{E}[I] = -0.0002$, with $p = 0.001$ over 999 permutations, and remains positive and significant under k -nearest-neighbor weights, which confirms the precondition for contiguity-constrained aggregation. Evaluated on the partition currently distributed with the study, (8), (9) and (10) give $\eta^2 = 0.83$, $CV = 0.48$ and $\Delta = 77.5\%$ nationally, against $\eta^2 = 0.23$ for a random partition with identical cluster sizes; in Belo Horizonte the 5,113 tracts are reduced to 1,063 units, with $\Delta = 78.8\%$, $CV = 0.40$, $\eta^2 = 0.57$ and 95.6% of the units within the ESF caseload reference.

4 Results

This section presents the main findings from the application of the proposed methodology, including the construction of the HVI and the spatially constrained territorial clustering. Analyses, conducted primarily on 2022 Brazilian Census tracts nationwide, were implemented in Python. Tabular processing and geoprocessing relied on `pandas`, `geopandas` and `shapely`, the last supplying the boundary predicates from which queen contiguity is derived, and the spatial statistics were computed with `libpysal` and `esda`. The regionalization itself was implemented directly, using a disjoint-set structure to merge clusters, rather than through an existing clustering package. With the edges of G sorted once by attribute dissimilarity, the procedure is near-linear in the number of adjacencies and therefore feasible over the whole country. Given the input data, the contiguity graph and τ , the sequence of merges is fully determined, so the reported partitions are exactly

reproducible. The code for index construction, contiguity, and clustering is available upon *acceptance/publication*, and the resulting clusters with HVI scores for every Brazilian municipality can be accessed at <http://ivs-cluster-na-aps.s3-website-sa-east-1.amazonaws.com/>.

4.1 Health Vulnerability Index (HVI)

The HVI was successfully constructed for all 458,772 census tracts by integrating the four domains (Urban Infrastructure, Human Capital, Income & Employment, and Demographic Vulnerability) using harmonized census variables. The index is scaled from 0 (lowest vulnerability) to 1 (highest vulnerability). Nationally, the mean HVI was 0.50, revealing clear regional disparities. As shown in Figure 2, the South region exhibited the lowest average vulnerability, while the North and Northeast presented both higher means and greater variability.

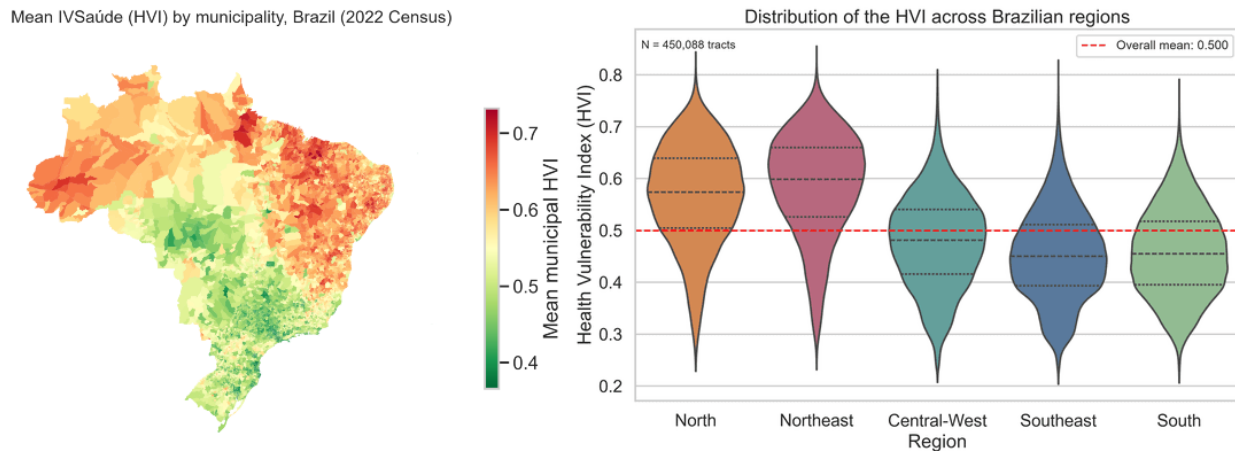


Figure 2: Distribution of the Health Vulnerability Index (HVI) across Brazilian regions.

4.2 Spatial Analysis and Territorial Clustering

The core innovation of this study is a spatial aggregation framework that converts high-resolution HVI data into operationally viable planning units for the ESF. By combining the HVI with a spatially constrained greedy hierarchical clustering algorithm based on queen contiguity, the method produces contiguous and socioeconomically homogeneous clusters while enforcing a population threshold of approximately 3,000 inhabitants per ESF team, as recommended by the Brazilian Ministry of Health (Figure 3).

Nationally, the framework reduced the number of territorial units by 77.1% without compromising spatial contiguity or vulnerability representation. In Belo Horizonte, the original 5,113 census tracts were aggregated into 1,063 clusters (a 78.8% reduction; Figure 4). Of the resulting clusters, 95.6% fall within the ESF caseload reference of 3,000 inhabitants, with a median of 2,411 residents and a maximum of 4,495. Since no census tract in the municipality exceeds the reference on its own, the residual excess is produced by the aggregation itself rather than by indivisible tracts. For comparison, the current municipal structure relies on approximately 3,800–4,000 micro-areas, which is the smallest SUS planning unit, each typically covering 150–200 families (roughly 450–750 inhabitants) under one Community Health Agent (ACS) and grouped under the 639 existing ESF teams [57, 58].

These results demonstrate the practical value of the approach: it substantially reduces operational planning complexity while preserving spatial coherence, internal homogeneity, and direct

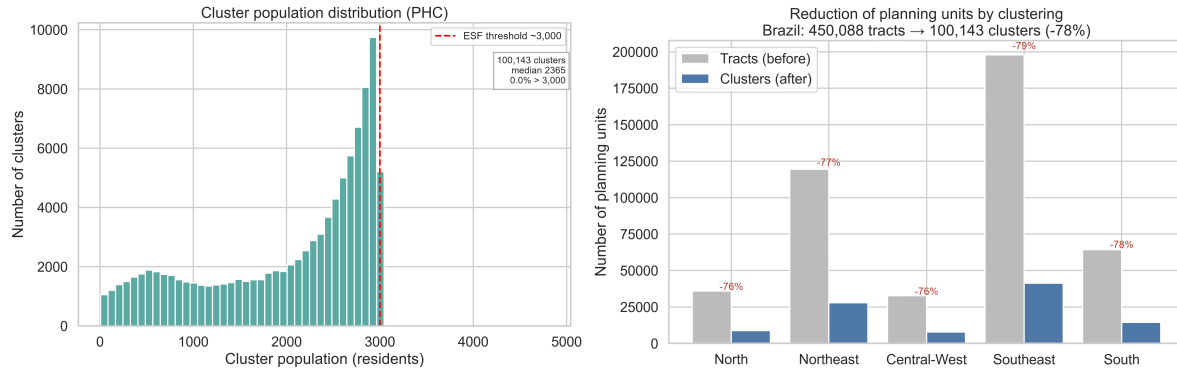


Figure 3: Population distribution of the clusters (ESF threshold $\sim 3,000$ residents) and the reduction in the number of planning units through clustering (-77.8%).

Territorial clustering in Belo Horizonte

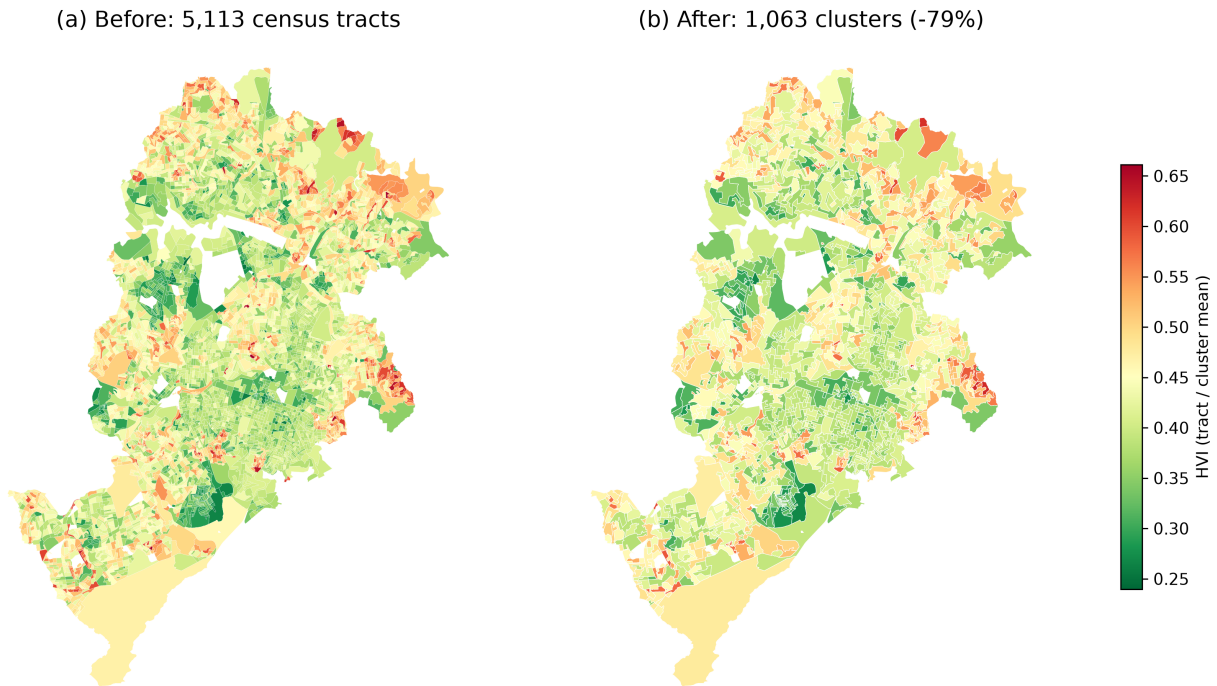


Figure 4: Visual comparison of spatial clustering in Belo Horizonte: (a) individual census tracts before grouping; (b) clusters formed after the grouping process based on IVS similarity, spatial contiguity, and a population limit of 3,000 inhabitants.

compatibility with ESF allocation. By generating larger yet still manageable units that are more homogeneous than both raw census tracts and existing micro-areas, the framework provides a scalable middle ground between excessive fragmentation and overly aggregated municipal districts, thereby supporting evidence-based microplanning and resource allocation within the SUS.

5 Discussion

5.1 Advancing Territorial Clustering for PHC Planning

The principal methodological contribution of this work is the empirical demonstration that spatially constrained clustering can reduce the number of planning units by approximately 77% while generating homogeneous, contiguous, and operationally relevant territories. This directly addresses two key limitations identified in the literature: the fragmentation introduced by administrative boundaries (the Modifiable Areal Unit Problem) and the difficulty of translating highly disaggregated vulnerability indicators into actionable PHC geographies [10, 12, 27]. By producing clusters that satisfy population thresholds compatible with ESF teams while preserving internal homogeneity in vulnerability conditions, the proposed methodology bridges the gap between diagnostic precision and managerial feasibility.

Beyond the territorial aggregation procedure itself, the proposed HVI advances existing Brazilian vulnerability frameworks. The Social Vulnerability Index (IVS) developed by IPEA constitutes an important milestone by combining indicators of urban infrastructure, human capital, and income and labour conditions into a composite measure [14]. However, the IVS was designed primarily to capture social vulnerability rather than health needs; it is typically disseminated at municipal or higher scales and does not explicitly incorporate demographic characteristics strongly associated with PHC needs, such as population ageing and dependency structures.

More recently, the Ministry of Health introduced the Equity and Sizing Index (IED) as part of the new federal PHC financing model [59, 60]. The IED combines municipal population size with categorical vulnerability classes to support equitable resource allocation. While suitable for financing purposes, it is not intended as a health-vulnerability measure, relies on ordinal categories that discard continuous information, and operates exclusively at the municipal level, thereby obscuring intra-urban inequalities that are often decisive for PHC planning.

The proposed HVI builds upon previous Brazilian vulnerability measures by addressing the gaps highlighted by [28] and extending nationwide indices such as the IVS, IED, and GeoSES. It combines census-tract resolution, an explicit demographic-health dimension, continuous indicators, transparent and reproducible computation, and spatially constrained clustering to produce contiguous and operational PHC planning territories. Table 1 compares the main characteristics of these approaches.

Table 1: Comparison of Brazilian vulnerability measures.

Feature	IVS [14]	IED [60]	GeoSES [29]	Review [28]	HVI
Census-tract data	No	No	No	Gap	Yes
Health-oriented	No	No	Partial	No	Yes
Continuous index	Yes	No	Yes	–	Yes
Transparent & reproducible	Yes	Partial	Yes	High	Yes
Arbitrary weights	Partial	Yes	No (PCA)	–	No
Spatial clustering	No	No	No	No	Yes
Operational PHC planning	No	Limited	No	No	Yes
Purpose	Social	Financing	Health research	Literature	Health + PHC

Rather than replacing existing indices, the proposed HVI complements them by addressing a different planning objective. The IVS measures social vulnerability, the IED supports equitable federal PHC financing, and GeoSES provides a multidimensional socioeconomic index for health research. By contrast, the HVI combines health-oriented vulnerability assessment with spatially constrained clustering to generate contiguous and operational planning territories, providing a

suitable analytical basis for ESF territorialization, workforce allocation, and service organization.

5.2 Decision-support layer for PHC allocation

The resulting spatial units function as a decision-support layer for PHC, providing operationally coherent and population-balanced territorial structures. Although this study does not implement facility location-allocation models, the derived regions are directly compatible with such optimization approaches to reduce inequities [61]. They can serve as structured inputs for workforce allocation, service localization, and catchment-area design in subsequent planning stages.

These findings have direct, practical implications for the Brazilian Unified Health System. By integrating a PHC-specific HVI with spatially constrained clustering, the framework generates actionable planning zones that support a better definition of ESF catchment areas and of team allocation. The framework’s scalability and replicability across municipal, state, and national levels bridge the gap between generic national indices and local micro-planning needs, substantially reducing planning complexity while preserving vulnerability patterns and strengthening equity-oriented territorial planning [3, 4].

Furthermore, the framework provides a practical mechanism for implementing current Brazilian PHC policies. It translates Ministry of Health recommendations for ESF territorial organization into transparent, reproducible planning units, enabling more consistent, needs-based planning across different administrative levels. While this study does not evaluate the impact of these policies on health service utilization, it offers a decision-support methodology that can facilitate their implementation and optimize PHC resource allocation.

6 Conclusion

The spatially constrained clustering approach applied to the proposed Health Vulnerability Index achieved a substantial reduction of approximately 77% in the number of health planning territories nationally, while generating contiguous, homogeneous, and population-balanced units aligned with Family Health Strategy requirements. In doing so, the framework overcomes key limitations associated with generic vulnerability indices and administrative boundaries that have long constrained evidence-based PHC planning [9, 10, 12]. More broadly, this work contributes a reproducible methodological framework for translating high-resolution population vulnerability into operational health planning regions, thereby connecting spatial epidemiology with evidence-based health system design.

From a practical standpoint, the methodology offers a transparent and scalable tool for identifying priority areas, reorganizing ESF coverage, and supporting more equitable resource allocation. By reducing operational complexity without sacrificing spatial coherence or vulnerability representation, it facilitates genuine microplanning within the SUS.

Theoretically, the study contributes to the literature on health vulnerability and territorial health planning by demonstrating how a purpose-built composite index, combined with spatially constrained clustering, can operationalize the social determinants of health paradigm at a scale relevant for service delivery [18, 21, 27].

The main limitation of the present work stems from the periodicity and static nature of demographic censuses, which are typically updated only once per decade. Rapid urban dynamics, migration, informal settlement growth, and socioeconomic transitions may outpace census updates. Continuous enrichment with more dynamic data sources will therefore be essential for sustained operational use.

Future research should incorporate annually updated service utilization data to better capture health demand alongside need, explore optimization-based PHC planning formulations aimed at maximizing accessibility and equity, and rigorously evaluate the impact of the newly defined planning units on actual PHC performance through prospective studies. Taken together, these advances would further strengthen the role of spatially explicit, needs-based approaches in the continuous improvement of primary health care systems.

Authored contribution (CRediT)

Misael de Souza Vieira Pinto: Conceptualization, Data Curation, Investigation, Methodology, Software, Visualization, Writing – Original Draft Preparation. **Gabriel Marcos de Moraes Nascimento:** Data Curation, Formal Analysis, Investigation, Methodology, Validation, Writing – Review & Editing. **Lygia Bueno Fragoso:** Conceptualization, Investigation, Project Administration, Resources, Writing – Review & Editing. **Elaine Leandro Machado:** Conceptualization, Formal Analysis, Investigation, Methodology, Validation, Writing – Review & Editing. **João Flávio de Freitas Almeida:** Conceptualization, Formal Analysis, Investigation, Project Administration, Supervision, Validation, Writing – Original Draft Preparation.

Declarations

Funding

This research was supported by Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) [Code: 001], the National Council for Scientific and Technological Development (CNPq) [Code: 444452/2023-0] - Brazil

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Research Data availability

The research data is available in a data repository: <https://github.com/misaelsvp/hvi-aps>
The results are available at <http://ivs-cluster-na-aps.s3-website-sa-east-1.amazonaws.com/>

Use of AI Technologies: Artificial intelligence tools were used for bibliographic search, organization, and improvement of the writing. The final content was reviewed by the authors, who take full responsibility for the work.

Ethics approval

We use aggregated data, anonymous in nature. Ethics approval is not required.

References

- [1] Marília Nascimento Carvalho et al. Necessidade e dinâmica da força de trabalho na atenção básica de saúde no brasil. *Ciência & Saúde Coletiva*, 23(1):295–302, 2018. doi:[10.1590/1413-81232018231.08702015](https://doi.org/10.1590/1413-81232018231.08702015).

- [2] Abdullah A Khan. An integrated approach to measuring potential spatial access to health care services. *Socio-economic planning sciences*, 26(4):275–287, 1992. doi:[10.1016/0038-0121\(92\)90004-O](https://doi.org/10.1016/0038-0121(92)90004-O).
- [3] James Macinko and Claunara Schilling Mendonça. Estratégia saúde da família, um forte modelo de atenção primária à saúde que traz resultados. *Saúde em Debate*, 42:18–37, 2018. doi:[10.1590/0103-11042018S102](https://doi.org/10.1590/0103-11042018S102).
- [4] Aylene Bousquat, Ligia Giovanella, Rosana Aquino, Ana Luiza Queiroz Vilasbôas, Elaine Thumé, Elaine Tomasi, Maria Helena Magalhães de Mendonça, Alaneir de Fátima da Silva, Patty Fidelis de Almeida, Paulo Henrique dos Santos Mota, et al. Primary health care assessment in brazil: conception and methodology of the 2024 national basic health unit census. *Cadernos de Saúde Pública*, 42:e00164625, 2026. doi:[10.1590/0102-311XEN164625](https://doi.org/10.1590/0102-311XEN164625).
- [5] Rosana Aquino and Ana Luiza Queiroz Vilasbôas. Impacts of the family health strategy on the health status of the brazilian population: a literature review. *Ciência & Saúde Coletiva*, 30:e13942025, 2026. doi:[10.1590/1413-812320253012.13942025EN](https://doi.org/10.1590/1413-812320253012.13942025EN).
- [6] María Eugenia Elorza et al. Estimating need, demand and supply in primary health care services: A local application in argentina. *MEDICC Review*, 20(3):36–44, 2018. URL <https://www.scielosp.org/article/medicc/2018.v20n3/36-44/>.
- [7] Nonglak Pagaiya et al. Forecasting imbalances of human resources for health in the Thailand health service system: application of a health demand method. *Human Resources for Health*, 17:4, 2019. doi:[10.1186/s12960-018-0336-2](https://doi.org/10.1186/s12960-018-0336-2).
- [8] Diana Fernandes Lopes, Ana Luísa Ramos, and Eduardo Anselmo de Castro. The health workforce demand: a systematic literature review. *Ciência & Saúde Coletiva*, 26(supl. 1), 2021. doi:[10.1590/1413-81232021266.1.40842020](https://doi.org/10.1590/1413-81232021266.1.40842020).
- [9] Olinda do Carmo Luiz, Luiza Sterman Heimann, Ruth C. Boaretto, Adriana G. Pacheco, Umberto C. Pessoto, Lauro C. Ibanhes, Ieda E. do Nascimento Castro, Jorge Kayano, Virginia Junqueira, José Luiz da Rocha, Carlos T. Cortizo, and Eurivaldo T. Junior. Differences in living conditions and health between cities: construction of a composite indicator. *Revista de Saúde Pública*, 43(1):115–122, 2009. doi:[10.1590/S0034-89102009000100015](https://doi.org/10.1590/S0034-89102009000100015).
- [10] Fabiana Brito Silva. Análise do índice de vulnerabilidade social das áreas de cobertura das equipes saúde da família e sua distribuição territorial em Uberaba–MG, 2010. Dissertação de mestrado em atenção à saúde, Universidade Federal do Triângulo Mineiro, Uberaba, MG, Brazil, 2011. URL <http://bdtd.ufbm.edu.br/handle/tede/149>.
- [11] Luiz Odorico Monteiro de Andrade, Alberto Pellegrini Filho, Orielle Solar, Félix Rígoli, Lígia Malagon de Salazar, Pastor Castell-Florit Serrate, Kelen Gomes Ribeiro, Theadora Swift Koller, Fernanda Natasha Bravo Cruz, and Rifat Atun. Social determinants of health, universal health coverage, and sustainable development: case studies from Latin American countries. *The Lancet*, 385(9975):1343–1351, 2015. doi:[10.1016/S0140-6736\(14\)61494-X](https://doi.org/10.1016/S0140-6736(14)61494-X).
- [12] Stan Openshaw. A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. *Transactions of the institute of british geographers*, pages 459–472, 1977. doi:[10.2307/622300](https://doi.org/10.2307/622300).

- [13] Instituto Brasileiro de Geografia e Estatística (IBGE). Censo demográfico 2022: Resultados preliminares e dicionário de variáveis. IBGE, 2023. URL <https://censo2022.ibge.gov.br/panorama/>.
- [14] Instituto de Pesquisa Econômica Aplicada (IPEA). Atlas da vulnerabilidade social nos municípios brasileiros, 2015. URL <https://ivs.ipea.gov.br>.
- [15] María Eugenia Elorza, Nebel Silvana Moscoso, and Anibal Manuel Blanco. Assessing performance in health care: A mathematical programming approach for the re-design of primary health care networks. *Socio-Economic Planning Sciences*, page 101454, 2022. doi:10.1016/j.seps.2022.101454.
- [16] Panagiotis Mitropoulos, Emmanuel Adamides, and Ioannis Mitropoulos. Redesigning a network of primary healthcare centres using system dynamics simulation and optimisation. *Journal of the Operational Research Society*, 74(2):574–589, 2023. doi:10.1080/01605682.2022.2096499.
- [17] João Flávio de Freitas Almeida, Fabricio Oliveira, Samuel Vieira Conceição, Virginie Goepp, and Francisco Carlos Cardoso de Campos. Designing a minimum-cost health system for countrywide universal coverage. *Health Systems*, 15(2):91–110, 2026. doi:10.1080/20476965.2025.2570686.
- [18] Susan L Cutter, Bryan J Boruff, and W Lynn Shirley. Social vulnerability to environmental hazards. In *Hazards vulnerability and environmental justice*, pages 115–132. Routledge, 2012. doi:10.4324/9781849771542.
- [19] Susan L Cutter. The origin and diffusion of the social vulnerability index (SoVI). *International journal of disaster risk reduction*, 109:104576, 2024. doi:10.1016/j.ijdrr.2024.104576.
- [20] Yung-Jaan Lee. Social vulnerability indicators as a sustainable planning tool. *Environmental Impact Assessment Review*, 44:31–42, 2014. doi:10.1016/j.eiar.2013.08.002.
- [21] Marcus Valerius Peixoto, Andrezza Marques Duque, Daya Devi Oliveira, Silvia Maria Voci, Allan Robert Silva, Marco Prado Nunes, et al. Construção e aplicação de um índice de vulnerabilidade em saúde. *Interfaces Científicas-Saúde e Ambiente*, 8(2):281–296, 2020. doi:10.17564/2316-3798.2020v8n2p281-296.
- [22] Raquel S. de S. Esteves, Juliana V. de M. Mambrini, Adriana C. Oliveira, and Márcia H. N. G. de Abreu. Performance of primary dental care services: an ecological study in a large brazilian city. *The Scientific World Journal*, 2013:176589, 2013. doi:10.1155/2013/176589. URL <https://doi.org/10.1155/2013/176589>.
- [23] Bruno Casaes Teixeira, Tatiana Natasha Toporcov, Francisco Chiaravalloti-Neto, and Alexandre Dias Porto Chiavegatto Filho. Spatial clusters of cancer mortality in brazil: A machine learning modeling approach. *International Journal of Public Health*, 68:1604789, 2023. doi:10.3389/ijph.2023.1604789. URL <https://doi.org/10.3389/ijph.2023.1604789>.
- [24] Daniela Contreras, Srirama Bhamidipati, and Sean Wilkinson. Social vulnerability and spatial inequality in access to healthcare facilities: The case of the santiago metropolitan region (rms), chile. *Socio-Economic Planning Sciences*, 90:101735, 2023. doi:10.1016/j.seps.2023.101735.

- [25] Francesco Vidoli and Monica Auteri. Health-care demand and supply at municipal level: A spatial disaggregation approach. *Socio-Economic Planning Sciences*, 84:101229, 2022. doi:[10.1016/j.seps.2022.101229](https://doi.org/10.1016/j.seps.2022.101229).
- [26] Saeed Yousefi, Reza Farzipoor Saen, Hadi Shabanpour, and Kian Ghods. An innovative patient clustering method using data envelopment analysis–discriminant analysis and artificial neural networks: A case study in healthcare systems. *Socio-economic planning sciences*, 95:102054, 2024. doi:[10.1016/j.seps.2024.102054](https://doi.org/10.1016/j.seps.2024.102054).
- [27] Juan C. Duque, Luc Anselin, and Sergio J. Rey. The Max-P-regions problem. *Journal of Regional Science*, 52(3):397–419, 2012. doi:[10.1111/j.1467-9787.2011.00743.x](https://doi.org/10.1111/j.1467-9787.2011.00743.x).
- [28] Maria Yury Travassos Ichihara, Dandara Ramos, Poliana Rebouças, Flávia Jôse Oliveira, Andrêa JF Ferreira, Camila Teixeira, Mirjam Allik, Srinivasa Vittal Katikireddi, Mauricio L Barreto, Alastair H Leyland, et al. Area deprivation measures used in Brazil: a scoping review. *Revista de saúde pública*, 52:83, 2018. doi:[10.11606/S1518-8787.2018052000933](https://doi.org/10.11606/S1518-8787.2018052000933).
- [29] Ligia Vizeu Barrozo, Michel Fornaciali, Carmen Diva Saldiva de André, Guilherme Augusto Zimeo Morais, Giselle Mansur, William Cabral-Miranda, Marina Jorge de Miranda, João Ricardo Sato, and Edson Amaro Junior. GeoSES: A socioeconomic index for health and social research in Brazil. *PloS one*, 15(4):1–17, 2020. doi:[10.1371/journal.pone.0232074](https://doi.org/10.1371/journal.pone.0232074).
- [30] Lorena Sampaio Almeida, Ana Lúcia Soares Cota, and Diego Freitas Rodrigues. Sanitation, arboviruses, and environmental determinants of disease: impacts on urban health. *Ciencia & saude coletiva*, 25:3857–3868, 2020. doi:[10.1590/1413-812320202510.30712018](https://doi.org/10.1590/1413-812320202510.30712018).
- [31] Diogo Cunha Ferreira, Ingrid Grazielle, Rui Cunha Marques, and Jorge Gonçalves. Investment in drinking water and sanitation infrastructure and its impact on waterborne diseases dissemination: The brazilian case. *Science of the Total Environment*, 779:146279, 2021. doi:[10.1016/j.scitotenv.2021.146279](https://doi.org/10.1016/j.scitotenv.2021.146279).
- [32] Fernanda Siqueira Malta and Eduarda Marques da Costa. Socio-environmental vulnerability index: an application to rio de janeiro-brazil. *International journal of public health*, 66:584308, 2021. doi:[10.3389/ijph.2021.584308](https://doi.org/10.3389/ijph.2021.584308).
- [33] Nancy D Berkman, Stacey L Sheridan, Katrina E Donahue, David J Halpern, and Karen Crotty. Low health literacy and health outcomes: an updated systematic review. *Annals of internal medicine*, 155(2):97–107, 2011. doi:[10.7326/0003-4819-155-2-201107190-00005](https://doi.org/10.7326/0003-4819-155-2-201107190-00005).
- [34] Federica Alfani, Fabio Clementi, Michele Fabiani, Vasco Molini, Enzo Valentini, et al. Once NEET, always NEET? A synthetic panel approach to analyze the Moroccan labor market. *World Bank Policy Research Working Paper*, 9238:1–33, 2020. URL <https://u-pad.unimc.it/handle/11393/265558>.
- [35] Mark J Van Ryzin, Diana Fishbein, and Anthony Biglan. The promise of prevention science for addressing intergenerational poverty. *Psychology, Public Policy, and Law*, 24(1):128, 2018. doi:[10.1037/law0000138](https://doi.org/10.1037/law0000138).
- [36] Thanya I Pathirana and Caroline A Jackson. Socioeconomic status and multimorbidity: a systematic review and meta-analysis. *Australian and New Zealand journal of public health*, 42(2):186–194, 2018. doi:[10.1111/1753-6405.12762](https://doi.org/10.1111/1753-6405.12762).

- [37] Mireia Utzet, Ferran Botías, Michael Silva-Peñaherrera, Aurelio Tobías, and Fernando G Benavides. Informal employment and poor self-perceived health in latin america and the caribbean: a gender-based comparison between countries and welfare states in a pooled analysis of 176,786 workers. *Globalization and health*, 17(1):140, 2021. doi:[10.1186/s12992-021-00792-3](https://doi.org/10.1186/s12992-021-00792-3).
- [38] Nathan Tefft and Andrew Kageleiry. State-level unemployment and the utilization of preventive medical services. *Health services research*, 49(1):186–205, 2014. doi:[10.1111/1475-6773.12091](https://doi.org/10.1111/1475-6773.12091).
- [39] Keylla Talitha Fernandes Barbosa, Fabiana Maria Rodrigues Lopes de Oliveira, and Maria das Graças Melo Fernandes. Vulnerability of the elderly: a conceptual analysis. *Revista brasileira de enfermagem*, 72(suppl 2):337–344, 2019. doi:[10.1590/0034-7167-2018-0728](https://doi.org/10.1590/0034-7167-2018-0728).
- [40] Clio Ciaschini and Luca Salvati. Age dependency ratio. In *Sustainable Development: Official Statistics and Thematic Indicators as a Compass to Understand Complexity*, pages 35–48. Springer, 2026. doi:[10.1007/978-3-032-12559-0_3](https://doi.org/10.1007/978-3-032-12559-0_3).
- [41] Lee J Cronbach. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334, 1951. doi:[10.1007/BF02310555](https://doi.org/10.1007/BF02310555).
- [42] Patrick AP Moran. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23, 1950. doi:[10.2307/2332142](https://doi.org/10.2307/2332142).
- [43] Andrew David Cliff and J Keith Ord. *Spatial processes: Models & Applications*, volume 44. Pion London, 1981. ISBN 9780850860818.
- [44] David C. Folch and Seth E. Spielman. Identifying regions based on flexible user-defined constraints. *International Journal of Geographical Information Science*, 28(1):164–184, 2014. doi:[10.1080/13658816.2013.848986](https://doi.org/10.1080/13658816.2013.848986).
- [45] Yongyi Kang and Amr Magdy. EMP: Max-P regionalization with enriched constraints. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pages 1914–1926. IEEE, 2022. doi:[10.1109/ICDE53745.2022.00189](https://doi.org/10.1109/ICDE53745.2022.00189).
- [46] Brittany L. Melix and Christopher K. Uejio. Delineating health enumeration areas in Florida. *Geographical Analysis*, 58(3), 2026. doi:[10.1111/gean.70048](https://doi.org/10.1111/gean.70048).
- [47] Mohsen Tavakol and Reg Dennick. Making sense of Cronbach’s alpha. *International Journal of Medical Education*, 2:53–55, 2011. doi:[10.5116/ijme.4dfb.8dfd](https://doi.org/10.5116/ijme.4dfb.8dfd).
- [48] Susan L. Cutter, Christopher T. Emrich, Daniel P. Morath, and Corey M. Dunning. Integrating social vulnerability into federal flood risk management planning. *Journal of Flood Risk Management*, 6(4):332–344, 2013. doi:[10.1111/jfr3.12018](https://doi.org/10.1111/jfr3.12018).
- [49] Liton Chakraborty, Horatiu Rus, Daniel Henstra, Jason Thistlethwaite, and Daniel Scott. A place-based socioeconomic status index: Measuring social vulnerability to flood hazards in the context of environmental justice. *International Journal of Disaster Risk Reduction*, 43:101394, 2020. doi:[10.1016/j.ijdr.2019.101394](https://doi.org/10.1016/j.ijdr.2019.101394).
- [50] Charles Spearman. The proof and measurement of association between two things. *The American journal of psychology*, 100(3/4):441–471, 1987. doi:[10.2307/1412159](https://doi.org/10.2307/1412159).

- [51] David N. Bernstein, David Shin, Rudolf W. Poolman, Joseph H. Schwab, and Daniel G. Tobert. Are commonly used geographically based social determinant of health indices in orthopaedic surgery research correlated with each other and with PROMIS Global-10 physical and mental health scores? *Clinical Orthopaedics & Related Research*, 482(4):604–614, 2024. doi:[10.1097/CORR.0000000000002896](https://doi.org/10.1097/CORR.0000000000002896).
- [52] Olivia L. Schultes, Maria Helena F. Morais, Maria da Conceição M. Cunha, Anderson Sobral, and Waleska T. Caiaffa. Spatial analysis of dengue incidence and *Aedes aegypti* ovitrap surveillance in Belo Horizonte, Brazil. *Tropical Medicine & International Health*, 26(2):237–255, 2021. doi:[10.1111/tmi.13521](https://doi.org/10.1111/tmi.13521).
- [53] Joe H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963. doi:[10.1080/01621459.1963.10500845](https://doi.org/10.1080/01621459.1963.10500845).
- [54] Marie Chavent, Vanessa Kuentz-Simonet, Amaury Labenne, and Jérôme Saracco. ClustGeo: an R package for hierarchical clustering with spatial constraints. *Computational Statistics*, 33(4):1799–1822, 2018. doi:[10.1007/s00180-018-0791-1](https://doi.org/10.1007/s00180-018-0791-1).
- [55] Yiqun Chen, Qiliang Liu, Jie Yang, and Xiaoqing Cheng. Spatially constrained statistical approach for determining the optimal number of regions in regionalization. *International Journal of Geographical Information Science*, 38(10):2108–2147, 2024. doi:[10.1080/13658816.2024.2372779](https://doi.org/10.1080/13658816.2024.2372779).
- [56] Jörg Kalcsics and Roger Z. Ríos-Mercado. Districting problems. In Gilbert Laporte, Stefan Nickel, and Francisco Saldanha da Gama, editors, *Location Science*. Springer, Cham, 2 edition, 2019. doi:[10.1007/978-3-030-32177-2_25](https://doi.org/10.1007/978-3-030-32177-2_25).
- [57] Prefeitura de Belo Horizonte. Saúde - prefeitura de belo horizonte, 2026. URL <https://prefeitura.pbh.gov.br/saude>.
- [58] Ministério da Saúde and Secretaria de Atenção à Saúde. Atenção primária à saúde no brasil: guia para organização e funcionamento das equipes de estratégia saúde da família. Technical report, Ministério da Saúde, Brasília, 2015. URL <https://docs.bvsalud.org/biblioref/coleciona-sus/2015/34424/34424-895.pdf>.
- [59] Ministério da Saúde. Portaria GM/MS nº 3.493, de 10 de abril de 2024, 2024. URL <https://www.gov.br/saude/pt-br/composicao/saps/esf/legislacao>. Institui a nova metodologia de cofinanciamento federal da Atenção Primária à Saúde.
- [60] Secretaria de Atenção Primária à Saúde. Nota técnica nº 122/2024/cgfap/saps/ms. Technical report, Ministério da Saúde, Brasília, 2024. URL <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/notas-tecnicas/2024/nota-tecnica-no-122-2024-cgfap-saps-ms.pdf>. Metodologia do Índice de Equidade e Dimensionamento (IED).
- [61] Özlem Karsu and Alec Morton. Inequity averse optimization in Operational Research. *European Journal of Operational Research*, 245(2):343–359, 2015. doi:[10.1016/j.ejor.2015.02.035](https://doi.org/10.1016/j.ejor.2015.02.035).

Appendix – Variables Used in the Health Vulnerability Index (HVI)

The proposed Health Vulnerability Index (HVI) was constructed exclusively from variables of the 2022 Brazilian Demographic Census (IBGE). This appendix describes every indicator used in the four dimensions of the HVI together with the corresponding IBGE variables, allowing complete reproducibility of the index. The HVI is composed of four dimensions: Urban Infrastructure (UI); Human Capital (HC); Income and Employment (IE); and Demographic Vulnerability (DV). All indicators were computed at the census-sector level. Variables expressed as proportions were normalized to national percentile ranks before aggregation, as described in Section 3.2.

Urban Infrastructure (UI)

The Urban Infrastructure dimension measures deficiencies in basic sanitation, housing conditions, and environmental services that directly affect population health.

Indicator	Definition
1) Inadequate water supply	$(V00112+V00113+...+V00118)/V00001$
2) Inadequate sewage system	$(V00311+V00312+...+V00316)/V00001$
3) No garbage collection	$(V00399+V00400+V00401+V00402)/V00001$
4) Households without bathroom	$V00495/V00001$
5) Overcrowded households	$(V00023+V00024+V00025+V00026)/V00001$

Table 2: Indicators composing the Urban Infrastructure dimension.

Code	Description
V00001	Occupied permanent private households (denominator)
V00112	Water supplied by deep or artesian well
V00113	Water supplied by shallow well, groundwater well or cistern
V00114	Water supplied by spring, fountain or natural source
V00115	Water supplied by water truck
V00116	Water supplied by stored rainwater
V00117	Water supplied by rivers, lakes, streams or reservoirs
V00118	Other water supply source
V00311	Sewage disposed into septic tank (not connected to sewer)
V00312	Sewage disposed into rudimentary pit
V00313	Sewage disposed into ditch
V00314	Sewage discharged into river, lake or sea
V00315	Other sewage disposal method
V00316	No sewage disposal facility
V00399	Garbage burned on the property
V00400	Garbage buried on the property
V00401	Garbage dumped on vacant lots or public areas
V00402	Other garbage disposal method
V00495	Household without private bathroom with shower and toilet
V00023	Household with 7 residents
V00024	Household with 8 residents
V00025	Household with 9 residents
V00026	Household with 10 or more residents

Table 3: IBGE Census 2022 variables used in the Urban Infrastructure dimension.

Human Capital (HC)

This dimension captures educational attainment and family characteristics associated with health literacy and social vulnerability.

Indicator	Definition
1) Illiteracy rate (15+)	$V00901/(V00900+V00901)$
2) Youth illiteracy (15–29)	$V00853/(V00852+V00853)$
3) Female-headed households without spouse	$V01204/V00001$
4) Illiterate household heads	$V00985/(V00984+V00985)$
5) Illiterate female household heads	$V00989/(V00988+V00989)$

Table 4: Indicators composing the Human Capital dimension.

Code	Description
V00900	Literate residents aged 15 years or older
V00901	Illiterate residents aged 15 years or older
V00852	Literate residents aged 15–29 years
V00853	Illiterate residents aged 15–29 years
V01204	Female household head without spouse, living with child(ren) and/or stepchild(ren)
V00984	Literate household head aged 15 years or older
V00985	Illiterate household head aged 15 years or older
V00988	Literate female household head aged 15 years or older
V00989	Illiterate female household head aged 15 years or older

Table 5: IBGE Census 2022 variables used in the Human Capital dimension.

Income and Employment (IE)

The Income and Employment dimension captures socioeconomic deprivation affecting access to health services.

Indicator	Definition
1) Low household-head income	Median income of household heads below one-half of the Brazilian minimum wage (variable V06006).
2) Adolescent household heads	$V01064/V00001$
3) Substandard housing	$(V00050+...+V00053+V00055+...+V00057)/(V00001+V00002)$

Table 6: Indicators composing the Income and Employment dimension.

Demographic Vulnerability (DV)

The Demographic Vulnerability dimension captures age structure and dependency characteristics associated with higher Primary Health Care needs.

Code	Description
V00001	Occupied permanent private households
V00002	Occupied makeshift private dwellings
V01064	Household head aged 12–17 years
V06006	Household head monthly income
V00050	Room-for-rent housing or tenement
V00052	Dilapidated or unfinished permanent residential structure
V00053	Tent or temporary shelter
V00055	Natural shelter or other makeshift structure
V00056	Makeshift structure on public land
V00057	Dilapidated or unfinished permanent non-residential structure

Table 7: IBGE Census 2022 variables used in the Income and Employment dimension.

Indicator	Definition
1) Population aged 65 years or older	$(V00654+V00655+V00656)/V01006$
2) Population aged 0–4 years	$V01031/V01006$
3) Households headed by older adults	$V01068/V00001$
4) Demographic dependency ratio	$V01031+...+V01033+V01040+V01041/(V01034+...+V01039)$
5) Premature mortality (30–69 years)	$V01234+...+V01237+V01245+...+V01248/(V01037+...+V01040)$

Table 8: Indicators composing the Demographic Vulnerability dimension.

Code	Description
V00001	Occupied permanent private households
V01006	Total resident population
V00654	Population aged 65–69 years
V00655	Population aged 70–79 years
V00656	Population aged 80 years or older
V01031	Population aged 0–4 years
V01032	Population aged 5–9 years
V01033	Population aged 10–14 years
V01034	Population aged 15–19 years
V01035	Population aged 20–24 years
V01036	Population aged 25–29 years
V01037	Population aged 30–39 years
V01038	Population aged 40–49 years
V01039	Population aged 50–59 years
V01040	Population aged 60–69 years
V01041	Population aged 70 years or older
V01068	Household head aged 60 years or older
V01234	Male deaths aged 30–39 years
V01235	Male deaths aged 40–49 years
V01236	Male deaths aged 50–59 years
V01237	Male deaths aged 60–69 years
V01245	Female deaths aged 30–39 years
V01246	Female deaths aged 40–49 years
V01247	Female deaths aged 50–59 years
V01248	Female deaths aged 60–69 years

Table 9: IBGE Census 2022 variables used in the Demographic Vulnerability dimension.

This preprint was submitted under the following conditions:

- The authors declare that the necessary Terms of Free and Informed Consent of participants or patients in the research were obtained and are described in the manuscript, when applicable.
- The authors declare that the preparation of the manuscript followed the ethical norms of scientific communication.
- The authors declare that they are aware that they are solely responsible for the content of the preprint and that the deposit in SciELO Preprints does not mean any commitment on the part of SciELO, except its preservation and dissemination.
- The authors declare that the data, applications, and other content underlying the manuscript are referenced.
- The deposited manuscript is in PDF format.
- The authors declare that the research that originated the manuscript followed good ethical practices and that the necessary approvals from research ethics committees, when applicable, are described in the manuscript.
- The authors declare that once a manuscript is posted on the SciELO Preprints server, it can only be taken down on request to the SciELO Preprints server Editorial Secretariat, who will post a retraction notice in its place.
- The authors agree that the approved manuscript will be made available under a [Creative Commons CC-BY](#) license.
- The submitting author declares that the contributions of all authors and conflict of interest statement are included explicitly and in specific sections of the manuscript.
- The authors declare that the manuscript was not deposited and/or previously made available on another preprint server or published by a journal.
- If the manuscript is being reviewed or being prepared for publishing but not yet published by a journal, the authors declare that they have received authorization from the journal to make this deposit.
- The submitting author declares that all authors of the manuscript agree with the submission to SciELO Preprints.