Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Geographic heterogeneity of recorded public diabetes care in Paraguay: a nationwide ecological study with Bayesian spatial analysis across harmonised administrative areas, 2023–2025

Raul Desvars

rauldesvars@unc.edu.py

Universidad Nacional de Concepción

Jorge Rojas Godoy

Universidad Nacional de Concepción

Landdy Gonzalez

Universidad Nacional de Concepción

Research Article

Keywords: Health geography, Health services utilisation, Diabetes mellitus, Spatial epidemiology, Indirect standardisation, BYM2, Administrative boundary harmonisation, Routine health data, Paraguay

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11540923/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Geographic heterogeneity of recorded public diabetes care in Paraguay: a nationwide ecological study with Bayesian spatial analysis across harmonised administrative areas, 2023–2025

Raul Desvars, Jorge Rojas Godoy, Landdy Gonzalez

Affiliation: Facultad de Medicina, Universidad Nacional de Concepción, Concepción, Paraguay

Correspondence: rauldesvars@unc.edu.py

Abstract

Background: Routine administrative records are increasingly mapped for health surveillance, but boundary changes, population composition and place-of-care attribution complicate interpretation. We examined the geography of recorded public diabetes care in Paraguay and the contribution of demographic standardisation, spatial dependence statistics, Bayesian disease mapping and cluster detection.

Methods: Nationwide ecological analysis of annual counts of people attended for diabetes (ICD-10 E10–E14) in outpatient services of the Ministry of Public Health, 2023–2025. The 263 current districts were aggregated to 246 polygons of the 2012 open cartography. Expected counts were indirectly standardised for sex and age with district projections and the 2022 census; crude normalisation was a sensitivity analysis. We applied empirical Bayes Moran statistics, Kulldorff's scan statistic and BYM2 models, including a referral-hospital adjustment.

Results: We analysed 510,221 records (26.7 per 1000 population annually); from 15 years of age, rates in women were 1.6–1.9 times those in men. Standardisation changed expected counts by factors of 0.68–1.94 and reduced the empirical Bayes Moran's I from 0.39 to 0.11

(p = 0.005); one local Moran cluster remained after false discovery rate control. The BYM2 latent standard deviation was 0.47 (95% credible interval 0.42–0.55) and the spatial mixing parameter 0.56 (0.21–0.86). Fifty-seven areas, mainly in the south-centre of the Eastern Region and often departmental capitals, had a posterior probability ≥0.80 of a standardised ratio above 1.25, and 64, notably in the Chaco, along the Brazilian border and around Asunción, of a ratio below 0.80. Areas with a referral hospital had higher recorded use (relative ratio 1.42; 1.21–1.64), but adjustment changed the pattern little. Annual rankings were stable (Spearman 0.77–0.88). Every candidate scan cluster reached the minimum Monte Carlo p-value.

Conclusions: Recorded public diabetes care in Paraguay is geographically heterogeneous, but much of the regional structure of crude rates reflects population sex and age composition. Because events are attributed to the place of care, standardised patterns describe service activity, not prevalence, coverage or unmet need. Demographic standardisation should precede spatial inference on routine data; with large counts, Bayesian models add uncertainty-based classification more than smoothing, and scan p-values discriminate poorly.

Keywords: Health geography; Health services utilisation; Diabetes mellitus; Spatial epidemiology; Indirect standardisation; BYM2; Administrative boundary harmonisation; Routine health data; Paraguay

Background

Diabetes mellitus is an expanding public health challenge, particularly in low- and middle-income countries [1]. Paraguay's second national survey of risk factors, undertaken in 2022, estimated that 10.6% of adults aged 18–69 years had diabetes [2]. Alongside survey data, routine administrative records provide a geographically detailed description of service activity, but not a direct measure of disease prevalence.

Health geography distinguishes potential spatial accessibility from realised service use. The location at which care is delivered can differ substantially from the residence of the person receiving it, and referral centres can concentrate activity from broad catchment areas [3,4]. Consequently, an administrative ratio that divides service events occurring in one location by that location's resident population is a measure of recorded service intensity, rather than a residence-based coverage rate.

Several related challenges arise when routine administrative data are mapped. Areas differ in the sex and age composition of their populations, which shapes the expected demand for care of a chronic disease. Small populations may produce unstable ratios. Spatial autocorrelation may reflect shared infrastructure, referral networks or reporting practices rather than aetiological clustering. Finally, administrative units can change over time; aggregation to a constant geography reduces temporal discontinuities at the cost of spatial detail and may introduce the modifiable areal unit problem [5]. Bayesian disease-mapping models, local indicators of spatial association and scan statistics address different aspects of these challenges [6–9].

Previous Latin American studies have identified geographic heterogeneity in diabetes-related hospitalisations and mortality [10,11], but these outcomes are distinct from the outpatient use of a single public provider network. To our knowledge, the geography of diabetes care across all districts of Paraguay has not been described. The availability of openly published Paraguayan records, district population projections, census tables and public geographic information creates an opportunity to investigate how these constraints affect the description of care delivery. The constraints are common to routine health information systems in many countries, so methods that address them with open data are of interest beyond Paraguay.

We aimed to characterise geographic differences in recorded diabetes care delivered through the network of Paraguay's Ministry of Public Health and Social Welfare (MSPBS) during

2023–2025, to assess how much of the apparent geographic structure reflects the sex and age composition of district populations, to quantify the contribution of spatially structured random effects, to compare complementary procedures for local cluster identification, and to provide a reproducible workflow, transferable to other settings, for harmonising changing administrative units and standardising place-of-care counts with open data. Particular attention was paid to the interpretation of place-of-care counts and to the use of a constant historical boundary system.

Methods

Design, setting and reporting

This nationwide ecological analysis used annually aggregated, publicly accessible administrative records. Paraguay consists of Asunción and 17 departments; the current district frame of the National Institute of Statistics (INE) contains 263 districts [12]. The principal spatial units in this analysis were 246 harmonised historical polygons, not the 263 current districts. The study is reported according to the REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) extension of the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [13]; the completed checklist is provided in Additional file 1.

Outcome and administrative source audit

The Directorate-General for Strategic Health Information (DIGIES) of the MSPBS publishes the number of people recorded as attended for diabetes mellitus in outpatient consultations, disaggregated by year, category of the International Classification of Diseases, 10th Revision (ICD-10; E10–E14), sex, age group, health region and district of consultation [14]. The outcome analysed was the annual aggregate count reported by district of care. Records for 2025 are provisional; partially reported 2026 data were excluded. As no patient-level

identifiers were available, sums over years represent annual administrative records and may count the same individual in more than one year. Duplication across facilities or diagnostic subgroups within a year cannot be excluded.

The original workbook was audited with a standard-library Python program that extracts the counts by source row and reconciles annual totals with those reported in the workbook (Additional files 2–4). The 267 distinct named region–district entries were reduced to 262 reporting labels after reconciling five pairs of alternative spellings or names of the same district used in different years. This is one fewer named unit than the 263-district national frame: San José del Rosario (San Pedro) has no entry in the source, and absence of an entry was not interpreted as zero activity. The 2025 extract contained 1,381 records without a district code in the single-district Capital health region and two in Caazapá; the former were assigned to Asunción and the latter excluded. All remaining district-year records were retained, giving an analytic total of 510,221.

Population denominators and relative-use measures

District populations for 2023, 2024 and 2025 came from the INE 2025 revision of district estimates and projections [12]. Because these estimates are not published by sex and age at district level, the total of each area was distributed across ten strata, defined by sex and by the five age groups of the DIGIES tables (under 15, 15–29, 30–49, 50–64 and 65 years or older), according to the population structure of the area in the 2022 National Population and Housing Census [15]. The census tables for the 263 districts were aggregated to the analytical areas, and their departmental totals were checked against the published totals. Expected counts were obtained by indirect standardisation: for year t, $E_{it} = \Sigma_k N_{itk} \times (\Sigma_j O_{jtk} / \Sigma_j N_{jtk})$, where N_{itk} is the estimated population of area i in stratum k and O_{jtk} is the recorded count in that stratum attributed to place of care. Pooled relative-use ratios were calculated as $\Sigma_t O_{it} / \Sigma_t E_{it}$, so that a ratio of 1 indicates the activity expected if the national sex- and age-specific rates of each

year applied to the resident population of the area. As a sensitivity analysis, all analyses were repeated with expected counts based on national crude rates, $E_{it} = N_{it} \times (\Sigma_j O_{jt} / \Sigma_j N_{jt})$. Annual crude rates were presented per 1000 resident population, and ratios were interpreted solely as recorded public-service intensity relative to the national benchmark. Exact Poisson intervals were calculated around observed-to-expected (O/E) ratios as a descriptive measure of counting uncertainty; these intervals do not incorporate error in population projections, inter-district care flows or heterogeneity of ascertainment.

Historical geographic harmonisation

The spatial analysis used the open geoBoundaries district geometry, which reproduces the 2012 administrative cartography [16]. A multipart feature merging Bella Vista (Itapúa) and Bella Vista Norte (Amambay) was separated, and two spurious polygons smaller than 0.3 km² were dissolved into San Juan del Paraná. Current district names were matched to historical areas using normalised strings and a reviewed list of aliases. The 17 districts created after 2012 were aggregated into the historical area containing the majority of their Family Health Units (USF), according to the coordinates of the national USF list [17] (Additional file 5). The result comprised 246 analytical polygons, to which recorded counts, population estimates and census populations were aggregated.

The 2012 geometry was used because it is the most recent district layer distributed under an open licence in a machine-readable form, which allows the complete workflow to be reproduced. Updated cartography has been developed for the 2022 census, but around 40% of districts lack legal instruments that define their limits [18], so contemporary district boundaries are themselves partly conventional. The results therefore refer to the 246 harmonised areas rather than to contemporary district polygons; this choice reduces geographic detail and is a possible source of zoning sensitivity. No inference is drawn about differences within aggregated parent areas.

Spatial dependence and local clusters

Neighbour relationships were defined by queen contiguity, with a mean of 5.4 neighbours per area, no isolated areas and a single connected component. Global autocorrelation was assessed using the population-adjusted empirical Bayes formulation of Moran's I, with 9,999 permutations [19]. Local empirical Bayes Moran statistics were assessed with Benjamini–Hochberg false discovery rate control at 5% [8,20]. High–high and low–low categories describe adjacency patterns in relative use, not geographic clustering of individuals with diabetes.

Circular spatial clusters were additionally assessed with Kulldorff's Poisson scan statistic [9], centred on representative points of the polygons, with a maximum window of 10% of the 2025 population and a 25% maximum in a sensitivity analysis. Statistical significance was assessed with 999 Monte Carlo replicates conditional on the observed total, so that the minimum attainable p-value was 0.001. Non-overlapping secondary clusters were considered, and results were interpreted by their relative-use contrast rather than by statistical significance alone.

Bayesian hierarchical model

For the pooled period, observed counts were analysed with the BYM2 model, a reparameterisation of the Besag–York–Mollié model [6,7]. Counts O_i were modelled as Poisson($E_i\theta_i$), with $\log \theta_i = \beta_0 + \sigma(\sqrt{\rho}\cdot u^*_i + \sqrt{(1-\rho)}\cdot v_i)$, where u* is an intrinsic conditional autoregressive component with a sum-to-zero constraint, scaled so that the geometric mean of its marginal variances equals one, and v is an independent standard normal component [7]. The parameter σ is the total standard deviation of the latent log-scale random effects, and ρ is the proportion of that variance allocated to the spatially structured component. Neither parameter estimates a causal effect of geography or the fraction of all observed variability explained by location.

A penalised-complexity prior with $Pr(\sigma > 1) = 0.01$, a Beta(0.5, 0.5) prior for ρ and a Normal(0, 5^2) prior for β_0 were specified [21]. Hamiltonian Monte Carlo with the No-U-Turn sampler used four chains with 2000 warm-up and 2000 retained iterations each. Convergence was assessed with the rank-normalised R-hat and the effective sample size [22]. Areas were described as substantially elevated when $Pr(\theta_i > 1.25 \mid \text{data}) \geq 0.80$ and substantially reduced when $Pr(\theta_i < 0.80 \mid \text{data}) \geq 0.80$ [23]. The same model was fitted separately for each year, with four chains of 1000 warm-up and 1000 retained iterations, and temporal persistence was quantified with Spearman rank correlations between annual posterior medians.

A secondary BYM2 model included an indicator of at least one regional, general or specialised hospital in the area, based on the 2025 MSPBS facility register [24]. Because this is an ecological indicator of service availability, its coefficient was interpreted as an association with place-of-care counts rather than as evidence about patient referral pathways.

Software and reproducibility

Analyses were conducted in Python 3 with pandas, GeoPandas, libpysal and esda (PySAL) [25], PyMC [26] and ArviZ. The analysis code, the corrected boundary layer, the district crosswalk, the census tables and the derived datasets are provided in Additional file 6. A large language model (Claude, Anthropic) was used to assist with drafting analysis code and with language editing; the authors reviewed the code, the outputs and the text, and take responsibility for the content.

Results

Population and recorded consultations

The DIGIES source contained 153,563 annual records in 2023, 175,624 in 2024 and 181,036 in the provisional 2025 data. After the exclusion of two 2025 records without an assignable district, the analysed total was 510,221 (153,563, 175,624 and 181,034 by year). The pooled

national crude rate was 26.7 records per 1000 inhabitants per year, and annual rates were 24.3, 27.6 and 28.2 per 1000. The median cumulative count per harmonised area was 820 (interquartile range 450–1991), and six areas had fewer than 100 records.

Recorded rates rose steeply with age, from 0.7 per 1000 person-years in boys and 0.8 in girls under 15 years to 86.3 in men and 140.1 in women aged 65 years or older, and were highest in women aged 50–64 years (108.6 per 1000). In each age group from 15 years onwards, rates in women were 1.6–1.9 times those in men.

Crude rates ranged from 8.1 per 1000 in Boquerón to 53.8 in Misiones (Table 1). High crude rates predominated in the southern and central Eastern Region, whereas low rates appeared in the Chaco, along much of the Brazilian border and in parts of the metropolitan area (Figure 1).

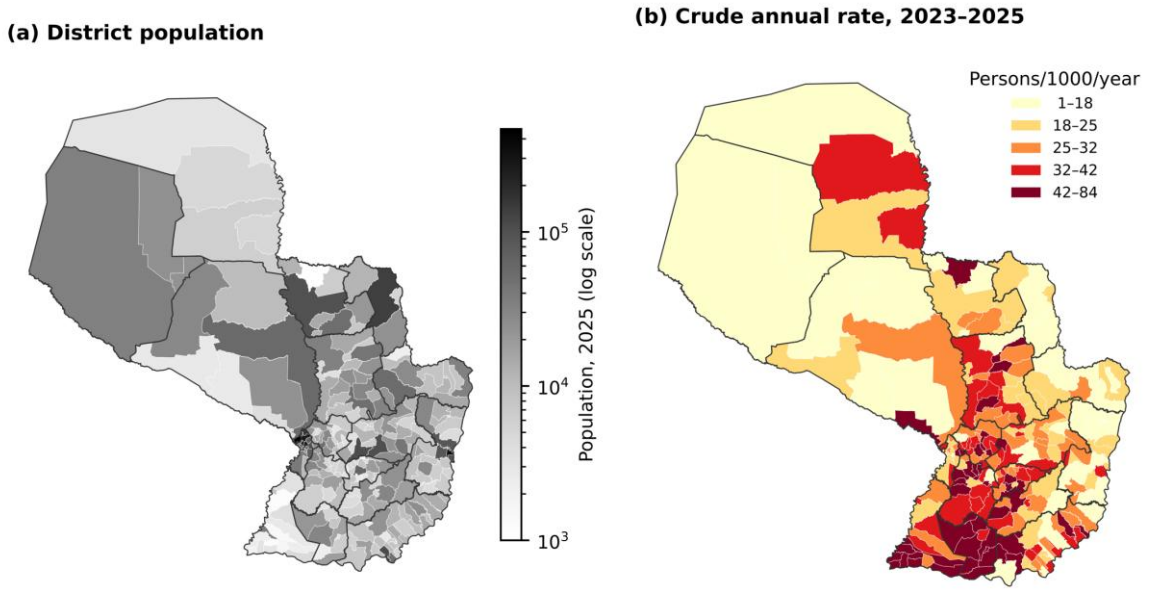


Figure 1. District population and crude rate of recorded public diabetes care, 2023–2025. (a) Population in 2025 on a logarithmic scale. (b) Annual rate of people recorded as attended for diabetes in the MSPBS network per 1000 resident population, 2023–2025, in quintiles. Thick lines are department boundaries; both panels use the 246 harmonised areas.

Table 1. Recorded public diabetes care, crude and sex- and age-standardised relative use by department, 2023–2025

Department	Areas	Records	Rate per 1000 per year	Crude O/E	Standardised O/E
Misiones	10	18346	53.8	2.02	1.81
Guairá	18	23226	42.0	1.57	1.38
Ñeembucú	16	10820	45.9	1.72	1.30
Alto Paraguay	4	1554	28.0	1.05	1.27
Paraguarí	17	25722	41.0	1.54	1.22
Asunción	1	53655	38.3	1.44	1.16
San Pedro	20	33652	29.8	1.12	1.15
Cordillera	20	27365	32.8	1.23	1.11
Caazapá	11	14044	31.2	1.17	1.10
Itapúa	30	38835	27.9	1.05	1.03
Caaguazú	22	37666	27.3	1.02	1.03
Canindeyú	12	11806	19.1	0.72	0.95
Central	19	142903	24.1	0.90	0.92
Concepción	9	15033	22.7	0.85	0.90
Amambay	4	9377	17.9	0.67	0.78
Presidente Hayes	8	6737	17.1	0.64	0.74
Alto Paraná	22	37542	15.9	0.60	0.70
Boquerón	3	1938	8.1	0.30	0.38
Paraguay	246	510221	26.7	1.00	1.00

Areas: harmonised historical polygons. Records: annual records pooled over 2023–2025. Crude O/E: observed records divided by the records expected from national year-specific crude rates. Standardised O/E: observed records divided by the records expected from national year-, sex- and age-specific rates (indirect standardisation). Population denominators refer to residence, whereas recorded activity refers to place of care. Departments are ordered by standardised O/E.

Effect of sex and age standardisation

Standardisation changed the expected count of each area by a factor of 0.68 to 1.94 (10th–90th percentiles 0.80–1.37) relative to crude normalisation (Figure 2a). Expected counts increased in departments with older populations, such as Ñeembucú (1.33), Paraguarí (1.26) and Asunción (1.24), and decreased in departments with younger populations, such as Canindeyú (0.76), Boquerón (0.79) and Alto Paraná (0.86). Standardised and crude ratios

were correlated (Spearman 0.89), but departmental contrasts narrowed: standardised O/E ranged from 0.38 in Boquerón to 1.81 in Misiones, compared with 0.30 to 2.02 for crude O/E (Table 1, Figure 2b).

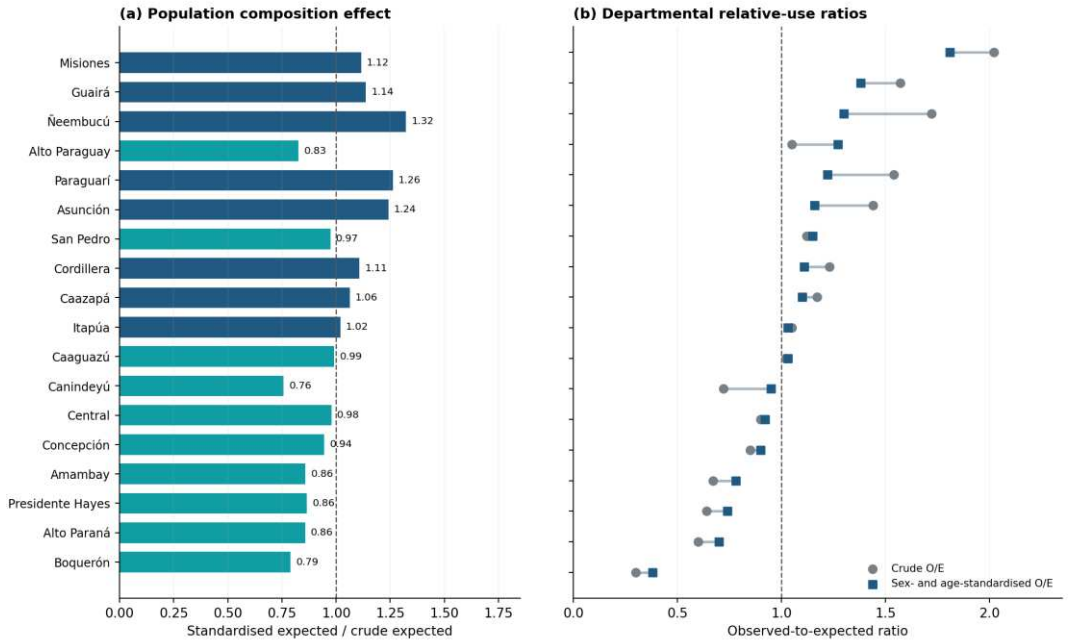


Figure 2. Effect of sex and age standardisation on recorded public diabetes care, 2023–2025. (a) Department-level ratio of expected records under indirect sex- and age-standardisation to those under national crude rates; a ratio above 1 indicates a larger adjusted expectation. (b) Crude and standardised observed-to-expected ratios at department level. Values are derived from Table 1 and must not be interpreted as residence-based coverage or disease prevalence.

Spatial autocorrelation and Bayesian smoothing

The pooled empirical Bayes global Moran's I of standardised ratios was 0.11 (null expectation −0.004; permutation $p = 0.005$), compared with 0.39 ($p < 0.001$) under crude normalisation. Annual values were 0.08 ($p = 0.017$), 0.11 ($p = 0.005$) and 0.12 ($p = 0.003$) for 2023, 2024 and 2025. The BYM2 model yielded a latent standard deviation σ of 0.47 (95% credible interval 0.42–0.55) and a spatial mixing parameter ρ of 0.56 (95% credible interval 0.21–0.86), compared with 0.80 (0.53–0.99) under crude normalisation; the largest R-hat was 1.01, the smallest bulk effective sample size was 297, and there were no divergent transitions. Posterior median ratios ranged from 0.04 to 2.77. Smoothing had little influence on point

estimates: the median absolute relative difference between O/E ratios and posterior medians was 0.13%, and the median widths of the respective 95% intervals were 0.13 and 0.13. Fifty-seven of the 246 areas, with a combined 2025 population of approximately 1.04 million, had a posterior probability of at least 0.80 of a standardised ratio above 1.25. Another 64 areas (approximately 2.13 million residents) met the corresponding criterion for ratios below 0.80 (Figure 3b). High-use areas were most frequent in Misiones, San Pedro, Caaguazú and Guairá and included several departmental capitals with a regional hospital, such as Paraguarí (2.77), San Juan Bautista (1.95), Villarrica (1.89) and Caazapá (1.77), as well as Itauguá (1.61), the seat of a national referral hospital. Low-use areas predominated in Alto Paraná, Itapúa and the metropolitan districts of Central department; the lowest ratios were observed in three neighbouring Chaco areas, Loma Plata (0.04), Teniente Irala Fernández (0.10) and Puerto Pinasco (0.12), a region where cooperative health services of the Mennonite colonies operate, and in Ciudad del Este (0.47). Under crude normalisation, 80 areas had been classified as high use and 68 as low use, of which 51 and 52, respectively, retained the same classification after standardisation.

After false discovery rate control, the local Moran statistic identified a single low–low area, Mariscal Estigarribia in Boquerón; 36 areas had local p-values below 0.05 before adjustment. Under crude normalisation, 11 high–high and 11 low–low areas had remained significant after adjustment, mainly in Misiones and Ñeembucú and in the Chaco and eastern border departments, respectively.

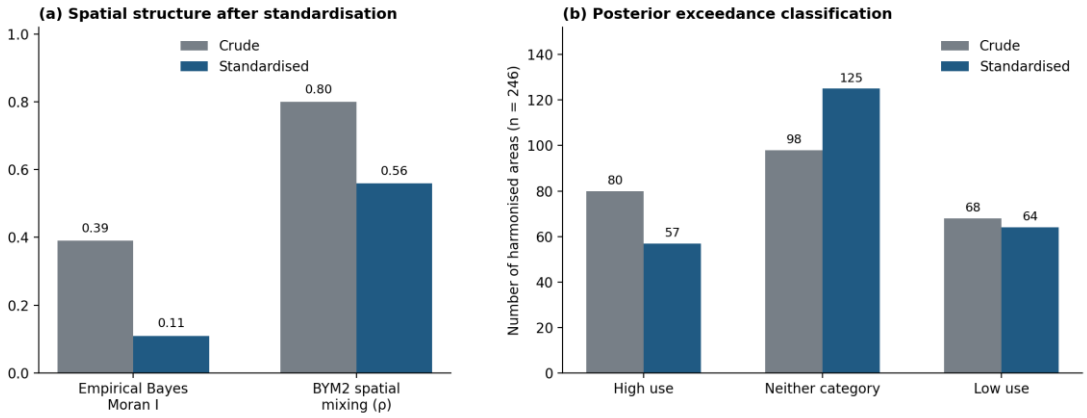


Figure 3. Comparison of spatial structure and Bayesian posterior classifications before and after indirect sex and age standardisation. (a) Global empirical Bayes Moran's I and the BYM2 spatial mixing parameter (ρ) for crude and standardised analyses. These represent different statistics and are not directly interchangeable. (b) Numbers of harmonised areas classified as having substantially high or low relative use using posterior exceedance probabilities ≥0.80; the remaining areas meet neither criterion. Results are derived from the reported spatial analyses.

Spatial scan clusters

All candidate clusters reached the minimum attainable Monte Carlo p-value of 0.001, so statistical significance alone did not discriminate between geographic contrasts. With the 10% population limit, the strongest high-use cluster encompassed 36 areas in the southern Eastern Region (relative ratio inside versus outside, 1.55), and the most marked low-use clusters were the three neighbouring Chaco areas (0.07) and Ciudad del Este (0.46) (Table 2, Figure 4). Expanding the maximum window to 25% left the high-use clusters unchanged and added larger low-use clusters with modest contrasts, comprising 11 metropolitan areas of Central department (relative ratio 0.83) and 20 areas along the eastern border (relative ratio 0.84).

Table 2. Non-overlapping spatial scan clusters of sex- and age-standardised recorded public diabetes care, 2023–2025

Cluster	Areas	Population 2025	Records	O/E	Relative ratio	Location
H1	36	403463	58451	1.49	1.55	Southern Eastern Region
H2	11	319841	35545	1.34	1.36	Cordillera, Central and Paraguarí

Cluster	Areas	Population 2025	Records	O/E	Relative ratio	Location
H3	3	26183	4267	1.93	1.93	Ayolas, Santiago and Yabebyry (Misiones)
H4	8	164827	15657	1.35	1.36	San Pedro and Concepción
H5	1	75666	8886	1.46	1.47	Villa Elisa (Central)
H6	7	50044	7584	1.43	1.44	Ñeembucú
H7	1	106470	11348	1.31	1.31	Coronel Oviedo (Caaguazú)
H8	5	76947	8605	1.32	1.33	Itapúa
L1	1	336105	10987	0.47	0.46	Ciudad del Este (Alto Paraná)
L2	3	61604	302	0.07	0.07	Loma Plata, Teniente Irala Fernández and Puerto Pinasco (Chaco)
L3	1	112615	6709	0.62	0.62	Fernando de la Mora (Central)
L4	2	111179	5597	0.62	0.61	Caaguazú and R.I. 3 Corrales (Caaguazú)
L5	1	129927	8700	0.72	0.71	Lambaré (Central)
L6	2	58086	1878	0.55	0.55	Filadelfia and Mariscal Estigarribia (Boquerón)

H and L identify elevated and reduced relative use; clusters with a relative ratio between 0.75 and 1.25 are not shown. Relative ratio: recorded use inside versus outside the cluster. All Monte Carlo p-values were 0.001 (999 replicates). Population: 2025 resident estimate. Records: pooled over 2023–2025. Maximum window: 10% of the population.

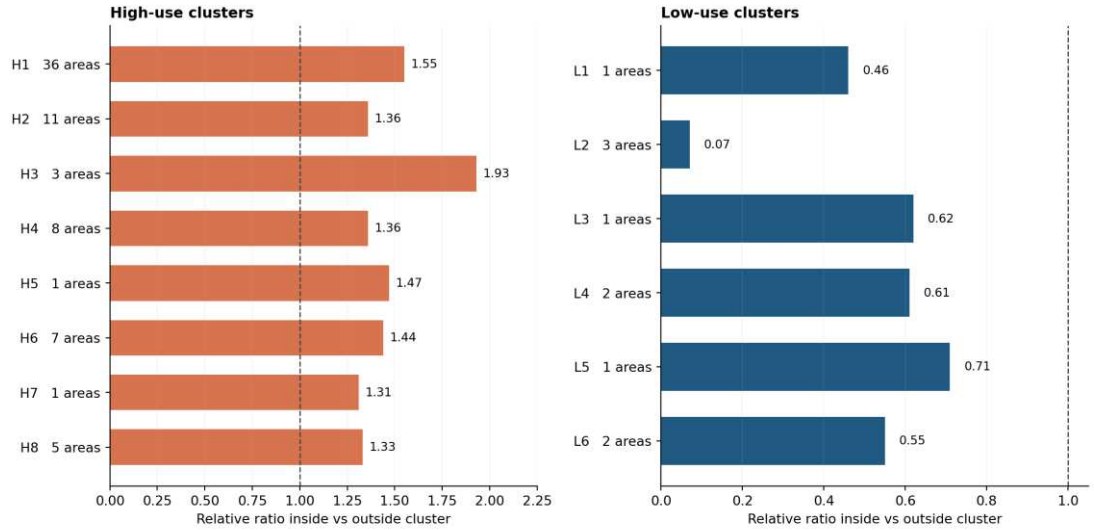


Figure 4. Relative-use contrasts for the non-overlapping high- and low-use clusters detected with the Poisson spatial scan statistic using sex- and age-standardised expected counts. Cluster labels H1–H8 and L1–L6 correspond to Table 2. Each bar is the relative ratio of recorded use inside versus outside the respective cluster. Cluster geography and constituent areas are detailed in Table 2. The minimum attainable Monte Carlo p-value was 0.001 with 999 replicates.

Temporal persistence and hospital sensitivity

Spearman correlations between annual BYM2 posterior medians were 0.85 (2023 versus 2024), 0.88 (2024 versus 2025) and 0.77 (2023 versus 2025), indicating persistent geographic rankings (Figure 5); in the annual models, the largest R-hat was 1.03 and there were no divergent transitions. Sixty-seven areas had a posterior probability of at least 0.95 of a ratio above 1 in each year, and 72 areas a probability of at most 0.05 in each year. In the secondary model, 26 areas contained at least one regional, general or specialised hospital. Recorded use was higher in these areas (relative ratio 1.42; 95% credible interval 1.21–1.64), but adjusted and unadjusted posterior medians correlated at 0.96. The spatial mixing parameter of the adjusted model was 0.75 (95% credible interval 0.37–0.99). The adjusted model classified 48 areas as substantially elevated and 69 as substantially reduced, of which 48 and 64, respectively, had been classified in the same way in the main model.

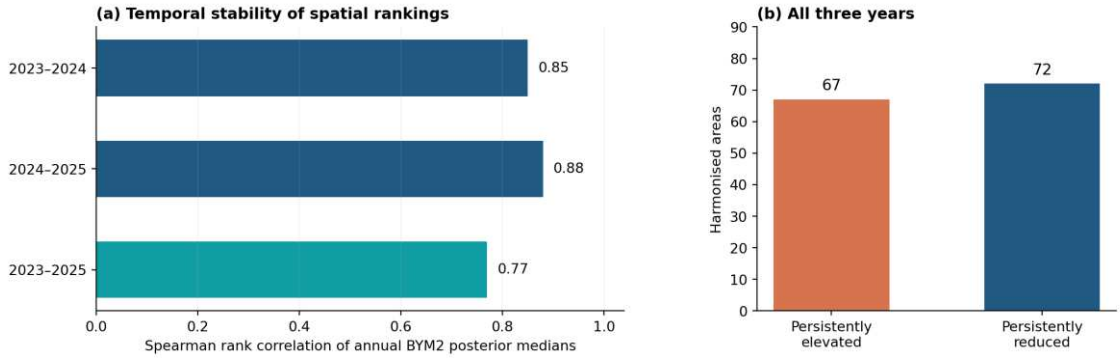


Figure 5. Persistence of spatial rankings over 2023–2025. (a) Spearman correlations between year-specific BYM2 posterior median relative-use ratios. (b) Number of harmonised areas consistently classified across all three years as having posterior probability ≥0.95 of a ratio above 1 or ≤0.05 of a ratio above 1. Results refer to sex- and age-standardised expected counts; 2025 records are provisional.

Discussion

This nationwide study shows that administrative records of public diabetes care in Paraguay are geographically heterogeneous and that the heterogeneity persists from year to year. A large part of the regional structure visible in crude rates, however, reflected the sex and age

composition of district populations: after indirect standardisation, global autocorrelation fell sharply, the spatially structured share of the BYM2 random effects decreased and local clusters largely disappeared, whereas the overall magnitude of between-area variation remained substantial. Each spatial procedure addressed a different question: global and local Moran statistics described adjacency patterns, BYM2 provided probability-based classifications, and the scan statistic identified contiguous sets of areas with relatively high or low recorded use.

The influence of standardisation is a central methodological finding. Recorded use rose steeply with age and was markedly higher in women, and the southern and central departments of the Eastern Region have older populations than the Chaco and the eastern border. Crude normalisation therefore attributed to geography differences that were largely demographic, and spatial statistics computed on crude ratios mainly recovered the geography of population ageing. Standardising for sex and age before spatial analysis is routine in disease mapping, but service counts are frequently mapped as crude rates; our results show that the choice can alter the substantive conclusions about clustering, not only the magnitude of the estimates.

A second interpretive finding is that the geography of service use must not be conflated with the geography of disease or of residence-based access. An administrative event allocated to the district of consultation can arise from any locality served by the facility. After standardisation, the dominant pattern was local rather than regional: high ratios in departmental capitals and in the seat of a national referral hospital, often adjacent to areas with low ratios. This hub-and-hinterland configuration is consistent with care recorded where it is provided, and it may itself reduce spatial autocorrelation, because the contrasts occur between neighbours. The positive association with hospital presence supports this

interpretation, although adjustment for it changed the overall pattern little; catchment boundaries and unmet need cannot be established from these data.

Geographic differences in provider mix further complicate comparisons. In areas served by other public subsystems, social security, private providers or cooperatives, as in the Mennonite colonies of the Central Chaco, low MSPBS counts need not imply poor access to diabetes care. Conversely, a high public-sector ratio may reflect concentration of services, programme organisation or administrative completeness rather than a greater prevalence of diabetes. The higher recorded rates in women are consistent with known sex differences in the use of primary and outpatient care and should not be read as a higher prevalence. Studies in Brazil and Mexico have likewise reported marked municipal heterogeneity and spatial clustering of diabetes hospitalisations and mortality [10,11], although those outcomes are closer to disease burden than outpatient activity. Linking provider systems and recording the district of residence would be necessary to distinguish these mechanisms.

The study also illustrates why smoothing, clustering and the treatment of administrative geography should be reported separately. Bayesian shrinkage changed relative-use estimates little in most areas because counts were large. The value of BYM2 lay mainly in expressing spatial dependence and uncertainty through posterior probabilities. The parameter ρ should be interpreted as a feature of the random-effect decomposition, not as the proportion of between-area differences caused or explained by location; its decrease after standardisation indicates that part of what had been allocated to the structured component was demographic.

Smoothing would matter more for rarer outcomes, such as complications or diabetes-related mortality, which could be analysed with the same workflow.

The scan analysis requires particular restraint. Nearly universal nominal significance is expected when large event totals coincide with substantial geographic heterogeneity and a Poisson reference distribution; the p-values neither establish practical importance nor

separate provider-driven clustering from differences in the underlying population. Relative contrasts, variation across window sizes and agreement with independently characterised local patterns are more informative than significance alone.

Boundary aggregation is a further substantive consideration. Harmonising post-2012 districts to 246 historical polygons places counts, denominators and census populations on a single geographic support, but it obscures differences within parent areas, alters the neighbourhood graph and changes the meaning of comparisons with contemporary administrative units. Because the limits of many districts are not legally defined [18], any district layer for Paraguay involves conventions, and matched-boundary analyses with the 2022 census cartography would be a useful extension. The maps should therefore be read as a coherent areal representation of recorded activity rather than a definitive depiction of contemporary district-by-district utilisation.

For health planning, the principal use of these maps is to identify geographic patterns of recorded MSPBS activity that warrant explanation, not to rank districts by diabetes risk or adequacy of care.

The methodological lessons apply wherever service activity is mapped from routine aggregate reporting. Health information platforms such as the District Health Information System (DHIS2), used in many low- and middle-income countries, store routine data largely as aggregate counts by facility and administrative unit [27]. In such settings, the workflow offers three transferable elements: harmonisation of changing administrative units to an openly licensed geometry using the coordinates of primary care facilities; indirect standardisation of place-of-care counts by combining small-area population projections with the census sex and age structure when stratified denominators are not published; and a side-by-side comparison of crude and standardised spatial statistics that makes the influence of population composition explicit before clustering is interpreted. Because the code relies only

on open-source software and published tables, it can be adapted to other countries, outcomes and boundary systems by changing the input files. In any such setting, the demographic composition and the data-generating geography must be made explicit before complex spatial inference is applied.

Strengths and limitations

Strengths include national coverage for three consecutive years, explicit aggregation to a constant spatial support, sex and age standardisation with census data for every district, independent verification of the annual source counts, a reproducible workflow and triangulation of three established spatial procedures. Limitations include the ecological design, restriction to the MSPBS network, possible duplicate recording within years, the absence of residence-based patient origin and unmeasured flows across district boundaries. Standardisation used five broad age groups and the 2022 census structure for all three years, so differences in age within groups and changes in composition after the census are not captured; moreover, it adjusts for the composition of the resident population, whereas the counts refer to place of care. Remaining variation therefore combines provider allocation, utilisation, care flows and data quality. The 2025 data were provisional, and the source contained one fewer labelled district than the official frame. The 246-area geography reduces resolution and remains susceptible to the modifiable areal unit problem, and the assignment of some post-2012 districts relied on a single Family Health Unit. Finally, the BYM2 random effects absorb extra-Poisson variation, whereas the Poisson reference distribution of the scan statistic does not, which explains its limited discrimination; overdispersion-robust scan methods could be explored in future work. These features limit causal, coverage and equity interpretations.

Conclusions

Place-of-care records reveal substantial and persistent geographic heterogeneity in MSPBS diabetes service activity across Paraguay, but much of the regional structure of crude rates reflects the sex and age composition of the population. After standardisation, the remaining pattern is dominated by local contrasts around departmental capitals and referral facilities. Bayesian modelling adds uncertainty-based classification, whereas spatial association and scan procedures describe different aspects of geographic concentration; with large administrative counts, statistical significance and smoothing add relatively little. Where boundaries change, providers are mixed and the place of care differs from the place of residence, spatial inference is most useful when demographic standardisation precedes it, when these measurement constraints are stated explicitly and when no conclusions are drawn about disease prevalence or unmet need. The open workflow can be transferred to other countries with changing districts and facility-based reporting.

List of abbreviations

BYM2: reparameterised Besag–York–Mollié model; DHIS2: District Health Information System; DIGIES: Directorate-General for Strategic Health Information; ICD-10: International Classification of Diseases, 10th Revision; INE: National Institute of Statistics; MSPBS: Ministry of Public Health and Social Welfare; O/E: observed to expected; RECORD: REporting of studies Conducted using Observational Routinely-collected health Data; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; USF: Family Health Unit.

Declarations

Ethics approval and consent to participate

This study analysed publicly available, aggregated administrative statistics, without individual-level data, contact with participants or intervention. Ethics committee approval and informed consent were therefore not required.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files. The audited district-year counts, the source audit program and the audit summary are provided in Additional files 2–4. Additional file 6 contains the original DIGIES workbook [14], the INE district population estimates [12], the 2022 census tables of population by sex and age group [15], the national list of Family Health Units [17], the MSPBS facility register [24], the geoBoundaries district layer [16], the corrected boundary layer, the district crosswalk, the derived datasets and the analysis code.

Project name: Geography of recorded public diabetes care in Paraguay. Project home page and archived version: Additional file 6. Operating system: platform independent. Programming language: Python 3. Other requirements: pandas, GeoPandas, libpysal, esda, PyMC, ArviZ, SciPy, matplotlib, mapclassify and openpyxl (versions listed in the requirements file). Licence: MIT. Any restrictions to use by non-academics: none.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no external funding.

Authors' contributions

All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes prevalence and treatment from 1990 to 2022: a pooled analysis of 1108 population-representative studies with 141 million participants. Lancet. 2024;404:2077–93. https://doi.org/10.1016/S0140-6736(24)02317-1.

2. Ministerio de Salud Pública y Bienestar Social. Segunda encuesta nacional muestra elevada prevalencia de factores de riesgo. 2023. https://www.mspbs.gov.py/portal/27507/segunda-encuesta-nacional-muestra-elevada-prevalencia-de-factores-de-riesgo.html. Accessed 26 Oct 2026.

3. Gao F, Languille C, Karzazi K, Guhl M, Boukebous B, Deguen S. Efficiency of fine scale and spatial regression in modelling associations between healthcare service spatial accessibility and their utilization. Int J Health Geogr. 2021;20:22. https://doi.org/10.1186/s12942-021-00276-y.

4. Stacherl B, Sauzet O. Gravity models for potential spatial healthcare access measurement: a systematic methodological review. Int J Health Geogr. 2023;22:34. https://doi.org/10.1186/s12942-023-00358-z.

5. Apparicio P, Gelb J, Dubé AS, Kingham S, Gauvin L, Robitaille É. The approaches to measuring the potential spatial access to urban health services revisited: distance types and aggregation-error issues. Int J Health Geogr. 2017;16:32. https://doi.org/10.1186/s12942-017-0105-9.

6. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math. 1991;43:1–20. https://doi.org/10.1007/BF00116466.

7. Riebler A, Sørbye SH, Simpson D, Rue H. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. Stat Methods Med Res. 2016;25:1145–65. https://doi.org/10.1177/0962280216660421.

8. Anselin L. Local indicators of spatial association—LISA. Geogr Anal. 1995;27:93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.

9. Kulldorff M. A spatial scan statistic. Commun Stat Theory Methods. 1997;26:1481–96. https://doi.org/10.1080/03610929708831995.

10. Mendes Júnior AA, de Sousa ÁFL, Santos GRS, Lima SVMA, dos Santos AD, Almeida VS, et al. Spatial and temporal analysis of hospitalizations due to primary care–sensitive conditions related to diabetes mellitus in a state in the northeast of Brazil. Int J Environ Res Public Health. 2024;21:1538. https://doi.org/10.3390/ijerph21111538.

11. Dávila Cervantes CA, Baptista EA. Mortality from type 2 diabetes mellitus across municipalities in Mexico. Arch Public Health. 2024;82:196. https://doi.org/10.1186/s13690-024-01432-z.

12. Instituto Nacional de Estadística. Estimaciones y proyecciones de la población distrital, 2000–2035. Revisión 2025. Fernando de la Mora: INE; 2025. https://www.ine.gov.py/resumen/284/estimaciones-y-proyecciones-de-la-poblacion-distrital-2000-2035-revision-2025. Accessed 26 Oct 2026.

13. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12:e1001885. https://doi.org/10.1371/journal.pmed.1001885.

14. Ministerio de Salud Pública y Bienestar Social (MSPBS), Dirección General de Información Estratégica en Salud (DIGIES). Número de personas atendidas en consultorio a causa de diabetes mellitus, por año, tipo, grupo de edad, sexo, región

sanitaria y distrito de ocurrencia, 2023–2026 [administrative dataset; XLSX]. Asunción: MSPBS/DIGIES; 2026. Source file: 1790267154_1_PERSONAS.ATENDIDAS.CONSUL.DIABETES.REG.OCURR.DISTRI.SEXO.TIPO.2023-2026P.xlsx. Institutional data portal: https://digies.mspbs.gov.py/datos-abiertos/. Accessed 29 Oct 2026.

15. Instituto Nacional de Estadística. Censo Nacional de Población y Viviendas 2022. Estructura de la población por grupos de edad y sexo, según departamento y distrito [dataset]. Fernando de la Mora: INE; 2024. https://www.ine.gov.py/publicacion/2/poblacion. Accessed 26 Oct 2026.

16. Runfola D, Anderson A, Baier H, Crittenden M, Dowker E, Fuhrig S, et al. geoBoundaries: a global database of political administrative boundaries. PLoS One. 2020;15:e0231866. https://doi.org/10.1371/journal.pone.0231866.

17. Ministerio de Salud Pública y Bienestar Social. Listado nacional de las Unidades de Salud de la Familia (USF) [dataset]. Asunción: MSPBS; 2025.

18. Instituto Nacional de Estadística. Cartografía para el Censo Nacional de Población y Viviendas 2022 de Paraguay [presentation]. Santiago: Comisión Económica para América Latina y el Caribe; 2023. https://cepal.org/sites/default/files/presentations/cartografia-para-cnpv2022-paraguay-ine-oct2023.pdf. Accessed 26 Oct 2026.

19. Assunção RM, Reis EA. A new proposal to adjust Moran's I for population density. Stat Med. 1999;18:2147–62.

20. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

21. Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH. Penalising model component complexity: a principled, practical approach to constructing priors. Stat Sci. 2017;32:1–28. https://doi.org/10.1214/16-STS576.

22. Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner PC. Rank-normalization, folding, and localization: an improved R-hat for assessing convergence of MCMC. Bayesian Anal. 2021;16:667–718. https://doi.org/10.1214/20-BA1221.

23. Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. Environ Health Perspect. 2004;112:1016–25. https://doi.org/10.1289/ehp.6740.

24. Ministerio de Salud Pública y Bienestar Social. Establecimientos de salud dependientes del MSPyBS, año 2025 [dataset]. Asunción: MSPBS; 2025.

25. Rey SJ, Anselin L. PySAL: a Python library of spatial analytical methods. Rev Reg Stud. 2007;37:5–27.

26. Abril-Pla O, Andreani V, Carroll C, Dong L, Fonnesbeck CJ, Kochurov M, et al. PyMC: a modern, and comprehensive probabilistic programming framework in Python. PeerJ Comput Sci. 2023;9:e1516. https://doi.org/10.7717/peerj-cs.1516.

27. Dehnavieh R, Haghdoost A, Khosravi A, Hoseinabadi F, Rahimi H, Poursheikhali A, et al. The District Health Information System (DHIS2): a literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. Health Inf Manag. 2019;48:62–75. https://doi.org/10.1177/1833358318777713.

Additional files

Additional file 1. Format: .docx. Title: RECORD checklist. Description: items of the STROBE statement and RECORD extension, with the section of the manuscript in which each item is reported.

Additional file 2. Format: .csv. Title: District-year counts analysed. Description: audited counts of people recorded as attended for diabetes by health region, reconciled district label and year, 2023–2025.

Additional file 3. Format: .py. Title: Source audit program. Description: standard-library Python 3 program that extracts the counts from the original DIGIES workbook, reconciles district labels and checks annual totals.

Additional file 4. Format: .txt. Title: Source audit summary. Description: provenance of the DIGIES data, annual totals, treatment of records without a district code and reconciled district labels.

Additional file 5. Format: .docx. Title: Districts created after 2012 and boundary corrections. Description: assignment of the 17 districts created after 2012 to their parent historical areas based on the locations of Family Health Units, and corrections to the 2012 boundary layer.

Additional file 6. Format: .zip. Title: Code and data. Description: original source files, including the 2022 census tables, corrected boundary layer, district crosswalk, derived datasets, district-level results of the main and crude-normalised analyses, and Python code for harmonisation, standardisation, spatial statistics, Bayesian models and figures.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- table2.jpg
- table1.jpg
- Figure5YearspecificBYM2relativeuseratios20232025.jpg
- Figure3BYM2relativeuseratiosandsubstantialdepartures.jpg
- Table2Spatialscanclusters.jpg
- Figure4Highandlowuseclustersdetectedbyspatialscan.jpg
- Figure2Effectofstandardisationandstandardisedrelativeuse.jpg
- exencionIJHG.pdf
- Table1Departmentalrecordedpublicdiabetescare.jpg
- Figure1Districtpopulationandcruderate.jpg