Civil Legal Support Index

Area-level Civil Legal Support Index for Scotland

Longitudinal Panel Data 2017–2024

Methodology: construction of the Civil Legal Support Index

The objective of this study is to construct a multi-dimensional panel dataset measuring social need and legal aid supply across the 32 Scottish Local Authorities between 2013 and 2024. The dataset integrates crime, housing, health, labour market, deprivation, demographic and barrier indicators into a unified LA_Name × Year structure suitable for econometric analysis.

The primary challenges in the data architecture included:

- aggregating datasets reported at the granular Data Zone level to the Local Authority level where policy operates
- integrating time-varying metrics with static 2011 Census measures
- standardising raw source files that varied drastically in structure and formatting, requiring programmatic extraction, regex parsing, and mathematical harmonisation.

How the Index was built

The CLSI is a composite measure that captures the unmet need for civil legal aid and advice services across Scotland's Data Zones. It integrates three theoretically distinct components - Need, Supply Deficit, and Barriers - each built from multiple indicators that are individually harmonised before being combined.

CLSI = SCORE_Need + SCORE_SupplyDeficit + SCORE_Barriers

All three component scores are on the same z-score scale. Higher CLSI = greater unmet need. Components are summed, not averaged, to preserve full variation across all three dimensions.

1.1 Need Score (SCORE_Need)

The Need score captures the socioeconomic conditions that generate demand for legal aid. It is structured into four sub-domains. All indicators point in the same direction: higher values indicate greater need.

1.1.1 What it measures

- Economic deprivation — unemployment, benefit claimancy, economic inactivity and income deprivation

- Health and disability — activity limitation, chronic illness, and mental health indicators
- Housing precarity — lack of central heating, overcrowding and social/private renting
- Social-legal vulnerability — crime exposure and asylum seeker status.

1.1.2 How it is calculated

- Count variables (e.g. JSA claimants, people without central heating) are converted to rates per 1,000 population using mid-year estimates (PD1.2024).
- Each rate and each non-count indicator is standardised into a z-score using the global mean and SD across all Data Zones and all years 2017–2024.
- Z-scores within each sub-domain are averaged with equal weight to produce a sub-domain score.
- The four sub-domain scores are averaged with equal weight to produce SCORE_Need.

Sub-Component	Variables Used	Computation
Economic	JSA claimant rate; unemployment rate (UR); economically inactive (EA12, EA24, EA28); no qualifications (EQ2); SIMD income & employment rates	Rate per 1,000 → z-score (global) → equal-weight average
Health & Disability	Activities limited a lot (DD2); SIMD: illness factor, alcohol, drugs, mortality, depression, emergency stays	Rate per 1,000 where count; z-score (global) → equal-weight average
Housing	No central heating (CH2); social renting (HT7); private renting (HT8); SIMD overcrowding & no-central-heat rates	Rate per 1,000 where count; z-score (global) → equal-weight average
Social-Legal	Crime count (CrimeCount); asylum seeker accommodation (AS1–AS7)	Rate per 1,000 → z-score (global) → equal-weight average

1.2 Supply Deficit Score (SCORE_SupplyDeficit)

The Supply Deficit score reflects the shortage of legal aid solicitors and advice providers relative to the population in each Data Zone. Because higher solicitor supply represents a better situation, this component requires a directional reversal so that higher scores always mean worse conditions — consistent with the rest of the index.

1.2.1 What it measures

- Annual solicitor counts (SC.2017–SC.2024), converted to a rate per 10,000 population
- Registered legal charities, advice providers accredited under the Scottish National Standards for Information and Advice Providers, and law centres (per 10,000 residents).

1.2.2 How it is calculated

- Solicitor counts are converted to a rate per 10,000 population (PD1.2024 as denominator). Multiplier is 10,000, distinct from the 1,000 multiplier used for Need variables.

- The rate is standardised into a z-score using the global mean and SD across all Data Zones and all years 2017–2024.
- The standardised value is multiplied by −1. Areas with fewer solicitors receive higher positive scores (greater supply deficit); areas with more solicitors receive lower or negative scores.
- If multiple supply variables are included, each is standardised and reversed individually. SCORE_SupplyDeficit is then their equal-weight average.

SCORE_SupplyDeficit = mean(−1 × z(solicitors per 10,000) [, −1 × z(other supply vars)])

Each supply variable standardised and reversed independently, then averaged.

Why multiply by −1?

Without reversal, a zone with many solicitors would have a high positive z-score which, when added to Need and Barriers, would incorrectly inflate the CLSI. Reversing the sign ensures every component is internally consistent: higher always signals worse conditions. The component is then added (not subtracted) in the formula.

1.3 Barriers Score (SCORE_Barriers)

The Barriers score captures structural impediments to accessing legal services. These barriers operate independently of both need and supply: even where need is high and solicitors exist, barriers can prevent people from reaching services.

1.3.1 What it measures

- Urban-rural remoteness — the Scottish Government's 8-category UR classification (UR1 = Large Urban through UR8 = Very Remote Rural)
- Geographic access — average drive times and public transport travel times to key services, drawn from SIMD 2020
- Digital exclusion — percentage of premises without superfast broadband (≥ 30 Mb/s), drawn from SIMD 2020.

1.3.2 How it is harmonised

- Remoteness: UR1–UR8 binary indicators are combined into a weighted composite (UR category × weight 1–8), summed, then globally z-scored. Higher = more remote.
- Geographic access: each drive-time and public-transport-time variable is globally z-scored. The resulting z-scores are averaged into a single sub-score. Higher = longer travel times.
- Digital exclusion: the broadband variable is globally z-scored directly. Higher = more premises without adequate broadband.
- The three sub-scores are averaged with equal weight to produce SCORE_Barriers.

Sub-Component	Variables Used	Computation
Remoteness	UR1–UR8 (2022 classification)	Weighted composite (category × weight 1–8) → z-score (global)
Geographic Access	SIMD2020 drive times (GP, PO, retail, schools); public transport times (GP, PO, retail)	z-score (global) per variable → equal-weight average
Digital Exclusion	SIMD2020 broadband (% without ≥ 30 Mb/s)	z-score (global)

1.4 Global standardisation and the panel

All z-scores are calculated globally — using the mean and standard deviation pooled across all Data Zones and all years 2017–2024 — rather than within each year separately. This makes temporal comparisons meaningful: CLSI = +1.5 in 2024 has the same absolute level of unmet need as CLSI = +1.5 in 2017.

The components are summed. Since each component is already a z-score-scaled average, the sum retains full variation across all three dimensions and remains comparable across years.

1.5 Temporal coverage and the Census proxy problem

Because most Need variables derive from Census 2022 and SIMD 2020, they are held constant across all years. Annual variation in the CLSI between 2017 and 2024 is therefore driven primarily by the time-varying components: solicitor supply (SC.2017–2024), JSA claimant flows, unemployment rates, and crime counts.

This is methodologically defensible: structural determinants of legal need (housing quality, educational attainment, chronic health) exhibit high persistence within zones over short-to-medium timeframes. Supply provision and economic conditions, by contrast, can shift rapidly year to year.

Index Normalization (0–100 Scale):
To facilitate interpretation, the raw results undergo Min-Max Normalization, computed explicitly for each year:

$$CLSI_{0-100} = \left(\frac{CLSI_{Raw} - \min(CLSI_{Raw})}{\max(CLSI_{Raw}) - \min(CLSI_{Raw})}\right) \times 100$$

This standardises the panel so that, in any given year, the least vulnerable council scores 0 and the most vulnerable council scores 100.

Mismatch Index

for Scotland

Longitudinal Panel Data 2017–2024

Methodology: construction of the Legal Aid Mismatch Index

2.1. Overview

Model: Gravity Model of Legal Aid Accessibility

We propose a spatial interaction model based on the gravity model framework, commonly used in transportation and service accessibility studies. The model estimates the expected flow of Civil Legal Aid users between demand locations (e.g. population centres) and supply locations (e.g. legal aid offices), based on size and distance.

Specification:

Let F_{ij} denote the intensity of access (or potential flow/flow of access) from a residential location i to a legal aid service location j (access score or accessibility measure based on proximity and supply; or like referrals or service uptake). To estimate F_ij we can use SLAB postcode-level usage data to estimate observed access from i to j (i.e., how many people from location i used services located at j):

$$F_{ij} = G \frac{P_i^{\alpha} S_j^{\beta}}{D_{ij}^{\gamma}}$$

Where: - P_i : population or demand proxy in origin *i* - S_j : supply proxy (e.g. number of legal aid providers, solicitor offices, hours available) in location *j* - D_{ij} : road or travel time distance between *i* and *j* - α, β, γ : parameters to be estimated - G : constant of proportionality (absorbed in log-linear form)

Thus, this study constructs an origin-level mismatch index to measure the extent to which observed legal aid application flows differ from the level of demand predicted by a structural gravity model. The purpose of the index is to identify places where realised use of Civil Legal Aid services is lower or higher than would be expected given local population size, the distribution of solicitor supply, travel distance to providers, and administrative geography. The analysis is undertaken at the level of Scottish Data Zones and distinguishes three legal subject groups: Adults with Incapacity, Family, and Other (covering all other topics).

2.2. Construction of origin–destination application flows

The analysis begins with case-level administrative data, where each record corresponds to a single Civil Legal Aid application. For each application, the applicant's 2011 Data Zone is treated as the origin and the branch's 2011 Data Zone is treated as the destination. Subject group and grant status are also retained.

Variable names were standardised across all source files so that applicant Data Zone, branch Data Zone, subject group, and grant outcome were recorded consistently. Records with unknown origin Data Zones, identified by the code S00000000, were excluded. Cases flagged as outwith Scotland were also removed, along with records with missing destination Data Zones.

The cleaned case-level dataset was then aggregated to origin–destination–group cells. For each origin i, destination j, subject group g, and year t, the total number of applications was calculated. This is the main flow variable used in the modelling stage and is denoted F_ijgt. Separate datasets were prepared for the different financial year files and then combined into a single panel covering 2017–2024.

2.3. Destination-side Supply

Supply was measured using administrative data on solicitor presence at the destination Data Zone level. For each destination j and year t, the preferred supply measure is the count of solicitors, denoted S_jt. These data were harmonised across files and collapsed to destination-year level before being merged into the modelling dataset. We incorporate in the supply measure also agencies offering legal support in the destination.

2.4. Population measures

Population was incorporated on both the origin and destination sides. Origin population, P_it, represents the population of origin Data Zone i in year t. Destination population, P_jt, represents the population of destination Data Zone j in year t. These variables were merged onto the relevant origin-year and destination-year records and used as scale controls in the gravity model.

2.5. Construction of the modelling dataset

Observed application flows represent only realised origin–destination matches. To estimate a gravity model, these observed flows were expanded into a broader set of potential origin–destination alternatives, including zero-flow observations.

Observed origin–destination–year combinations were first extracted from the flow data so that all empirically observed pairs would be retained. Origins were defined using the full Data Zone-by-year population base. Destinations were defined as the union of destination-year combinations appearing either in the supply data or in the observed flow data. Where solicitor supply was missing for an included destination-year, it was set to zero.

The origin and destination panels were then cross-joined by year to generate all feasible origin–destination combinations. Geographic coordinates were merged onto both origins and destinations using a Scottish Data Zone latitude–longitude lookup, and straight-line

distance between them was calculated in kilometres. This distance measure is denoted D_{ijt}. For same-zone observations or cases where the calculated distance was zero, a small positive adjustment was applied so that the variable could be logged.

To create a plausible destination choice set while keeping the dataset computationally manageable, destination options were restricted for each origin-year. All observed origin–destination pairs were retained. In addition, all destinations within 80 kilometres were included. Where fewer than ten destinations fell within this radius, the nearest destinations were added until at least ten candidate destinations were available. A sensitivity test was undertaken to investigate how relaxing and modifying these thresholds affect the results, which remain consistent.

The dataset was then expanded across the three legal subject groups: Adults With Incapacity, Family, and Other. Destination–group combinations that never recorded any applications in any year were removed. Observed application counts were then merged back onto the expanded panel, and all unobserved origin–destination–group–year cells were assigned a flow value of zero.

2.6. Local authority linkage and home-bias measure

Each origin and destination Data Zone was linked to its corresponding local authority using a Data Zone-to-local-authority lookup. A binary indicator, $sameLA_{ij}$, was created to capture whether the origin and destination belong to the same local authority. This variable is included in the model to account for possible within-authority preference or administrative affinity in destination choice.

2.7. Gravity model estimation

Expected application flows were estimated using a Poisson pseudo-maximum likelihood gravity model with high-dimensional fixed effects.

The model was estimated using ppmlhdfe, with standard errors clustered by origin and destination Data Zone. Fitted values are denoted $\hat{F}_{ijgt}$. Additional local socio-economic variables were included to test the results' consistency.

2.8. Aggregation of observed and expected flows

$$Expected_{igt} = \Sigma_j \hat{F}_{ijgt}$$
$$Observed_{igt} = \Sigma_j F_{ijgt}$$

$$Expected_{it} = \Sigma_g \Sigma_j \hat{F}_{ijgt}$$
$$Observed_{it} = \Sigma_g \Sigma_j F_{ijgt}$$

These totals summarise, for each origin, how much demand the model predicts should occur and how much demand is actually observed.

2.9. Mismatch Index

Group-specific mismatch:

$$M_{igt} = \ln[(Expected_{igt} + \varepsilon)/(Observed_{igt} + \varepsilon)]$$

where ε = 0.5.

Overall mismatch:

$M_{it} = \ln[(Expected_{it} + \varepsilon)/(Observed_{it} + \varepsilon)]$

Interpretation:
• M = 0: observed demand equals predicted demand.
• M > 0: predicted demand exceeds realised applications.
• M < 0: observed applications exceed predicted demand.

2.10. Standardised Mismatch scores

Group-specific z-score:

$Mz_{igt} = (M_{igt} - \text{mean}(M_{gt})) / \text{sd}(M_{gt})$

Overall z-score:

$Mz_{it} = (M_{it} - \text{mean}(M_{t})) / \text{sd}(M_{t})$

These standardised scores preserve the relative ordering of areas while expressing mismatch in standard deviation units.