Comparing NTG Region and Health District Population Estimates

Background

Population estimation at regional levels forms the foundation of demographic analysis, health planning, and policy evaluation. In the Northern Territory (NT), two geographic systems currently provide small-area population estimates, the NT Government (NTG) Region and the Health District (HD). Both systems derive from the same Australian Bureau of Statistics (ABS) Census and Estimated Resident Population (ERP) data and use iterative proportional fitting (IPF) to ensure that estimates align with official demographic totals. However, they differ substantially in their methodological design, spatial purpose, and analytical suitability. This paper guides the use of population data, examines the characteristics of NTG Region and Health District datasets, explains the IPF methodologies used to generate ERP estimates for each geography. Health Statistics and Informatics recommend that there are analytical advantages in adopting Health District data for demographic and health related analyses in the Northern Territory.

Rules for guiding use of population data in NT:

I. **Health District (HD) population estimates are the preferred and most valid dataset** for demographic and health research by NT Health.

- For demographic and epidemiological analyses, particularly those involving age standardisation, disease prevalence, fertility, mortality, and migration rate modelling, the method used to derive population estimates for HD offers a more stable and interpretable population than the NTG Region estimates.
- The method for deriving HD population estimates is preferred to the NT Regions because the IPF and proportional interpolation produce realistic annual population trajectories, reduce volatility in small populations, and ensure compatibility with health service geographies.
- These regions align with the ABS Indigenous Regions (IREGS). However, to support reporting aligned with services delivery in NT, we encourage labelling consistent with service regions (Appendix, Table-A1).
- In remote areas or small and uneven populations, HD estimates should be used for analysis in preference to NTG Region estimates, as HD based ABS ERP population estimates provide more reliable, stable, and demographically realistic population denominators through proportional interpolation (see Standard statement, page 9 for reference purposes).

II. **NTG Region estimates can be used for service and program planning** but should not be used for health statistical reporting. These estimates are not publicly available and will only be available on request.

- Potential uses of NTG Region estimates include providing population counts and percentage distributions for designing targeted services, programs or policies. The estimates align closely with ABS NT population totals but may show abrupt year to year changes due to linear interpolation.
- HD estimates, with their proportional interpolation and smoother annual trends, are more suitable for calculating prevalence rates and applying demographic or health models that require stable denominators.

III. **Revised and updated population estimates maintained by Health Statistics and Informatics (HSI)** will be available following the ABS December release of Estimated Resident Population.

IV. **Limiting NTG Region estimates at older ages to minimise bias and instability.**

- To reduce bias and instability arising from zero values and very small cell counts, age-group distributions by NTG Region should be presented up to the 65+ age group rather than extending to 85+ years.
- For analyses involving older age groups, HD estimates should be used as the primary population base, as proportional allocation provides a more reliable and coherent age structure.
- Where HD estimates are used in very sparsely populated areas with frequent single-digit cell counts, results should likewise be presented up to 65+ years to reduce noise and improve statistical robustness.

V. **Consistent labelling of small areas in the analysis produced by NT Health.** Population estimates and analytical outputs should use the recommended Health District based labelling aligned with contemporary service delivery boundaries (see Appendix, Table-A1). Where alternative geographic labels are used in source data, results should be re-labelled to ensure consistency, comparability, and policy relevance across NT Health analyses.

1. Differences Between NTG and Health District Data

Although both datasets originate from the same ABS Census and Estimated Resident Population (ERP) sources, the NTG Region and Health District series differ in their geographic basis, constraint structure, and purpose.

Feature	**NTG region data**	**Health district data**
Population coverage	ERP by region, indigenous status, sex, and age	ERP by district, Indigenous status, sex, and age
Update frequency	Annual updates tied to ABS ERP June series	Same update rule but smoother interpolation across census cycle
Census year inputs	Indigenous status × sex × age constraints	District × Indigenous status × sex × age constraints (age 65+ open)
Time series	1986-2025	1971-2025 (longer series)

2. Differences Between the Methods Used for NTG and Health District Estimates

Both systems use IPF followed by rounding to integer values, but the inputs, constraints, and interpolation methods vary.

> ***Constraint***
>
> *Is an external control total that each estimated population cell must match exactly after adjustment. These constraints are derived from reliable ABS data, such as totals by sex, age, region/district, and Indigenous status.*

(a) NTG Region Method

- ***IPF Constraints***

  I. Region x Sex x Age

  II. Region x Indigenous status

  III. Indigenous status x Sex x Age

  These three constraints ensure exact consistency with ABS ERP tables at multiple marginal levels.

  IPF iteratively scales cell estimates $E_{s,a}^{r,i}$ until all constraints are satisfied.

- ***Interpolation***

  Linear interpolation using a 1/4 fraction for the four intercensal years (e.g., 2017–2020).

$$\tilde{E}_{s,a}^{r,i}(t) = \lambda_{s,a}^{r,i}(t) E_{s,a}^{r,+}(t)$$

  where the weight applied to year t is given by the expression

$$\lambda_{s,a}^{r,i}(t) = \lambda_{s,a}^{r,i}(\underline{t}) + \frac{1}{4}\left(\lambda_{s,a}^{r,i}(\overline{t}) - \lambda_{s,a}^{r,i}(\underline{t})\right)$$

Only the *between-census* years are interpolated; census endpoints are fixed.

Example

Suppose for Big Rivers region, Males aged 25–29 years, Aboriginal population:

Year	$\lambda_{M,25-29}^{BigRivers,\text{Aboriginal}}$	$E_{M,25-29}^{BigRivers,+}$
2016 ($\underline{t}$)	0.40	2,500
2021 ($\overline{t}$)	0.50	3,000

$$E_{M,25-29}^{BigRivers,+}(2018) = 2500 + \frac{1}{4}(3000 - 2500) = 2500 + 125 = 2{,}625$$

Where $E_{M,25-29}^{BigRivers,+}(2018)$ is the total number of males aged 25–29 (all Indigenous groups combined) in the Big Rivers region in 2018.

To find the Aboriginal ERP for 2018:

$$\lambda(t) = 0.40 + \frac{1}{4}(0.50 - 0.40) = 0.425$$

$$\tilde{E}_{M,25-29}^{BigRivers,Aboriginal}(2018) = 0.425 \times 2{,}625 = 1{,}116$$

Thus, the Aboriginal male population aged 25–29 years in the Big Rivers region for 2018 is estimated at 1,116. The same process repeats for all intercensal years (2017–2020). In essence, linear interpolation ensures the population share evolve steadily between censuses.

(b) Health District Method

- ***IPF Constraints***

I. District × Indigenous status × Sex × Age
II. Sex × Age (for the whole NT)

The two-constraint system avoids over-fitting in small areas with sparse data, ensuring convergence and stability in estimates.

- ***Interpolation***

Continuous proportional interpolation using a 1/5 fraction across the full five-year census interval (2016–2021).

Every year, including census years, lies on the same smooth trajectory.

$$\tilde{E}_{s,a}^{d,i}(t) = \lambda_{s,a}^{d,i}(t) E_{s,a}^{+,+}(t)$$

where the weight $\lambda_{s,a}^{d,i}(t)$ is given by the expression

$$\lambda_{s,a}^{d,i}(t) = \lambda_{s,a}^{r,i}(\underline{t}) + \frac{1}{5}\left(\lambda_{s,a}^{d,i}(\overline{t}) - \lambda_{s,a}^{d,i}(\underline{t})\right)(t - \underline{t})$$

Here $\lambda_{s,a}^{d,i}(\underline{t})$ and $\lambda_{s,a}^{d,i}(\overline{t})$ are the weights

$$\lambda_{s,a}^{d,i}(\underline{t}) = \frac{E_{s,a}^{d,i}(\underline{t})}{E_{s,a}^{+,+}(\underline{t})}$$

$$\lambda_{s,a}^{d,i}(\overline{t}) = \frac{E_{s,a}^{d,i}(\overline{t})}{E_{s,a}^{+,+}(\overline{t})}$$

Symbol	Description
$\tilde{E}_{s,a}^{r,i}(t)$	Interpolated (estimated) ERP for sex (s), age (a), region (r), and Indigenous status (i) in year t
$E_{s,a}^{r,+}(t)$	ERP total for region × sex × age, summed over Indigenous groups
$\lambda_{s,a}^{r,i}(t)$	Proportion or share weight of the total ERP belonging to Indigenous group i for year t
$\underline{t}$	The earlier census year (e.g., 2016)
$\overline{t}$	The later census year (e.g., 2021)
$\frac{1}{4}$	Fractional step between censuses, since there are four intercensal years (2017–2020)
$\tilde{E}_{s,a}^{d,i}(t)$	Interpolated ERP for sex (s), age group (a), district (d), and Indigenous status (i) in year *t*.
$\frac{1}{5}$	Fraction for interpolation across the full 5-year census period.
$(t - \underline{t})$	Offset in years from the first census year (e.g., 2018 → offset = 2)

Example

Suppose we are estimating the Aboriginal female population aged 30–34 years in the Katherine district.

Year	$E_{F,30-34}^{+,+}(t)$ $(NT\ Total)$	$E_{F,30-34}^{Katherine,\text{Aboriginal}}(t)$	Calculated λ
2016 ($\underline{t}$)	10,000	1000	$\lambda(\underline{t})$=1000/10000=0.10
2021 ($\overline{t}$)	11,000	1430	$\lambda(\overline{t})$= 1430/11000=0.13

We want to find the interpolated value for 2018.

1. Compute λ for 2018:

$$\lambda(2018) = 0.10 + \frac{1}{5}(0.13 - 0.10)\ (2018 - 2016) = 0.10 + 0.012 = 0.112$$

2. Compute interpolated ERP:

$$E^{+,+}_{F,30-34}(2018) = 10000 + \frac{1}{5}(11000 - 10000)(2018 - 2016) = 10{,}400$$

$$\tilde{E}^{Katherine,Aboriginal}_{F,30-34}(2018) = 0.112 \times 10400 = 1{,}165$$

So, the Aboriginal female population aged 30-34 years in Katherine for 2018 is estimated at 1,165 persons.

3. Why These Methodological Differences Exist

Reason	**Explanation**
Constraint consistency	Regional-level constraints are applied for NTG because Indigenous-status data are provided at that scale, rather than at smaller geographies. HD omits this because small area Indigenous counts are too small or volatile. Omitting prevents instability from small or volatile population counts. This ensures numerical stability, temporal consistency, and analytical reliability for demographic and health analyses.
Interpolation scope	NTG Region interpolates discrete ERP counts for 4 inter-census years (hence 1/4). HD interpolate rates continuously for all 5 years (hence 1/5).
Analytical objective	NTG data should serve population accounting only; HD data has a preferred demographic analysis, and is best suited for analytical and statistical reporting, including direct or indirect standardisation of mortality rates.
Change assumption	Linear interpolation (for NTG Region) assumes steady change between two censuses; Proportional interpolation (for HD) assumes each district keeps a stable share of NT totals, scaled each year.

4. Advantages of Proportional Allocation in Health District Data

The use of proportional allocation in generating HD population estimates offers several analytical and demographic advantages. Unlike linear interpolation, which distributes change in fixed absolute amounts between census years, proportional allocation distributes population change relative to each group's share of the NT total. This approach is particularly suitable for Health District data, where age, sex, and Indigenous structures vary significantly across regions.

- **Smooth Annual Transitions for Health Indicators**

  Proportional allocation ensures that each year's population estimate follows a continuous and realistic trajectory between censuses. Because Health District estimates are used as denominators in calculating morbidity, mortality, and service utilisation rates, having a smooth progression over time prevents artificial spikes or dips in trend analyses.

  Example: A gradual 2–3% annual increase in the 45–64 age group yields consistent rates of diabetes or heart disease across years, making it easier to detect genuine epidemiological changes rather than statistical artefacts.

- **Stability in Small or Sparse Populations**

  Health Districts, particularly in remote areas, often have small or unevenly distributed populations. Proportional allocation stabilises these estimates by scaling each subgroup's change relative to the NT wide total rather than applying fixed numeric increments. This prevents unrealistic fluctuations in small subgroups, such as Indigenous males aged 65+, which are especially sensitive to rounding and sampling variability.

  Example: If an HD has only 200 Indigenous males aged 65+, proportional allocation might increase this group by 2–3 people per year instead of abrupt 10–20 person jumps seen with linear interpolation.

- **Maintains Demographic Structure Across Years**

  By adjusting each subgroup proportionally, the overall age-sex and Indigenous distribution within each HD remains demographically coherent. This means the shape of the population pyramid evolves naturally, preserving consistent age gradients and Indigenous shares across time.

  Example: If Aboriginal people represented 60% of the population in 2016 and 63% in 2021, proportional allocation ensures that this change occurs gradually across five years, maintaining realistic population balance each year.

- **Enhanced Accuracy for Age-Standardisation and Modelling**

  For health and demographic analyses such as age standardisation, life table construction, and disease prevalence modelling, HD data need to provide continuous and coherent denominators across age and time. Proportional allocation achieves this by ensuring that all years, including census years lie on the same smooth curve. This allows reliable comparison between districts and over time.

Example: When calculating standardised prevalence of hypertension by HD from 2016–2021, proportional allocation prevents rate distortion that could arise from stepwise, uneven denominators.

- **Alignment with Total NT Population**

  Annual HD estimate remains consistent with the NT wide ERP by age and sex. This internal balance ensures that the sum of all HD populations equals the NT total at every time point, supporting both local and territory level analyses without double counting or residual discrepancies.

5. Advantages of Health District Estimates for analysis

While both datasets are valuable, HD estimates have particular strengths for applied demography and public health analytics:

Demographic Objective	**Why Health District Estimates Perform Better**
Age-standardisation and rate calculation	Provide smooth, proportionally interpolated age-sex-Indigenous distributions across years, avoiding artificial jumps seen in NTG Region. This continuity yields more reliable age standardised morbidity or mortality rates.
Small-area health analysis	Correspond to real health service catchments. Analysts can compute district specific fertility, mortality, and disease prevalence directly aligned with operational boundaries.
Temporal trend modelling	The proportional interpolation ensures a linear annual trajectory in rates, making regression or time series models (e.g., ARIMA or panel models) statistically valid.
Cross-district comparison	Because all districts share identical sex-age totals (constraint $E_{s,a}^{(+,+)}(t)$), differences in rates reflect genuine demographic variation, not residual rounding or inconsistent constraints.
Stability in small populations	HD IPF distributes totals using proportional rather than absolute adjustments, avoiding negative or erratic changes in small Indigenous or older-age groups.
Integration with health indicators	Health district ERPs can be directly paired with NT Health administrative data (hospitalisations, chronic condition registries), producing coherent numerators and denominators for disease and service utilisation rates.

6. Use of Population Estimates in Remote Areas and Older Age Groups

Population estimation becomes particularly challenging in settings where population is small, unevenly distributed, or likely to suffer from small cell counts in older age groups and remote areas/regions of the NT. In these contexts, the choice of population estimation has important implications for the reliability and interpretation of demographic and health statistics.

i. Preference for Health District Estimates in Small or Uneven Populations

In areas where populations are very small or unevenly distributed, such as remote communities and older age groups, population estimates are particularly sensitive to the interpolation method used. In these settings, even small numerical changes can result in large apparent year to year fluctuations, which can distort rates, trends, and comparisons. Although NTG Region data are often presented for reporting purposes, the underlying population estimates used for analysis in these contexts should be drawn from HD estimates because they provide more stable and demographically realistic results.

HD estimates use proportional interpolation, which distributes population change gradually over time in proportion to the NT wide population structure. This approach reduces artificial jumps caused by small numbers, rounding effects, or fixed linear increments. The benefit of proportional interpolation is most evident in remote areas and older age groups, where cell counts are small and traditional linear approach can amplify volatility. If NTG Region estimates alone were used in these settings, artificial fluctuations may be introduced, making changes in health outcomes appear larger or smaller than they truly are. Using HD based estimates therefore improves the reliability, interpretability, and comparability of population-based indicators by ensuring that observed trends reflect real demographic change rather than statistical noise.

Standard statement

For analyses involving remote areas or small and uneven population groups in the Northern Territory (NT), Health District based on ABS ERP population estimates are used. This data has been transformed using proportional interpolation method producing the most stable and interpretable population denominators compared with other geographical small area population data for the NT. (Suggested citation: Chowdhury, P. & Wright, A. (2025). *Comparing NTG Region and Health District Population Estimates*. Health Statistics and Informatics, NT Health, Darwin.*)*

ii. Limiting NTG Region estimates at Older Ages and in Sparse Areas

NTG Region estimates are also more susceptible to zero values and very small cell counts in older age groups and remote areas. These issues can adversely affect statistical analyses and introduce bias, particularly in population projections, regression modelling, and rate calculations. In small populations, the combination of a complex constraint structure and linear interpolation can lead to abrupt changes, overestimation or underestimation in specific age groups, and increased instability when input data are sparse. Zero counts at very old ages, such as 85 years and over, are especially problematic because they can inflate rates, increase variance, and violate modelling assumptions used in demographic and epidemiological analysis.

In contrast, HD estimates use proportional allocation, which stabilises estimates for small subgroups, including Indigenous population aged 65 years and over that are particularly sensitive to rounding and sampling variability. While NTG Region estimates remain appropriate for presenting broad population counts and percentage distributions, their use at very old ages can result in misleading patterns driven by random variation rather than real demographic change. To minimise bias and instability, age group distributions by NTG Region at the regional level should therefore be presented up to the 65+ age group rather than extending to 85+ years. Therefore, in remote settings and older age groups, HD estimates provide the most reliable and coherent age structure and should be used as the primary population base for analysis.

Note: When calculating prevalence, proportions, or conducting statistical analysis using HD estimates in very sparsely populated areas, results should also be presented up to the 65+ age group rather than 85+ years. Although HD estimates are methodologically more robust and stable than NTG Region estimates, very small numbers (high frequency of single digit cell counts in very sensitive areas only) at extreme ages can still introduce noise and reduce interpretability. Aggregating older ages improves statistical robustness while preserving meaningful demographic patterns.

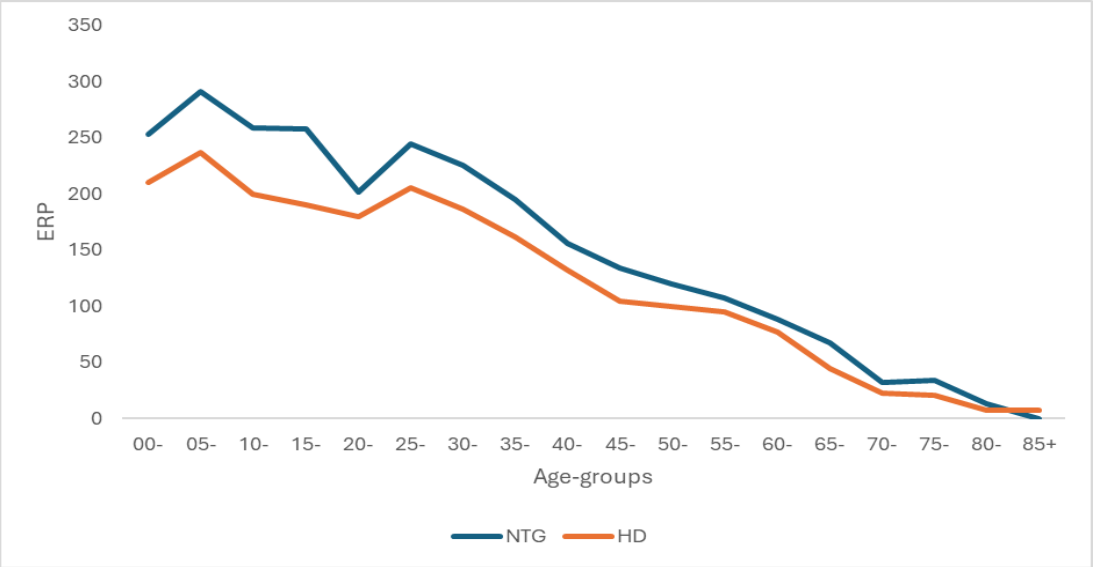
ERP Line Diagrams

The line diagram for the Aboriginal male population in Barkly illustrates the sensitivity of NTG Region estimates to small populations and the use of linear interpolation. At older ages, particularly 85 years and over, NTG Region estimates produce zero count, while HD estimates produce small but non-zero value (shown in Figure-1) which is statistically more plausible and suitable for analysis. Barkly has a very small total population and a highly uneven distribution

between Tennant Creek and surrounding remote communities, making it especially vulnerable to instability in age-specific population estimates.

The smoother age trajectory observed in the HD estimates reflects the use of proportional interpolation, which produces more stable age-specific population patterns in small populations. In contrast, NTG Region estimates exhibit greater fluctuation across adjacent age groups, reflecting the effects of linear interpolation combined with sparse input data. These abrupt changes can result in overestimation or underestimation in specific age groups and increased instability when population counts are small, as observed in Figure-1.

Figure-1: ERP for Aboriginal males by age group in Barkly/Tennant Creek, comparing HD and NTG Region estimates, 2025.

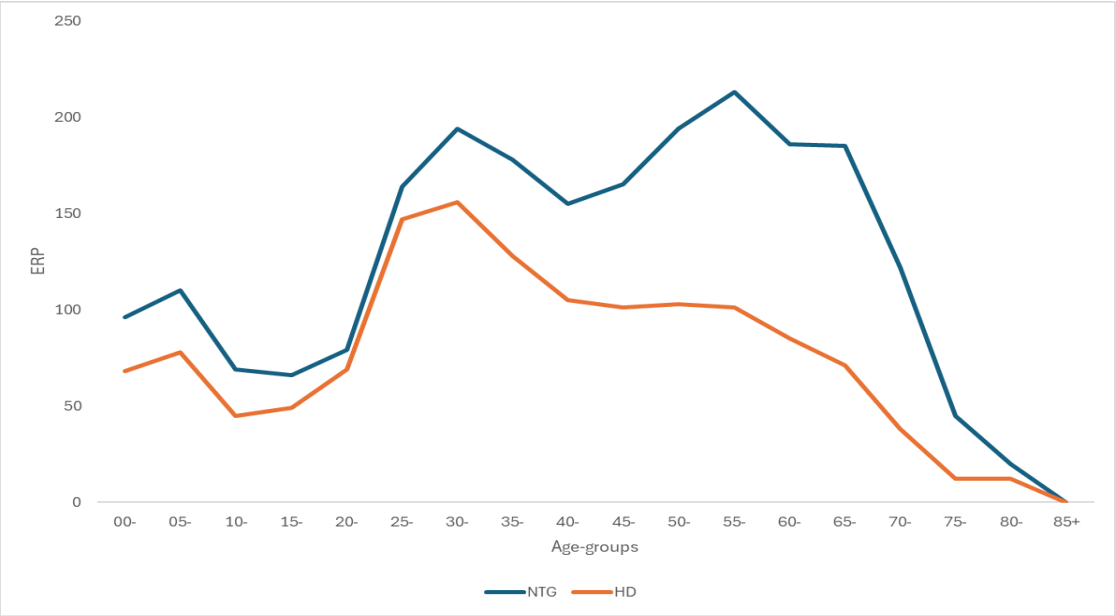


A similar pattern is observed in the ERP for the non-Aboriginal female population in the Top End (Figure 2). The NTG Region estimates display pronounced fluctuations across adjacent age groups, reflecting the effects of linear interpolation combined with multiple constraints in sparsely populated settings. As shown in Figure 2, these fluctuations become more evident from around 45 years and over, where population counts are smaller and more sensitive to fixed numeric adjustments.

In contrast, HD estimates exhibit smoother and more consistent age profiles that better reflect gradual demographic change. This stability arises from proportional interpolation, which distributes population change more evenly across age groups and reduces artificial variation caused by small cell counts. Similar patterns are observed in other sparsely populated and remote regions, including Barkly, Big Rivers, and parts of the Top End and Central Australia.

Together, these examples reinforce the analytical advantage of HD estimates in remote and small-population settings, particularly for age-specific analysis and rate calculations.

Figure-2: ERP for non-Aboriginal females by age groups in Top End/Jabiru-Tiwi, comparing HD and NTG Region estimates, 2025.



7. Summary

This report compares NTG Region and HD population estimates and explains why HD estimates are more methodologically robust for demographic and population health analysis in the Northern Territory. Both systems use ABS Census and the Estimated Resident Population data, and both rely on iterative proportional fitting to match official totals. However, they use different constraints and different interpolation methods, which affects stability and reliability of the estimates.

NTG Region estimates use multiple constraints and apply linear interpolation between census years. This method works well for producing population counts and percentage distributions for government planning and for designing programs and services. But the linear approach can introduce abrupt year to year changes and amplifies inconsistencies, particularly in small or sparsely populated groups.

HD estimates use a simpler and more stable set of constraints and apply proportional interpolation across the full census cycle. This produces smooth and steady year to year changes that more accurately reflect real population movement. It also avoids the artificial jumps that appear in NTG Region estimates and is more reliable in remote areas and in small Indigenous and older age groups where numbers fluctuate easily.

For demographic and health analysis, including age standardisation, disease prevalence, mortality and fertility studies, and small area trend analysis, HD estimates provide clearer and more dependable population denominators. Overall, HD estimates give NT Health a more consistent, understandable, and practical population base for future analysis and planning.

Appendix: Labelling of small areas in the NT

The use of Health Districts does not align with contemporary service-delivery boundaries and whilst the population discrepancies between the core datasets are small at disaggregated level in the NT, there is a need to ensure consistent labelling and definitions in the analysis produced by NT Health. The Table below aligns labelling of data to help ensure consistency in terminology.

The recommended labelling consolidates terminology to contemporary service-delivery boundaries. The approach supports more meaningful interpretation of results and enables consistent comparison across administrative, service-delivery, and population-health frameworks. It also enhances policy relevance by ensuring reported outcomes correspond with the regions responsible for planning, funding, and service oversight.

Table-A1: Recommended Small Area Labelling to Align Population Data with NT Service Delivery Boundaries.

Health Districts (Business Intelligence)	**Indigenous Regions labels**	**Recommended labelling to align with service delivery**	**Binary regions – Top End and Central Australia**	**Binary – NT urban/remote***
Darwin urban	Darwin	Darwin	Top End	Urban
Darwin - rural	Jabiru-Tiwi	Top End		Remote
East Arnhem	Nhulunbuy	East Arnhem		Remote, except Gove township
Katherine	Katherine	Big Rivers		Remote, except Katherine township
Tennant Creek	Tennant Creek	Barkly	Central Australia	Remote, except Tennant Creek township
Alice Springs rural	Apatula	Central Australia		Remote

Alice Springs urban	Alice Springs	Alice Springs (or combine with Central Australia)		Urban

*Not based on ARIA but based on NT specific urban and remote classification. All urban areas have common public services: domestic airport, public hospital, high school.

Information compiled by:

Poulomi Chowdhury, Alyson Wright

Suggested citation: Chowdhury, P. & Wright, A. (2025). *Comparing NTG Region and Health District Population Estimates*. Health Statistics and Informatics, NT Health, Darwin.

Enquiries to:

Health Statistics and Informatics

Floor-6, Manunda Place, Darwin

PO Box 40596, Casuarina NT 0811

Email: Statistics.Health@nt.gov.au