Journal of
Computational Mathematics and Algorithms

Research Article

Volume: 1 | Isssue: 1 | July 2026 | Pages 01–04

# Quantifying Estimate Saturation through Mathematical Reliability Theory: An Application to Medical Scheme Member-Movement Data in South Africa, 2014–2024

Michael Mncedisi Willie

Policy Research and Monitoring, Council for Medical Schemes, Pretoria, South Africa

***Corresponding Author:** Michael Mncedisi Willie, Policy Research and Monitoring, Council for Medical Schemes, Pretoria, South Africa.

**Submitted:** 29 June 2026 **Accepted:** 04 July 2026 **Published:** 13 July 2026

**Citation:** Willie, M.M. (2026). Quantifying Estimate Saturation through Mathematical Reliability Theory: An Application to Medical Scheme Member-Movement Data in South Africa, 2014–2024. Journal of Computational Mathematics and Algorithms, 1(1), 1-4.

**Abstract**

**Background:** "Data saturation" the point at which additional data cease to alter a study's conclusions is widely invoked but rarely defined mathematically, leaving the question "how much data is enough?" to subjective judgement. We recast saturation as a property of estimator reliability and apply it to a large administrative dataset.

**Methods:** Reliability theory implies that the relative standard error (RSE) of an aggregate estimator decays as a square-root law, $RSE(n) \approx c/\sqrt{n}$, the continuous analogue of the Spearman–Brown formula. Saturation is then defined as the sample size $n^*$ at which RSE first falls below a tolerance $\tau$. Using Council for Medical Schemes member-movement records (2014–2024; 70 schemes; 26,959 usable stratum-year observations), we estimated a system turnover proportion and a dependant exit-share, characterised reliability decay by bootstrap resampling, and validated the square-root law.

**Results:** Pooled beneficiary turnover was 46.8% and the dependant share of exits 51.8%. Exit intensity rose monotonically with age from 16.4% in infants to 84.7% at 85+ years, a classical reliability wear-out profile. The bootstrap RSE followed the predicted law ($c = 1.08$), giving saturation at $n^* \approx 467$, 2,919 and 11,673 observations for tolerances of 5%, 2% and 1%. Saturation was parameter-specific: the dependant exit-share required ≈ 2,778 observations at 5%.

**Conclusions:** Framing saturation as a reliability threshold yields an explicit, reproducible stopping rule and a design tool for deciding how much administrative data is sufficient for a target precision.

**Keywords:** Data Saturation, Reliability Theory, Spearman–Brown, Bootstrap, Relative Standard Error, Medical Schemes, Administrative Data, South Africa.

## Introduction

The notion of data saturation originated in grounded-theory research as the moment when additional observation yields no new categories or insight [1,2]. It has since become one of the most cited and most contested justifications for sample size, precisely because it is usually asserted rather than measured [3]. Recent work has sought to operationalise it, for example through the proportion of new information contributed by successive data increments [4,5], but these advances remain largely confined to qualitative coding and are seldom connected to the statistical theory that governs how estimates stabilise as data accumulate.

Reliability theory offers exactly that connection. In classical test theory the reliability of a composite increases with the number of constituent measurements according to the Spearman–Brown prophecy formula [6-8], while in statistical estimation the standard error of an average shrinks in proportion to the inverse square root of the sample size [9,10]. Both express the same underlying mechanism: information accumulates with diminishing marginal returns. Saturation, understood this way, is not a mysterious qualitative endpoint but the sample size at which an estimator becomes reliable enough for its intended use.

This paper formalises that idea and demonstrates it on a substantial real-world dataset: the member-movement records collated by the Council for Medical Schemes (CMS), the statutory regulator of private health financing in South Africa [11,12]. These records document, for each scheme, financial year and age band, the beneficiaries joining, transferring and leaving a natural setting for reliability theory, in which a member exit is the analogue of a component "failure." Our objectives are threefold: (i) to define data saturation as a reliability threshold and derive the implied stopping rule; (ii) to estimate the reliability-decay curve empirically and test the square-root law; and (iii) to illustrate the substantive insights a membership turnover level, an age-related exit lifecycle, and a temporal trend that the framework yields along the way.

## Theoretical Framework

### Reliability of an Aggregate Estimator

Let each observation $i$ contribute a numerator $b_i$ (events) and a denominator $e_i$ (exposure). The pooled ratio estimator after $n$

observations is $\hat{\theta}_n = \Sigma b_i / \Sigma e_i$. Its sampling variability, relative to its own magnitude, is summarised by the relative standard error $RSE(n) = SE(\hat{\theta}_n)/\hat{\theta}_n$. Under broad conditions the central limit theorem gives $SE(\hat{\theta}_n) \propto n^{(-½)}$, so that

$$RSE(n) \approx c / \sqrt{n} ,$$

where the constant c captures the heterogeneity of the per-observation contributions: noisier, more skewed data inflate c and slow convergence [10,13].

**The Spearman–Brown Analogue**

Classical test theory describes the reliability $\rho_n$ of a composite of n equivalent measurements through the Spearman–Brown formula [6,7], $\rho_n = n\rho_1 / [1 + (n-1)\rho_1]$. Both this expression and the square-root law in Section 2.1 are monotone, concave functions of n that approach their ceiling with diminishing returns. Reliability theory thus predicts not only that estimates improve with more data, but that the rate of improvement falls away the formal signature of saturation.

**Saturation as a Reliability Threshold**

Given a tolerance τ (the largest relative error a user will accept), we define the saturation point as the smallest sample size at which the estimator meets that tolerance:

$$n^* = \min \{ n : RSE(n) \le \tau \} \approx (c / \tau)^2 .$$

Three properties follow immediately. First, n* scales with the square of the desired precision, so halving the tolerance quadruples the data requirement. Second, n* is parameter-specific: quantities with larger c (rarer or more dispersed) saturate later. Third, the rule is fully reproducible it replaces a subjective judgement of “enough” with an explicit, pre-registrable criterion. Beyond n*, the marginal information per additional observation is, by construction, smaller than the analyst has declared material.

## Data and Methods

**Data Source**

We analysed CMS member-movement returns for the financial years 2014 to 2024. Each record reports, for a given scheme (identified by a reference number), scheme type (open or restricted) and five-year age band, the number of members transferring in, members not transferring, new dependants joining, and members, dependants and beneficiaries leaving. The file contained 27,545 records across 70 registered schemes; after removing 586 records with zero beneficiary throughput, 26,959 stratum-year observations remained. An internal consistency check confirmed that beneficiaries leaving equalled members leaving plus dependants leaving in every record.

**Estimands**

The primary estimand was the beneficiary turnover proportion θ = (beneficiaries leaving) / E, where the exposure E is the beneficiary throughput of the record, defined as members transferring in plus members not transferring plus new dependants joining plus beneficiaries leaving. A secondary estimand, the dependant exit-share φ = (dependants leaving)/(beneficiaries leaving), depends only on the unambiguous exit columns and was used to show that saturation is parameter-specific. Age-band exit intensity, θ evaluated within each age band, served as a reliability lifecycle profile.

**Saturation Estimation**

Two complementary procedures characterised convergence. First, to visualise stabilisation, records were placed in random order and the running pooled estimate $\hat{\theta}_n$ computed as observations accumulated; the procedure was repeated over 300 independent orderings to form a 95% envelope. Second, to quantify reliability, the RSE was estimated by nonparametric bootstrap [10]: at each of 60 sample sizes spaced geometrically between 5 and 26,959, we drew 400 resamples with replacement, recomputed $\hat{\theta}$, and took RSE = SD/mean. The square-root law was fitted by least squares on the log scale, and saturation points n* were obtained from the fitted c for tolerances τ = 5%, 2% and 1%. All analyses used Python (NumPy, pandas, SciPy); resampling used a fixed seed for reproducibility.

## Results

**Descriptive Overview**

Across the eleven years the records captured very large aggregate flows of the order of $2.2 \times 10^8$ beneficiary exits, $1.3 \times 10^8$ new dependant joins and $7.0 \times 10^7$ in-transfers in cumulative stratum-year terms. Pooled beneficiary turnover was $\hat{\theta}^* = 46.8\%$ and the dependant exit-share $\hat{\varphi}^* = 51.8\%$. Open schemes showed higher turnover (47.8%) than restricted schemes (37.9%), consistent with the greater stability of employer-based membership. The temporal trend (Figure 2b) was broadly flat but punctuated by a marked peak in 2020 (49.7%), the first pandemic year, and a trough in 2017 (44.0%).

**Reliability Lifecycle by Age**

Exit intensity traced a clear lifecycle (Figure 2a). After an “infant exception” beneficiaries under one year had the lowest exit intensity (16.4%), reflecting newborn registrations that rarely terminate within the year the profile rose almost monotonically across the lifespan, from roughly 47–50% in working ages to 53.1% at 55–59, 57.7% at 70–74 and 84.7% at 85 years and older. This increasing-failure-intensity shape is the demographic counterpart of the wear-out region in reliability engineering [11,12], and indicates that retention risk is concentrated at the oldest ages.

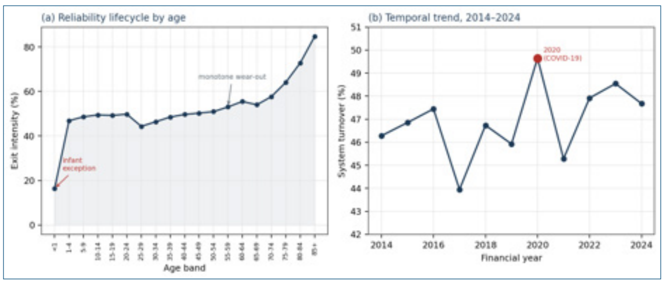


**Figure 1:** Reliability Lifecycle and Temporal Trend. (a) Beneficiary Exit Intensity by Age Band, Showing the Infant Exception and a Monotone Wear-Out Gradient; (b) System Turnover by Financial Year, with the 2020 Pandemic-Year Peak Highlighted.

**Estimate Saturation and the Reliability Law**

The running estimate converged rapidly toward $\hat{\theta}^*$ (Figure 2a, below): the central trajectory entered and remained within 1% of the final value after only about 69 accumulated observations, although the 95% envelope across orderings narrowed more slowly. Quantifying this with the bootstrap, the RSE fell along the predicted square-root law with constant c = 1.08 (Figure 2b, below), confirming that reliability theory describes the convergence well. Solving $n^* \approx (c/\tau)^2$ gave saturation at approximately 467 observations for a 5% tolerance, 2,919 for 2% and 11,673 for 1%. Saturation was parameter-specific: the dependant exit-share, being more dispersed (c = 2.64), required about 2,778 observations to reach 5% nearly six times the primary estimand's requirement.

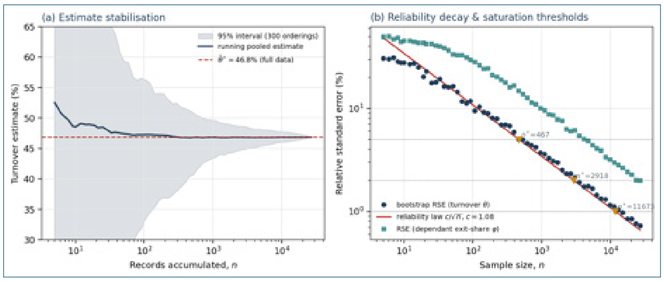


**Figure 2:** Saturation Diagnostics. (a) Running Pooled Turnover Estimate Under 300 Random Orderings, converging to $\hat{\theta}^* = 46.8\%$; (b) Bootstrap Relative Standard Error Versus Sample Size on Log–Log Axes, Following the Reliability Law $c/\sqrt{n}$ (c = 1.08), with Saturation Points n* Marked for 5%, 2% and 1% Tolerances. The Dependant Exit-Share (teal) Saturates More Slowly.

### Saturation Requirements

**Table 1: Saturation Sample Sizes $n^* \approx (c/\tau)^2$ by Estimand and Tolerance $\tau$, with the Share of the 26,959-Observation Dataset Required at $\tau$ = 5%. Larger Dispersion (c) Implies Later Saturation.**

| Estimand | c | n* (5%) | n* (2%) | n* (1%) | % of dataset (5%) |
|---|---|---|---|---|---|
| Beneficiary turnover (θ) | 1.08 | 467 | 2,919 | 11,673 | 1.7% |
| Dependant exit-share (φ) | 2.64 | 2,778 | 17,355 | 69,422 | 10.3% |

## Discussion

### Principal Findings

Recasting data saturation as a reliability threshold converts an often-subjective claim into an explicit, reproducible quantity. Applied to a decade of South African medical-scheme movement data, the framework recovered the classical square-root reliability law (c = 1.08) and translated it directly into sample-size requirements fewer than 500 stratum-year observations for 5% precision on system turnover, rising to roughly 11,700 for 1%. Along the way it surfaced substantive results: high beneficiary turnover (46.8%), a stark age gradient in exit intensity, and a pandemic-year spike in 2020.

### Methodological Contribution

The central contribution is the stopping rule $n^* \approx (c/\tau)^2$. Because c is estimable from a modest pilot sample, an analyst can forecast how much data a target precision will demand before committing to full extraction useful when assembling administrative or routinely collected data is costly. The rule also makes transparent two facts that qualitative saturation arguments tend to obscure: precision improves only with the square root of effort, and different estimands from the same dataset saturate at very different points, so a single "saturated" label for a study is rarely warranted.

### Substantive Interpretation

The age lifecycle mirrors a reliability wear-out curve: exit intensity climbs steadily with age and accelerates beyond 75 years, consistent with mortality and end-of-life disenrolment, while the low infant value reflects within-year newborn additions. The open-versus-restricted gap accords with the tighter retention of employer-sponsored cover, and the 2020 peak is compatible with pandemic-related disruption to employment and contributions. These patterns are descriptive and ecological, but they demonstrate that the reliability lens is not merely a sample-size device it also organises interpretation around where "failures" concentrate.

### Strengths and Limitations

Strengths include the scale and regulatory completeness of the data, a theory-driven rather than ad hoc convergence criterion, and bootstrap estimation that makes no parametric assumption about the sampling distribution. Several limitations temper the findings. The exposure denominator E is a beneficiary-throughput proxy that combines principal-member and dependant counts, so turnover should be read as a churn index rather than a life-table disenrolment probability; the unit heterogeneity is an acknowledged simplification. The resampling treats stratum-year records as exchangeable draws, which understates dependence among records from the same scheme; a clustered bootstrap would yield slightly larger c and later n*. Estimates are aggregate and unadjusted for covariates, and the temporal trend is not modelled formally. Finally, the framework quantifies statistical saturation stability of an estimate and not the separate question of whether the right quantity is being measured.

## Conclusions

Data saturation need not remain an impressionistic stopping point. Treating it as the sample size at which an estimator's relative error meets a declared tolerance gives a simple, defensible and pre-registrable rule, $n^* \approx (c/\tau)^2$, grounded in reliability theory. Applied to medical-scheme member movement, it both calibrated data requirements and revealed a coherent reliability lifecycle of membership, suggesting the approach is a practical addition to the analysis of large administrative datasets [14-16].

**Funding:** This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

**Conflict of Interest:** None

**Author Contributions:** MMW conceived the study, performed the analysis and wrote the manuscript.

## References

1. Glaser, B.G., Strauss, A.L. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine.
2. Guest, G., Bunce, A., Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. Field Methods, 18(1), 59-82.
3. Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., et al. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. Qualitative & Quantitative, 52(4), 1893-1907.
4. Guest, G., Namey, E., Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research.

PLOS ONE, 15(5), e0232076.
5. Lowe, A., Norris, A.C., Farris, A.J., Babbage, D.R. (2018). Quantifying thematic saturation in qualitative data analysis. Field Methods, 30(3), 191-207.
6. Spearman, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3(3), 271-295.
7. Brown, W. (1910). Some experimental results in the correlation of mental abilities. British Journal of Psychology, 3(3), 296-322.
8. Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297-334.
9. Lord, F.M., Novick, M.R. (1968). Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley.
10. Efron, B., Tibshirani, R.J. (1993). An Introduction to the Bootstrap. New York: Chapman & Hall.
11. Council for Medical Schemes. (2024). Annual Report 2023/24. Pretoria: Council for Medical Schemes.
12. Republic of South Africa. (1998). Medical Schemes Act 131 of 1998. Pretoria: Government Printer.
13. Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
14. Barlow, R.E., Proschan, F. (1965). Mathematical Theory of Reliability. New York: Wiley.
15. Rausand, M., Høyland, A. (2004). System Reliability Theory: Models, Statistical Methods, and Applications (2nd ed.). Hoboken, NJ: Wiley.
16. van Belle, G. (2008). Statistical Rules of Thumb (2nd ed.). Hoboken, NJ: Wiley.