ACAC
AUSTRALIAN CONSORTIUM FOR AGED CARE

How should long-term care facility outliers be identified?

Quality indicators are increasingly used to monitor the performance of long-term care facilities and support ongoing quality improvement efforts. One way in which indicators are used to improve quality is to identify performance outliers – that is, long-term care facilities that show unusually good or unusually poor results compared to what is expected. Highlighting long-term care facilities with very good performance provides opportunities to learn from their practices. Identifying low-performing long-term care facilities can facilitate timely action to improve care delivery and minimise potential harms to residents.

Several statistical methods for identifying potential outliers exist, but there is limited agreement on the most appropriate to use, particularly in the setting of long-term care. As for any statistical test, each method has a certain rate of false positives and false negatives, which may depend on the characteristics of the indicator that is being used. Understanding how these rates differ across methods is important when choosing which method to use.

ACAC work in this area

We examined several methods to detect potential outliers for four indicators routinely used to monitor the quality of care in long-term care facilities to understand how the different methods performed across varying facility and indicator characteristics (i.e., facility size, quality indicator prevalence, and degree of overdispersion).

Our approach

We applied existing outlier detection methods to risk adjusted indicators of emergency department presentations, antipsychotic use, fall-related hospitalisations, and pressure injury related hospitalisations. Data from 2,744 long-term care facilities in Australia from the Registry of Senior Australians (ROSA) National Historical Cohort for the 2019 calendar year was used. The four indicators are routinely reported in Australia as part of the ROSA Residential Care Outcomes Monitoring System (Inacio et al., 2020), and were selected because they are widely used internationally and cover a wide range of prevalences of occurrence.

For each indicator, we calculated how many low or high outlier long-term care facilities were identified when using different types of detection limits (i.e., confidence interval-based or prediction interval-based) and when adopting different strategies for accounting for overdispersion (i.e., no adjustment, a multiplicative adjustment, or an additive adjustment using random effects).

To understand the reasons for differences between methods, we created simulation scenarios that were similar to the real data and included known outliers. Using these simulations, we determined how the prevalence of an indicator and its level of overdispersion affected the false positive and false

negative rates of each method. By extrapolating back to the real data, where the true outliers are not known, we inferred which methods are appropriate for particular uses and particular indicators.

What we found

- Identification of facility outliers is sensitive to the choice of method and the characteristics of the indicator.
- For moderate prevalence indicators, such as emergency department presentations (where 37.9% of residents experienced the event), antipsychotic use (21.5%), and fall-related hospitalisations (13.6%), confidence interval-based and prediction interval-based methods performed similarly in outlier detection.
- For low prevalence indicators, such as pressure injury related hospitalisations (3.4%), prediction interval-based methods were slightly more sensitive than confidence interval-based methods. Detection limits based on confidence intervals around the benchmark identified fewer potential outliers with higher-than-expected event rates compared to detection limits using prediction intervals.
- When overdispersion was present, methods that did not account for it had very high false positive rates, whereas those that did account for it were able to control the false positive rate at an acceptable level.
- Considerations for the use of different methods is given in Table 1 and illustrated in Figure 1.

Implications and recommendations based on our findings

- Control limits based on prediction intervals should generally be preferred, as they are less susceptible to misinterpretation and exhibit less sensitivity to the properties of the specific indicator used.
- Overdispersion is ubiquitous in quality indicator data, and quality monitoring programs need to use methods that appropriately account for this when identifying facility outliers.

For full details, access our peer-reviewed publication:

Harries, D., Schwabe, J., Jorrissen, R.N., Pulling, B.W., Caughey, G.E., Comans, T., Inacio, M.C. on behalf of the ACAC QMET Collaborators. (2026). Comparing methodological approaches for outlier detection: a cross-sectional study of older persons in 2,744 Australian long-term care facilities. *Int J Qual Health Care*, *mzag144*. https://doi.org/10.1093/intqhc/mzag144

For full details, access our deidentified analysis code:

https://osf.io/mzy4h

Other references

Inacio, M.C., Lang, C., Caughey, G.E., Bray, S.C.E., Harrison, S.L., Whitehead, C., Visvanathan, R., Evans, K., Corlis, M., Cornell, V., & Wesselingh, S. (2020). The Registry of Senior Australians outcome monitoring system: quality and safety indicators for residential aged care. *Int J Qual Health Care*, *32*(8), 502-510. https://doi.org/10.1093/intqhc/mzaa078

ACAC
AUSTRALIAN CONSORTIUM FOR AGED CARE

Table 1: Considerations for the use of different methods for the studied indicators when one may prioritise higher sensitivity or higher specificity

	Quality Indicator			
	Pressure-injury Related Hospitalisations	**Falls**	**Antipsychotic Use**	**Emergency Department Presentations**
Indicator characteristics	Low prevalence (3.4%), modest overdispersion	Moderate prevalence (13.6%), modest overdispersion	Moderate prevalence (21.5%), limited overdispersion	High prevalence (37.9%), high overdispersion
Choice of limits				
Higher sensitivity	PI method with score test.	PI method with score test.	PI method with score test.	PI method with score test.
Higher specificity	PI methods with score test (with higher confidence level if necessary). CI method with Wilson, Jeffreys, or exact may be conservative for this indicator.	PI methods with score test (with higher confidence level if necessary). CI method with Wilson, Jeffreys, or exact may be conservative for this indicator.	PI methods with score test (with higher confidence level if necessary). CI method with Wilson, Jeffreys, or exact may be conservative for this indicator.	PI methods with score test (with higher confidence level if necessary). CI method with Wilson, Jeffreys, or exact may be conservative for this indicator.
Adjustment for overdispersion				
Higher sensitivity	Necessary, consider robust methods.	Necessary, consider robust methods.	May be necessary, consider robust methods.	Necessary, consider robust methods.
Higher specificity	Necessary, consider robust methods.	Necessary, consider robust methods.	Necessary, consider robust methods.	Necessary, consider robust methods.
Multiplicity adjustment				
Higher sensitivity	FDR control may be considered.	FDR control may be considered.	FDR control may be considered.	FDR control may be considered.
Higher specificity	FDR control may be considered, or FWER for more stringent control.	FDR control may be considered, or FWER for more stringent control.	FDR control may be considered, or FWER for more stringent control.	FDR control may be considered, or FWER for more stringent control.
Overall recommended suitable methods for:				
Higher sensitivity	PI method with score test and handling of overdispersion (multiplicative or additive).	PI method with score test and handling of overdispersion (multiplicative or additive).	PI method with score test and handling of overdispersion (multiplicative or additive).	PI method with score test and handling of overdispersion (multiplicative or additive).
Higher specificity	PI methods, with higher confidence level if necessary. Adjustment for overdispersion required.	PI methods, with higher confidence level if necessary. Adjustment for overdispersion required.	PI methods, with higher confidence level if necessary. Adjustment for overdispersion recommended.	PI methods, with higher confidence level if necessary. Adjustment for overdispersion required.

Abbreviations: CI – confidence intervals, FDR – false discovery rate, FWER – family-wise error rate, PI – prediction intervals

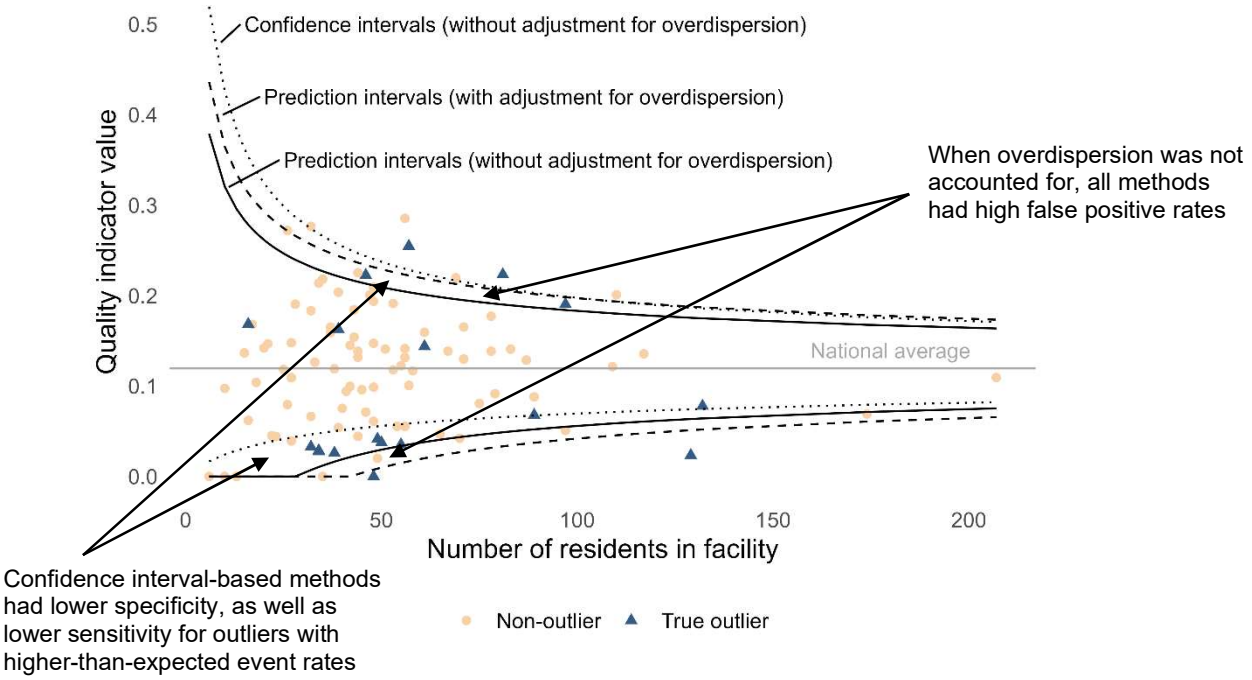


Figure 1: Example of outlier detection using different approaches for constructing control limits and accounting for overdispersion