



Australian Bureau of Statistics

Locked Bag 10 | Belconnen ACT 2616

Ref: EC24-000107

Mr Dennis Trewin AO, FASSA
Former Australian Statistician

Dear Dennis,

Thank you for providing this report on your Statistical Review of the process for providing regional enrolment projections for electoral redistribution purposes.

I would like to formally accept the report and its findings, and to thank you for undertaking the review. I intend to implement the full set of recommendations.

The ABS has now improved the quality management procedures for producing enrolment projections in line with your report. Your timely and expert advice reassures me that these improvements will ensure similar errors are not repeated in the future.

I welcome the recommendations and insights you have made in Section 5 of the report to strengthen and reinforce the importance of the ABS' quality management arrangements more generally. I am actively working with ABS management to ensure these insights are embedded across all areas of ABS statistical production to further improve our quality management practices.

Thanks again for undertaking this work which will support the ABS efforts to ensure the quality of its outputs, and the confidence and trust of data users.

Yours sincerely,

A handwritten signature in black ink that reads 'David Gruen'.

Dr David Gruen AO
Australian Statistician

26 March 2024

Statistical review of process for providing regional enrolment projections for electoral redistribution purposes: Summary of recommendations

5.1 Recommendations for regional enrolment projections for electoral purposes

Recommendation 5.1.1

Extend the set of output edits performed at the main gate to ensure they are more likely to identify possible errors.

Recommendation 5.1.2

Population Statistics Branch should introduce arrangements for review of the face validity of the estimates prior to AEC making the enrolment projections public, including by the Regional Population Unit (RPU) and AEC Divisional Managers.

Recommendation 5.1.3

The ABS should monitor the Suggestions on the AEC website for criticisms of the enrolment projections. AEC should be encouraged to contact the ABS if they hear or observe criticisms.

Recommendation 5.1.4

There should be a review of Population Statistics Branch resources to ensure it has all the resources it requires, given increased staff turnover and demand for population statistics.

5.2 General lessons

Recommendation 5.2.1

Standards should be produced for systems developed using user-oriented tools. These would have a lighter touch than a full Systems Development Methodologies but would include requirements to test the accuracy of the system and provide sufficient documentation to enable someone else to use the system.

Recommendation 5.2.2

Reinforce the importance of quality gates as part of the ABS Quality Management arrangements, especially for non-regular statistical outputs.

Recommendation 5.2.3

Consider the need for updates to the Quality Management Framework given it has been in place for nearly 15 years. In particular consider:

- updating the ABS Quality Management Framework,
- preparing some additional supporting resource materials,
- promoting the revised framework,
- providing some experts to support statistical areas revising their quality management frameworks, and
- assigning responsibility for leadership of quality management improvement across the ABS (noting in most official statistical agencies, this is assigned to Methodology).

Recommendation 5.2.4

A face validity or 'sniff test' should be applied to all ABS outputs but especially those of high importance. Where possible, this should be someone not closely involved in the compilation ('a fresh set of eyes'). Most often it will be a more senior officer involved in the publication sign-off process but there may be others who have sufficient knowledge of the subject matter to provide some insights.

Recommendation 5.2.5

Negative feedback from sophisticated users on accuracy issues should be taken seriously, and ABS should provide an explanation (even if the concerns are invalid) to maintain confidence in the figures.

Statistical review of process for providing regional enrolment projections for electoral redistribution purposes

Dennis Trewin AO, FASSA

Executive summary

Context

Enrolment projections provided to the Australian Electoral Commission (AEC) for redistributions currently underway in Victoria and Western Australia contained errors in the projected estimates of enrolments at the regional level. This was discovered during the AEC public consultation. Following investigation, the ABS resolved the errors and supplied corrected projections to the AEC.

The error was made in the SAS¹ program written to undertake the task but was not detected until after the AEC notified the ABS of concerns they had received about the accuracy of the enrolment projections. The objective of the review was to look at the processes, not the methodology, for compiling regional population projections for electoral redistribution purposes. In essence, the report is about why the error was not picked up earlier and the steps that should be undertaken to avoid similar errors in the future.

Findings

My main findings were as follows. My quantitative comments are based on Victoria as the error had more severe consequences there than in Western Australia.

1. The SAS program prepared to compile the regional enrolment projections was inadequately tested.
2. Consistent with good practice, a variety of output edits² were undertaken before delivery to AEC. However, an important output edit, that would have clearly identified the presence of an error, was not undertaken. For Victoria, it showed that improbably large number of SA1s³ had a projected change in the electoral population of 8 to 10 per cent over the 5 years to 2028. Virtually no SA1s had a lower projected change, and none had a higher projected change. Population Statistics Branch included this additional output edit in their revised electoral projections and will do so for future projections made for electoral purposes.
3. One of the output edits involved correlative analysis at the SA1 level between the projections of the population aged 18 years and over and the enrolment projections. This was also undertaken at the SA2⁴ and Electoral Division levels. This showed lower than usual correlations, but this was incorrectly attributed to the impact of COVID on the Medicare data which is the major data source used for estimating change in regional populations.
4. There were no face validity checks⁵ of outputs by someone familiar with sub-state population trends in the States. For example, I know the population trends in Victoria well enough to have identified that there was a problem. The Regional Population Unit within the

¹ SAS is a computer programming language designed for statistical analysis.

² Output edits are validation checks performed on aggregate data outputs.

³ SA1s or Statistical Area Level 1s are small geographic areas containing populations of between 200 and 800 people (as at July 2021).

⁴ SA2s or Statistical Area Level 2s are geographic areas containing populations of 10,000 people on average (as at July 2021).

⁵ Face validity checks involve looking at summary outputs to see whether they align with real world knowledge.

ABS believe they have sufficient knowledge of population trends in the States and Territories to play this role in the future. Furthermore, Population Statistics Branch plans to recommence briefings (similar to those conducted up until 2020) of AEC Divisional Managers before they are publicly released.

5. A number of the Suggestions on the AEC website raised concerns with the Victorian enrolment projections, but these were not monitored by the ABS, nor did AEC raise concerns with the ABS until a week or so before the letter from the Australian Electoral Commissioner on 21 December. The AEC first published the enrolment projections on 25 October and Suggestions started flowing shortly afterwards, some of which raised concerns about the projections. (The concerns were mentioned in Antony Green's blog of 27 November). It is likely that the error could have been identified in mid-November.
6. Once the ABS became aware of the error, it responded reasonably quickly (10 working days for Victoria and 12 working days for Western Australia).
7. I believe human resource issues are a partial explanation. The resources available to Population Statistics Branch have remained stable although turnover has been high because of the increased demand for demography skills. Furthermore, the workload has increased significantly over this time. It is one of areas of the ABS where errors can cause significant damage to the ABS reputation.

Recommendations for regional enrolment projections for electoral purposes

The main quality gate⁶ is at the right place in the process for compiling electoral enrolment projections. However, the set of output edits performed at this gate should be extended to ensure they are more likely to identify possible errors. This was done for the revised estimates and confirmed that the error would have been found with the additional output edits.

Furthermore, Population Statistics Branch should introduce arrangements for reviewing the face validity of the estimates. This needs to be someone who is familiar with the sub-State population trends in the State or Territory. The Regional Population Unit believes it has this capability. Also, as proposed by Population Statistics Branch, briefings of AEC Divisional Managers should recommence prior to AEC making the enrolment projections public.

For future re-distributions, the ABS should monitor the Suggestions on the AEC website for criticisms of the enrolment projections. AEC should be encouraged to contact the ABS if they hear or observe criticisms.

Population Statistics Branch resources are stretched. Although resource levels have remained constant, staff turnover has increased, and the demand for statistical population statistics has also increased. It is a very important area of the ABS and there should be a review to ensure it has all the resources it requires.

General lessons

The ABS and other official statistical agencies rely a lot on user developed systems using user-oriented tools (e.g. SAS, spreadsheets). Systems Development Methodologies (SDM) are used for larger systems using IT professionals. Standards should also be produced for systems developed using user-oriented tools. These would have a lighter touch than a full SDM but would include

⁶ A quality gate is a pre-defined assessment point where the quality or accuracy of intermediate or final outputs are 'signed off' before the data can pass through to the next stage of processing.

requirements to test the accuracy of the system and provide sufficient documentation to enable someone else to use the system.

The ABS Quality Management arrangements involve a system of quality gates where intermediate and final outputs must be 'signed off' before they pass the next gate. This is a good arrangement especially if the quality gates are placed where they will be most cost effective. However, their importance needs to be reinforced, especially for non-regular statistical outputs, and there may need to be some updates to the Quality Management Framework given it has been in place for nearly 15 years.

A face validity or 'sniff test' should be applied to all ABS outputs but especially those of high importance. I would include electoral projections for redistribution in that category. Where possible, this should be someone not closely involved in the compilation ('a fresh set of eyes'). Most often it will be a more senior officer involved in the publication sign-off process but there may be others who have sufficient knowledge of the subject matter to provide some insights.

In my experience, negative feedback from sophisticated users on accuracy issues should be taken seriously. They are right more often than not. This is certainly the case for electoral projections where several of the political parties and a number of psephologists study these numbers in great detail. (Even if their concerns are invalid, they should be provided with an explanation to maintain their confidence in the figures.)

Conclusion

Population Statistics Branch has or plans to make several adjustments to the quality management procedures for the processes to develop enrolment projections for electoral redistribution purposes. If they are successfully implemented, similar errors should not be repeated in the projections provided for future redistributions.

1. Introduction

1.1 Context

Enrolment projections provided to the Australian Electoral Commission (AEC) for redistributions currently underway in Victoria and Western Australia contained errors in the projected estimates of enrolments at the regional level. This was discovered during the AEC public consultation. Following investigation, the ABS resolved the errors and supplied corrected projections to the AEC. The error did not affect the number of seats allocated to Victoria or Western Australia.

1.2 Review objectives

The ABS asked me to identify approaches to mitigate similar errors occurring in the release of ABS statistics and to inform the ABS on potential areas of risk exposure in current processes.

Furthermore, I was asked to review:

- the ABS response to the initial feedback about the projections through to the supply of corrected projections.
- the technical cause of the error.
- potential contributors to the error occurring. These could include capability, technology, compilation and quality assurance processes, capacity or any other circumstances of relevance during the period this work was being undertaken.

I was not asked to review the methodology for producing regional enrolment projections.

1.3 Conduct

To conduct the Review, I studied the documentation I was provided or requested or accessed from the website. I had several meetings with Population Statistics Branch and senior ABS staff who were involved with the management of the error. I also examined the relevant Suggestions on the AEC website. However, to a large extent I am relying on my prior experience at studying errors and quality management processes in the ABS, Statistics New Zealand and Statistics Sweden.

In this Report, I deal first with the errors in the electoral enrolment projections (section 2) before attempting to extract some general lessons for the ABS (section 3).

2. Enrolment projections

2.1 The error

The error occurred in data provided for both Victoria and Western Australia. The impact was much greater in Victoria so most of my analysis below refers to Victorian data.

The enrolment projections involve several iterative steps involving rather large data sets at the SA1 level. This is the lowest level at which the projections are produced but it is the building block for higher levels of aggregation. The detail of the error is shown in Box 1. It was due to an error in the SAS program written by a staff member who was a statistical analyst, that is not an IT expert. This is often referred to as 'User Computing' which is the term I will use in this report. In simple terms, the error occurred because the wrong denominator was used when compiling the enrolment ratios used for the projections.

User computing is used extensively in the ABS and other official statistical agencies using tools such as SAS and spreadsheets. They could not function without this capability. However, errors are common often because the level of formal testing is not as great when compared with IT developed systems which usually use a formal Systems Development Methodology (SDM). It is an area of high

risk especially when user computing is deployed for high profile statistics such as the CPI⁷ and changes are made (or even change in personnel).

Box 1

Detail of the error

The ABS process involved calculating an 'enrolment ratio' at the date of the enrolment count (09/08/2023) for each SA1, using the following formula:

$$\text{enrolment ratio} = \frac{\text{number of electors @ 09/08/2023}}{\text{projected 18 + population @ 09/08/2023}}$$

The numerator was supplied by AEC as enrolment counts by SA1. The denominator was calculated by the ABS using its population projections system (at the SA2 level) and also interpolated SA1 level population estimates from 2016 to 2022 (produced by ABS Regional Population Unit). This enrolment ratio is then assumed to be constant and is applied to each SA1's projected population aged 18 years and over, to create a projection of enrolled persons at the target date.

However, the enrolment ratio calculation was incorrectly calculated using this formula (note wrong denominator):

$$\text{enrolment ratio} = \frac{\text{number of electors @ 09/08/2023}}{\text{projected 18 + population @ 17/04/2028}}$$

Statistics Sweden made a spreadsheet error in their monthly CPI in the late 2000s. It became serious because it was not discovered for some time and the error accumulated in the same direction each month. After it was discovered, there was a lot of criticism as it had impacted on contracts indexed to the CPI and the Central Bank even said their monetary policy decisions had been impacted.

The official enquiry into the CPI error recommended that there be an independent review by international experts. A US colleague and I were asked to undertake an independent review of the processes for the CPI and other important statistics. The two main causes of the CPI error were inadequate testing of the accuracy of the spreadsheet compilations and inadequate documentation which meant the new compiler of these components of the CPI index could not use previously written programs. These risky behaviours were not limited to the CPI!

I mention this example to reinforce the importance of ensuring the right quality management processes are in place when user computing is used.

2.2 ABS response

How quick was the ABS response to the initial feedback about the projections through to the supply of corrected projections? The timeline is shown in Box 2.

⁷ Consumer Price Index

Box 2

Timeline – Victoria redistribution

13 October 2023: ABS provided data to AEC

25 October 2023: AEC published enrolment projections to be used in the redistribution process

25 October – 24 November 2023: Time period for Suggestions to be made on the redistribution

20 November 2023 (approx.): Date of first Suggestion stating that the enrolment projections look wrong

27 November: Antony Green’s blog refers to possible errors in enrolment projections

21 December 2023: AEC (Electoral Commissioner) first formally advised the ABS of concerns about the projections

12 January 2024: ABS alerted AEC of the error

16 January 2024: ABS provided corrected enrolment projections

The ABS responded reasonably quickly once formally advised by AEC on 21 December of the criticisms, particularly given the disruptions caused by the Christmas New Year period. The revised projections for Victoria were provided on 16 January and for Western Australia on 18 January, 10 and 12 working days respectively after AEC formally advised the ABS of the error.

I believe the error should have been detected earlier. The AEC published the electoral projections for Victoria on 25 October at the beginning of the consultation period from 25 October until 24 November. (It should be noted that this corresponded with the peak workload for ABS official population projections.) Suggestions would have flowed shortly afterwards, several of which made strong criticisms of the projections. I cannot identify the precise dates and when they were loaded to the AEC web site, but it appears the first significant criticism (Suggestion 32) appeared around 20 November. Certainly, Antony Green’s blog noted the criticisms on 27 November. The ‘danger signs’ would have evident earlier if the ABS had been monitoring the Suggestions. There is a website called Tally Room (www.tallyroom.com.au) which has a blog on the Victorian redistribution and scepticism about the enrolment projections started at the end of October.

I have looked at the Suggestions for Western Australia and did not find criticisms of the enrolment projections probably because the impact of the error was much smaller than in Victoria.

2.3 Process and quality management

All processes are subject to error. All humans make errors from time to time. You could try to a risk avoidance approach where every step is ‘double checked’. However, this is expensive, and studies show that in practice it is not so effective at picking up errors.

The ABS has a system of ‘quality gates’. This is a sensible approach – the key questions to be answered are to determine where the quality gate is placed and what quality checks should be undertaken at each gate. The main ABS quality gate is after the enrolment projections are produced, which I think is the most appropriate place, but there are other checks that could be undertaken, some of which were undertaken when the revised enrolment projections were produced.

I would suggest the following as the main components of a quality management program for enrolment projections.

1. Programs developed through user computing should be tested for accuracy using parallel running or test data packs that have been created for that purpose.
2. As at present, there should be output edits to assess the accuracy of the enrolment projections.
3. Arrangements should be made to check their face validity before they are published.
4. Monitoring and analysing feedback from informed users.

Each is discussed in turn in the following sections.

2.4 User computing

The ABS has a SDM for the development of large-scale systems using IT professionals. However, a lot of user computing is deployed to generate statistics or for customised user requests like enrolment projections. It may even be used within large scale systems. There should be ABS standards for systems developed by users. These should be proportionate and not on the scale of a SDM for larger systems. However, at a minimum, I think the standards should ensure:

- There is sufficient documentation to enable someone else to successfully operate the system if the original developer is no longer in the position or unavailable.
- Acceptance testing is undertaken to ensure the system meets business and user requirements.
- The system is tested for accuracy using parallel running or a 'test pack' that enables the compilations to be validated.

2.5 Output edits and validation

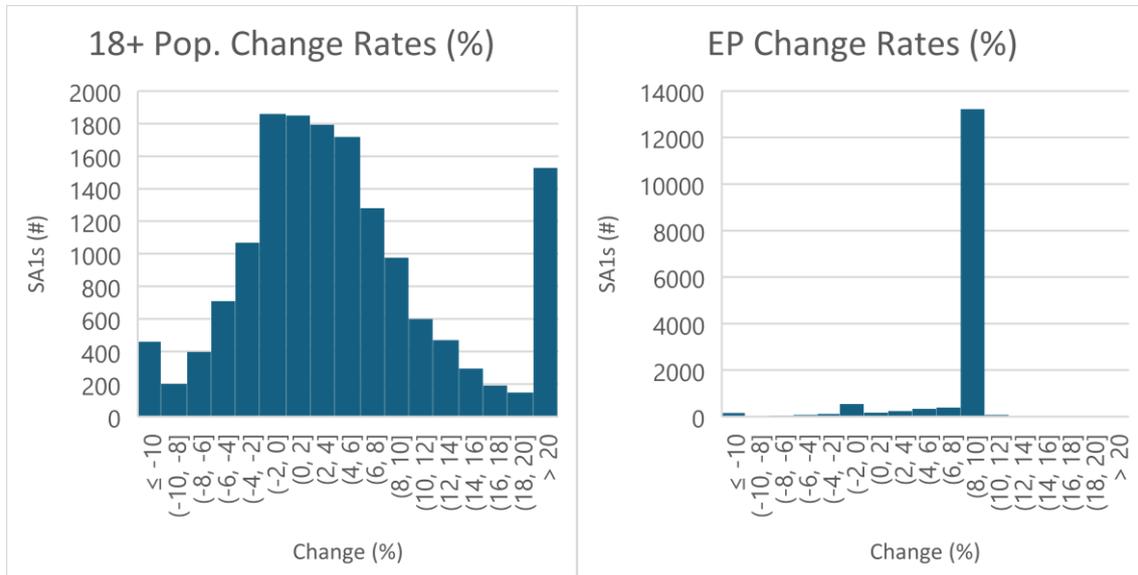
Population Statistics Branch conducts a series of output edits on the enrolment projections. This is consistent with good practice but unfortunately the tests used for initial enrolment projections did not pick up the error. They have revised the set of output edits which were then used for the revised estimates.

Another factor was the change in the behaviour of the Medicare data source because of the COVID-19 pandemic. This is the main data source for the population estimates which the enrolment projections are based on. One of the important tests is a correlative analysis of 'projected electors' versus 'projected population'. The R squared for the line of best fit was 0.4205 at the SA1 level, much lower than it had been for previous redistributions. This was explained away by the Covid induced changes in the behaviour of the Medicare data. The corresponding R squared for the corrected data was 0.6269 consistent with past experience. The relevant graphs are shown in Appendix 1.

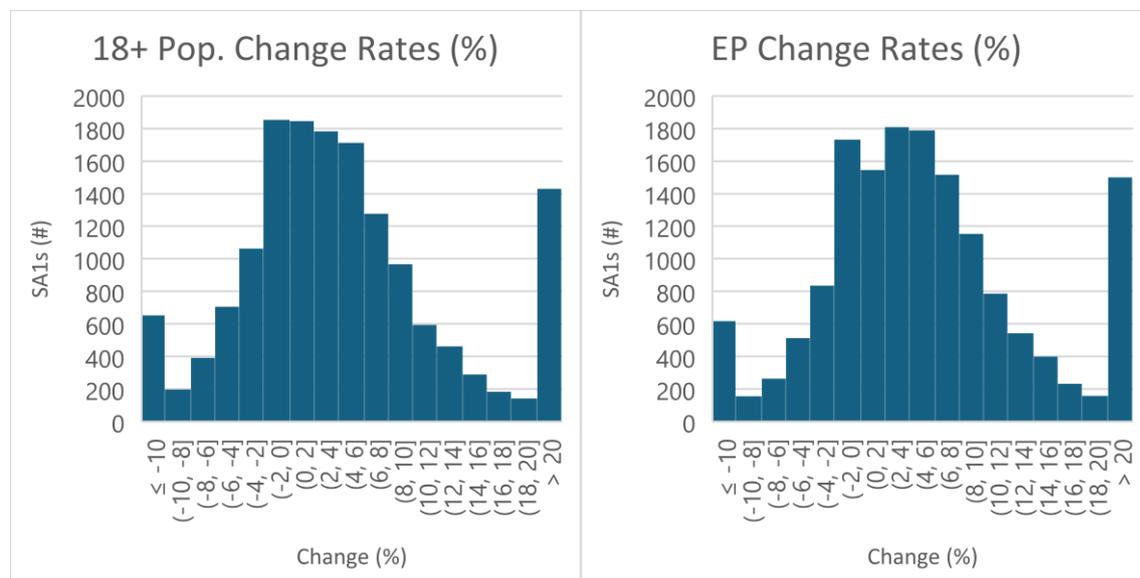
The following graphs shows the outcomes of the new test that was undertaken on the corrected data compared with outcomes if the same test had been applied to uncorrected data. These graphs were extracted from material provided by Population Statistics Branch. The comparison of the two graphs on the left hand clearly shows that something was wrong. Virtually all the SA1s showed a growth of 8-10% which was clearly unrealistic. The corrected population growth rates on the lower right hand side graph. This test (and others) will become part of the suite of output edits undertaken for future redistributions.

Graphical Representation of Population aged 18 years and over and Enrolled Population Change Rates Between Delivered and Re-Run Data.

Change in projected population aged 18 years and over and enrolled persons by number of SA1s in projections delivered on 13 October 2023.



Change in projected population aged 18 years and over and enrolled persons by the number of SA1s in projections re-run on 10 January 2024.



Are there other tests that should be undertaken? The correlative analysis described in Appendix 1 readily identifies outliers (e.g. very low enrolment to population ratio) that should be investigated. Often, there will a ready explanation e.g. high international student population. Outliers are currently examined in the compilation process. These were followed up by Population Statistics Branch.

The key data inputs into the process are (i) the regional population estimates (down to the SA1 level), (ii) 2016-22 growth rates, and (iii) number of electors at the SA1 level at the date of the enrolment count (9 August 2023 for Victoria). In order to estimate, the projected estimates of the number of

electors, there are a number of rather complex compilations and adjustments (which could potentially be incorrect) that provide estimates of (iv) the projected population and (v) enrolment ratios at the SA1 level.

The recently revised set of output checks should detect if there are any issues with (i), (ii), (iii) or the compilations although it would require further investigation to determine the source of the error. Other checks that might be undertaken are:

- The distribution of enrolment ratios. These are currently used to identify outliers for further investigation or amendment but there would be merit in also comparing with past distributions. This analysis would have identified that the median enrolment ratio was too low before the forcing to add to higher level estimates increased them.
- The SA1 enrolment projections are forced to add up to State/Territory enrolment for the target date. There will be a discrepancy and the size of that discrepancy (before forcing) should provide some insight into the validity of the enrolment projections. If the discrepancy is too large, it is a signal that something may be wrong.

2.6 Face validity checks

Population Statistics Branch acknowledges that these should be undertaken. They propose two checks.

1. The Regional Population Unit should review the estimates prior to them being provided to AEC. This could be done at both the Electoral Division and SA2 levels.
2. There should be a presentation to the Electoral Division managers as part of the package provided to AEC. The Division Managers should have a good feel for population growth characteristics of their electorates.

Presentation like (2) were provided up until the pandemic in 2020.

I strongly support both these checks.

2.7 Monitoring and analysing feedback from informed users

A lot of people look closely at electoral redistribution proposals and the supporting data. This includes the major political parties. Many of these are well informed users and any criticisms of electoral enrolments are worth analysing. These may come through the formal Suggestions submitted following the AEC invitation. In the Victorian Distribution, the first Suggestion raising doubts about the accuracy of the enrolment numbers was Dr Mark Mulcair around 20 November. Other criticisms followed shortly afterwards, including from the Victorian Branch of the ALP. There was even earlier chatter on blogs such as Tally Room.

The ABS should monitor suggestions and encourage the AEC to contact the ABS if they hear of well-informed criticism.

3. Approaches for mitigating similar errors

The ABS asked me to identify approaches to mitigate similar errors occurring in the release of ABS statistics and to inform the ABS on potential areas of risk exposure in current processes. In the limited time available for the review, I can only make a few observations based on my past quality management experience at the ABS and Statistics New Zealand, as well as 10 years of annual quality reviews at Statistics Sweden. I have not talked to any other area at the ABS except Population Statistics Branch.

The amount of effort put into quality management depends both on the importance of the statistical outputs and the likelihood of risk. This is a judgement that needs to be made by statistical managers. For very important statistics like population projections used for electoral purposes and with a fairly high level of risk (or the CPI, for example), effective quality management arrangements are crucial.

It is also important to recognise that statistical areas should be designing a risk management rather than a risk avoidance process. The latter are expensive and, in practice, have been shown to rarely avoid all risks.

The ABS has a Quality Management Framework that uses quality gates (Quality Management of Statistical Processes Using Quality Gates, Dec 2010, catalogue 1540.0). This was developed in 2010 and was adapted from similar approaches used in some other official statistical agencies. It may need updating in light of experience since then. I am not advocating an alternative approach. However, the existing system does need to be used and used consistently. It may well be time to,

- update the ABS Quality Management Framework,
- prepare some additional supporting resource materials,
- promote the revised framework, and
- provide some experts to support statistical areas revising their quality management frameworks.

It may well be time to assign responsibility for leadership of quality management improvement across the ABS. In most official statistical agencies, Methodology is assigned this responsibility.

When designing/reviewing the quality management arrangements, the first step is to document the process and understand the risks at each step of the process. The ABS uses the Generic Statistical Business Process Model (GBSPM)⁸ as the standard for documenting processes. I make some comments below on those risks which might require more careful consideration.

After documentation of the process, the quality gates should be determined including their placement, design and follow-up actions including the sign-off process.

What are the areas of higher risk? For 10 years from 2010, I undertook an annual quality review of some of Statistics Sweden's most important outputs. Like the ABS, Statistics Sweden is a well-developed statistical office with an excellent reputation. I also have had a lot of exposure to past ABS and Statistics New Zealand quality issues (fortunately, not too many). An area of major risk in all these Offices was user computing largely because the extent of testing is not as great as with larger systems developed using IT professionals. There should be standards for where user computing is deployed to produce publishable statistics. I recommend that standards should be produced for systems developed using user-oriented tools. These would have a lighter touch but would include requirements to test the accuracy of the system (e.g. parallel running or specially designed test packs) and provide sufficient documentation to enable someone else to use the system.

Source data is another area of higher risk. I discuss administrative data separately below. For data collected by the ABS, non-response is the greatest risk. This is increasing even though the ABS is doing better with response rates than most of its peer organisations in other countries. Whilst it remains important to minimise non-response, with higher levels of non-response, there is a greater risk of non-response bias because of the resulting sample being unrepresentative. Special efforts

⁸ The GSBPM is an international standard for describing and defining the set of business processes needed to produce official statistics.

need to be devoted to ensure the sample is appropriately balanced (a priority for non-response follow-up) and weighting systems exist to mitigate residual non-response bias.

Coverage can also be a concern for both surveys of households and businesses. For household surveys it can often be 'hidden' such as the case with the lack of representation of young adults in many surveys. Education (or a suitable proxy) is becoming an increasingly important weighting variable. For many household surveys, the better educated are over-represented and this could be an issue if the survey variables in question are related to education (e.g. literacy surveys). For business surveys, there is a need to understand the impact of under-coverage (e.g. new businesses not yet on the register) and over-coverage (e.g. duplications, defunct units). Any lack of coherence between data sources can be an issue for compilations such as the national accounts. In the past, this has often been due to source data areas using different approaches to adjusting quality concerns with non-response and coverage.

Administrative data is not collected for statistical purposes, so it is not surprising that there are a number of areas of risk. Coverage is probably the biggest area of risk. Often administrative data systems will not cover all the population of interest by design. In addition to this under-coverage and that caused by members of the population of interest not being covered, there can be duplication and over-coverage (e.g. defunct units) because of shortcomings in the arrangements for maintaining the administrative systems. The impacts on the statistical outputs should be analysed and, if sufficiently important, mitigation steps should be devised. These may include surveys especially designed to adjust for the shortcomings in the administrative data.

Specification error can be another issue (i.e. the variable provided in the administrative data does not match the ideal statistical concept). Surveys may also be needed to provide the necessary adjustments although not necessarily for every time period.

Discontinuities are often prevalent in administrative data. For example, these may be due to policy or program changes which require adjustments to the administrative data that was collected is to support the program. It can be indirect as was the impact of the COVID pandemic on the Medicare data source especially when the vaccination program commenced. There may also be delays in supply of the data from the administering authority from time to time.

The main point I am trying to make in the preceding paragraphs is that there are numerous risks involved in the data sources for official statistics. These need to be understood, their potential impact analysed, potential mitigation determined and the need for a quality gate determined.

4. Findings

My main findings were as follows. My quantitative comments are based on Victoria as the error had more severe consequences there than in Western Australia.

1. The SAS program prepared to compile the regional enrolment projections was inadequately tested.
2. Consistent with good practice, a variety of output edits were undertaken before delivery to AEC. However, an important output edit, that would have clearly identified the presence of an error, was not undertaken. For Victoria, it showed that improbably large number of SA1s had a projected change in the electoral population of 8 to 10 per cent over the 5 years to 2028. Virtually no SA1s had a lower projected change, and none had a higher projected change. Population Statistics Branch included this additional edit in their revised electoral projections and will do so for future projections made for electoral purposes.

3. One of the output edits involved correlative analysis at the SA1 level between the projected population aged 18 years and over and enrolment projections. It was also undertaken at the SA2 and Electoral Division levels. This showed lower than usual correlations, but this was incorrectly attributed to the impact of COVID on the Medicare data which is the major data source used for estimating change in regional populations.
4. There were no face validity checks of outputs by someone familiar with sub-State population trends in the States. For example, I know the population trends in Victoria well enough to have identified that there was a problem. The Regional Population Unit within the ABS believe they have sufficient knowledge of population trends in the States and Territories to play this role in the future. Furthermore, Population Statistics Branch plans to recommence briefings (similar to those conducted up until 2020) of AEC Divisional Managers before they are publicly released.
5. A number of the Suggestions on the AEC website raised concerns with the Victorian enrolment projections, but these were not monitored by the ABS, nor did AEC raise concerns with the ABS until a week or so before the letter from the Australian Electoral Commissioner on 21 December. The AEC first published the enrolment projections on 25 October and Suggestions started flowing shortly afterwards, some of which raised concerns about the projections. (The concerns were mentioned in Antony Green's blog of 27 November). It is likely that the error could have been identified in mid-November.
6. Once the ABS became aware of the error, it responded reasonably quickly (10 working days for Victoria and 12 working days for Western Australia)
7. I believe human resource issues are a partial explanation. The resources available to Population Statistics Branch have remained stable although turnover has been high because of the increased demand for demography skills. Furthermore, the workload has increased significantly over this time. It is one of areas of the ABS where errors can cause significant damage to the ABS reputation.

5. Recommendations

5.1 Recommendations for regional enrolment projections for electoral purposes

The main quality gate is at the right place in the process for compiling electoral enrolment projections. However, the set of output edits performed at this gate should be extended to ensure they are more likely to identify possible errors. This was done for the revised estimates and confirmed that the error would have been found with the additional output edits.

Furthermore, Population Statistics Branch should introduce arrangements for reviewing the face validity of the estimates. This needs to be someone who is familiar with the population trends in the State or Territory. The Regional Population Unit believes it has this capability. Also, as proposed by Population Statistics Branch, briefings of AEC Divisional Managers should recommence prior to AEC making the enrolment projections public.

For future re-distributions, the ABS should monitor the Suggestions on the AEC website for criticisms of the enrolment projections. AEC should be encouraged to contact the ABS if they hear or observe criticisms.

Population Statistics Branch resources are stretched. Although resource levels have remained constant, staff turnover has increased, and the demand for statistical population statistics has also increased. It is a very important area of the ABS and there should be a review to ensure it has all the resources it requires.

5.2 General lessons

The ABS and other official statistical agencies rely a lot on user developed systems using user-oriented tools (e.g. SAS, spreadsheets). Systems Development Methodologies (SDM) are used for larger systems using IT professionals. Standards should also be produced for systems developed using user-oriented tools. These would have a lighter touch than a full SDM but would include requirements to test the accuracy of the system and provide sufficient documentation to enable someone else to use the system.

The ABS Quality Management arrangements involve a system of quality gates where intermediate and final outputs must be 'signed off' before they pass the next gate. This is a good arrangement especially if the quality gates are placed where they will be most cost effective. However, their importance needs to be reinforced, especially for non-regular statistical outputs, and there may need to be some updates the Quality Management Framework given it has been in place for nearly 15 years.

A face validity or 'sniff test' should be applied to all ABS outputs but especially those of high importance. I would include electoral projections for redistribution in that category. Where possible, this should be someone not closely involved in the compilation ('a fresh set of eyes'). Most often it will be a more senior officer involved in the publication sign-off process but there may be others who have sufficient knowledge of the subject matter to provide some insights.

In my experience, negative feedback from sophisticated users on accuracy issues should be taken seriously. They are right more often than not. This is certainly the case for electoral projections where several of the political parties and a number of psephologists study these numbers in great detail. (Even if their concerns are invalid, they should be provided with an explanation to maintain their confidence in the figures.)

6. Conclusion

Population Branch has or plans to make several adjustments to the quality management procedures for the processes to develop enrolment projections for electoral redistribution purposes. If they are successfully implemented, similar errors should not be repeated in the projections provided for future redistributions.

29 FEBRUARY 2024

Dennis Trewin AO, FASSA is the former Australian Statistician (2000-07), Australian Electoral Commissioner (2000-07), Deputy Government Statistician, New Zealand (1992-95) and President of the International Statistical Institute (2001-03) and the International Association of Survey Statisticians (1995-97).

Appendix 1: correlative analysis

The graphs below show the strength of the relationship between the projected enrolment counts and the projected total population. They are extracted from material provided by Population Statistics Branch. The closer the R-squared value is to 1, the stronger the relationship. If the roll count matched the population the R-squared value would be 1. However, we wouldn't expect that, as there will always be parts of the population who are not enrolled, and this will vary from SA1 to SA1. There is a noticeably weaker relationship for Victoria between enrolment projections and the projected population in comparison to other jurisdictions; the ABS understands this to be a combination of several factors, including the unique growth profile of Victoria, as well as the effects of sub-state data affected by the Covid-19 pandemic being carried forward over the projection horizon. There is also increasing error within population projections with increasing time due to a greater degree of uncertainty in component assumptions.

Figure 1: Corresponding roll count and projected population at 17 April 2028, by Statistical Area Level 1 (SA1) (before correction)

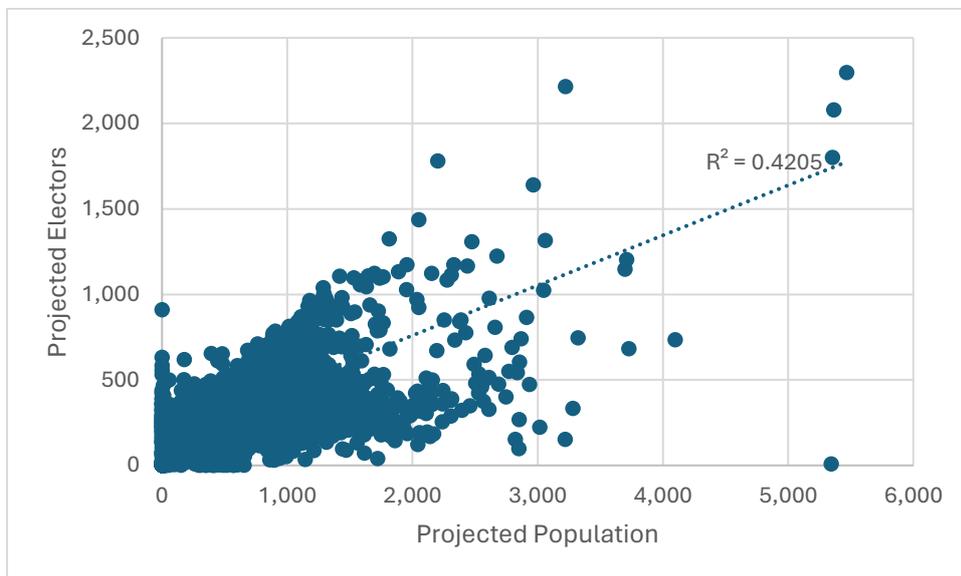


Figure 2: Corresponding roll count and projected population at 17 April 2028, by Statistical Area Level 1 (SA1) (corrected data)

