Health New Zealand
Te Whatu Ora

MINISTRY OF HEALTH
MANATŪ HAUORA

Briefing for information

The Sapere report and re-weighting primary care capitation funding

Date due to MO:	12 December 2024	**Action required by:**	N/A
Security level:	IN CONFIDENCE	**Reference:**	H2024057558
To:	Hon Dr Shane Reti, Minister of Health		
Consulted:	Health New Zealand: ☒		
Proactive release:	This **title** is proposed by the Ministry of Health for proactive release: ☐		

Contact for telephone discussion

Name	Position	Telephone
Maree Roberts	Deputy Director-General, Strategy, Policy and Legislation	s 9(2)(a)
Emma Prestidge	Group Manager, Primary, Family and Community Health Policy, Strategy, Policy and Legislation	s 9(2)(a)
Martin Hefford	Director, Living Well, Te Whatu Ora | Health New Zealand	s 9(2)(a)

Minister's office to complete:

☐ Noted ☐ Seen

☐ Needs change ☐ Withdrawn

☐ See Minister's Notes ☐ Overtaken by events

Comment:

Briefing for information

The Sapere report and re-weighting primary care capitation funding

Security level:	IN CONFIDENCE	**Date:**	12 December 2024
To:	Hon Dr Shane Reti, Minister of Health		

Purpose of report

1. This briefing:
   a. outlines the key findings of the Sapere analysis for updating capitation funding in primary care and provides you with the latest draft of the report, *Rebalancing the scales: Revising capitation weights for primary care funding*, and

   b. identifies further work required to inform detailed advice on re-weighting capitation.

Summary

2. The current capitation model was put in place in the early 2000's. In 2024, Health New Zealand (Health NZ) commissioned Sapere to analyse primary care data to support advice on how the funding formula could better reflect health needs.

3. The draft of the Sapere report is attached for your information. The report is still subject to some technical review and change.

4. The work undertaken is a technical analysis of variables needed to improve the effectiveness of the capitation formula. It has not considered fundamental changes to the overall funding model for primary care.

5. While the Sapere analysis is a key input into potential changes to capitation, further work is needed to inform this, including additional analysis, funding required, and the implications for some existing primary care funding streams.

6. The Ministry and Health NZ will provide you with the final Sapere report and provide you with further advice on options for re-weighting the capitation formula in early 2025. This advice will also be reflected in the March 2025 Cabinet report-back as part of the Strategic Plan for Primary Care.

Recommendations

We recommend you:

a)	**Note** this briefing and the Sapere report *Rebalancing the scales: Revising capitation weights for primary care funding*	**Yes/No**
b)	**Note** that the review of the capitation model has identified that, in addition to the existing age and sex variables, deprivation, multimorbidity, rurality, and ethnicity are all statistically significant predictors of primary care	**Yes/No**

activity and costs.

c) **Note** that further work is required to develop options for re-weighting capitation and timeframes for implementation **Yes/No**

Maree Roberts
Deputy Director-General

Strategy, Policy, and Legislation
Ministry of Health
Date: 12/12/2024

Debbie Holdsworth
Co-Director of Funding for Community and Mental Health

Planning, Funding and Outcomes
Te Whatu Ora | Health New Zealand
Date: 12 / 12 / 2024

Hon Dr Shane Reti
Minister of Health

Date:

The Sapere report and re-weighting primary care capitation funding

Background

7. Capitation funding for primary care was introduced in 2002. The funding formula was developed based on use patterns at the time and has not been fundamentally changed in the interim despite changes in both use of care and life expectancy.

8. The capitation formula is particularly limited in that it only takes age and sex variables into account when determining the level of capitation funding. This is an overly simplistic model that results in inadequate funding of care for some individuals and can cause financial and workforce pressures in areas of need.

9. Various ad hoc changes have been introduced to compensate for weaknesses in the existing capitation formula. These changes have complicated the funding model and created some perverse outcomes without resolving the core issues.

10. Health NZ engaged Sapere to analyse primary care data and provide updated weightings as key inputs for strengthening the capitation formula.

11. The Sapere analysis in 2024 builds on the methodology used in the review completed in 2022. It was supported by a Technical Advisory Group (TAG) comprised of sector representatives. The role of the TAG was to provide expertise from a range of sector perspectives including but not limited to general practice, urgent care, nursing-Led care, rural health, Māori health, Pacific health, PHO leadership and management, practice management and academia. The TAG helped ensure that proposals for changing the formula reflect real world operational considerations.

The capitation formula can be better weighted to reflect health needs…

12. The hypothesis which Sapere tested was that the accuracy of the capitation weights could be improved by expanding the formula to include factors beyond age and sex (the current capitation weighting criteria). Factors which they tested (based on existing research) were rurality, socio-economic deprivation, ethnicity, and multi-morbidity.

13. The methodology which Sapere used included:

    a. Sampling primary care appointment data for over 2 million patients from across 18 Primary Health Organisations (PHOs) for the year 2023.

    b. Determining the annual clinical usage of GPs, nurses, and nurse practitioners for each patient.

    c. Assigning costs to the identified FTEs based on collective agreements to calculate the total annual cost per patient.

d. Using multivariate regression analysis to determine the relevance and impact of various chosen characteristics as predictors on health care costs.

14. While the report is still subject to a final peer review, Sapere's findings suggest that the predictors align with expectations. For example, individuals from more deprived areas, those with more comorbidities, and older individuals appear to have increased healthcare use and costs. While age and sex continue to be the strongest predictors of primary care activity, other factors like the P3 score, ethnicity, and Geographical Classification of Health (GCH) also explain significant variation in primary care costs.

15. As such, the findings of the review likely supports the concept that primary care funding could be more accurately distributed by reallocating capitation funding on the basis of a set of refreshed factors and weights.

16. The draft Sapere report, *Rebalancing the scales: Revising capitation weights for primary care funding*, is attached.

17. Transitioning to a formula that recognises the factors identified in the analysis would result in a greater proportion of funding being directed to practices in rural areas, in more deprived areas, with more Māori or Pacific patients, older patients, or with more clinically multi-morbid patients.[1] This is likely to encourage greater supply of primary care in these areas and address some access issues.

Things to consider in relation to re-weighting capitation…

18. More work is required to explore the implications for capitation following the Sapere analysis and provide options on funding any changes to the formula. Some of the issues which need further consideration include:

i. s 9(2)(f)(iv)

[1] Note that the proposal to incorporate ethnicity as part of the overall formula does not breach the direction set out in Cabinet Circular CO(24)5 on 'needs-based service provision' as capitation is not a targeted service (it is universally available with no ethnicity-based eligibility restrictions), and there is robust analysis underpinning the inclusion of ethnicity as a minor part of the overall capitation formula.

ii. s 9(2)(f)(iv)

Next steps

19. Health NZ will continue to work with Sapere to have the report finalised.

20. The Ministry and Health NZ will work together to develop options for implementing revised weights considering the identified variables within existing funding baselines. This includes incrementally implementing this over time as funding allows. You will be kept appraised of proposed developments both directly and through the Strategic Plan for Primary Care.

ENDS.

Rebalancing the scales: Revising capitation weights for primary care funding

David Moore, Douglas Yee, Hamish Hann,
Dr Julius Ohrnberger, Leané de Beer

24 December 2024

Contents

Statement from the Technical Advisory Group v

Executive summary vi

1. Reassessing weights in New Zealand's capitation formula 1
   - 1.1 Patient groups were defined to reflect the changing demographics of New Zealand 2
   - 1.2 We used multivariate regression analysis to identify relevant measures of activity 3
   - 1.3 Sensitivity testing for model validity 4
2. Patients and the activity they generate 6
   - 2.1 The enrolled population 6
   - 2.2 A representative sample 7
   - 2.3 Patient consultation data 10
   - 2.4 A PHO's data was excluded 11
   - 2.5 Calculating FTE and costs 12
   - 2.6 Nurse activity was underrepresented 12
   - 2.7 GP clinical non-contact time was under-recorded 13
3. Including multimorbidity 16
   - 3.1 The P3 index is based on pharmaceutical dispensing 16
   - 3.2 The M3 is based on hospital admissions 17
   - 3.3 Why multimorbidity is important 17
   - 3.4 Comparing the two measures 20
   - 3.5 Imputing missing P3 scores 22
4. Results 24
   - 4.1 Regression results 24
   - 4.2 Cohort weights were based on the mean of patient cost 31
   - 4.3 Estimating clinician costs from regression results 38
5. Implications for practice funding 40
   - 5.1 Estimating effects on practice revenue 40
   - 5.2 Redistributing capitation using the current formula and revised weights 46
   - 5.3 Practice level effects on revenue 47
   - 5.4 A simplified approach to applying cost weights 49
6. Incorporating supply-side effects 55
   - 6.1 Including supply-side effects can allocate resources more efficiently, but reinforce inequities in utilisation 55
   - 6.2 Using practice-level fixed effects to address time-invariant confounding of supply-side factors 56
   - 6.3 Fixed effects results 57

References 64

Appendices

Appendix A Method for calculating GP, nurse, and nurse practitioner FTE 68
Appendix B Creating variate groupings 78
Appendix C R-script for analysis 95

Tables

Table 1: Population cohort variables 3
Table 2: Tested models and specifications....4
Table 3: Validity tests of model specifications....5
Table 4: Summary statistics of the population of interest 11
Table 5: Ratio of GP non-contact to contact time by PHO....14
Table 6: Effects of predicted cost variables by model specification....25
Table 7: Regression estimates from specified models with cost as the dependent variable....25
Table 8: Predicted capitation weights, assuming deprivation quintile 1, GCH R2, and P3 score < 0.2 ...34
Table 9: Consultation statistics by Geographical Classification of Health (GCH)....38
Table 10: First contact capitation revenue changes to practices from applying the revised weights to different proportions of the total funding available for capitation....47
Table 11: Population statistics of six practices with different population characteristics used for comparison 47
Table 12: Regression results from the AGIM model....51
Table 13: Regression results from fixed effects models with cost as the dependent variable 60
Table 14: Description of the four template types....69
Table 15: Conditions defining template type....70
Table 16: Conditions to define provider role 71
Table 17: Conditions to exclude templates used to manage after hours and accident and medicine services 72
Table 18: Base table used in the FTE analysis....73
Table 19: Example of a duplicate appointment record for an extended appointment that would be excluded....74
Table 20: Conditions used to flag patient contacts 75
Table 21: Summary statistics of the population of interest prior to grouping and including inactive patients....78
Table 22: Regression estimates from specified models with cost as the dependent variable 80
Table 23: Predicted capitation weights, assuming deprivation quintile 1, GCH R1, and P3 score < 0.2.93

Figures

Figure 1: Variables used to define population cohorts 2
Figure 2: Sample selection and exclusion process 7
Figure 3: Distribution of enrolled patients in the sample population of the practices by five-year age band....8
Figure 4: Distribution of enrolled patients in the sample population by gender across age bands....9

Figure 5: Distribution of enrolled patients in the sample population compared to the total enrolled population by prioritised ethnicity 9
Figure 6: Distribution of enrolled patients in the sample population by deprivation 10
Figure 7: Calculated vs. actual practice-level nurse FTE before adjustments 13
Figure 8: Predicted vs. actual practice-level nurse FTE after adjustments 13
Figure 9: Mean non-contact to contact time ratio for GPs by total number of hours worked 14
Figure 10: Mean M3 score by age and gender 18
Figure 11: Mean P3 score by age and gender 18
Figure 12: M3 and P3 scores by ethnicity 19
Figure 13: M3 and P3 scores by deprivation quintile 19
Figure 14: M3 and P3 scores by geography 20
Figure 15: Scatterplot of M3 scores and FTE usage 21
Figure 16: Mean FTE by M3 score 21
Figure 17: Scatterplot of P3 scores and FTE usage 21
Figure 18: Mean FTE by P3 score 21
Figure 19: Distribution of population with missing P3 score vs. total enrolled population by age band 22
Figure 20: Distribution of population with missing P3 score vs. total enrolled population by ethnicity 23
Figure 21: Distribution of population with missing P3 score vs. total enrolled population by socio-economic deprivation 23
Figure 22: Mean weights by patient characteristics 33
Figure 23: Age–gender distribution of FTE per 1,000 patients across age bands 35
Figure 24: Predicted additional weight over deprivation quintile 1, attributed to socio-economic deprivation 35
Figure 25: Distribution of FTE by socio-economic deprivation per 1,000 patients 36
Figure 26: Predicted additional weight over P3 score < 0.2, attributed to multimorbidity 37
Figure 27: FTE distribution by P3 score per 1,000 patients 37
Figure 28: Predicted additional weight over U1, U2, and R1, attributed to Geographical Classification of Health (GCH) 38
Figure 29: Distribution of first contact capitation revenue changes for general practices applying revised population cohort weights (1 per cent increments) 41
Figure 30: Proportion of a practice's enrolled population over 65 compared to per cent change in capitation revenue applying revised weights 42
Figure 31: Proportion of a practice's enrolled population over 65 compared to per cent change in capitation revenue applying revised weights with ARC practices excluded 42
Figure 32: Proportion of a practice's enrolled population in deprivation quintile 5 compared to per cent change in capitation revenue applying revised weights 43
Figure 33: Proportion of Māori/Pacific peoples in a practice's enrolled population compared to per cent change in capitation revenue applying revised weights 44
Figure 34: Proportion of practice population under five years compared to the proportion of Māori/Pacific peoples 44
Figure 35: Proportion of practice population over 65 years compared to the proportion of Māori/Pacific peoples 45

Figure 36: Proportion of a practice's enrolled population who have an R3 GCH compared to the per cent change in capitation revenue applying revised weights 45
Figure 37: Density graph showing the per cent change in first contact capitation revenue if 20%, 50%, 80%, and 100% of capitation revenue were determined by the revised weights 46
Figure 38: Per cent change in practice revenue from first contact capitation 48
Figure 39: Mean weights by characteristic (fixed-effects model)....58
Figure 40: Predicted additional weight over R1, attributed to Geographical Classification of Health (GCH) 94

Statement from the Technical Advisory Group

The analysis described in this report derives a series of relative weights for the funding of general practice care in proportion to estimated workload. This work is based upon data for typical current primary health care teams, covering general practitioner, nurse practitioner, and practice nurse roles. The analysis is based upon a large and diverse dataset, covering over half the population of New Zealand, and health services ranging from mainstream general practice to hauora Māori providers. The process to derive these weightings has been monitored throughout by our group in order to ensure to the best of our ability that it is open and robust, and that it addresses any questions we have raised during the development of the new weightings. We are satisfied that the development of these weightings has been completed in a rigorous and considered manner.

It is important to understand what these results describe, and what they do not. They are relative weightings for workload, reflecting the varying clinical cost of delivering general practice services to different demographic groups. The weightings reflect the relative clinical workload and labour cost of delivering general practice services as they are typically provided to people in New Zealand in 2023. These weightings do not reflect either (a) future developments in the configuration of the primary care workforce, or (b) existing levels of unmet need for care that are not addressed by primary care services as they operate today.

These weightings are therefore an important component that will contribute to decisions about the future funding of general practice services, but will necessarily have to be implemented in ways that reflect policy priorities and projected future developments in the way that care is provided. They provide an empirical starting point from which to allocate resources in a manner that better matches actual service workload than the current capitation formula. In practice these may require modification to reflect policy priorities, incentives, changed workforces or other aspects of primary care where change is expected.

We are sensitive to the fact that future funding decisions must take into account not only the mechanism for distributing funds as indicated by the new weightings, but also the total quantum of funding that is necessary to ensure accessible and sustainable primary and community healthcare services. As the new weightings result in a redistribution of funds compared to the status quo funding distributions, they would cause a reduction of funding to most general practices and thus disruption to some services if they are implemented without full consideration of the level of funding required to sustain the sector.

Our advisory group has canvassed the issue of unmet needs on several occasions throughout the duration of this work. While there is not a straightforward answer to the challenge of estimating levels of unmet need, which is inherently difficult to observe in activity data, we strongly encourage Health New Zealand to monitor measures that may indicate unmet need for primary health care (such as low immunisation coverage, stated lack of access, ED presentations, absence of open service enrolment and population health outcomes). Analysis of such measures should aim to form the basis for incorporating unmet need factors into future resource allocation mechanisms, including capitation.

Executive summary

Health New Zealand | Te Whatu Ora (Health NZ) commissioned us to revisit the capitation weight calculation for primary care funding allocation in New Zealand. This work builds on and extends the methodology used in our previous capitation review for Health NZ in 2022.

Purpose of the review

The current capitation calculation method is based on age and gender, which may not adequately capture the variation in population health needs. We tested a model that includes only age and gender and found that this method explains less variation compared to the inclusion of more variables. Research indicates that other factors such as rurality, socio-economic deprivation, ethnicity, and multimorbidity also significantly impact health needs. This review aims to improve the accuracy of capitation weight calculation by expanding the formula to include these additional dimensions, thereby potentially enhancing the efficiency in allocation of resources according to the observed needs of the enrolled practice population. We took an activity-based approach for the capitation weight calculation.

Data used in the review

Our estimation samples consisted of primary care appointment data for 2,481,605 patients from 469 practices across 18 primary health organisations (PHOs) for the year 2023. The sample included enrolled patients from practices who provided data to us, representing more than half of New Zealand's enrolled practice population. The sample is representative of the overall enrolled population, ensuring the external validity of our findings. The large sample size provides sufficient statistical power and flexibility to perform detailed multivariate and interaction analyses across demographic, health, and socio-economic groups.

Methodological approach

We approached the analysis in five steps.

1. **FTE calculation**

   We determined each patient's annual clinical full-time equivalents (FTE) for GPs, nurses, and nurse practitioners. These FTEs were based solely on patient-facing clinical time and non-patient-facing clinical tasks, as derived from our appointments dataset. To enhance accuracy, nurse FTEs were adjusted using data from a practice managers' survey, while GP non-contact clinical FTEs were refined using results from a representative diary study by the Royal New Zealand College of General Practitioners (RNZCGP). Non-clinical activities, such as administrative duties or overheads, were not included in these FTE calculations.

2. **Costing**

   We assigned costs to the identified nurse, nurse practitioner, and GP FTEs based on collective agreements. These costs were then aggregated to calculate the total annual cost per patient.

3. **Multivariate regression analysis**

   We used multivariate regression analysis to determine the relevance and impact of various predictors on health care costs. The total annual patient cost is the outcome variable, with the individual patient as the unit of analysis. Predictors include age, gender, rurality (measured by the Geographical Classification of Health), socio-economic deprivation (measured by NZ deprivation quintiles), ethnicity (European, Asian, Māori, Pacific peoples, and Other), and multimorbidity (measured by the P3 score). These variables were identified as relevant determinants of health needs through a rapid literature review. We tested four different model specifications and selected the one with the best goodness of fit—i.e. the model that explains the most variation in costs. The chosen model incorporates interaction terms for age and gender, age and ethnicity, and rurality and multimorbidity, while avoiding additional interaction terms to prevent overfitting.

4. **Prediction**

   Using the findings from the multivariate regression analysis, we predicted costs for each of the 2,160 permutations of characteristics.

5. **Capitation weights**

   We calculated relative capitation weights using the predicted costs. This involved creating a sample of hypothetical patients, each with a unique combination of characteristics, including age, gender, ethnicity, socio-economic deprivation, multimorbidity, and geography. The patient with the lowest predicted cost was set as the baseline, with their predicted cost serving as the denominator to calculate the weights for other patients. We then identified and grouped patients with similar predicted weights based on their characteristics.

Findings

Our multivariate regression results showed that the predictors align with expectations. For example, individuals from more socio-economically deprived areas, those with more comorbidities, and older individuals exhibited increased healthcare use and costs. Age and gender were the strongest predictors of primary care activity in terms of coefficient size, while other factors like the P3 score, ethnicity, and Geographical Classification of Health (GCH) also explained significant variation in primary care costs. People in isolated rural areas (rural two and three) had notably higher predicted costs compared to urban and rural (one) residents, primarily due to increased clinical non-contact time. Although the impact of socio-economic deprivation on FTE costs was less pronounced when considering multimorbidity and rurality, it remained a crucial factor, capturing additional aspects like the ability to invest in health.

Our predicted weights indicated that higher costs (above the sample average) were closely linked with high health needs among:

- Māori, Pacific peoples and people of Other ethnicities
- infants and the elderly
- those experiencing socio-economic deprivation
- those with more comorbidities
- those in isolated rural residency.

Policy implications

We examined the effects of reallocating capitation funding using the new weights. If 100 per cent of the first contact revenue was distributed according to our revised weights, practices with a higher proportion of patients with high needs would receive increased funding. However, if services aimed at improving access for high-need groups (SIA funding) were also redistributed based on these weights, practices with a high proportion of Māori/Pacific peoples and more socio-economically deprived populations might experience a decrease in capitation funding. This is because the SIA funding pool would be reallocated among the entire population according to the new relative weights.

A simplified approach to applying weights

We developed a model that incorporates age-gender interactions, with additional additive adjusters to represent socio-economic deprivation, multimorbidity and rurality. We find this model only explains slightly less variation than our preferred model and is therefore a good substitute for its predictive ability. We can use a simplified model in the form of 24 age-sex cohorts, and 13 additive adjusters representing other patient characteristics.

Incorporating supply-side effects

We conducted an exploratory analysis whereby we utilised practice-level fixed effects to address time-invariant confounding of supply-side factors. Our statistical tests showed evidence in favour of this approach over alternative approaches, suggesting this approach may be more efficient in the context of allocating capitation funding according to activity. However, the resulting estimated coefficients from this approach showed a reduced effect arising from a patient being Māori, Pacific or of Other ethnicity relative to prior approaches, as well as if a patient was from a more rural area. This finding was consistent with the literature which shows that incorporating supply-side effects may reinforce existing healthcare inequalities amongst groups who historically underutilise primary healthcare.

Next steps

In our previous work, *A Future Capitation Funding Approach* (2022), we estimated the funding needed to sustainably deliver care and address historical unmet needs. We considered that analysis as an initial step. This analysis is one step further. However, revising the population weights is only one of several crucial steps needed to develop an equitable and effective capitation funding formula that appropriately compensates practices for the care they provide to their patients. Other factors that could be considered include:

- decisions on how to implement calculated capitation weights
- where there is additional loading on the primary care system because of insufficient capacity in secondary care
- advanced roles in primary care, and where clinicians operate beyond their usual scope
- flexibility for team-based care
- payments for specific pathways, particularly where there is a shortfall in secondary care, such as treatment for eating disorders, gender dysphoria, schizophrenia, etc.
- extra support for practices managing patients with long-term conditions, such as auto-immune disorders where more intensive care is often required.

1. Reassessing weights in New Zealand's capitation formula

Sapere was asked to review the weights that might be used in a capitation formula, as one of the funding mechanisms of primary care. A nationally consistent capitation funding formula was introduced in 2002. This was implemented as a mechanism to replace a spectrum of needs-based formulae in use in some regions in New Zealand, while formulae based on organisations' historical fee-for-service were used in other parts of the country (Crampton et al., 2002). The capitation formula served four primary objectives (Crampton et al., 2002):

1. Reduce the inequitable distribution of primary care resource across the country, arising from the use of varying formulae.
2. Better target resources to populations with higher health needs.
3. Promote enhanced roles for nurses and other members of a primary health care team, which was disincentivised under a fee-for-service model.
4. Promote the continuity of care of populations by providing incentives around understanding the enrolled population and their health needs.

The capitation formula focuses primarily on age and sex, and their relationship with healthcare activity, to allocate aspects of primary care funding. Including other variables potentially better aligns the relationship between observed activity and, possibly, primary funding. For example, Martínez-Pérez et al. (2024) showed that the inclusion of a multimorbidity and complexity measure reduces the predictive power of age on healthcare expenditure. Other factors recognised as important but not explicitly included in our modelling were socio-economic determinants to health and health inequalities in New Zealand such as income, employment, education, housing, ethnicity, access to facilities, and social support (National Health Committee, 1998; Toi Te Ora Public Health, 2024).

We attempted to proxy these determining factors with observable general practice data to align capitation weights more closely with healthcare activity. The inclusion of more characteristics in a capitation formula captures a more robust set of characteristics that take up more general practice time.

The analysis largely built on the previous Sapere report, *A Future Capitation Funding Approach*, undertaken in 2021. Although broadly in line with the results of that work, we used a wider definition of activity, had a much larger dataset more representative of different kinds of general practice, and identified P3 as a better explanator of activity attributed to multimorbidity than M3. We also included the Geographical Classification of Health in our calculations, as this proved to be an important predictor of primary care activity.

1.1 Patient groups were defined to reflect the changing demographics of New Zealand

We organised our patient groups to give as refined a view as possible of primary care activity. Following are the population variables we worked with:

- The population of New Zealand is steadily ageing, yet we do not currently accurately estimate the differences in healthcare need for older populations (Parr-Brownlie et al., 2020). Our analysis regrouped the population into more granular age bands than the status quo.
- Children under five were split into more granular age bands due to the strong variation in healthcare use in the first years of life. The different approach to defining age bands was to capture differences in healthcare need across different age groups.
- The ethnic composition of the country is also changing. Migrants as a proportion of the population are increasing. There will also be significantly larger populations of Māori, Asians and Pacific peoples relative to Europeans in the future, compared to now (Statistics New Zealand, 2024).

While the current capitation system has a mechanism in place to improve access for high-need groups, we included factors such as ethnicity and socio-economic deprivation at a more granular level within our weight calculations. This intended to capture the differences in predicted cost that are directly attributable to these factors.

Other characteristics included in population cohorts were multimorbidity, which we captured using P3 scores, and the Geographical Classification of Health (GCH). The GCH is a geographic index that distinguishes an area into urban (U1 and U2 based on the population of the area), and rural (R1, R2 and R3 based on the distance from an urban centre). Figure 1 outlines the variables used in our analysis.

Figure 1: Variables used to define population cohorts

A hypothetical patient was defined for each unique combination of variables outlined in Table 1. This resulted in 2,160 permutations of characteristics covering the entire enrolled population, for which each permutation had an associated predicted cost and cost weight.

Table 1: Population cohort variables

Variable	Values
Ethnicity	European and Asian; Māori, Pacific peoples and Other ethnicities
Gender	Male; Female; Other
Age	In age bands: 0; 1; 2–4; 5–14; 15–24; 25–34; 35–49; 50–64; 65–69; 70–74; 75–79; 80+
Deprivation	Using quintiles from NZDep18: 1 (least deprived), 2, 3, 4, 5 (most deprived)
Multimorbidity	Grouped using P3 Score terciles: <0.2; 0.2–0.6; 0.6+
GCH	U1, U2 and R1; R2; R3

The rationale behind these groupings is discussed in Appendix B. Data is available at a more granular level and presented in section 2.2 to show its representativeness.

1.2 We used multivariate regression analysis to identify relevant measures of activity

Multivariate regression analysis allows us to consider multiple factors that might influence primary care activity simultaneously. This allows for a more robust and comprehensive understanding of the complex relationships driving primary care activity.

We used a simple but effective model to capture differences in primary care activity generated across characteristics. To predict costs associated with delivering primary care resource, we employed the following regression model:

$$Cost_i = \beta_1 + \beta_2 X_i + \epsilon_i$$

Where $Cost_i$ is the cost associated with an individual patient i determined by the providers they used in 2023. β_1 is an intercept term, and X_i is a vector of individual patient i's personal characteristics:

$$\mathrm{X_i} = \mathrm{AgeBand_i} + \mathrm{Gender_i} + \mathrm{Ethnicity_i} + \mathrm{Multimorbidity_i} + \mathrm{DeprivationQuintile_i} + +\mathrm{GCH_i} + \mathrm{AgeBand_i} \times \mathrm{Gender_i} + \mathrm{AgeBand_i} \times \mathrm{Ethnicity_i}$$

β_2 is the coefficient associated with each of the variables, and ϵ_i denotes the error term.

We accounted for age and gender, and age and ethnicity interactions due to the complex relationship these variables might have with each other. Similar interactions were used in other capitation calculations, for example by NHS England (Anselmi et al., 2022).

We motivated the interactions based on our observed distributions of FTE for gender, age groups, and ethnicity. For instance, we expect females to have different outcomes than males from their teen years through to their child-bearing years. The impact of age on health outcomes might also vary based on ethnicity. For example, certain health risks may increase with age more sharply in one ethnic group compared to another.

We calculated how much of a clinician's time is spent with an individual patient based on appointment data, and then aggregated patients by age group, gender, ethnicity, socio-economic deprivation, multimorbidity, and rurality. This shows how much care is required by the average person within a demographic cluster. Our underlying assumption was that time spent with a patient is a valid proxy for primary care resource use.

1.3 Sensitivity testing for model validity

To verify that our model specification is one that accurately captures primary care cost differences between groups, we tested several different model specifications. We compared the models' fit by comparing their adjusted R^2.

We further used the model specification separately for five primary health organisations (PHOs):
s 9(2)(ba)(i), s 9(2)(b)(ii)
and provide a large sample such that we were able to generate robust estimates for their enrolled populations. We generally expected the FTE to be similar for people of the same characteristics on a per capita basis.

We generated a random sample of 1,000 patients from the entire dataset and predicted the total FTE using each of the models.[1] The model with a higher adjusted R^2 and the least variation of predicted FTE across PHOs will reflect a better fit and therefore will be the model with which we proceed.[2] The models tested are presented in Table 2.

Table 2: Tested models and specifications

Model name	Specification ($X =$)
Age–gender model	AgeBand + Gender
Simple linear model (SLM)	AgeBand + Gender + Ethnicity + Deprivation + P3
Age–interactions model (AIM)	AgeBand + Gender + Ethnicity + AgeBand × Gender + AgeBand × Ethnicity + Deprivation + P3
AIM with rurality (AIM-R)	AgeBand + Gender + Ethnicity + AgeBand × Gender + AgeBand × Ethnicity + Deprivation + P3 + GCH

[1] We used FTE as the outcome variable rather than cost, as we note there are differences in the workforce across PHOs.

[2] Comparing AIC and BIC figures for the different models, presented in Table 7, also shows that adding more variables and interaction effects decreases the AIC and BIC estimates.

Model name	**Specification ($X =$)**
AIM with rurality + AgeBand × P3 (AIM-R-P3)	AgeBand + Gender + Ethnicity + AgeBand × Gender + AgeBand × Ethnicity + Deprivation + P3 + AgeBand × P3 + GCH

Table 3: Validity tests of model specifications

Model	**Adj R^2**	**Predicted FTE for a random sample of 1,000 patients**				
		s 9(2)(ba)(i), s 9(2)(b)(ii)				
Age – Gender	0.086	1.05	0.91	1.09	1.09	1.23
SLM	0.103	1.04	0.92	1.09	1.09	1.15
AIM	0.108	1.07	0.94	1.12	1.11	1.17
AIM-R	0.109	1.02	0.91	1.05	1.08	1.12
AIM-R-P3	0.116	0.99	0.90	1.02	1.06	1.10

Table 3 shows that the adjusted R^2 marginally increased with the introduction of new variables and new interactions. This suggests that the additional variables and interactions help to explain more of the variation in the model. We predicted weights with the AIM-R (age–interactions model with rurality) as it:

- includes all observable variables in our scope that contribute to explaining additional variation in the model
- is a relatively simple model compared to the AIM-R-P3 model
- captures similar interactions highlighted in other capitation calculations (for example, Anselmi et al., 2022).

The variation between the predicted FTE across the five specified PHOs was also relatively minimal compared to other model specifications. While the AIM-R-P3 model marginally outperformed the others based on adjusted R^2 and with minimising of predicted FTE variation between PHOs, it created very small sample sizes for some age band–P3 combinations. This could produce biased estimates and risks overfitting the data.

2. Patients and the activity they generate

We created a large dataset that covers approximately 47 per cent of the enrolled population. This dataset was based on data from practices' management systems (PMS) and the National Enrolment Service (NES). We proxied the clinician resource required by patients with GP, nurse, and nurse practitioner consultations. The sample population was representative, although with slight variations between the demographics of the sample population versus New Zealand.

Practice management systems under-record nurse time and we adjusted nurse time with data from a practice manager's survey.

2.1 The enrolled population

We conducted an analysis to determine how much clinician time each patient used based on appointment data. The clinician time was then converted to an equivalent clinician FTE. The rationale behind this analysis was to determine how much clinician resource was used to deliver care to different groups of patients based on their characteristics. The methodology used to calculate GP, nurse, and nurse practitioner FTE is outlined in Appendix A.

Patients who attended appointments or interacted with a GP, nurse, or nurse practitioner from the 469 practices in 2023 were our sample population. The enrolled population of these practices with one appointment or more in 2023 (at September 2023) was 2,470,252 patients, representing 49 per cent of the total GP-enrolled population of New Zealand.

There were an additional 158,187 patients who had at least one appointment but could not be matched to the enrolled population. This inability to match was due to data inconsistencies or because they were not enrolled. It was not possible to see when unenrolled patients may have visited multiple practices—since they had not been matched to an NHI number, we could not determine when someone who visited more than one practice was the same person.

We excluded "outliers" in the data. The following exclusions were made:

- patients with an unrealistically high number of consults throughout the year[3]
- provider time where there were no contact consults throughout the day
- consults with a provider that was not a GP, nurse, or nurse practitioner.

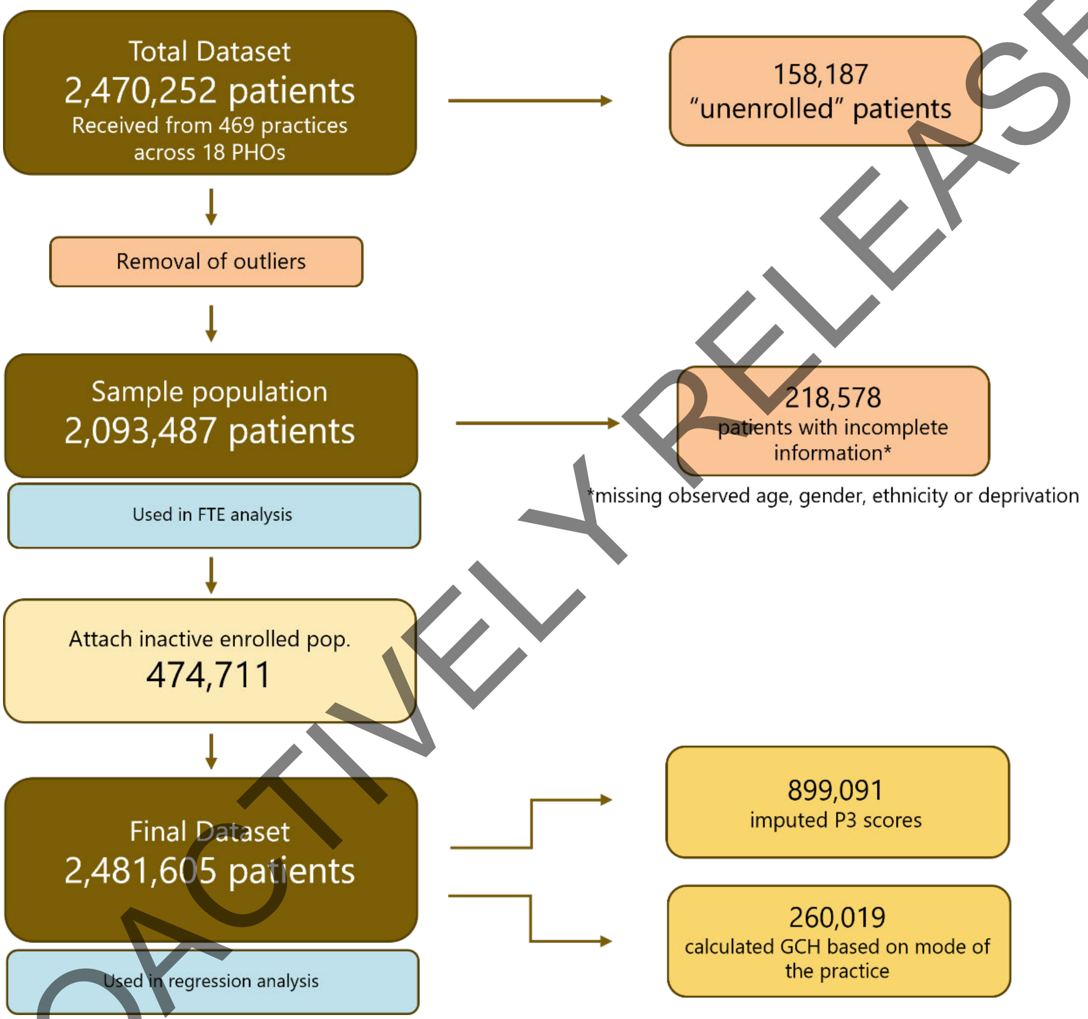
These exclusions removed 86,593 enrolled patients from our dataset, resulting in a final dataset of 2,093,487 enrolled patients. 218,578 patients with incomplete information were also removed for the calculation of cost weights; their exclusion was based on the absence of one or more data points but does not imply that these patients were missing all variables. We then attached the 474,711 inactive patients that would have been enrolled in the practices that provided data, but did not have an

[3] They were not excluded from the FTE analysis, but considered "dummy" patients, whereby providers used a designated patient ID to perform non-conduct tasks. These dummy patients were, however, removed from the regression analysis to predict cost weights.

appointment during the year. There were 2,481,605 patients eligible for the regression analysis used to predict cost weights.

We also attached a P3 score to each patient's encrypted NHI number, as well as a GCH index to the domicile of the patient's residence. To address the problem of missing data in these fields, we used multiple imputation to calculate the values of missing P3 scores and inferred missing GCH indices using the mode of the GCH index at the practice level.

Figure 2: Sample selection and exclusion process

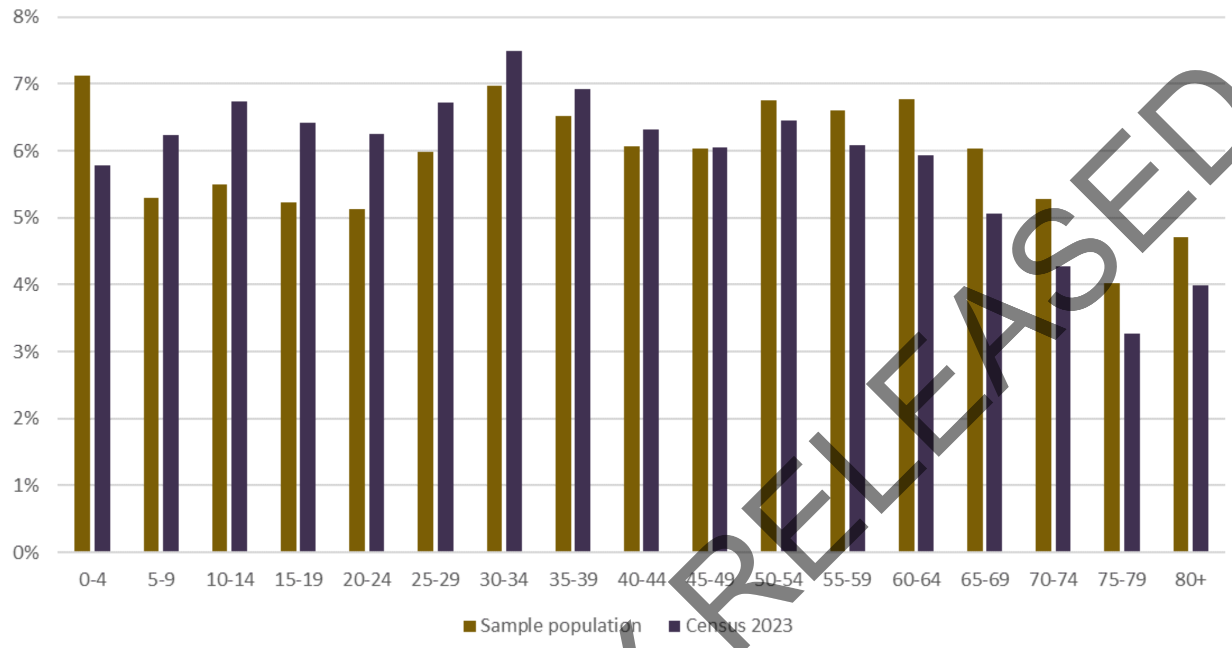


2.2 A representative sample

We relied on the enrolled sample population's representativeness of the population of New Zealand to make valid inferences on primary care activity and how this is driven by patients' characteristics.

We were confident that the sample data was representative when comparing our sample population to Census 2023 data and the total GP-enrolled population. Figure 3 shows the sample population was representative across age groups for New Zealand. In the sample, there was a relatively even distribution of patients across the age groups from zero to 69. The 30–34 age group represents the highest proportion of patients (7.2 per cent). The youngest age group, zero to four years, also had a significant proportion (6.9 per cent). There was a noticeable decline in patients from ages 70–74 onwards.

Figure 3: Distribution of enrolled patients in the sample population of the practices by five-year age band



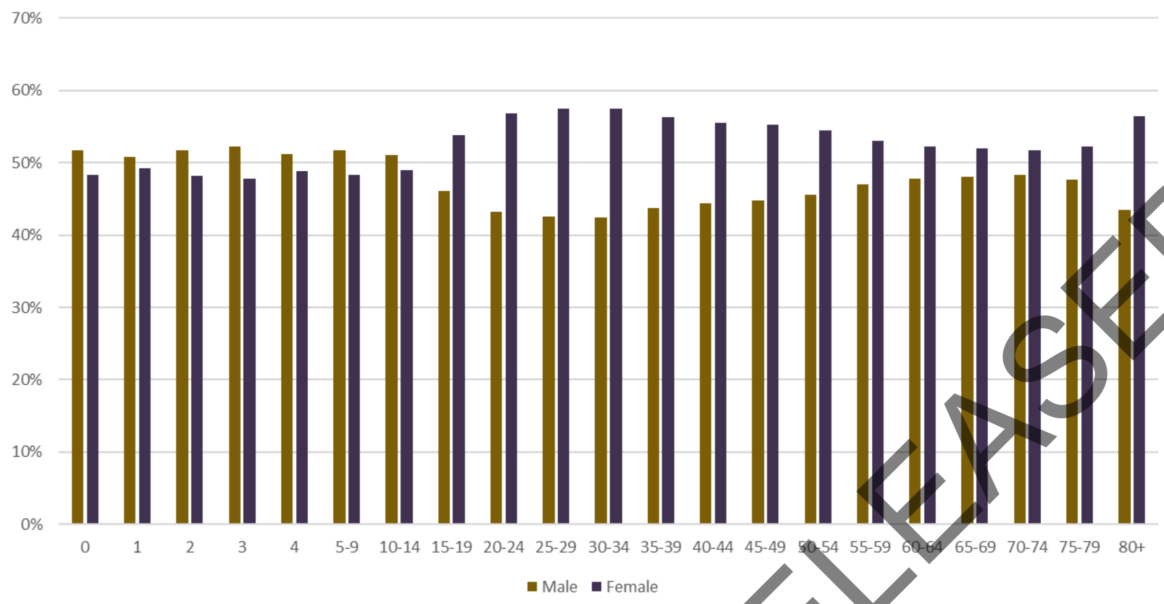
Source: Sapere analysis based on PMS/NES data, and data from Stats NZ retrieved from stats.govt.nz
Note: Census data did not disaggregate age groups zero to four into one-year age bands.

We made the following additional observations:

- Representation of females in the sample population was slightly higher (53.8 per cent) than in New Zealand's population (50.3 per cent) (Statistics New Zealand, 2024).
- Representation of males was slightly lower (46.1 per cent) than in New Zealand's population (49.7 per cent). This aligned with what we know about males generally being less likely to go to the doctor than females, reflecting broader trends in healthcare utilisation (Ministry of Health, 2023; Santosh & Crampton, 2009). The underrepresentation of males in the enrolled patient population could be attributed to factors such as societal norms and lower health awareness.
- The total sample population who identified with another gender besides male or female represented 0.04 per cent of the population of interest. They are underrepresented compared to the population proportion of 0.8 per cent according to Statistics New Zealand (2021). We speculated two possible reasons for this observation: (1) there may be relative underutilisation of primary care services amongst gender-diverse populations, or (2) people may be predisposed to identify with perceived conventional gender identities such as male and female within administrative processes.

Figure 4 reveals that females were consistently more represented than males from the 15–19 age band onwards. The disparity in representation was more prevalent in late teen, early adult years and in older age groups such as in the 80+ age group, which is unsurprising given the median life expectancy of females (84.0 years) compared to males (80.5 years).

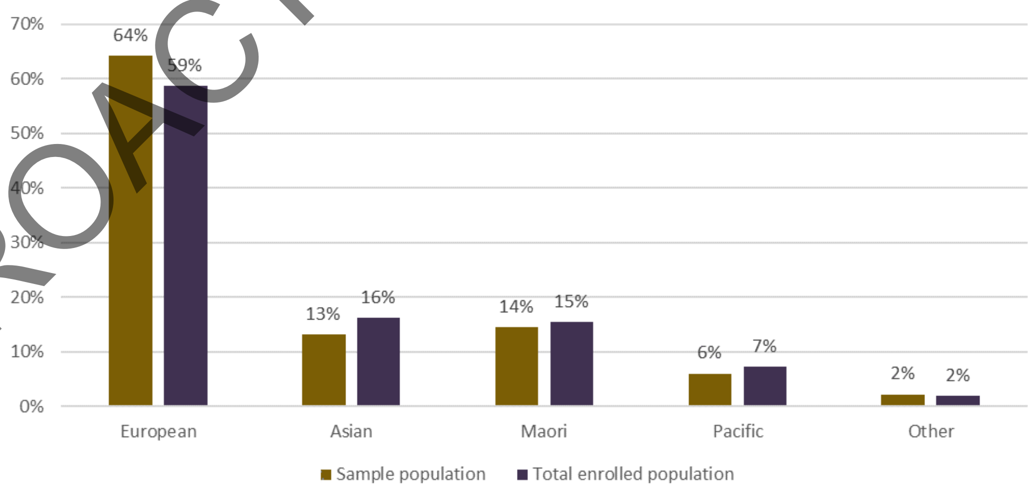
Figure 4: Distribution of enrolled patients in the sample population by gender across age bands



Source: Sapere analysis based on PMS/NES data

The enrolled sample population had slightly higher representation of Europeans compared to the total enrolled population. 64 per cent of the patients in our sample were European, Māori were 14 per cent, Asians 13 per cent and Pacific peoples were 6 per cent. According to the 2023 Census, 17.8 per cent of New Zealand's population identified as Māori, highlighting that Māori are underrepresented in the enrolled population. Asians, Māori, and Pacific peoples were slightly underrepresented in the sample dataset when compared to the total enrolled population. Figure 5 compares our sample with the total enrolled population.[4]

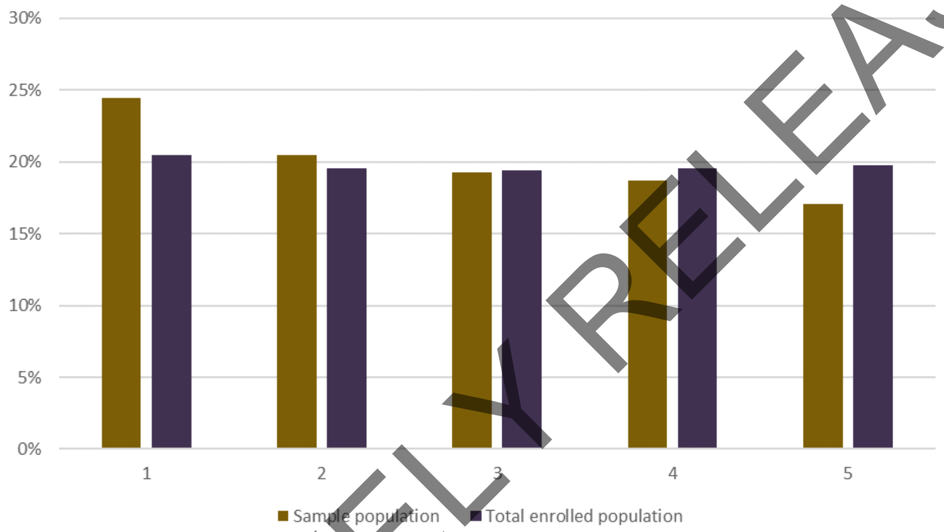
Figure 5: Distribution of enrolled patients in the sample population compared to the total enrolled population by prioritised ethnicity



[4] We did not compare this to census data as it no longer uses prioritised ethnicity.

The distribution of socio-economic deprivation was slightly unbalanced, with the proportion of the enrolled population decreasing with increasing socio-economic deprivation. However, each quintile made up close to 20 per cent of the sample respectively. This distribution suggests a well-rounded, but not a perfectly representative sample in terms of socio-economic status. Comparing the sample population with the total enrolled population showed that the direction of the distribution is the same for those in quintiles 1 to 3. The population in quintile 1 was slightly overrepresented, while populations in quintiles 2 to 5 are slightly underrepresented. Figure 6 depicts the distribution of the enrolled sample population by socio-economic deprivation quintile, with quintile 1 representing the least deprived and quintile 5 representing the most deprived.

Figure 6: Distribution of enrolled patients in the sample population by deprivation



2.3 Patient consultation data

We received the following data for 469 practices across 18 PHOs. The dataset collected was as follows:

- **Appointments**: appointment data at the patient level, with appointment date and time, physical consultation date and time, and the duration of the consultation.
- **Encounters**: notes that a clinician has written for a patient that can be matched to an appointment.
- **Facilities**: a list of practice names that can be matched to a practice ID.
- **Invoices**: invoice data for appointments, including the relevant funder for appointments.
- **Patient demographic data.**
- **Prescriptions**: prescriptions written by a clinician that can be matched to an appointment.
- **Providers**: a list of providers and their role that can be matched to an appointment.

Data on appointments, prescriptions and patient notes was limited to the 2023 calendar year. Provider and patient data were not time bound.

Table 4 shows summary statistics of the population of interest as used in the multivariate regression analyses, including the number of patients observed in each group, their mean predicted cost, and mean annual FTE utilisation.

Table 4: Summary statistics of the population of interest

Variable	Number of patients	Mean annual FTE	Mean predicted costs ($)
Age band			
0	36,218	0.0019	392.78
1	30,690	0.0015	313.47
2–4	91,151	0.0008	175.61
5–14	307,237	0.0005	101.83
15–24	291,736	0.0006	132.13
25–34	337,380	0.0007	147.66
35–49	471,156	0.0007	163.51
50–64	479,699	0.0009	206.42
65–69	135,588	0.0012	258.53
70–74	116,241	0.0013	292.17
75–79	86,248	0.0016	344.74
80+	98,261	0.0019	425.11
Gender			
Female	1,283,027	0.0010	212.67
Male	1,197,786	0.0008	169.44
Other	792	0.0015	337.14
NZDep2018			
1	578,789	0.0008	182.59
2	504,902	0.0008	185.44
3	477,061	0.0009	192.14
4	479,237	0.0009	194.35
5	441,616	0.0009	208.26
Ethnicity			
European and Asian	1,892,916	0.0009	191.58
Māori, Pacific and Other ethnicities	588,689	0.0009	192.68
GCH			
R2	142,025	0.0010	211.14
R3	26,013	0.0014	299.95
U1, U2 and R1	2,313,567	0.0009	189.44

Source: Sapere analysis based on PMS and NES data

2.4 A PHO's data was excluded

In the context of our study focusing on capitation of enrolled patients, a methodological decision was made to exclude one PHO, s 9(2)(ba)(i), s 9(2)(b)(ii) from our dataset. s 9(2)(ba)(i), s 9(2)(b)(ii) offers extended opening hours as well as 24-hour clinics and same day consultations. The exclusion of s 9(2)(ba)(i), s 9(2)(ba)(i), s 9(2)(b)(ii) data was based on several considerations related to the nature of the patient population and s 9(2)(ba)(i), s 9(2)(b)(ii) model of care:

- **Data completeness and consistency**: practices that use the s 9(2)(ba)(i), s 9(2)(b)(ii) s 9(2)(ba)(i), s 9(2)(b)(ii) do not collect appointment data. There is a queuing system whereby patients can identify a practice with the shortest queue and visit any one of several practices for a consult. Thus, we didn't have a 'duration' field, necessary for us to estimate consultation time.
- **Casual patients and enrolment status**: a high number of casual patients. Casual patients might be more likely to be enrolled in other practices, leading to their consults being captured incorrectly. Including such patients could potentially introduce inconsistencies in the data and skew results.

Many Asian and Pacific peoples were enrolled with s 9(2)(ba)(i), s 9(2)(b)(ii) and, therefore, this exclusion meant that these groups became underrepresented in our analysis. We did, however, still have significant numbers of Pacific and Asian patients throughout other PHOs.

2.5 Calculating FTE and costs

We started by taking each appointment and its allocated duration from the appointments data. Each row included a provider ID, which we matched to their clinical role (GP, nurse, or nurse practitioner). The duration of each appointment was then converted into full-time equivalents (FTE) based on the number of clinical working days per year for each role, as specified in their respective collective agreements. Clinical non-contact time, such as time spent on patient-related tasks outside of direct appointments, was accounted for and applied at the practice and role level.

We also used the collective agreements to determine salaries for each role. The FTE was then multiplied by salaries at the appointment level to determine the provider cost associated with each consultation.

The full details of the method used to calculate nurse, nurse practitioner and GP FTE is outlined in Appendix A.

2.6 Nurse activity was underrepresented

We adjusted our estimates to address the underrepresentation of nurse work in the data.

The data received from the 469 practices included appointments involving 9,861 clinicians. This was broken down as follows:

- 7,204 general practitioners
- 143 nurse practitioners
- 2,514 nurses.

We compared a sub-sample of our calculated FTEs to available practice manager survey data. We underestimated nurse FTE, with the disparity between calculated and actual FTE increasing as the size of the practice increased (Figure 7). We know nurse activity tends to be underestimated in PMS data, for example, multiple nurses may share a single nurse ID. We scaled up nurse FTE in our sample by reweighting nurse FTEs. The results from the scaling-up exercise are presented in Figure 8.

Figure 7: Calculated vs. actual practice-level nurse FTE before adjustments

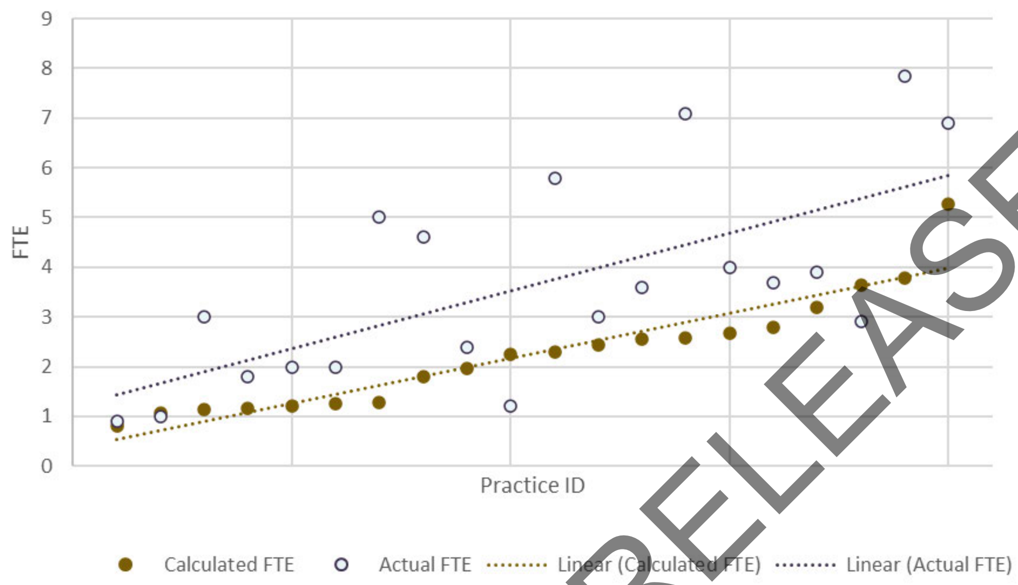
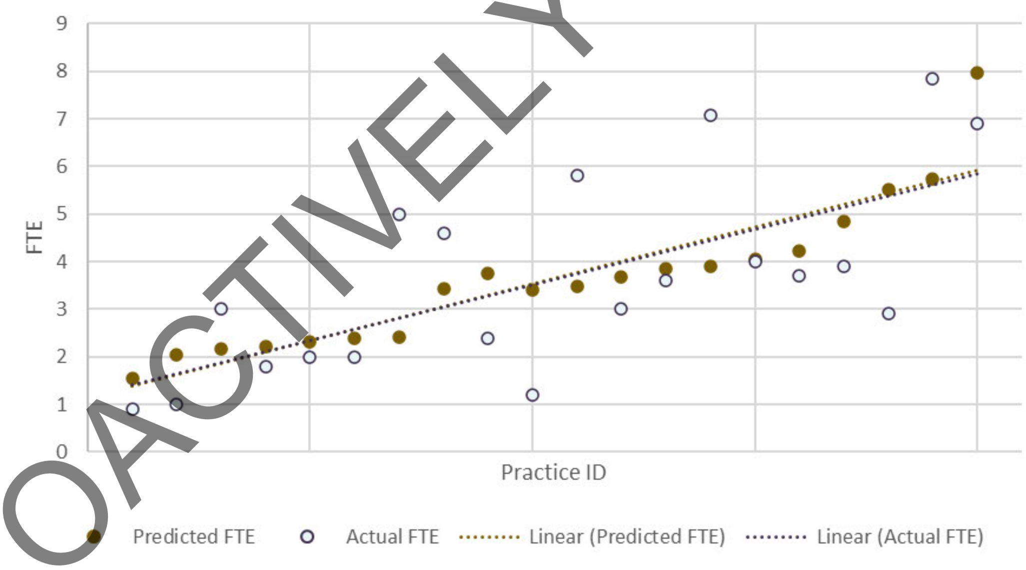


Figure 8: Predicted vs. actual practice-level nurse FTE after adjustments



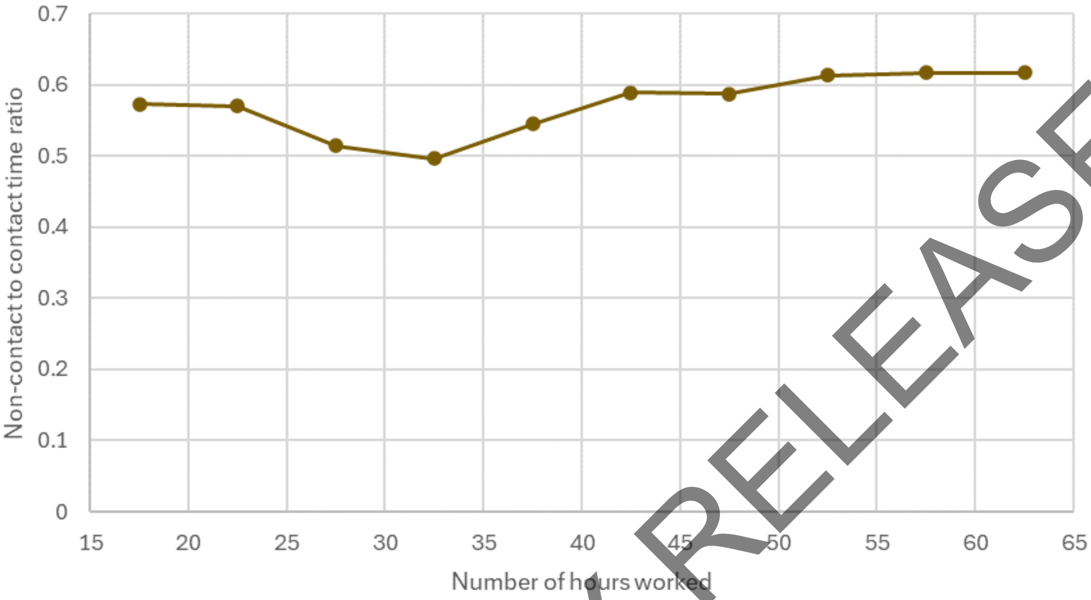
We did not adjust the nurse practitioner FTE due to lack of information about their workload in the practice survey.

2.7 GP clinical non-contact time was under-recorded

GPs appeared to be less inclined to record non-contact activity in the PMS. We understood that the non-contact time recorded by GPs is only a fraction of their actual non-contact time. To better inform our analysis of the amount of non-contact time spent with patients, we received data from the 2024 'Your Work Counts' diary study (see Royal New Zealand College of General Practitioners, 2024).

According to results from the study, the proportion of non-contact to contact time remains relatively consistent across the total number of hours a GP works. We calculated the weighted mean of the non-contact to contact time ratio to be 0.56.

Figure 9: Mean non-contact to contact time ratio for GPs by total number of hours worked



Source: Sapere analysis based on data provided by the RNZCGP.
Note: Sapere was provided data for survey respondents who worked 5–15 hours, and 65+ hours, but these were excluded due to the low number of responses. Data for total number of hours worked was provided in ranges; this analysis took the median value of the range.

We then calculated the same non-contact to contact time ratio for GPs recorded for our sample dataset and compared this result with the RNZCGP diary study.

We assumed a level of homogeneity in how GPs record clinical time within a PHO and so we calculated the mean non-contact to contact time ratio of GPs at the PHO level. The non-contact time recorded in practice management systems was typically underreported compared to the ratio of 0.56, as shown in Table 5. Therefore, we derived scaling factors for each PHO and then applied the scaling factor to inflate GP non-contact time so that it matched the RNZCGP ratio of 0.56, effectively implying 8.4 minutes of clinical non-contact time for every 15 minute consultation, which would be dispersed equally amongst all patients the GP consulted.

Table 5: Ratio of GP non-contact to contact time by PHO

s 9(2)(ba)(i), s 9(2)(b)(ii)

	Non-contact to contact time ratio	Scaling factor
	0.15	3.71
	0.12	4.81
	0.10	5.76
	0.19	2.90
	0.19	2.86

s 9(2)(ba)(i)

Non-contact to contact time ratio	Scaling factor
0.13	4.29
0.20	2.85
0.32	1.76
0.14	3.89
0.18	3.06
0.03	17.22
0.20	2.83
0.70	0.79
0.15	3.63
0.16	3.47
0.16	3.38
0.10	5.48
0.17	3.28

3. Including multimorbidity

We reviewed two multimorbidity indices to use in the revised capitation model—the P3 and the M3 index. The M3 index includes 61 conditions and is observed at secondary care. The P3 includes 30 conditions and is based on community pharmaceutical dispensing (Stanley & Sarfati, 2017). Both M3 and P3 measure various illnesses and conditions, which carry different weights based on their severity. The M3 and P3 indexes were then calculated as the sum of the various conditions' weights.

The P3 index was more relevant to primary care services because it is derived from pharmaceutical dispensing authorised by general practice and other services.

M3 and P3 are important measures of complexity; this section highlights which populations face higher levels of multimorbidity. To do this, we used data from practice management systems linked to NES data and M3 and P3 indices for the enrolled population where possible.

3.1 The P3 index is based on pharmaceutical dispensing

The pharmaceutical prescribing profile (P3) was developed and validated using pharmaceutical dispensing records. The index includes 30 medication categories, where the weight of each category is derived from coefficients of a Cox Proportional Hazards Model for mortality (Tiruye et al., 2024). The sum of the coefficients from each of the categories is reported as the P3 score.

The list of conditions used in the calculation of the P3 score is as follows (Stanley et al., 2020):

- Gastric acid disorder
- Cardiovascular diseases with medication from only one medication category (CVD1)
- CVD2
- CVD3
- Depression
- Reactive airway disease
- Anxiety and tension
- Steroid responsive conditions
- Diabetes
- Hypothyroidism
- Congestive heart failure
- Anaemias
- Angina
- Psychotic illness
- Epilepsy
- Anticoagulation
- Osteoporosis
- Migraine
- Arrythmias
- Malnutrition
- Rheumatoid arthritis
- Parkinson disease
- Transplant autoimmune disorder
- Dementia
- Hepatitis B/C
- Pancreatic insufficiency
- Tuberculosis
- Human immunodeficiency viruses
- Multiple sclerosis
- Pulmonary hypertension.

3.2 The M3 is based on hospital admissions

> **The weighted multimorbidity measure (M3 index)** was designed for comparing health outcomes at the population level (Stanley & Sarfati, 2017). The M3 index uses ICD-10 diagnostic codes from recorded public and private hospital admissions and the New Zealand Cancer Registry to identify and define a list of health conditions that were used to calculate the index.

Conditions were selected based on:

a) impact on quality/quantity of life
b) requiring complex healthcare management or coordination
c) being likely to last more than three months.

The resulting 61 health conditions (see Stanley & Sarfati, 2017, for a full list of conditions) were used to determine a binary indicator for each person. Component scores for the M3 index were then calculated as the beta coefficients for each condition using Cox proportional hazards regression for one-year mortality from the index date. The model provided an estimate of mortality risk for each health condition, controlling for age and the presence of the other health conditions. The M3 index was calculated as the sum of the beta coefficients that were above zero (55 conditions).

3.3 Why multimorbidity is important

Both M3 and P3 showed that the prevalence of multimorbidity increases with age, and men generally have higher scores of multimorbidity, particularly when over 70 (as illustrated in Figure 10). Males in their late teens and early adult years also tended to have higher M3 scores on average. However, Figure 11 suggests that females in the teen and early adult age bands had higher levels of multimorbidity than males, on average. Males and females in other age bands generally had similar P3 scores, except for males in the 95+ age band who had higher scores than females. While both measures showed the same relationship between multimorbidity and age, the shape and magnitude of the distribution varied slightly, with the gender differences in scores varying by the measure. This difference likely comes down to the conditions that are measured by each of the indices.

Figure 10: Mean M3 score by age and gender

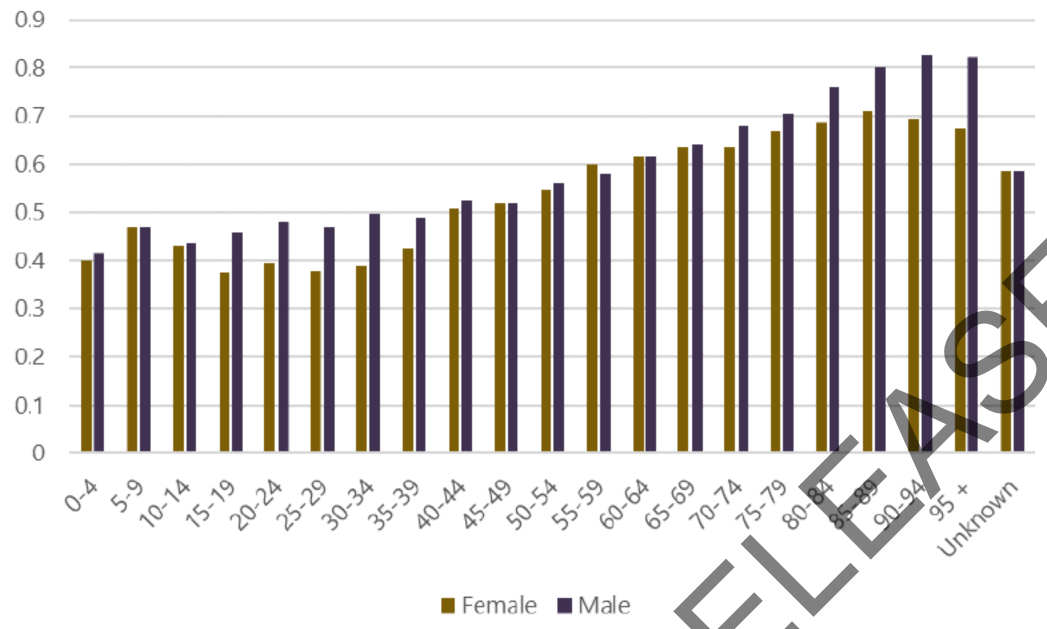
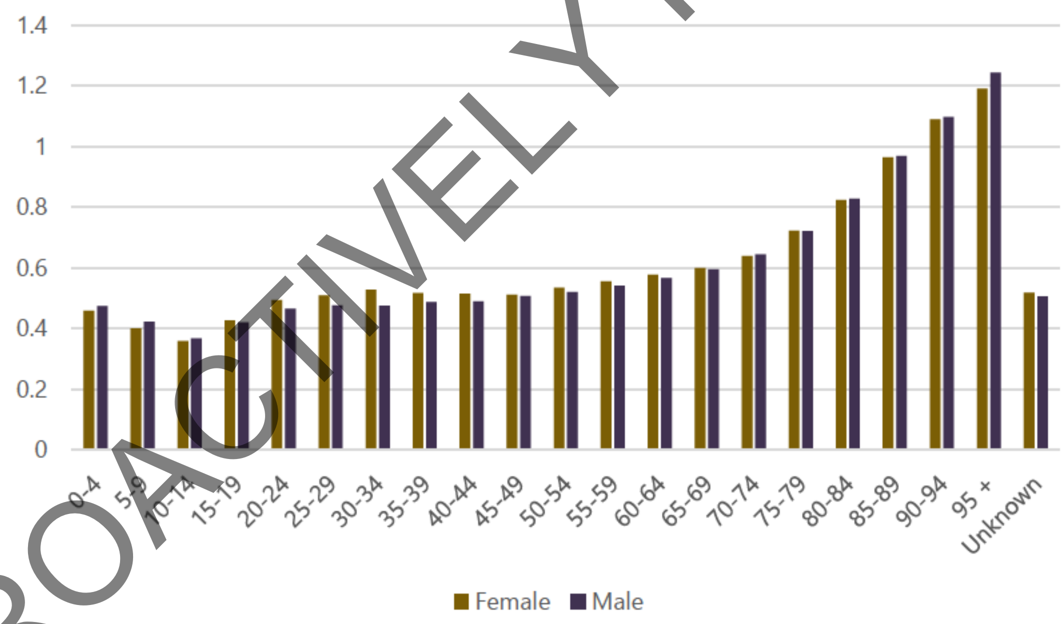
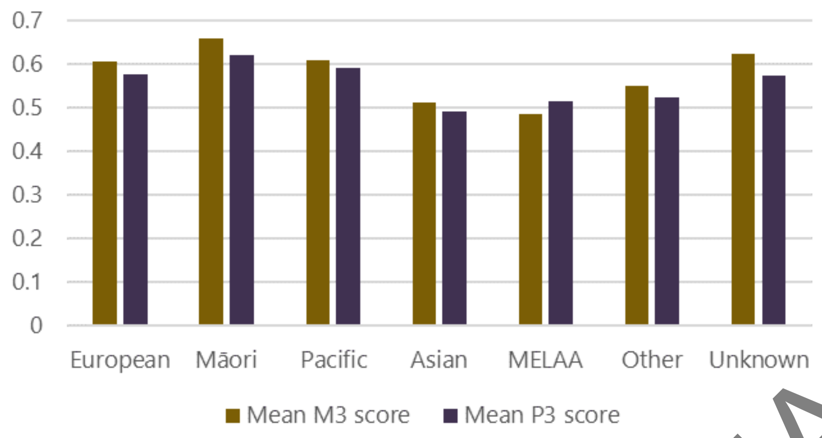


Figure 11: Mean P3 score by age and gender



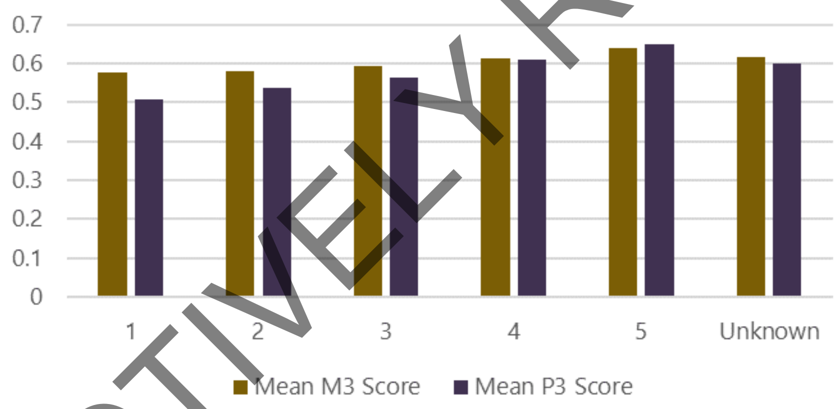
Both multimorbidity indices suggested that Māori and Pacific peoples in the sample population had more complex health needs when compared to Other ethnicities, on average (Figure 12).

Figure 12: M3 and P3 scores by ethnicity



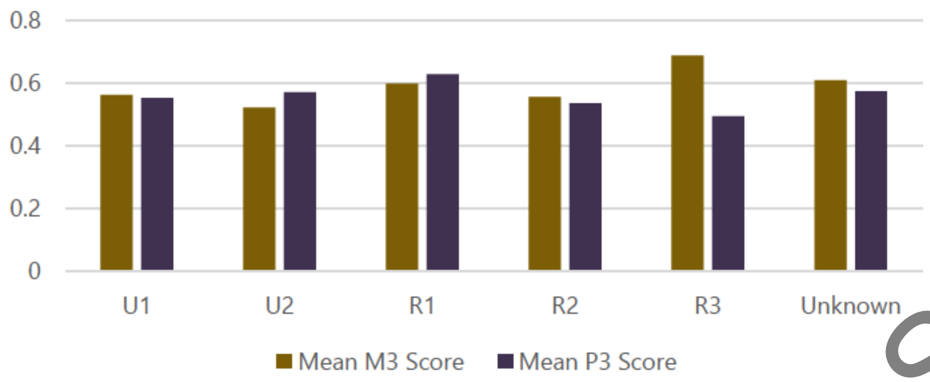
There was a clear relationship between socio-economic deprivation and the prevalence of multimorbidity. However, the difference in complexity between deprivation quintiles was more apparent when using the P3 index compared to the M3 index.

Figure 13: M3 and P3 scores by deprivation quintile



Rurality presented a complicated interrelationship between barriers to access, socio-economic deprivation and unmet need. Plotting M3 and P3 scores against GCH indices presented no clear relationship between rurality and the prevalence of multimorbidity in the sample population. M3 suggested those in the most rural areas (R3) had more severe multimorbidity while the P3 score indicated less severe multimorbidity.

Figure 14: M3 and P3 scores by geography



3.4 Comparing the two measures

M3 and P3 can predict utilisation to an extent.

M3 is a satisfactory predictor of primary care use up to a point before the relationship becomes questionable. We identified a clear correlation for lower M3 scores, but more volatile relationships for higher M3 scores. Figure 15 shows that the relationship between M3 and FTE increased until an M3 score of around two, before there was a significant amount of variation. Figure 16 rounds a patient's M3 score to the nearest quarter and plots the mean FTE used in 2023 for every rounded M3 score.

P3 was a more reliable measure for predicting primary care use, as:

- the tail end of the distribution accounted for a very small proportion of the population, which largely explains the volatile distribution
- a larger proportion of the population of interest can be matched to a P3 score
- being derived from pharmaceutical dispensing, P3 is more relevant to primary care services.

Figure 15: Scatterplot of M3 scores and FTE usage

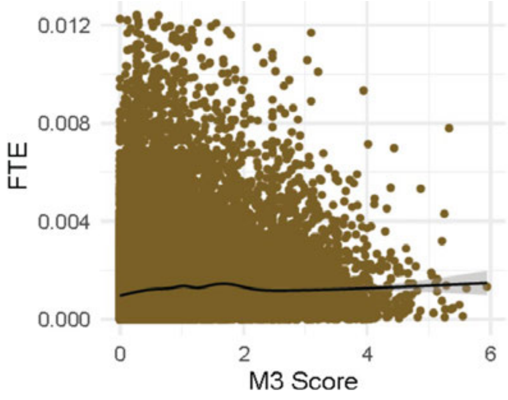


Figure 16: Mean FTE by M3 score

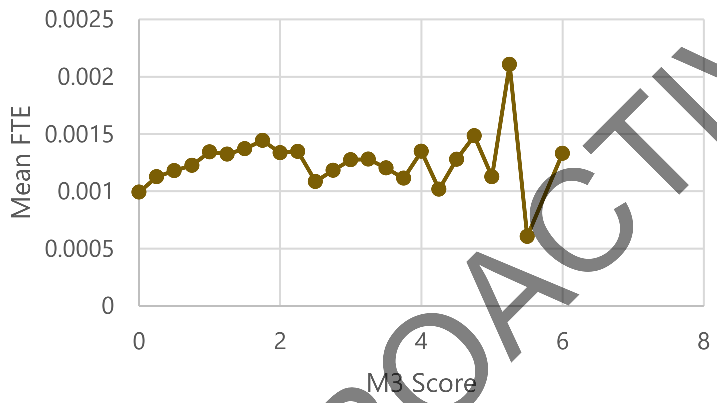


Figure 17: Scatterplot of P3 scores and FTE usage

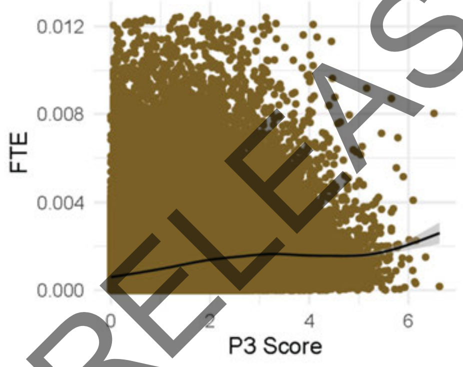
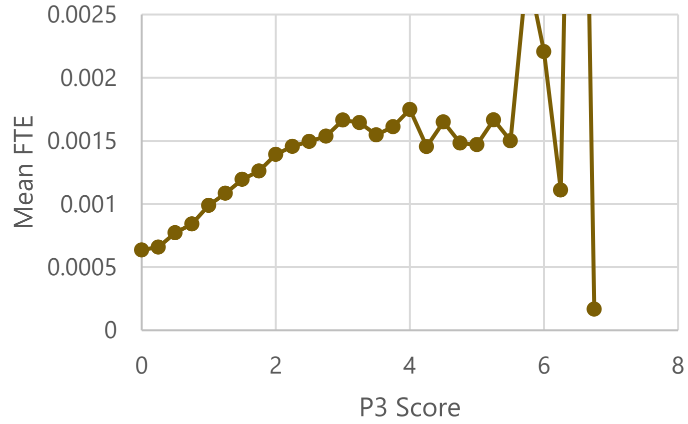


Figure 18: Mean FTE by P3 score



3.5 Imputing missing P3 scores

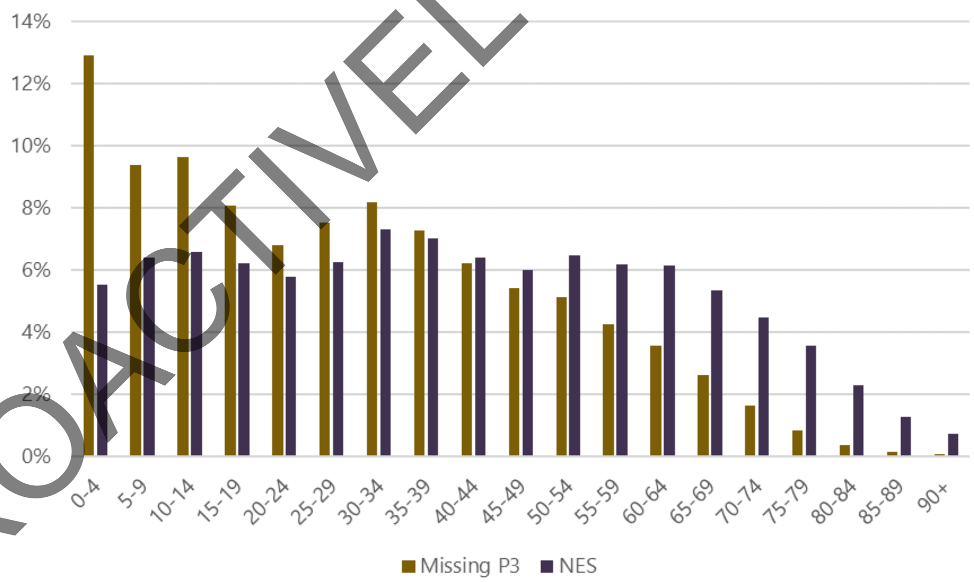
Using P3 presented a challenge of missing values. 899,091 patients in our dataset were missing a P3 score, which could have arisen for various reasons, including:

- They did not present with a condition from the P3 list of conditions[5]
- We were unable to match the patient to a P3 score due to data issues.
- No pharmaceutical dispensing was observed for the patients.

We used a multiple imputation method called predictive mean matching (PMM) which is widely used in statistical analysis (Morris et al., 2014)[6] to include these patients. This allowed us to retain our original sample size while also using plausible values of P3 scores where they are missing.

This adjustment was important. We chose the PMM method because it may be that those with missing values simply have a P3 score of zero. If this were the case, it would violate a critical assumption of multiple imputation, in that values are missing at random (MAR). Figure 19–Figure 21 show the distribution of populations by various characteristics compared to the total enrolled population. These figures showed that patients with missing P3 scores were less likely to be older (35+), less likely to be European and less likely to be in higher levels of socio-economic deprivation. This was a strong suggestion that a large proportion of these patients may have had a P3 score of zero.

Figure 19: Distribution of population with missing P3 score vs. total enrolled population by age band



[5] Noting that there are instances of patients recorded with a score of zero and no conditions.
[6] Imputation by predictive mean matching (PMM) borrows an observed value from a donor with a similar predictive mean (Morris et al., 2014).

Figure 20: Distribution of population with missing P3 score vs. total enrolled population by ethnicity

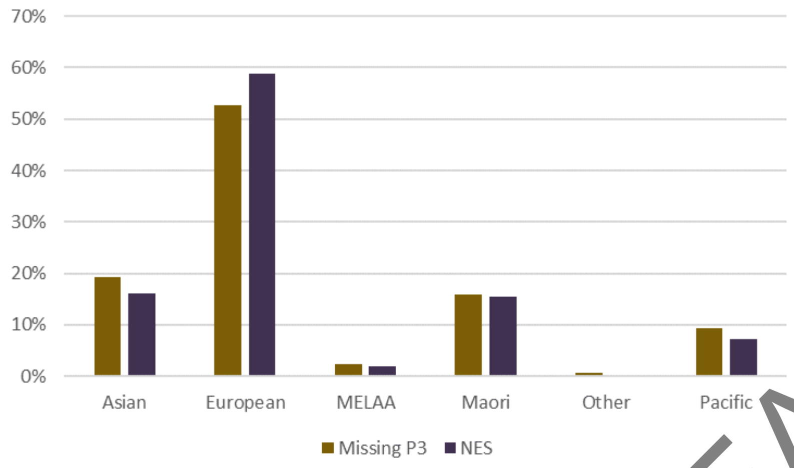
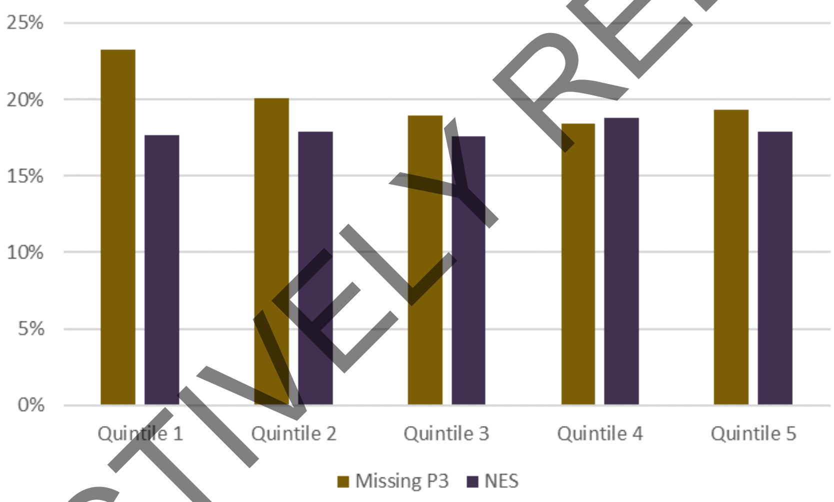


Figure 21: Distribution of population with missing P3 score vs. total enrolled population by socio-economic deprivation



PMM produces imputed values that are like the 'non-missing' dataset, as imputed values under PMM use 'borrowed' values from other patients with the same or similar characteristics. Imputed P3 values are therefore bounded by the observed P3 values.

4. Results

Our analysis showed primary care activity can be predicted by several demographic factors, mostly in line with expectation. Age was a strong predictor of activity, and women tended to utilise primary healthcare more than their male counterparts. Ethnicity predicted activity, but the degree and direction of its predictive power were also dictated by age. There were additional layers of complexity explained by other variables, including multimorbidity, socio-economic deprivation and rurality.

Our results showed higher general practice costs and activity can be attributed to:

- the elderly and infants
- Māori, Pacific peoples, and people of Other ethnicities
- those residing in more socio-economically deprived areas
- those with more comorbidities
- females in teen and adult years
- those in isolated rural areas (R3).

4.1 Regression results

The simple linear model (SLM) most easily visualised the distribution of predicted costs, as the coefficients could be interpreted in isolation. The SLM showed estimated coefficients for all variables were statistically significant at the 1 per cent level. We present the results from our regression analyses in Table 7. Age, gender, ethnicity, socio-economic deprivation and multimorbidity collectively played a role in explaining the variation observed in primary care activity.

The addition of rurality in the AIM-R (age-interactions with rurality) models showed R2 and R3 were statistically and economically significant predictors of higher associated general practice costs.

All other models incorporated interaction terms which separated the estimated coefficients into two categories:

- Main effects: shows the average effect of the variable on the outcome variable relative to the reference category. For instance, an estimated coefficient for an age band category shows its effect on cost relative to zero-year-olds.
- Interaction terms: indicates how the effect of age differs based on the interactor, i.e. gender and ethnicity.

The inclusion of interactive terms with granular age groupings can create some volatile estimates in both the main effects and interaction effects. Therefore, interpreting main effects and interaction effects in isolation can sometimes become meaningless as the model might attribute significant effect to the interaction in some age groups. We interpreted all main effects and interaction effects together to avoid misinterpreting the effect of age, gender, and ethnicity on costs. We extended the joint interpretation to estimated P3 coefficients in AIM-R-P3 results. Table 6 shows the effect of the independent variables on the distribution of predicted costs across age and gender.

Table 6: Effects of predicted cost variables by model specification

Variable	Effect on distribution							
	SLM		AIM		AIM-R		AIM-R-P3	
	Scale	Shape	Scale	Shape	Scale	Shape	Scale	Shape
Ethnicity	✓		✓	✓	✓	✓	✓	✓
P3 Group	✓		✓		✓		✓	✓
Deprivation	✓		✓		✓		✓	
GCH	Not included		Not included		✓		✓	

Age–gender interactions were significantly different from the reference age group during teen and adult years; these interactions reflect different activity for males and females during child-bearing years.

Ethnicity also had significant predictive power. Despite accounting for other variables, the estimated coefficients associated with ethnicity, and ethnicity–age interactions, were largely significant. This suggests the ethnicity variable controls for underlying drivers not accounted for by socio-economic deprivation, multimorbidity, or geography, including cultural barriers to accessing primary care services, or some other socio-economic differences between ethnicities.

Estimated age-P3 coefficients were also significant for those aged 15+.

Table 7: Regression estimates from specified models with cost as the dependent variable

	AIM-R	SLM	AIM	AIM-R-P3
Age band				
1	-76.054***	-81.074***	-75.984***	-75.035***
	(3.747)	(2.415)	(3.747)	(4.691)
2–4	-219.577***	-219.374***	-219.405***	-210.094***
	(2.839)	(1.81)	(2.84)	(3.544)
5–14	-294.357***	-290.206***	-294.133***	-282.614***
	(2.663)	(1.693)	(2.664)	(3.332)
15–24	-233.976***	-261.867***	-234.028***	-229.247***
	(2.709)	(1.714)	(2.71)	(3.373)
25–34	-226.098***	-248.031***	-226.251***	-230.526***
	(2.691)	(1.711)	(2.692)	(3.352)
35–49	-217.329***	-230.565***	-217.191***	-229.484***
	(2.677)	(1.705)	(2.678)	(3.336)
50–64	-191.689***	-188.578***	-191.153***	-220.672***
	(2.687)	(1.72)	(2.688)	(3.347)

	AIM-R	SLM	AIM	AIM-R-P3
65–69	-145.484***	-138.242***	-144.655***	-187.320***
	(2.878)	(1.868)	(2.88)	(3.545)
70–74	-115.899***	-106.635***	-115.038***	-170.547***
	(2.956)	(1.931)	(2.958)	(3.626)
75–79	-64.062***	-58.642***	-63.484***	-138.719***
	(3.182)	(2.099)	(3.183)	(3.871)
80+	2.251	10.822***	2.471	-79.630***
	(3.248)	(2.202)	(3.251)	(4.215)
Gender				
Male	13.731***	-40.127***	13.878***	13.324***
	(3.301)	(0.335)	(3.301)	(3.284)
Other	-358.972***	172.248***	-360.771***	-356.850***
	(16.582)	(24.791)	(16.164)	(8.02)
Ethnicity				
Māori, Pacific and Other ethnicities	-39.726***	20.321***	-38.802***	-38.669***
	(3.438)	(0.471)	(3.438)	(3.419)
NZDep2018 (quintiles)				
2	0.812*	1.512***	1.601***	1.217***
	(0.441)	(0.444)	(0.443)	(0.442)
3	5.954***	6.408***	6.532***	5.919***
	(0.471)	(0.473)	(0.472)	(0.47)
4	6.372***	7.004***	7.282***	6.160***
	(0.492)	(0.492)	(0.492)	(0.49)
5	21.010***	22.572***	22.953***	21.520***
	(0.592)	(0.591)	(0.59)	(0.587)
P3 score				
0.2 - 0.6	16.283***	16.596***	16.235***	20.312***
	(0.324)	(0.325)	(0.324)	(3.62)
> 0.6	89.445***	90.679***	89.261***	10.908**

	AIM-R	SLM	AIM	AIM-R-P3
	(0.477)	(0.48)	(0.477)	(4.632)
GCH				
R3	83.497***			
	(2.706)			
U1, U2 and R1	-11.047***			
	(0.791)			
Age band × male				
1	-2.416		-2.448	-1.507
	(4.783)		(4.783)	(4.764)
2–4	-9.715***		-9.829***	-8.239**
	(3.587)		(3.587)	(3.568)
5–14	-12.613***		-12.791***	-12.146***
	(3.356)		(3.357)	(3.338)
15–24	-78.991***		-79.049***	-79.077***
	(3.396)		(3.397)	(3.379)
25–34	-84.455***		-84.514***	-84.680***
	(3.389)		(3.389)	(3.371)
35–49	-76.851***		-77.008***	-76.633***
	(3.375)		(3.375)	(3.358)
50–64	-54.444***		-54.538***	-53.758***
	(3.4)		(3.4)	(3.383)
65–69	-41.845***		-41.662***	-41.093***
	(3.696)		(3.697)	(3.679)
70–74	-36.754***		-36.455***	-35.841***
	(3.824)		(3.825)	(3.805)
75–79	-38.358***		-38.127***	-38.249***
	(4.163)		(4.165)	(4.139)
80+	-17.577***		-17.369***	-17.301***
	(4.4)		(4.402)	(4.381)

	AIM-R	SLM	AIM	AIM-R-P3
Age band × other (gender)				
1	27.051		26.934	72.493***
	(16.94)		(16.532)	(9.324)
2–4	356.119***		379.789***	379.979***
	(117.453)		(108.317)	(118.903)
5–14	528.271***		528.614***	526.803***
	(78.226)		(78.129)	(78.136)
15–24	572.718***		573.817***	572.195***
	(47.406)		(47.254)	(45.098)
25–34	473.368***		474.642***	471.455***
	(28.005)		(27.726)	(23.992)
35–49	414.399***		414.929***	411.771***
	(31.984)		(31.763)	(28.641)
50–64	625.540***		626.438***	621.310***
	(209.22)		(209.033)	(207.783)
65–69	416.709***		416.391***	414.398***
	(129.855)		(129.833)	(128.887)
70–74	219.665***		225.106***	244.933***
	(31.27)		(34.454)	(19.521)
75–79	103.358***		102.744***	120.637***
	(17.721)		(17.335)	(10.223)
80+	306.251***		312.222***	271.781***
	(92.7)		(88.825)	(87.468)
Age band × Māori, Pacific and Other ethnicities				
1	-9.453*		-9.602*	-8.653*
	(5.039)		(5.039)	(5.013)
2–4	12.781***		12.420***	12.907***
	(3.759)		(3.759)	(3.737)
5–14	27.493***		26.964***	27.435***

	AIM-R	SLM	AIM	AIM-R-P3
	(3.506)		(3.506)	(3.486)
15–24	29.373***		28.755***	29.604***
	(3.561)		(3.561)	(3.541)
25–34	54.321***		53.831***	54.483***
	(3.593)		(3.593)	(3.574)
35–49	75.725***		75.153***	75.640***
	(3.599)		(3.599)	(3.58)
50–64	108.121***		108.088***	106.176***
	(3.738)		(3.738)	(3.716)
65–69	106.902***		107.157***	102.370***
	(4.812)		(4.814)	(4.775)
70–74	112.171***		112.617***	104.617***
	(5.518)		(5.522)	(5.473)
75–79	99.765***		100.258***	90.443***
	(7.101)		(7.109)	(7.052)
80+	42.210***		42.768***	38.454***
	(7.897)		(7.909)	(7.871)
Age band × P3 score 0.2–0.6				
1				-7.079
				(5.306)
2–4				-26.288***
				(3.931)
5–14				-28.596***
				(3.678)
15–24				-23.455***
				(3.72)
25–34				-12.195***
				(3.702)

	AIM-R	SLM	AIM	AIM-R-P3
35–49				-2.902 (3.686)
50–64				12.184*** (3.702)
65–69				23.056*** (3.949)
70–74				31.708*** (4.053)
75–79				44.545*** (4.362)
80+				49.054*** (4.8)
Age band × P3 score > 0.6				
1				12.745* (6.624)
2–4				11.956** (5.024)
5–14				-1.334 (4.712)
15–24				26.357*** (4.785)
25–34				52.135*** (4.78)
35–49				70.168*** (4.755)
50–64				114.503*** (4.775)
65–69				141.796*** (5.146)

	AIM-R	SLM	AIM	AIM-R-P3
70–74				166.147***
				(5.245)
75–79				194.535***
				(5.534)
80+				183.246***
				(5.686)
Constant	377.944***	373.156***	367.363***	381.844***
	(2.75)	(1.71)	(2.643)	(3.293)
Observations	2,481,605	2,481,605	2,481,605	2,481,605
AIC	34,740,652	34,757,884	34,744,081	34,726,528
BIC	34,741,377	34,758,164	34,744,780	34,727,508
R^2	0.112	0.105	0.11	0.117
Adjusted R^2	0.112	0.105	0.11	0.117
F Statistic	5,665.073*** (df = 55; 2481549)	14,608.830*** (df = 20; 2481584)	5,806.013*** (df = 53; 2481551)	4,367.517*** (df = 75; 2481529)
Note:				*p<0.1; **p<0.05; ***p<0.01
				Robust standard errors in parentheses.

4.2 Cohort weights were based on the mean of patient cost

We created a dataset with every possible combination of age, gender, ethnicity, socio-economic deprivation, P3 Group and GCH to calculate cohort weights. This dataset had 2,160 unique permutations describing the enrolled population.[7] We then predicted costs for this dataset of hypothetical patients and recentred those costs to the patient closest to the mean cost.

We refer to this patient as "Neo." The predicted cost for Neo became the denominator used to calculate the weight for all other patients, where:

$$\text{weight}_{\text{i}} = \frac{\text{predicted cost}_{\text{i}}}{\text{predicted cost}_{\text{Neo}}}$$

Therefore, the weight calculated for Neo = 1, and the weight for patient i will determine the weighting they would receive relative to Neo.

[7] We exclude patients that identify with another gender from weighting calculations as their small sample size leads to volatile predicted costs when using the specified regression model (AIM-R).

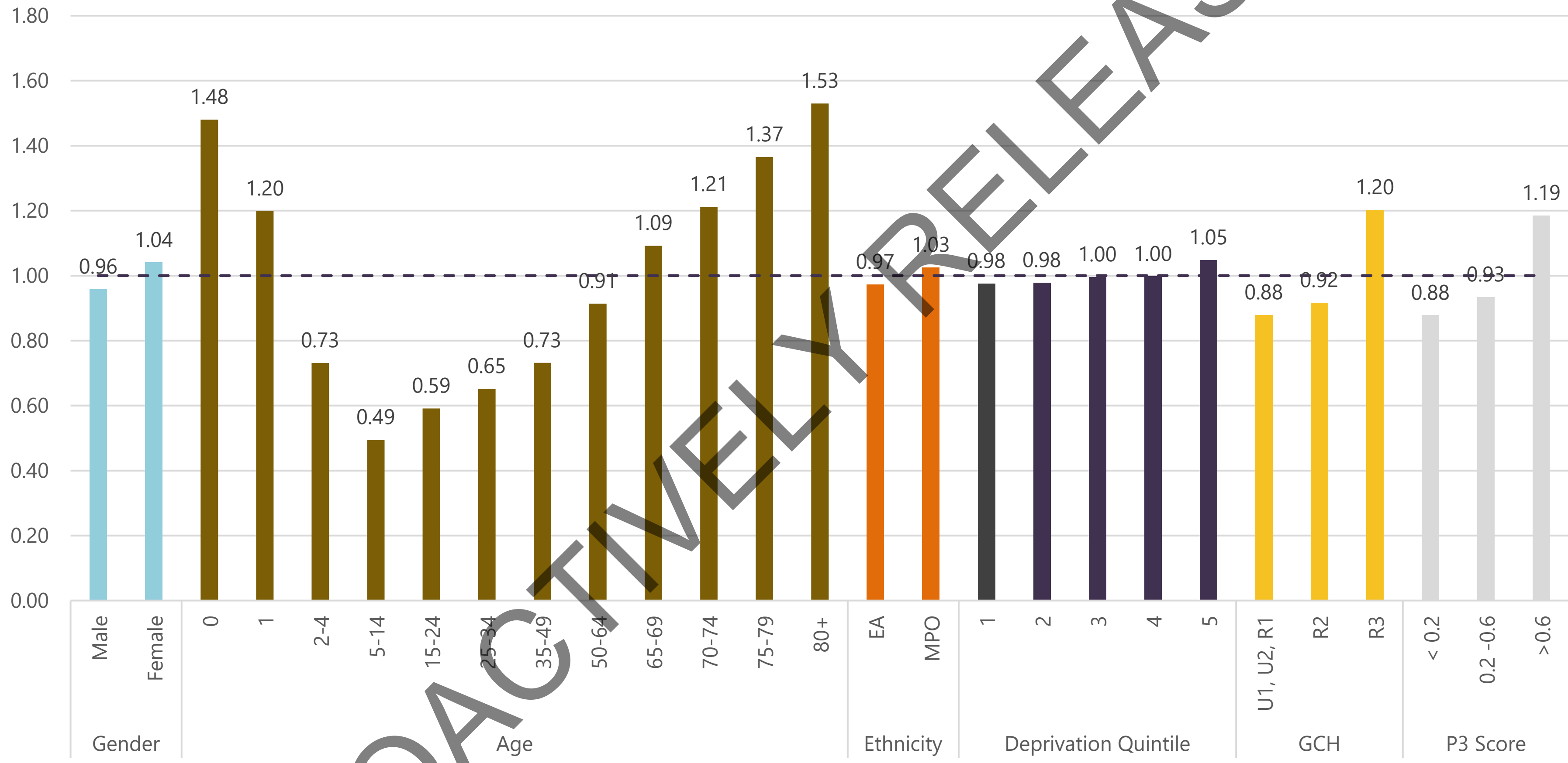
Figure 22 below shows the mean weights for each characteristic, relative to the hypothetical patient dataset mean weight of 1. Higher than average weights were calculated for the following characteristics:

- Females
- Ages zero to one, and 65+
- Māori, Pacific peoples, and people of Other ethnicities
- Socio-economic deprivation quintile 5
- People residing in R3 areas
- People with a P3 score > 0.6

While females and Māori, Pacific peoples, and people of Other ethnicities had higher than the mean cost weight on average, we highlight the importance of the age–gender and age–ethnicity interactions. Higher cost weights associated with females were driven by those in teen and child-bearing years, while higher cost weights associated with Māori, Pacific and Other ethnicities are driven by those aged 25+. The counterintuitive characteristics of Neo likely also reflect underutilisation for these groups.

Comparing the individual bars to the mean of 4.3 illustrates which variables play a stronger role in driving the weights. Figure 22 shows that the variation in predicted costs between age bands, GCH and P3 scores were substantial compared to the variation between genders, ethnicities and socio-economic deprivation levels. However, section 4.2.1 describes the importance of gender and ethnicity in explaining more variation in different age groups.

Figure 22: Mean weights by patient characteristics



Note: EA = European and Asian, MPO = Māori, Pacific peoples, and Other ethnicities

4.2.1 The relationship between age, gender and ethnicity

The relationship between age, gender and ethnicity predicted varying patterns of general practice cost.

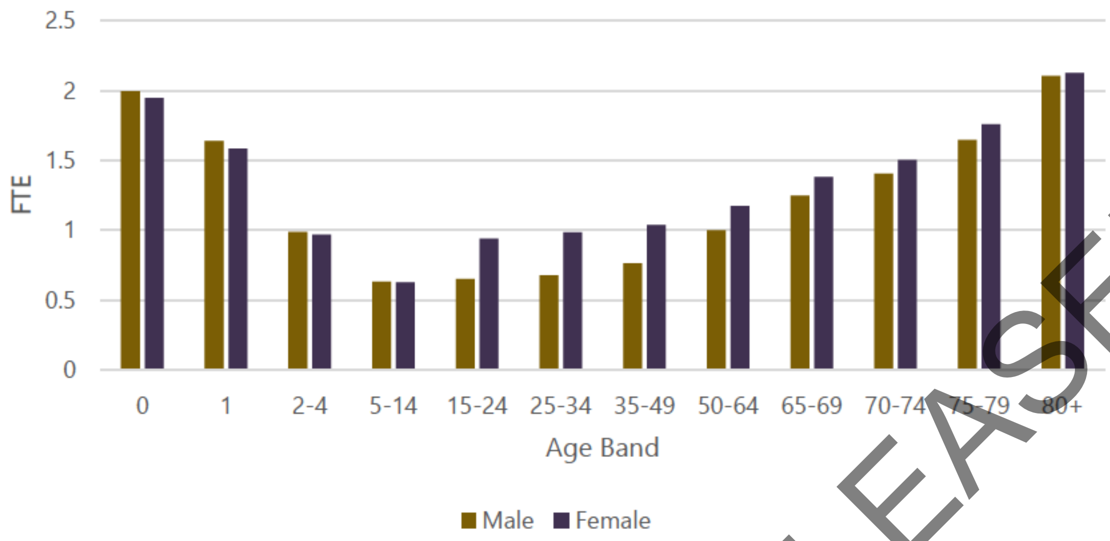
- **Predicted cost weights were the highest amongst infants and the elderly**. Table 8 shows that patients on the tail-ends of the distribution, i.e., zero-year-olds and 80+ year olds, faced the highest predicted cost weights for both ethnic groupings. Cost weights remained relatively high until around four years of age, before dropping in the 5–14 age band.
- **There was a gender disparity in predicted cost weights during teen and child-bearing years**. A consistent finding across all ethnic groups was the sudden uptick in predicted costs for females relative to males at the 15–24 age band. The significantly higher predicted costs amongst females were persistent until the sample aged into their elderly years. This likely reflects the higher need for primary care amongst females during these years, and potentially the reduced propensity for males to utilise primary care services relative to females.
- **Māori, Pacific peoples and people of Other ethnicities had higher predicted cost weights for adult age groups**. Predicted cost weights for Māori, Pacific peoples and Other ethnicities aged zero to 24 were lower relative to their Asian and European counterparts, but higher for those aged 25+.

Table 8: Predicted capitation weights, assuming deprivation quintile 1, GCH R2, and P3 score < 0.2

		0	1	2-4	5-14	15-24	25-34	35-49	50-64	65-69	70-74	75-79	80+
European and Asian	M	1.34	1.07	0.56	0.29	0.27	0.28	0.33	0.50	0.70	0.82	0.99	1.29
	F	1.30	1.04	0.54	0.29	0.49	0.52	0.55	0.64	0.80	0.90	1.08	1.30
Māori, Pacific and Other	M	1.21	0.91	0.46	0.25	0.23	0.33	0.46	0.73	0.93	1.07	1.20	1.30
	F	1.16	0.87	0.45	0.24	0.46	0.57	0.67	0.87	1.03	1.15	1.28	1.31

Our FTE analysis showed a positive correlation between age and clinician FTE after early childhood years (Figure 23). In line with our observations in Figure 3, very young children (aged zero to four) used a significant amount of FTE. FTE usage then dropped in prepubescent years and rose from there on. The gender difference in primary care use was particularly stark from teen years until the sample population reached their 60s, with females using more resources. FTE usage began increasing exponentially for age groups 65+.

Figure 23: Age–gender distribution of FTE per 1,000 patients across age bands

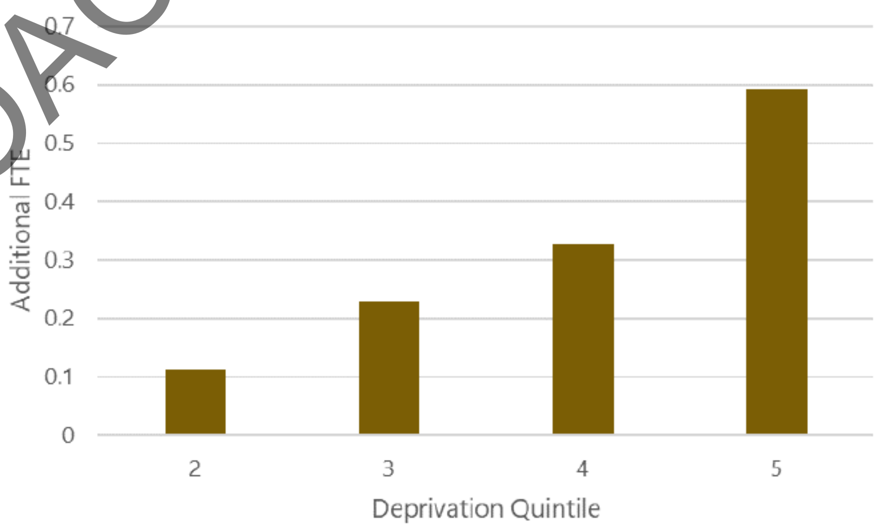


Looking across ethnicity, Māori, Pacific peoples and Other ethnicities used more FTE on average than Europeans at around 1.14 FTE per 1,000 patients, compared to 1.05 (see Table 4).

4.2.2 A clear link between socio-economic deprivation and GP activity

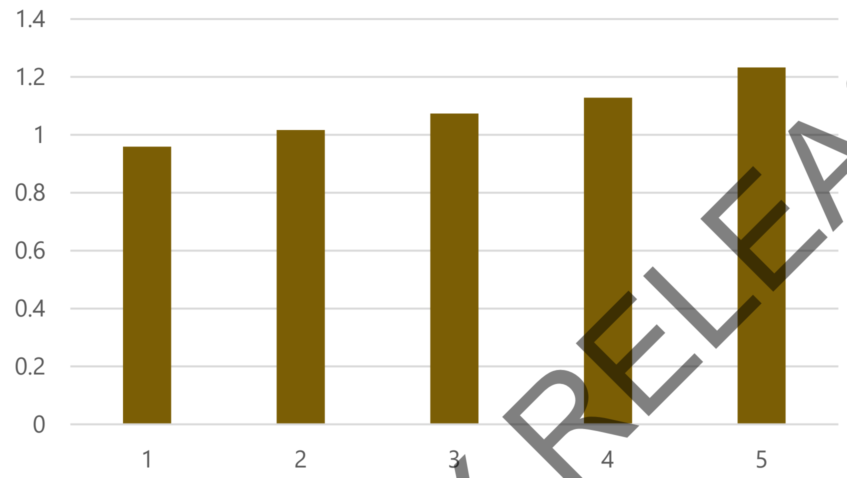
Socio-economic deprivation captures patient characteristics such as income, employment outcomes, homeownership, and standard of living. Our model showed that increasing levels of socio-economic deprivation are associated with higher costs. As the level of deprivation increased, the predicted additional cost weight also increased. There was also a significant jump between deprivation quintile 4 to 5 of 0.26, compared to 0.09–0.12 between deprivation quintiles 1 and 4. Socio-economic deprivation, however, may not capture factors such as rurality and multimorbidity.

Figure 24: Predicted additional weight over deprivation quintile 1, attributed to socio-economic deprivation



We observed a positive relationship between socio-economic deprivation and the calculated clinician FTE (Figure 25). Those in deprivation quintile 1 (least deprived) used the least clinician time at under one FTE per 1,000 patients. The distribution of FTE increased as the enrolled population sample became more deprived, with those in quintile 5 using more than 1.2 FTE per 1,000 patients. This observation was in line with findings such as Barlow et al. (2021) and Covvey et al. (2014), who show increasing level of GP utilisation in more deprived areas.

Figure 25: Distribution of FTE by socio-economic deprivation per 1,000 patients



4.2.3 Multimorbidity meant more clinician time

We observed that increased levels of multimorbidity associated with higher levels of individual GP cost weights using the P3 index of multimorbidity (Figure 24). A higher P3 score indicated increased multimorbidity of a patient. Patients with a P3 score of greater than 0.6 used far more FTE than patients with a lower multimorbidity score at more than 1.4 FTE per 1,000 patients. Comparatively, those with a P3 score of less than 0.2 used a mere 0.86 FTE per 1,000 patients, on average. Figure 27 highlights clinician FTE usage by level of multimorbidity. Our conclusion was that a multimorbidity index is important when considering patient interactions with primary care consultations.

P3 had significant predictive power, one which is more material than the additional weight attributed to socio-economic deprivation. This might be because of:

- increased consultation frequency to manage complex health needs and medication management. Those with a P3 score above 0.6 had an average of 3.3 visits in 2023, compared to 2.9 for P3 scores between 0.2 and 0.6, and 2.7 for P3 scores less than 0.2
- longer consultation times for clinicians to address multiple issues in a single visit
- enhanced monitoring and follow-up consultations to prevent complications.

The literature is complex, however, and we note that socio-economic deprivation and multimorbidity need to be looked at through both an activity lens, and an unmet need lens as patients are less likely to receive continuity of care despite its benefits (Salisbury et al., 2011).

Figure 26: Predicted additional weight over P3 score < 0.2, attributed to multimorbidity

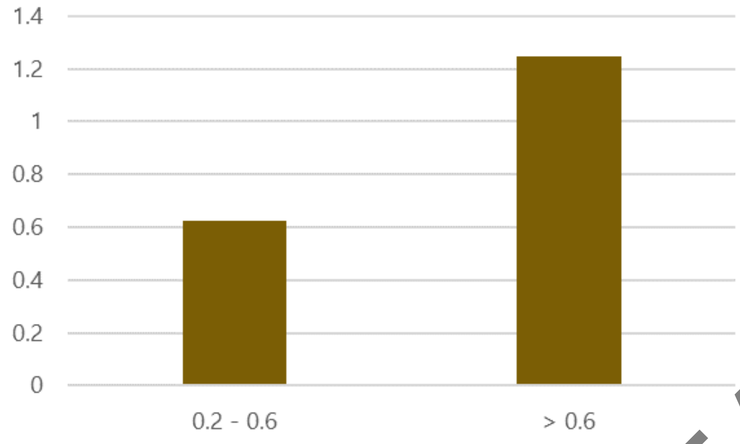
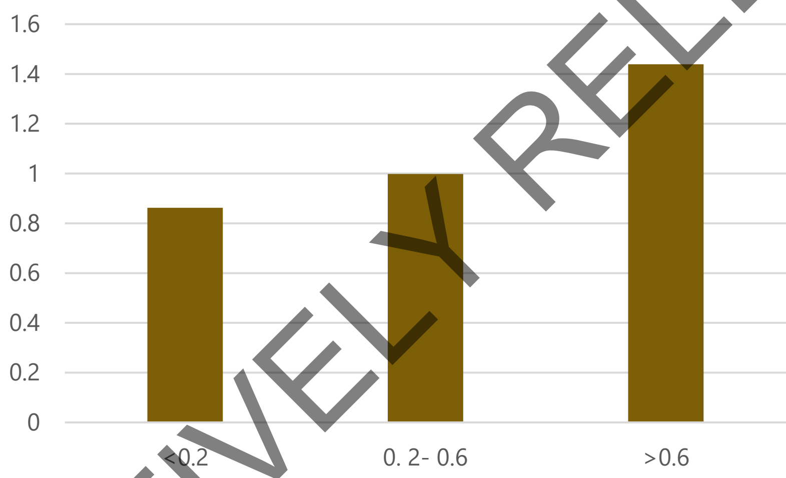


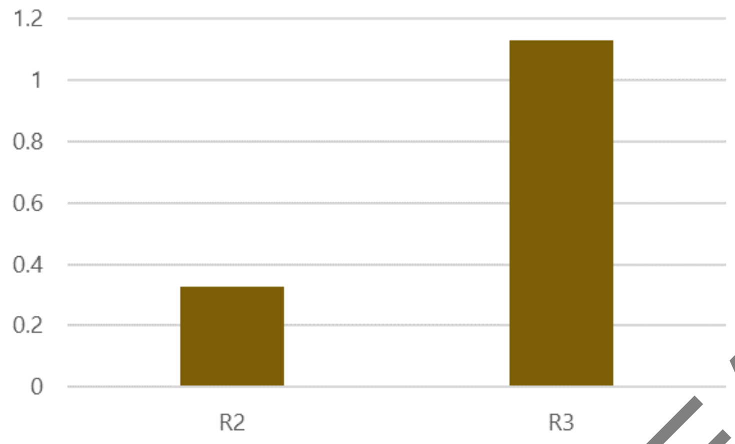
Figure 27: FTE distribution by P3 score per 1,000 patients



4.2.4 Activity-based weights were higher in isolated, rural areas

Our regression model predicted significant cost weights associated with those in more rural and isolated areas (R2 and R3), shown in Figure 28. Moving from R2 to R3 presented an additional cost weight of 0.8, which was the largest marginal increase from any of the non-interacted variables.

Figure 28: Predicted additional weight over U1, U2, and R1, attributed to Geographical Classification of Health (GCH)



Notably, the additional cost weights associated with more rural and isolated areas did not arise from more frequent consultations or longer consultations, but rather the impact of non-contact time. Table 9 shows that R3 patients had the second to least mean number of consultations in 2023, and the third to least mean contact time. However, the non-contact services being provided by practices that catered to predominantly R3 patients drastically scaled up the associated clinician FTE (Appendix A outlines how we used non-contact time to scale FTE).

Table 9: Consultation statistics by Geographical Classification of Health (GCH)

	Mean		
GCH	**Number of contacts**	**Yearly total consultation duration**	**FTE per 1,000 patients**
R1	2.83	28.86	1.11
R2	2.64	27.65	1.28
R3	2.67	29.79	1.57
U1	3.00	31.29	1.02
U2	2.89	30.72	1.11

4.3 Estimating clinician costs from regression results

Using the predicted costs, we estimated the salary costs of providing care to the enrolled population of New Zealand would be approximately $1.27 billion.

We calculated this amount using the predicted costs for GPs, nurses, and nurse practitioners from the regression models and added costs for other salaried staff using the following salaries:

- GP: $262,080 – using step 14 of the Senior Medical and Dental Officers' Collective Agreement.[8]

[8] Senior Medical and Dental Officers' Collective Agreement.

- Nurse: $106,739 – using the upper end of the scale of the Nursing and Midwifery Collective Agreement.[9]
- Nurse practitioner: $151,079 – using the second to highest step for a nurse practitioner salary from the Nursing and Midwifery Collective Agreement.

We used the salaries of the staff not included in our FTE analysis to estimate a cost per GP FTE which we then applied on a per capita basis to the enrolled population. The outcome was a staff cost per enrolled patient which we summed to get total salary costs.

Our estimated salary costs were less than what was estimated by Love et al. (2022). They estimated salary expenses of $1.35 billion which is around $77 million more than our estimate. The difference could be explained by the difference in the regression models and data that were used to predict FTE costs.

[9] Nursing and Midwifery Collective Agreement.

5. Implications for practice funding

We applied the revised population cohort weights derived from the preferred AIM-R model to the NES data for December 2023 to calculate the first contact capitation revenue practices would receive if 100 per cent of this revenue for patients were activity-based funding determined by the revised population cohorts and weights. We assumed no change in total capitation funding, i.e. how would $514.3 million be distributed among practices for providing services to the patients who received care in 2023 using the revised population cohorts and weights?

We observed the following:

- On average, practices with at least 50 per cent of the enrolled population over the age of 65 years could expect an increase of s 9(2)(j) in core capitation funding if 100 per cent of core capitation were determined by the revised population cohort weights.
- Re-weighting redistributed funding towards practices that had a higher proportion of elderly, Māori/Pacific peoples, or highly deprived populations.

We explored the impact on revenue if services to improve access (SIA) funding were also redistributed based on the revised weights. However, we recommend not to include SIA funding for redistribution using the revised weights. Co-payment 'buy out' funding such as Very Low Cost Access (VLCA) funding, zero fees under-14s funding, and Community Services Card (CSC) holder funding were not included in this stage of the analysis.

5.1 Estimating effects on practice revenue

We identified the amount paid to practices for enrolled patients:

- who visited their general practice in 2023, and
- for whom there is enough information recorded in the NES to categorise them into the revised population cohorts.

General practices received approximately $863.1 million[10] in first contact capitation revenue in 2023. Over 2.84 million of the total 4.97 million patient records were used for the comparison, which meant we had a first contact capitation funding pool of $514.3 million for distribution.

5.1.1 Some practices would see a significant increase in first contact capitation

s 9(2)(j)

s 9(2)(j)

s 9(2)(j)

s 9(2)(j)

5.1.2 Age appeared to have a strong effect on first contact capitation

s 9(2)(j)

s 9(2)(j)

s 9(2)(j)

5.1.3 Socio-economic deprivation and ethnicity had a strong impact on revenue

s 9(2)(j)

s 9(2)(j)

s 9(2)(j)

s 9(2)(j)

s 9(2)(j)

s 9(2)(j)

s 9(2)(j)

5.1.4 Rural practices would see increased funding

s 9(2)(j)

s 9(2)(j)

5.2 Redistributing capitation using the current formula and revised weights

s 9(2)(j)

s 9(2)(j)

s 9(2)(j)

s 9(2)(j)

5.3 Practice level effects on revenue

s 9(2)(j)

s 9(2)(j)

s 9(2)(j)

s 9(2)(j)

s 9(2)(j)

5.4 A simplified approach to applying cost weights

Understanding the implications on practice funding involves applying the weights to the enrolled population using the 2,160 cohorts. This can be difficult for practices to implement, particularly where there may not be enough information available for a patient. Box 1 below highlights our investigation into a simplified model that can be easily utilised by a practice or PHO on their enrolled population. The estimated coefficients from this simplified model are presented in Table 12.

Box 1: A simplified approach to applying cost weights

By utilising only age-gender interactions in our model, the resulting capitation weights could be easily expressed as **24 age-sex cohorts, with 13 additive adjusters** representing ethnicity, multimorbidity, socio-economic deprivation and rurality. We refer to this model as the Age-Gender Interaction Model (AGIM).[12]

The 24 age-sex cohorts were as follows:

	0	1	2-4	5-14	15-24	25-34	35-49	50-64	65-69	70-74	75-79	80+
Female	1.23	0.96	0.49	0.25	0.46	0.52	0.57	0.67	0.83	0.93	1.10	1.31
Male	1.28	1.00	0.50	0.25	0.24	0.27	0.35	0.53	0.73	0.85	1.01	1.30

While the adjusters representing other personal characteristics were:

Variable	Characteristic	Adjuster (+)
Ethnicity	European and Asian	0.000
	Māori, Pacific Peoples and Other ethnicities	0.068
NZDep	1	0.000
	2	0.003
	3	0.021
	4	0.022
	5	0.072
P3 Score	< 0.2	0.000
	0.2–0.6	0.057
	> 0.6	0.313
GCH	R2	0.000
	R3	0.293
	U1, U2 and R1	-0.038

The resulting capitation weights for each granular cohort were like the AIM-R model, with goodness-of-fit measures showing that the simplified model explained only marginally less variation than the AIM-R model and was therefore a sufficient substitute. The variation in predicted cost weights from this model were driven primarily by age, geography and multimorbidity, consistent with prior specifications.

[12] The estimated coefficients from this simplified model are presented in columns 1 and 2 of Table 23.

Table 12: Regression results from the AGIM model

	(AGIM)	(AGIM)[1]
Age band		
1	-79.569***	-84.457***
	(3.349)	(3.331)
2–4	-214.359***	-222.845***
	(2.539)	(2.526)
5–14	-283.600***	-294.067***
	(2.379)	(2.365)
15–24	-221.831***	-236.387***
	(2.426)	(2.412)
25–34	-205.788***	-222.640***
	(2.417)	(2.402)
35–49	-191.932***	-210.226***
	(2.404)	(2.387)
50–64	-160.968***	-183.173***
	(2.416)	(2.399)
65–69	-117.137***	-142.124***
	(2.64)	(2.622)
70–74	-88.100***	-115.394***
	(2.722)	(2.705)
75–79	-38.793***	-68.706***
	(2.967)	(2.949)
80+	22.476***	-11.608***
	(3.015)	(3.004)
Gender		
Male	13.284***	12.134***
	(3.315)	(3.291)
Other	-367.003***	-361.279***
	(10.89)	(15.981)
Ethnicity		

	(AGIM)	(AGIM)[1]
Māori, Pacific and Other	19.668***	18.705***
	(0.471)	(0.469)
NZDep2018 (quintiles)		
2	0.820*	-0.097
	(0.442)	(0.439)
3	6.012***	4.447***
	(0.471)	(0.469)
4	6.433***	3.657***
	(0.492)	(0.49)
5	20.877***	17.622***
	(0.592)	(0.589)
P3 Score		
0.2–0.6	16.553***	14.410***
	(0.324)	(0.393)
> 0.6	90.440***	117.285***
	(0.479)	(0.595)
GCH		
R3	84.553***	87.146***
	(2.71)	(2.696)
U1, U2 and R1	-11.063***	-12.189***
	(0.791)	(0.789)
Age band × male		
1	-1.949	-2.738
	(4.808)	(4.784)
2–4	-9.586***	-9.909***
	(3.603)	(3.582)
5–14	-12.271***	-12.394***
	(3.37)	(3.349)
15–24	-78.453***	-76.727***
	(3.41)	(3.388)

	(AGIM)	(AGIM)[1]
25–34	-83.971***	-80.900***
	(3.402)	(3.379)
35–49	-76.381***	-74.180***
	(3.388)	(3.365)
50–64	-54.241***	-52.604***
	(3.413)	(3.389)
65–69	-41.713***	-40.658***
	(3.71)	(3.683)
70–74	-36.623***	-35.275***
	(3.837)	(3.809)
75–79	-38.361***	-37.345***
	(4.175)	(4.147)
80+	-16.956***	-15.636***
	(4.41)	(4.386)
Age band × other		
1	-9.206	-59.248***
	(11.165)	(16.175)
2–4	345.371***	323.241***
	(126.501)	(119.642)
5–14	539.671***	531.512***
	(77.284)	(78.107)
15–24	584.844***	572.707***
	(45.725)	(47.131)
25–34	481.706***	470.281***
	(25.049)	(27.651)
35–49	421.920***	411.896***
	(29.253)	(31.31)
50–64	639.295***	632.792***
	(211.168)	(209.988)

	(AGIM)	(AGIM)[1]
65–69	418.127***	422.250***
	(129.265)	(126.335)
70–74	221.743***	221.143***
	(28.569)	(31.177)
75–79	148.269***	165.731***
	(11.052)	(16.091)
80+	315.066***	293.623***
	(91.876)	(92.363)
Constant	355.913***	381.282***
	(2.49)	(2.47)
AIC	34,747,913	34,728,216
BIC	34,748,498	34,728,801
Observations	2,481,605	2,481,605
R^2	0.109	0.116
Within R^2		
Adjusted R^2	0.109	0.116
Residual Std. Error (df = 2481560)	265.64	264.59
Note:	*p<0.1; **p<0.05; ***p<0.01	

[1]The model in column 2 assumes that null P3 scores are equal to zero

6. Incorporating supply-side effects

This section examines the potential for incorporating supply-side effects into funding models for primary care. Supply-side factors at the practice level can significantly influence the utilisation and cost of healthcare. These include workforce availability, workload per GP, workforce mix, and practice characteristics such as rurality and patient demographics, all of which impact patient access, efficiency of service delivery, and costs of care (Abel et al., 2020).

Insofar as our focus has been on patient-level characteristics, this exploratory analysis aims to demonstrate how integrating practice-level factors into funding models could work and to assess the implications for resource allocation and equity outcomes. By highlighting the interplay between practice characteristics and funding needs, this work seeks to inform future refinements in capitation funding approaches.

6.1 Including supply-side effects can allocate resources more efficiently, but reinforce inequities in utilisation

Incorporating supply-side effects in resource allocation models aims to help isolate and account for factors that influence the availability and provision of healthcare services. However, if supply variables are simply included in models of healthcare utilisation, this can lead to biased estimates of their impacts on utilisation, impact the statistical reliability of socio-economic variables, and lead to errors in identifying which variables truly represent patient needs (Gravelle et al., 2003).

Abel et. al (2020) aimed to develop a predictive model to identify general practices in England that may face a workforce supply–demand imbalance in the future. The authors used routine data on workforce, patient experience, and registered populations, alongside a census of general practitioners' career intentions. The model revealed that practices at high risk of a future supply–demand imbalance often had larger patient lists, employed more nurses, and served more deprived populations. These findings suggest that understanding supply-side effects is important for accurate predictions in healthcare planning, though they note that predictions are inherently limited by the data available and reliance on relative metrics in the absence of direct measures to capture supply-side factors.

Supply factors may be linked to unobserved patient needs, e.g. local healthcare supply might be influenced by these needs, or unobserved needs could increase local demand, leading to longer waiting times. Rice et. al (2000) notes supply-side factors may be influenced by historic need and utilisation, creating a feedback loop between supply-side factors and utilisation. If demand-side barriers have led to an underutilisation of primary healthcare, that has in turn influenced supply-side factors, incorporating those supply-side impacts in an allocation model may inadvertently reinforce existing healthcare inequalities. Anselmi et. al (2022) also caution that inconsistent reporting can affect need predictions, and populations served by providers who under-report would thus be penalised. This is consistent with our findings in section 6.3, which show a reduced effect of rurality and ethnicity on predicted cost weights relative to the results in section 4.1.

Gravelle et. al (2003) note that in previous analyses, supply-side effects have been excluded from analysis so that the remaining need-based variables would better reflect actual healthcare needs.

However, this can lead to a situation where need variables end up reflecting supply factors instead, which goes against the principle that healthcare allocations should be based on need rather than supply (Gravelle et al., 2003). Separating need from supply-side factors is thus important, and ideally will help to ensure that healthcare resource allocation reflects population needs rather than existing patterns of service provision (Anselmi et al., 2022).

The effect of isolating factors that account for the provision of healthcare services was demonstrated by a study which, in the context of developing a weighted capitation formula for mental health services in England, incorporated supply variables in the study to control for differences in access to care across regions, using a method to "sterilise" the effects of supply variables when estimating individual need (Anselmi et al., 2020). This process involved fixing the supply variables at their population average values during the estimation of individual needs, ensuring that any variation in predictions reflected differences in need rather than access to care.

6.2 Using practice-level fixed effects to address time-invariant confounding of supply-side factors

We re-conducted our analysis using fixed effects at the practice level to account for confounding of unobserved time-invariant supply-side effects. To do this, we incorporated practice-level fixed effects into our analysis. This allowed us to control for unobservable, practice-specific characteristics (such as skill mix, organisational culture, management approaches, approach to patient management etc.) that might influence costs but remain constant over time. Including fixed effects aimed to isolate the impact of the exploratory variables on costs while accounting for these practice-specific, time-invariant unobserved factors.

The fixed-effects model we used is specified as follows:

$$\mathrm{cost}_{\mathrm{it}} = \mathbf{X}_{\mathrm{it}}\beta + \mu_{\mathrm{j}} + t + \epsilon_{\mathrm{it}}$$

Where:

- $\mathrm{X}_{\mathrm{it}} = \mathrm{AgeBand}_{\mathrm{it}} + \mathrm{Gender}_{\mathrm{it}} + \mathrm{Ethnicity}_{\mathrm{it}} + \mathrm{Multimorbidity}_{\mathrm{it}} + \mathrm{DeprivationQuintile}_{\mathrm{it}} + \mathrm{GCH}_{\mathrm{it}} + \mathrm{AgeBand}_{\mathrm{it}} \times \mathrm{Gender}_{\mathrm{it}}$
- i refers to the individual, j refers to the practice and t represents the month of a patient's appointment(s).
- μ_{j} captures the unobserved, time-invariant practice-specific factors (practice-level fixed effects)
- ϵ_{it} represents the idiosyncratic error term
- t are time fixed effects by month of year.

6.2.1 Hausman test for model selection

To verify the appropriateness of the fixed-effects specification, we conducted a Hausman test comparing this model to the AGIM model. The test evaluates whether the practice-specific effects (μ_{j}) are correlated with the explanatory variables. The Hausman test produced a chi-squared test statistic of 114,964 ($p < 0.01$), providing evidence in favour of the fixed-effects model over the alternative specification (random effects).

6.3 Fixed effects results

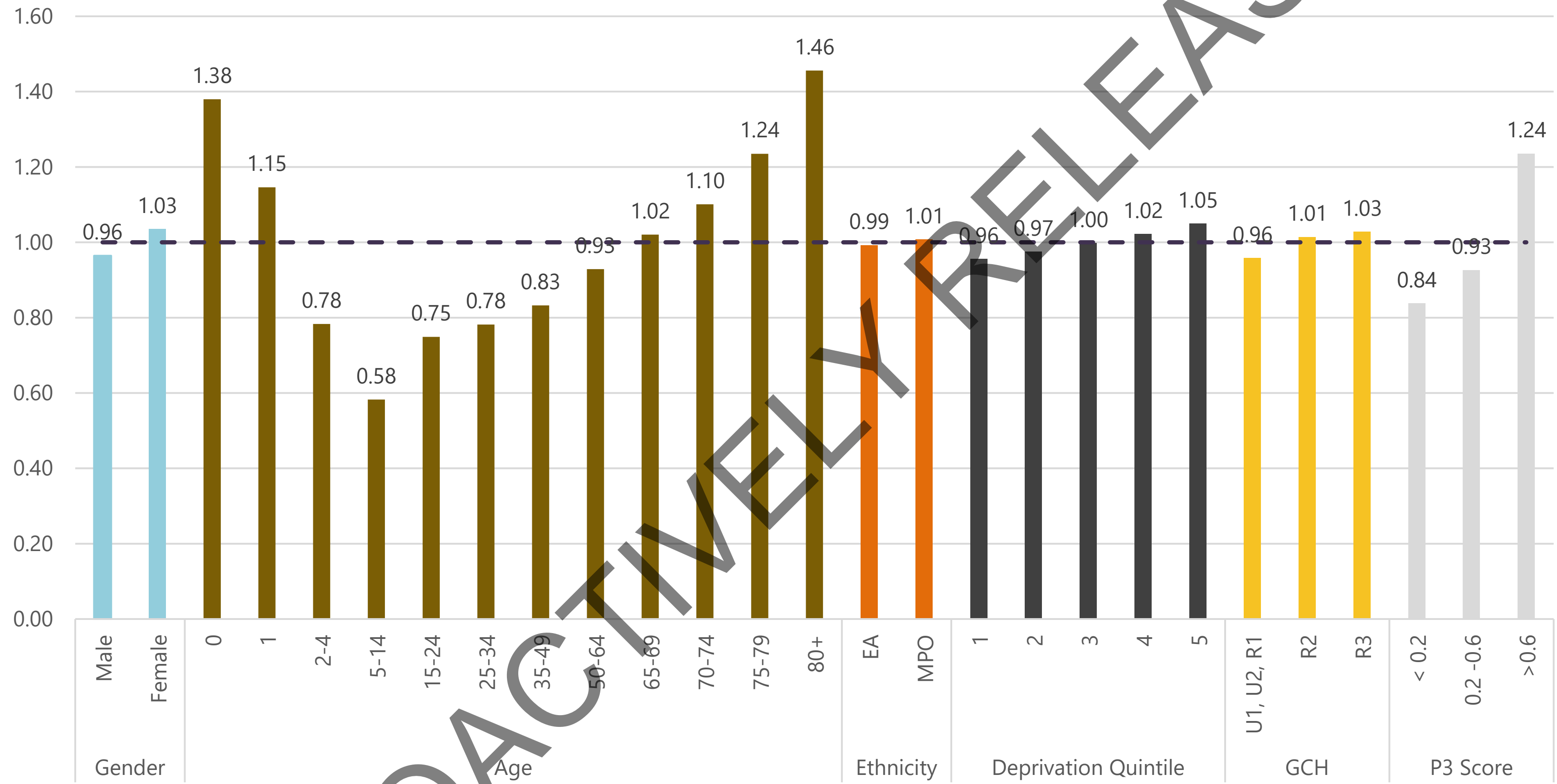
The estimated coefficients of the fixed effects model are reported in Table 13. The mean predicted cost weights for each explanatory characteristic used are shown in Figure 39. The model inevitably led to different cost weights for different individuals:

1. The patient with the **lowest** predicted cost weight became:
   - female, aged 5–14, European or Asian ethnicity, socio-economic deprivation quintile 1, P3 score < 0.2, lives in a U1, U2 or R1 area
2. The patient with the **highest** predicted cost weight became:
   - female, aged 80+, Māori, Pacific or Other ethnicity, socio-economic deprivation quintile 5, P3 score > 0.6, lives in an R3 area
3. The patient with **predicted cost weight = 1** (baseline individual) became:
   - male, aged 35–49, Māori, Pacific or Other ethnicity, socio-economic deprivation quintile 3, P3 Score > 0.6, lives in an R2 area.

We note that the characteristics of the baseline individual changed with the model, along with the following two key observations under the fixed effects model:

- **There was reduced variation within certain characteristics under the fixed effects model:** the fixed effects model produced significantly less variation in predicted cost weights within certain characteristics, such as rurality, age (infants and older patients), and ethnicity. While these groups still exhibited higher predicted weights, the magnitude of these differences was smaller compared to the AIM-R and AGIM models.
- **There was increased variation in multimorbidity under the fixed effects model**: by contrast, the fixed effects model revealed greater variation in cost weights for multimorbidity and socio-economic deprivation compared to the AGIM model. Patients with higher levels of multimorbidity were assigned relatively higher weights, while patients with lower levels of multimorbidity were assigned relatively lower weights. Furthermore, there is increased variation within socio-economic deprivation measured by NZDep quintiles, with a relatively 'cleaner' relationship between NZDep and assigned cost weights relative to the AIM-R and AGIM models.

Figure 39: Mean weights by characteristic (fixed-effects model)



Note: EA = European and Asian, MPO = Māori, Pacific and Other ethnicities

We can explain the differences in the modelling as follows:

- **Fixed effects removed between-practice variation:** the fixed effects model accounted for unobserved, time-invariant practice-level confounders (μ_j) by demeaning the data within each practice. This demeaning process removed variation caused by structural differences across practices, such as geographic location, infrastructure, or persistent demographic compositions. Characteristics such as age or ethnicity may correlate with the unobserved practice-level factors, and therefore showed less variation in predicted mean cost weights under the fixed effects model compared to the AIM-R and AGIM models, i.e., the AIM-R and AGIM models may overestimate the effect of these characteristics due to confounding. The most notable observation was the change in variation attributable to geography. In prior models, we have consistently seen that predicted cost weights for those in isolated rural areas were substantially higher than their more urban counterparts. While the predicted cost weights for R3 patients were still typically higher, using fixed effects substantially mitigated their explanatory power.

- **Multimorbidity and socio-economic deprivation could be within-practice drivers of costs:** by adding fixed effects, we controlled for average levels of health or deprivation at practices. In doing so, we removed confounding, resulting in the increase in magnitude of the explanatory power for multimorbidity and socio-economic deprivation.

In short, fixed effects might reveal a better statistical result, but the literature points to inequities being reinforced through a feedback loop between under-utilisation of primary care and the supply-side factors.

Table 13: Regression results from fixed effects models with cost as the dependent variable

	Fixed Effects
Age band	
1	-89.210***
	(5.982)
2–4	-232.995***
	(7.944)
5–14	-313.601***
	(17.728)
15–24	-205.923***
	(9.856)
25–34	-193.139***
	(9.237)
35–49	-175.806***
	(10.15)
50–64	-150.240***
	(10.604)
65–69	-117.050***
	(10.117)
70–74	-88.728***
	(10.281)
75–79	-30.123***
	(10.103)
80+	46.989***
	(10.275)
Gender	
Male	21.716***
	(4.429)
Other	-434.142***
	(34.565)
Ethnicity	

	Fixed Effects
Māori, Pacific and Other	6.267***
	(2.082)
NZDep2018 (quintiles)	
2	7.785***
	(1.299)
3	17.001***
	(1.636)
4	26.791***
	(1.897)
5	38.144***
	(2.619)
P3 Score	
0.2–0.6	36.068***
	(1.088)
> 0.6	161.550***
	(4.16)
GCH	
R3	5.719
	(22.196)
U1, U2 and R1	-22.607**
	(9.699)
Age band × male	
1	-11.444*
	(6.406)
2–4	-18.279***
	(4.831)
5–14	-19.742***
	(4.598)
15–24	-100.204***
	(5.413)

	Fixed Effects
25–34	-99.408*** (5.656)
35–49	-92.735*** (5.308)
50–64	-65.740*** (4.98)
65–69	-57.800*** (5.229)
70–74	-49.090*** (5.574)
75–79	-57.270*** (5.835)
80+	-32.205*** (6.409)
Age band × other	
1	
2–4	557.237*** (103.968)
5–14	844.136*** (143.113)
15–24	756.476*** (113.052)
25–34	578.117*** (52.409)
35–49	475.837*** (44.826)
50–64	1,301.555** (529.746)

	Fixed Effects
65–69	466.664***
	(126.579)
70–74	154.795**
	(69.132)
75–79	221.235***
	(34.791)
80+	86.425
	(167.102)
Constant	
AIC	104,480,132
BIC	104,487,125
Observations	7,231,501
R^2	
Within R^2	0.121
Adjusted R^2	0.215
Residual Std. Error (df = 2481560)	331.93
Note: *p<0.1; **p<0.05; ***p<0.01	

References

Abel, G. A., Gomez-Cano, M., Mustafee, N., Smart, A., Fletcher, E., Salisbury, C., Chilvers, R., Dean, S. G., Richards, S. H., Warren, F., & Campbell, J. L. (2020). Workforce predictive risk modelling: Development of a model to identify general practices at risk of a supply−demand imbalance. *BMJ Open*, *10*(1), e027934. https://doi.org/10.1136/bmjopen-2018-027934

Anselmi, L., Everton, A., Shaw, R., Suzuki, W., Burrows, J., Weir, R., Tatarek-Gintowt, R., Sutton, M., & Lorrimer, S. (2020). Estimating local need for mental healthcare to inform fair resource allocation in the NHS in England: Cross-sectional analysis of national administrative data linked at person level. *The British Journal of Psychiatry*, *216*(6), 338–344. https://doi.org/10.1192/bjp.2019.185

Anselmi, L., Lau, Y.-S., Sutton, M., Everton, A., Shaw, R., & Lorrimer, S. (2022). Use of past care markers in risk-adjustment: Accounting for systematic differences across providers. *The European Journal of Health Economics*, *23*(1), 133–151. https://doi.org/10.1007/s10198-021-01350-9

Barlow, P., Mohan, G., Nolan, A., & Lyons, S. (2021). Area-level deprivation and geographic factors influencing utilisation of General Practitioner services. *SSM - Population Health*, *15*, 100870. https://doi.org/10.1016/j.ssmph.2021.100870

Covvey, J. R., Johnson, B. F., Elliott, V., Malcolm, W., & Mullen, A. B. (2014). An association between socioeconomic deprivation and primary care antibiotic prescribing in Scotland. *The Journal of Antimicrobial Chemotherapy*, *69*(3), 835–841. https://doi.org/10.1093/jac/dkt439

Crampton, P., Sutton, F., & Foley, J. (2002). Capitation funding of primary care services: Principles and prospects. *The New Zealand Medical Journal*, *115*(1155), 271–274.

Gravelle, H., Sutton, M., Morris, S., Windmeijer, F., Leyland, A., Dibben, C., & Muirhead, M. (2003). Modelling supply and demand influences on the use of health care: Implications for deriving a needs-based capitation formula. *Health Economics*, *12*(12), 985–1004. https://doi.org/10.1002/hec.830

Love, T., Peck, C., & Watt, D. (2022). *A Future Capitation Funding Approach*. Sapere Research Group. https://srgexpert.com/wp-content/uploads/2022/12/FINAL-capitation-analysis-following-peer-review.pdf

Martínez-Pérez, J.-E., Quesada-Torres, J.-A., & Martínez-Gabaldón, E. (2024). Predicting healthcare expenditure based on Adjusted Morbidity Groups to implement a needs-based capitation financing system. *Health Economics Review*, *14*(1), 33. https://doi.org/10.1186/s13561-024-00508-4

Ministry of Health. (2023). *Annual Data Explorer 2022/23: New Zealand Health Survey*. https://minhealthnz.shinyapps.io/nz-health-survey-2022-23-annual-data-explorer/

Morris, T. P., White, I. R., & Royston, P. (2014). Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology*, *14*(1), 75. https://doi.org/10.1186/1471-2288-14-75

National Health Committee. (1998). *The Social, Cultural and Economic Determinants of Health in New Zealand: Action to Improve Health*. https://www.health.govt.nz/system/files/documents/publications/det-health.pdf

Parr-Brownlie, L. C., Waters, D. L., Neville, S., Neha, T., & Muramatsu, N. (2020). Aging in New Zealand: Ka haere ki te ao pakeketanga. *The Gerontologist*, *60*(5), 812–820. https://doi.org/10.1093/geront/gnaa032

Rice, N., Dixon, P., Lloyd, D. C. E. F., & Roberts, D. (2000). Derivation of a needs based capitation formula for allocating prescribing budgets to health authorities and primary care groups in England: Regression analysis. *BMJ : British Medical Journal*, *320*(7230), 284–288. https://doi.org/10.1136/bmj.320.7230.284

Royal New Zealand College of Practitioners. (2024). *Your Work Counts*. https://www.rnzcgp.org.nz/resources/our-voice/your-work-counts/

Salisbury, C., Johnson, L., Purdy, S., Valderas, J. M., & Montgomery, A. A. (2011). Epidemiology and impact of multimorbidity in primary care: A retrospective cohort study. *The British Journal of General Practice*, *61*(582), e12–e21. https://doi.org/10.3399/bjgp11X548929

Santosh, J., & Crampton, P. (2009). Gender differences in general practice utilisation in New Zealand. *Journal of Primary Health Care*, *1*(4), 261–269. https://doi.org/10.1071/HC09261

Stanley, J., Doughty, R. N., & Sarfati, D. (2020). A pharmaceutical dispensing–based index of mortality risk from long-term conditions performed as well as hospital record–based indices. *Medical Care*, *58*(2), e9. https://doi.org/10.1097/MLR.0000000000001217

Stanley, J., & Sarfati, D. (2017). The new measuring multimorbidity index predicted mortality better than Charlson and Elixhauser indices among the general population. *Journal of Clinical Epidemiology*, *92*, 99–110. https://doi.org/10.1016/j.jclinepi.2017.08.005

Statistics New Zealand. (2021). *LGBT+ population of Aotearoa: Year ended June 2020*. https://www.stats.govt.nz/reports/lgbt-plus-population-of-aotearoa-year-ended-june-2020

Statistics New Zealand. (2024, July 22). *National ethnic population projections, characteristics, 2018(base)–2043 update*. https://nzdotstat.stats.govt.nz/wbos/Index.aspx?DataSetCode=TABLECODE8614&_ga=2.260721994.610389.1710197299-1690322182.1708997154#

Tiruye, T., Roder, D., FitzGerald, L. M., O'Callaghan, M., Moretti, K., & Beckmann, K. (2024). Utility of prescription-based comorbidity indices for predicting mortality among Australian men with prostate cancer. *Cancer Epidemiology*, *88*, 102516. https://doi.org/10.1016/j.canep.2023.102516

Toi Te Ora Public Health. (2024, March). *Determinants of Health & Health Equity*.

https://toiteora.govt.nz/health-topics/determinants-of-health-and-health-equity

Appendix A Method for calculating GP, nurse, and nurse practitioner FTE

Assigning provider template types and roles

To calculate FTE and activity using provider templates we first needed to classify the type of template and the role of the provider.

Provider templates are a name given to the appointment books used in general practice. There is usually a template for every GP and nurse practitioner within a practice. These templates allow staff to book patients in with a provider, set the duration of the appointment, write basic notes about the appointment and track key milestones such as when a patient arrives, when they're in consult and when they have been invoiced.

The data showed us that in many practices, nurses tended not to have their own named appointment book. Generic provider templates were instead used to manage different roles and work within a practice e.g. "acute nurse". Generic templates appeared to be a pragmatic way for practices to simplify booking processes, manage nurse activity and allow for the seamless movement of nurses between different roles without the need to reshuffle templates.

Finally, provider templates were also used as a way of managing room bookings and resources within the practice. We wanted to exclude this data from the analysis where this occurred.

Determining provider template type

There were four template types defined for this analysis. These were Provider, Generic, Services, Admin. The definition was important because it:

a. helped to determine the provider role for the template, i.e. GP, NP, nurse
b. defined how FTE was calculated for that template.

The table below describes the different template types, how FTE was calculated for the different types, and the reason why it was calculated that way.

Table 14: Description of the four template types

Template type	Description	Template examples	Default provider role[1]	Is contact time or non-contact time included in FTE calcs?		
				Contact	Non-contact	Reason
Provider	Person-specific templates	Dr Sam Smith, Nurse Jill Jones	Other	Yes	Yes	
Generic	A template for a job or role, not person-specific	Acute Nurse, Clinical Practice Nurse, Admin Nurse, Script Nurse	Nurse	Yes	Yes	The assumption was that only one person would be assigned to work from a generic template at any given time. These templates were treated the same as a Provider template for the purpose of calculating FTE.
Services	A template for a specific service or procedure	GP Triage, Flu clinic Respiratory clinic, Care plus clinic, B4 School Checks, Nurse Phone Consultations	Nurse	Yes	No	These templates operated like a task list or queue so only contact time was included in FTE calculations.
Admin	Templates used to manage administrative tasks, resource bookings, or anything else templated in practice	Scripts, Admin phone, Clinic room 3, Car 1	Other	No	No	These templates were not used in FTE calculations. The assumption was that if any of these templates were used to manage work done by GPs, NPs or nurses, the time allocated to these tasks was captured in the other templates. For example, a GP had 15 minutes allocated in their Provider template for repeat prescriptions – in that time they would bring up the “script” template and work through the repeat prescription on that list. We therefore excluded Admin templates from FTE calculations to avoid double counting.

1. If the role (either GP, NP or nurse) cannot be determined from the template name or role code, then the default role is used.

How the template type was determined

Data source: providerName and providerType were used to determine the template type using a set of logic statements (conditions). The template was categorised the first time a condition was met so the order of the condition was important. For example, the template "Triage Phone Nurse" would be categorised as a service template because it met condition one, i.e. contained the word "triage". If condition two came before one, then this template would have been categorised as "Generic" as it would have met the condition that it contained the phrase "phone nurse".

The following table lists conditions that define the template type.

Table 15: Conditions defining template type

Order	Condition (conditions are not case sensitive)	Template type
1	providerName contains: • Triage	Services
2	providerName contains any of the following: • acute nurse • clinical nurse • admin nurse • practice nurse • phone nurse • script nurse • prescription nurse • rx nurse • registrar • student	Generic
3	providerName contains: • admin	Admin
4	providerName contains any of the following: • phone • triage • immunisation • vaccine • diabetes • respiratory • menz • hpv • b4 • cervical • smear • ltc • queue • immigration • aclasta OR starts with any of the following: • flu • imms	Services

	OR ends with: • imms	
5	providerName contains any of: • medical • practice	Services
6	providerName contains any of: • room • clinic 1 – clinic 6 • outreach • rx • script • prescription • medical • the doctors • txt • sms • messages OR starts with: • practice	Admin
7	providerName contains: • clinic	Services
8	Anything that did not match the above conditions	Provider

Examples: "Rx Nurse Jane Doe" would be classified as "Generic" as in condition two, not in condition six. However, "Rx Note" would be classified as "Admin".

How the provider role was determined

Data source: the same data source as that used to determine the template type.

There were four provider roles defined:

General practitioner (GP)
Nurse practitioner (NP)
Nurse
Other.

Due to inconsistencies in how staff were set up in the various practice management systems and the extensive use of provider templates that were not person-specific, the role for each provider template needed to be further refined using a providerName and the template type. This has been done using the following logic:

Table 16: Conditions to define provider role

Order	Condition (conditions are not case sensitive)	Role

1	Marked as General Practitioner, Nurse Practitioner or Nurse (as per DataCraft specifications)	GP, NP, or Nurse respectively
2	NAMEEXTERN begins with either: • dr • doctor OR ends with: • gp	GP
3	NAMEEXTERN contains: • nurse practitioner OR starts with • "np " (note space after p) OR ends with " np" (note space before n)	Nurse Practitioner
4	Name contains "nurse"	Nurse
5	Role code = NZMC	GP
6	Role code = NZNC	Nurse
7	Template type is "Generic" or "Services"	Nurse
8	Anything that did not match the above conditions	Other

Templates excluded

Templates used to manage after hours or accident and medicine services were excluded. The logic to exclude these was as follows:

Table 17: Conditions to exclude templates used to manage after hours and accident and medicine services

Order	Condition (conditions are not case sensitive)
1	providerName contained any of: • after hours • a&m • accident & medical • accident and medical

Calculating FTE

Base table

The calculations were done using a base table of data. The base table is very granular, and each row represents an appointment slot on a provider template. The base table is the appointment data with additional columns added with information required to calculate FTE or Contacts. The following table describes the base dataset, including the columns calculated during this analysis.

Table 18: Base table used in the FTE analysis

Fields	Description	Example	Source
phoName	PHO Name	s 9(2)(ba)(i), s 9	DataCraft
facilityId	An internal practice ID that can be used to match to the HPI code	52	DataCraft
hpiCode	Health Practitioner Index code used to identify a practice	F3S777-K	PHO Register
practiceName	Practice name	s 9(2)(ba)(i), s 9(2)(b)(ii)	PHO Register
firstContact	When the appointment was made	2018-03-12 15:04:24	APPLINE
appoint	Date and time of the appointment	2018-04-12 09:30:00	APPLINE
arrived	When the "arrived" button is ticked	2018-04-12 09:26:13	APPLINE
consult	When the in-consult button is ticked	2018-04-12 09:31:10	APPLINE
invoice	When an invoice is created for the patient	2018-04-12 09:43:29	APPLINE
duration	Duration the appointment is booked for in minutes; duration = avgDuration where dblBooked = 1	15	APPLINE
providerId	Unique provider code within the practice	JS	APPLINE
providerName	Name of the provider template	Dr John Smith	APPLINE
patientId	Internal patient identifier	A012345	APPLINE
nhiEncypted	Encrypted NHI	0x0BAF6F71625D0D	DataCraft
encounter	1 if notes were written for the patient by the provider on the same day, else 0	1	ENCOUNTERS
prescription	1 if a prescription was written for the patient by the provider on the same day, else 0	0	SCRIPT
providerType	GP, NP, Nurse or Other	GP	DataCraft / Calculated
templateType	Provider, Generic, Services, Admin	Provider	Calculated
excludedService	Used to flag included/excluded services such as templates for A&M services	0	Calculated
dummyPat	Flag = 1 if the patient is identified as not a real person	0	Calculated

patientContact	1 if patient contact, else 0	1	Calculated
include	1 if the row is to be included in the FTE and Contacts calculations	1	Calculated
dailyContacts	Total contacts for the provider for the day	18	Calculated
annualWorkdays	Workdays less weekends, public holidays and one anniversary day	209	Calculated
Annualfte	= Apptduration / (60 x 8 x Annualworkdays)	0.000062751004	Calculated
ageBand	The age band of the patient	30–34	Calculated from DataCraft /NES
ethnicity	The high-level ethnic group of the patient	Māori	DataCraft/NES
deprivationQuint	The socio-economic deprivation quintile of the patient's residence	3	DataCraft /NES
gender	The gender that the patient identifies with	Male	DataCraft /NES
p3score / p3Group	The P3 (multimorbidity) score of the patient, these are later grouped	0.18 / < 0.2	National Collections / Calculated
gch	Geographical Classification of Health	R2	University of Otago

Exclude descriptive rows created for extended bookings

When an extended booking is made for a patient, s 9(2)(ba)(i), s 9(2)(b)(ii creates a duplicate row for the appointment. This duplicate was created as a way of showing the user the appointment has been extended and its length (in the patient name column the user will see something like "(30 MINUTES DURATION)"). These rows needed to be excluded to avoid double-counting the time.

A duplicate was identified where there was more than one row for a patient at the same appointment date with the same provider. The maximum duration of all the appointments per patient, per provider, per day was used. Any rows that came after the initial row for the patient were excluded.

Example:

Table 19: Example of a duplicate appointment record for an extended appointment that would be excluded

providerId	Appoint	patientId	Name[1]	duration	Row index	include
JS	2023-07-23 09:30:00	A012345	John Smith	30	1	1
JS	2023-07-23 09:30:02	A012345	(DURATION 30 MINUTES)	30	2	0

JS	2023-07-23 10:00:00	NULL	Morning tea	15	3	1

1. The patient's name was not included in the dataset; it is only included here to demonstrate the reason for the exclusion

Excluded provider time if no contacts in a day

It is common practice for templates to be disabled if a provider is on leave or not rostered on as a way of ensuring practice staff do not accidentally book a patient with the provider. However, there were some examples where a note was written in the template instead. To correct for this issue, and avoid counting this time, we excluded days where there were no patient bookings on a provider's template.

How: Counted patientId by day and provider. If the count was zero, then excluded the entire day for that provider in the final FTE calculations.

Identified probable dummy patient IDs and made the time non-contact

Practices have dummy patient IDs for testing new tools within the PMS, for demonstration purposes or to have a standard name for regular meetings or activities. Dummy patient IDs shouldn't be counted as a patient contact.

How dummy patients were identified:

- In the dataset we did not have a patient name or a list of dummy patients. To identify the majority of these patients we assumed they would be used regularly within a practice and have an extremely high number of appointments in a year. Therefore, the following definition was used:
    - If the number of annual appointments for a patientId was > 400 then it was flagged as a dummy patient.

Further, we divided the population into 10 quintiles based on the number of appointments they had in a year. If the patient had no encrypted NHI number attached to them AND were in the 10th quintile of appointment numbers, then they were also flagged as a dummy patient. The underlying rationale behind this was that if the patient were high-needs, they would be more likely to be enrolled and therefore have an NHI number attached to their record.

Determined patient contact

A flag to indicate patient contact was added for each row. A patient contact was any row where a patient ID was entered and there was either a note written, script written, or some other indication that a patient had been in consultation such as a time stamp for when they arrived, were in consult or were invoiced. Exclusions to this rule were that we did not count dummy IDs or "TEMPPAT" (which usually signified something written in by staff, e.g. "meeting" rather than a patient booking) as a patient contact. Specifically, the logic was as follows:

Table 20: Conditions used to flag patient contacts

When	Then	Else

patientId is not blank AND patientId does not equal "TEMPPAT" or "TMPPATI" AND dummyPatient=0 AND (arrived (arrival time) is not blank OR consult (consultation time) is not blank OR invoice (time of invoice) is not blank OR encounter = 1 OR prescription = 1)	patientContact = 1	patientContact = 0

Summary of which rows were excluded in the calculations

Rows were excluded as follows:

- They were descriptive rows created for extended bookings
- The provider had no bookings during the day
- The provider template was for an excluded service, e.g. A&M
- The template type was "Admin"
- The template type was "Service" and patientContact = 0

The reasons for these exclusions have been described earlier.

After making these exclusions, our dataset comprised close to 1.2 million non-contact rows, and around 14.8 million contact rows.

FTE calculations

FTE = duration / (60 x 8 x annualWorkdays)

Where: annualWorkdays = the number of weekdays in the 2023 calendar year less public holidays and expected leave days. There were 11 public holidays that fell on a weekday in 2023:

New Year's Day	1 January 2023 (observed 3 January)
Day after New Year's Day	2 January 2023
Waitangi Day	6 February 2023
Good Friday	7 April 2023
Easter Monday	10 April 2023
Anzac Day	25 April 2023
King's Birthday	5 June 2023
Labour Day	23 October 2023
Christmas Day	25 December 2023
Boxing Day	26 December 2023

Plus one regional anniversary day.

For GPs, the number of leave days was determined by the Senior Medical Officers' Collective Agreement. This agreement is used to encourage a sustainable GP workforce whereby GPs get similar benefits to SMOs. The number of leave days was therefore calculated as 30 days of annual leave plus a provision of 10 days to account for expected sick leave (5 days) and professional development (5 days). annualWorkdays = 209 for GPs.

We used the New Zealand Nurses Organisation/Health New Zealand | Te Whatu Ora collective agreement for nursing to determine the number of leave days for nurses and nurse practitioners, calculated as 20 days of annual leave plus 4 days of professional development and 5 days of sick leave.

Apportioned non-contact time

Non-contact time rarely had a patient ID associated with it. Therefore, we apportioned non-contact time across all patients with an NHI by practice and role.

Converting time to cost

We converted contact and non-contact time to a cost. This helped determine the relative differences in time used by different groups, weighted by the role of the clinician who provided the service. To do this, we first converted the FTE calculated for GPs, nurses and nurse practitioners to minutes, and assigned an average salary to each clinician type. This was assigned as:

- $226,680 for GPs[13]
- $106,739 for nurses[14]
- $143,946 for nurse practitioners[15]

The duration of each appointment was then calculated to an FTE equivalent of the provider, and multiplied by the salary to determine a cost.

[13] Based on the midpoint of the Senior Medical and Dental Officers' Collective Agreement.
[14] Based on the upper end of the scale of the Nursing and Midwifery Collective Agreement, the assumption was most nurses have been in practice for a long time.
[15] Based on the midpoint of the Nursing and Midwifery Collective Agreement.

Appendix B Creating variate groupings

This section shows the results of our multivariate regression analysis using disaggregated groups. Table 21 is presented to illustrate the demographic make-up of the population of interest at a granular level.

Table 21: Summary statistics of the population of interest prior to grouping and including inactive patients

Variable	Number of patients	Mean annual FTE	Mean predicted cost ($)
Age band			
0	35,685	0.0020	383.53
1	29,219	0.0016	313.86
2	24,562	0.0010	199.06
3	23,676	0.0010	196.14
4	29,343	0.0010	191.92
5–9	108,792	0.0006	131.39
10–14	114,219	0.0006	130.55
15–19	106,965	0.0008	161.16
20–24	103,048	0.0009	183.04
25–29	119,687	0.0009	180.37
30–34	139,084	0.0009	180.97
35–39	130,549	0.0009	185.01
40–44	121,390	0.0009	192.02
45–49	121,322	0.0010	203.24
50–54	135,673	0.0010	214.97
55–59	132,551	0.0011	228.01
60–64	135,929	0.0012	246.34
65–69	120,851	0.0013	273.81
70–74	105,298	0.0015	304.41
75–79	79,059	0.0017	355.18
80+	89,992	0.0021	439.82
Gender			
Other	726	0.0016	339.80
Female	1,074,438	0.0011	238.87
Male	931,730	0.0010	204.98
Ethnicity			
Other	43,244	0.0009	195.51
Asian	259,443	0.0008	178.04
European	1,292,779	0.0011	229.15
Māori	291,250	0.0011	231.29
Pacific	120,178	0.0012	246.63
Deprivation quintile			

Variable	Number of patients	Mean annual FTE	Mean predicted cost ($)
1	493,442	0.0010	201.17
2	414,172	0.0010	213.79
3	386,289	0.0011	224.59
4	373,383	0.0011	235.09
5	339,608	0.0012	251.87
GCH			
R1	223,054	0.0011	229.12
R2	112,006	0.0013	253.46
R3	23,459	0.0016	321.14
U1	1,285,011	0.0010	216.10
U2	363,364	0.0011	228.87
P3 score			
< 0.2	728,843	0.0009	179.95
0.2–0.6	722,013	0.0010	207.47
> 0.6	556,038	0.0014	300.21

Results from disaggregated regression analysis

The inclusion of interactive terms with granular age groupings can create some volatile estimates in both the main effects and interaction effects. While there was a strong basis to include interactive effects, we observed:

- Ages 60–64: the estimated coefficient was extremely high relative to other age bands, but this was offset by extremely low estimated coefficients with the gender interactions for the same age band.
- Ages zero, two and four: the interaction model specifications did not produce estimated coefficients for male interactions.

These observations likely occurred due to sparse data, and the model specification attributing too much effect to the interaction term. Regrouping the population cohorts as in section 1.1 largely mitigated this issue.

Table 22 shows that:

- Age–gender interactions were significantly different from the reference age group during teen and adult years; these interactions reflected different activity for males and females during child-bearing years.
- Age–ethnicity interactions were also significant for adult age groups, particularly for Māori and Pacific peoples. The significant interaction effects for these groups supported grouping Māori, Pacific peoples and Other (the baseline category) into one ethnic grouping for a simpler model.

- Estimated Age–P3 coefficients were also statistically significant for those aged 10+.

Table 22: Regression estimates from specified models with cost as the dependent variable

	AIM-R (specified model)	SLM	AIM	AIM-R-P3
Age band				
0	Reference category			
1	-81.259***	-71.013***	-81.482***	-85.348***
	(-105.477, -57.040)	(-75.056, -66.970)	(-105.718, -57.246)	(-111.477, -59.218)
2	-169.603***	-184.554***	-169.785***	-177.598***
	(-195.823, -143.383)	(-188.802, -180.305)	(-196.025, -143.546)	(-205.680, -149.515)
3	187.942	-186.764***	214.673	178.572
	(-229.605, 605.488)	(-191.059, -182.469)	(-203.171, 632.517)	(-237.554, 594.699)
4	-173.290***	-190.722***	-173.854***	-172.471***
	(-198.095, -148.485)	(-194.760, -186.683)	(-198.676, -149.031)	(-198.995, -145.948)
5–9	31.206	-248.853***	30.448	1.372
	(-330.121, 392.532)	(-251.979, -245.727)	(-331.137, 392.034)	(-358.662, 361.406)
10–14	276.803	-247.878***	275.654	270.044
	(-27.318, 580.924)	(-250.988, -244.769)	(-28.685, 579.993)	(-33.037, 573.125)
15–19	327.540*	-220.861***	327.848*	298.354*
	(29.790, 625.289)	(-223.997, -217.725)	(29.885, 625.811)	(1.618, 595.091)
20–24	369.689*	-203.795***	369.825*	322.107*
	(72.045, 667.333)	(-206.945, -200.645)	(71.968, 667.683)	(25.471, 618.743)
25–29	272.217	-206.605***	272.484	214.121
	(-26.296, 570.730)	(-209.698, -203.512)	(-26.243, 571.211)	(-83.375, 511.617)
30–34	257.872	-204.813***	258.126	193.646
	(-43.458, 559.202)	(-207.857, -201.770)	(-43.420, 559.672)	(-106.651, 493.944)
35–39	217.951	-198.117***	217.964	150.544
	(-89.242, 525.144)	(-201.182, -195.052)	(-89.449, 525.378)	(-155.593, 456.681)
40–44	271.916	-190.190***	271.689	195.951
	(-45.657, 589.490)	(-193.279, -187.100)	(-46.112, 589.490)	(-120.523, 512.425)
45–49	265.154	-180.020***	265.086	173.445

	AIM-R (specified model)	SLM	AIM	AIM-R-P3
	(-67.760, 598.068)	(-183.111, -176.929)	(-68.066, 598.239)	(-158.307, 505.197)
50–54	335.735*	-169.290***	337.203*	230.778
	(2.753, 668.717)	(-172.346, -166.234)	(3.982, 670.423)	(-101.036, 562.591)
55–59	256.199	-157.514***	258.296	138.494
	(-96.552, 608.951)	(-160.578, -154.450)	(-94.708, 611.300)	(-213.005, 489.994)
60–64	2,272.549***	-139.953***	2,272.295***	2,130.224***
	(1,806.173, 2,738.925)	(-143.011, -136.895)	(1,805.585, 2,739.006)	(1,665.566, 2,594.881)
65–69	266.357	-113.075***	265.574	117.418
	(-150.945, 683.658)	(-116.175, -109.976)	(-152.027, 683.175)	(-298.365, 533.202)
70–74	96.431	-84.742***	104.729	-60.191
	(-370.128, 562.991)	(-87.896, -81.588)	(-362.164, 571.623)	(-525.024, 404.642)
75–79	117.753	-38.540***	119.58	-87.492
	(-471.777, 707.284)	(-41.828, -35.253)	(-470.373, 709.533)	(-674.808, 499.823)
80+	252.409	36.603***	260.999	-13.295
	(-214.397, 719.216)	(33.371, 39.836)	(-206.140, 728.139)	(-478.410, 451.821)
		Gender		
Female	356.371*	-143.327***	357.788*	336.985*
	(61.687, 651.055)	(-162.359, -124.296)	(62.893, 652.683)	(43.442, 630.528)
Male	369.817*	-176.970***	371.437*	350.184*
	(75.135, 664.499)	(-196.003, -157.937)	(76.544, 666.330)	(56.642, 643.725)
Other		Reference category		
		Ethnicity		
Asian	5.966	-22.107***	5.496	6.549
	(-11.164, 23.096)	(-24.769, -19.445)	(-11.646, 22.639)	(-10.515, 23.613)
European	23.314**	3.562**	23.975**	24.204**
	(6.927, 39.702)	(1.044, 6.080)	(7.576, 40.374)	(7.880, 40.528)
Māori	-31.856***	20.663***	-29.750***	-31.729***
	(-48.607, -15.104)	(18.007, 23.319)	(-46.512, -12.987)	(-48.415, -15.042)
Pacific	-16.026	34.960***	-16.918	-15.515
	(-34.557, 2.504)	(32.068, 37.851)	(-35.462, 1.626)	(-33.974, 2.943)

	AIM-R (specified model)	SLM	AIM	AIM-R-P3
Other	Reference category			
Deprivation quintile				
2	9.645***	10.324***	10.275***	9.226***
	(8.566, 10.724)	(9.242, 11.405)	(9.197, 11.352)	(8.152, 10.301)
3	18.393***	19.030***	18.914***	17.671***
	(17.286, 19.500)	(17.924, 20.135)	(17.811, 20.016)	(16.568, 18.774)
4	24.792***	25.985***	25.726***	23.504***
	(23.661, 25.923)	(24.860, 27.111)	(24.604, 26.849)	(22.378, 24.630)
5	40.901***	43.085***	42.926***	39.218***
	(39.685, 42.117)	(41.876, 44.293)	(41.721, 44.132)	(38.007, 40.430)
1	Reference Category			
P3				
P3 Group	44.705***	45.327***	44.565***	6.669***
	(44.245, 45.164)	(44.866, 45.787)	(44.105, 45.024)	(2.992, 10.347)
Geographical Classification of Health				
R1	Reference category			
R2	20.347***			20.640***
	(18.473, 22.220)			(18.774, 22.507)
R3	83.435***			84.237***
	(79.921, 86.950)			(80.736, 87.737)
U1	1.454*			1.118
	(0.265, 2.643)			(-0.066, 2.303)
U2	-1.683*			-1.840**
	(-3.058, -0.308)			(-3.210, -0.470)
Age band × female				
1	3.909		4.063	2.875
	(-4.147, 11.966)		(-3.999, 12.125)	(-5.151, 10.901)
2	10.316*		10.543*	9.583*
	(1.848, 18.784)		(2.069, 19.017)	(1.148, 18.019)

	AIM-R (specified model)	SLM	AIM	AIM-R-P3
3	-347.76		-374.775	-345.223
	(-764.528, 69.009)		(-791.841, 42.290)	(-760.379, 69.933)
4	8.631*		8.852*	7.052
	(0.582, 16.679)		(0.798, 16.905)	(-0.966, 15.070)
5–9	-271.429		-270.999	-241.106
	(-632.341, 89.483)		(-632.170, 90.171)	(-600.618, 118.407)
10–14	-516.379***		-515.590***	-501.898**
	(-819.956, -212.801)		(-819.385, -211.795)	(-804.300, -199.496)
15–19	-522.218***		-522.850***	-504.506***
	(-819.368, -225.067)		(-820.214, -225.486)	(-800.506, -208.506)
20–24	-541.275***		-541.693***	-524.517***
	(-838.321, -244.229)		(-838.952, -244.435)	(-820.413, -228.621)
25–29	-448.416**		-448.918**	-429.757**
	(-746.387, -150.445)		(-747.102, -150.734)	(-726.574, -132.940)
30–34	-436.521**		-436.855**	-417.637**
	(-737.371, -135.671)		(-737.921, -135.790)	(-717.322, -117.952)
35–39	-386.053*		-386.105*	-367.860*
	(-692.774, -79.333)		(-693.046, -79.165)	(-673.392, -62.327)
40–44	-430.812**		-430.868**	-412.283*
	(-747.911, -113.713)		(-748.194, -113.543)	(-728.154, -96.413)
45–49	-409.352*		-409.341*	-389.776*
	(-741.800, -76.904)		(-742.028, -76.655)	(-720.936, -58.616)
50–54	-476.761**		-478.256**	-456.556**
	(-809.208, -144.314)		(-810.941, -145.570)	(-787.715, -125.397)
55–59	-386.608*		-388.764*	-366.631*
	(-738.819, -34.398)		(-741.226, -36.301)	(-717.476, -15.786)
60–64	-2,401.074***		-2,400.714***	-2,372.345***
	(-2,867.000, -1,935.148)		(-2,866.974, -1,934.454)	(-2,836.462, -1,908.227)
65–69	-376.83		-375.898	-357.065
	(-793.562, 39.902)		(-792.928, 41.132)	(-772.179, 58.049)

	AIM-R (specified model)	SLM	AIM	AIM-R-P3
70–74	-172.807		-181.355	-175.106
	(-638.722, 293.108)		(-647.603, 284.894)	(-639.212, 289.001)
75–79	-149.871		-151.354	-141.258
	(-739.668, 439.926)		(-741.573, 438.866)	(-728.763, 446.247)
80+	-269.098		-277.559	-209.251
	(-735.012, 196.817)		(-743.807, 188.688)	(-673.359, 254.856)
		Age band × male		
1				
2				
3	-358.416		-385.372	-354.22
	(-775.180, 58.348)		(-802.434, 31.689)	(-769.371, 60.931)
4				
5–9	-281.59		-281.39	-250.737
	(-642.500, 79.320)		(-642.558, 79.779)	(-610.247, 108.774)
10–14	-531.042***		-530.494***	-516.708***
	(-834.617, -227.466)		(-834.287, -226.701)	(-819.109, -214.308)
15–19	-585.643***		-586.277***	-568.067***
	(-882.794, -288.493)		(-883.640, -288.913)	(-864.067, -272.067)
20–24	-620.142***		-620.638***	-603.858***
	(-917.189, -323.095)		(-917.898, -323.379)	(-899.755, -307.961)
25–29	-522.706***		-523.298***	-504.574***
	(-820.678, -224.735)		(-821.482, -225.113)	(-801.392, -207.756)
30–34	-506.689***		-507.062***	-488.454**
	(-807.539, -205.839)		(-808.128, -205.996)	(-788.139, -188.769)
35–39	-454.441**		-454.663**	-436.485**
	(-761.161, -147.720)		(-761.603, -147.722)	(-742.018, -130.952)
40–44	-495.681**		-495.867**	-477.003**
	(-812.780, -178.582)		(-813.193, -178.541)	(-792.873, -161.132)

	AIM-R (specified model)	SLM	AIM	AIM-R-P3
45–49	-477.030**		-477.104**	-457.186**
	(-809.478, -144.582)		(-809.790, -144.418)	(-788.346, -126.026)
50–54	-536.332**		-537.925**	-515.843**
	(-868.778, -203.885)		(-870.610, -205.241)	(-847.001, -184.685)
55–59	-435.274*		-437.606*	-414.959*
	(-787.484, -83.064)		(-790.068, -85.144)	(-765.803, -64.115)
60–64	-2,439.508***		-2,439.199***	-2,410.450***
	(-2,905.434, -1,973.583)		(-2,905.459, -1,972.940)	(-2,874.567, -1,946.333)
65–69	-416.545		-415.488	-396.534
	(-833.276, 0.186)		(-832.518, 1.541)	(-811.647, 18.580)
70–74	-206.883		-215.127	-209.006
	(-672.797, 259.032)		(-681.374, 251.121)	(-673.112, 255.100)
75–79	-186.788		-188.05	-178.779
	(-776.586, 403.009)		(-778.271, 402.170)	(-766.285, 408.726)
80+	-287.032		-295.28	-227.631
	(-752.947, 178.883)		(-761.528, 170.968)	(-691.739, 236.477)
		Age band × Asian		
1	13.97		14.117	11.205
	(-11.823, 39.763)		(-11.694, 39.929)	(-14.488, 36.899)
2	-7.578		-7.526	-8.704
	(-35.334, 20.179)		(-35.302, 20.251)	(-36.353, 18.945)
3	4.587		4.913	3.153
	(-23.141, 32.316)		(-22.835, 32.661)	(-24.468, 30.774)
4	-4.109		-3.602	-4.707
	(-30.389, 22.171)		(-29.901, 22.697)	(-30.885, 21.471)
5–9	-2.855		-2.28	-3.548
	(-22.697, 16.988)		(-22.136, 17.577)	(-23.314, 16.217)
10–14	-9.132		-8.472	-11.609
	(-29.299, 11.036)		(-28.653, 11.710)	(-31.698, 8.481)
15–19	-29.622**		-29.134**	-31.429**

	AIM-R (specified model)	SLM	AIM	AIM-R-P3
	(-50.556, -8.688)		(-50.083, -8.185)	(-52.282, -10.576)
20–24	-27.068*		-26.781*	-28.549**
	(-47.994, -6.141)		(-47.722, -5.840)	(-49.394, -7.704)
25–29	-30.390**		-30.041**	-31.075**
	(-50.114, -10.666)		(-49.779, -10.303)	(-50.723, -11.428)
30–34	-26.914**		-26.737**	-27.217**
	(-45.784, -8.045)		(-45.619, -7.854)	(-46.013, -8.422)
35–39	-30.841**		-30.581**	-31.448***
	(-49.612, -12.070)		(-49.366, -11.797)	(-50.146, -12.750)
40–44	-34.807***		-34.178***	-35.460***
	(-53.862, -15.752)		(-53.246, -15.109)	(-54.440, -16.479)
45–49	-34.646***		-34.377***	-35.091***
	(-54.441, -14.851)		(-54.186, -14.568)	(-54.809, -15.373)
50–54	-33.319**		-33.062**	-33.518**
	(-53.932, -12.707)		(-53.689, -12.435)	(-54.050, -12.985)
55–59	-42.352***		-42.012***	-41.682***
	(-63.653, -21.051)		(-63.328, -20.696)	(-62.901, -20.464)
60–64	-32.698**		-32.622**	-32.381**
	(-54.911, -10.485)		(-54.851, -10.394)	(-54.507, -10.254)
65–69	-34.278**		-34.265**	-31.738**
	(-57.760, -10.796)		(-57.764, -10.766)	(-55.130, -8.347)
70–74	-56.908***		-56.651***	-53.379***
	(-83.143, -30.673)		(-82.905, -30.398)	(-79.512, -27.245)
75–79	-69.389***		-69.727***	-65.855***
	(-99.614, -39.164)		(-99.973, -39.480)	(-95.963, -35.747)
80+	-76.152***		-76.326***	-69.993***
	(-106.882, -45.421)		(-107.079, -45.574)	(-100.605, -39.381)
		Age band × European		
1	8.541		8.773	7.003
	(-16.100, 33.181)		(-15.885, 33.431)	(-17.542, 31.548)

	AIM-R (specified model)	SLM	AIM	AIM-R-P3
2	-39.439**		-39.170**	-41.134**
	(-66.078, -12.800)		(-65.828, -12.512)	(-67.670, -14.598)
3	-39.797**		-39.365**	-40.864**
	(-66.423, -13.171)		(-66.010, -12.720)	(-67.387, -14.342)
4	-42.810***		-42.139**	-43.167***
	(-68.023, -17.597)		(-67.371, -16.908)	(-68.283, -18.052)
5–9	-23.339*		-22.706*	-24.358*
	(-42.309, -4.368)		(-41.690, -3.722)	(-43.255, -5.461)
10–14	-16.198		-15.575	-19.542*
	(-35.393, 2.996)		(-34.784, 3.633)	(-38.662, -0.421)
15–19	-6.811		-6.551	-8.484
	(-26.601, 12.979)		(-26.355, 13.253)	(-28.197, 11.229)
20–24	-7.892		-7.881	-9.728
	(-27.701, 11.916)		(-27.703, 11.942)	(-29.460, 10.004)
25–29	-12.615		-12.59	-13.972
	(-31.437, 6.208)		(-31.425, 6.245)	(-32.722, 4.777)
30–34	-12.612		-12.619	-13.233
	(-30.656, 5.432)		(-30.676, 5.437)	(-31.207, 4.741)
35–39	-20.227*		-20.004*	-21.131*
	(-38.172, -2.282)		(-37.962, -2.047)	(-39.006, -3.255)
40–44	-25.665**		-25.260**	-26.634**
	(-43.852, -7.478)		(-43.460, -7.060)	(-44.751, -8.517)
45–49	-31.843***		-31.648***	-32.866***
	(-50.662, -13.025)		(-50.480, -12.816)	(-51.611, -14.120)
50–54	-29.760**		-29.484**	-30.704**
	(-49.360, -10.161)		(-49.097, -9.870)	(-50.227, -11.181)
55–59	-35.356***		-34.796***	-35.433***
	(-55.596, -15.116)		(-55.051, -14.542)	(-55.594, -15.271)
60–64	-24.194*		-23.673*	-24.590*

	AIM-R (specified model)	SLM	AIM	AIM-R-P3
	(-45.352, -3.036)		(-44.846, -2.500)	(-45.666, -3.514)
65–69	-10.61		-9.945	-8.421
	(-33.061, 11.842)		(-32.413, 12.522)	(-30.786, 13.944)
70–74	-17.752		-16.756	-13.605
	(-42.869, 7.364)		(-41.890, 8.378)	(-38.625, 11.415)
75–79	-11.193		-11.174	-7.85
	(-39.985, 17.598)		(-39.985, 17.638)	(-36.530, 20.830)
80+	48.728**		48.540**	51.907***
	(19.384, 78.072)		(19.175, 77.905)	(22.677, 81.138)
		Age band × Māori		
1	6.638		6.805	5.676
	(-18.578, 31.854)		(-18.428, 32.039)	(-19.442, 30.794)
2	-6.309		-6.11	-6.847
	(-33.583, 20.965)		(-33.404, 21.183)	(-34.015, 20.321)
3	-18.54		-18.036	-18.24
	(-45.813, 8.733)		(-45.328, 9.256)	(-45.407, 8.926)
4	-6.821		-6.4	-6.848
	(-32.669, 19.027)		(-32.267, 19.466)	(-32.595, 18.900)
5–9	19.106		19.806*	18.566
	(-0.327, 38.538)		(0.360, 39.253)	(-0.791, 37.923)
10–14	26.079**		26.883**	23.854*
	(6.402, 45.756)		(7.191, 46.574)	(4.253, 43.455)
15–19	38.103***		38.244***	37.734***
	(17.820, 58.385)		(17.947, 58.541)	(17.530, 57.938)
20–24	40.209***		39.952***	40.666***
	(19.892, 60.526)		(19.620, 60.283)	(20.428, 60.904)
25–29	51.863***		51.628***	52.563***
	(32.497, 71.229)		(32.249, 71.007)	(33.272, 71.853)
30–34	54.696***		54.666***	55.755***
	(36.078, 73.315)		(36.035, 73.297)	(37.209, 74.301)

	AIM-R (specified model)	SLM	AIM	AIM-R-P3
35–39	58.750*** (40.159, 77.342)		58.652*** (40.047, 77.256)	59.770*** (41.250, 78.290)
40-44	59.580*** (40.713, 78.446)		59.896*** (41.016, 78.776)	60.400*** (41.606, 79.194)
45–49	63.618*** (44.146, 83.090)		63.809*** (44.323, 83.295)	63.620*** (44.223, 83.017)
50–54	77.427*** (57.213, 97.640)		78.319*** (58.091, 98.546)	76.684*** (56.548, 96.819)
55–59	75.947*** (55.091, 96.804)		77.353*** (56.481, 98.225)	75.136*** (54.360, 95.912)
60–64	93.505*** (71.731, 115.279)		95.065*** (73.276, 116.855)	90.978*** (69.288, 112.669)
65–69	99.337*** (76.165, 122.509)		101.240*** (78.051, 124.428)	97.963*** (74.881, 121.045)
70–74	100.934*** (74.988, 126.880)		103.710*** (77.746, 129.674)	98.610*** (72.765, 124.455)
75–79	103.233*** (73.259, 133.206)		105.179*** (75.184, 135.174)	96.941*** (67.083, 126.800)
80+	120.846*** (90.149, 151.543)		122.643*** (91.924, 153.362)	117.898*** (87.321, 148.476)
		Age band × Pacific		
1	3.164 (-24.570, 30.898)		3.245 (-24.509, 30.999)	1.837 (-25.790, 29.463)
2	-1.457 (-31.359, 28.444)		-1.502 (-31.425, 28.421)	-1.791 (-31.576, 27.994)
3	-14.58 (-44.450, 15.291)		-14.506 (-44.398, 15.386)	-14.617 (-44.371, 15.137)
4	-9.001 (-37.253, 19.250)		-8.692 (-36.964, 19.579)	-8.723 (-36.865, 19.419)

	AIM-R (specified model)	SLM	AIM	AIM-R-P3
5–9	23.220*		23.539*	23.341*
	(1.826, 44.615)		(2.129, 44.948)	(2.029, 44.652)
10–14	8.844		9.276	6.833
	(-12.774, 30.463)		(-12.358, 30.910)	(-14.702, 28.367)
15–19	9.605		9.885	9.259
	(-12.698, 31.909)		(-12.434, 32.204)	(-12.957, 31.476)
20–24	21.673		21.994	21.474
	(-0.546, 43.891)		(-0.240, 44.228)	(-0.658, 43.606)
25–29	33.150**		33.393**	33.365**
	(11.871, 54.429)		(12.099, 54.687)	(12.169, 54.562)
30–34	52.577***		52.586***	53.306***
	(31.987, 73.167)		(31.982, 73.191)	(32.796, 73.816)
35–39	53.074***		53.210***	53.758***
	(32.460, 73.688)		(32.582, 73.839)	(33.224, 74.292)
40–44	68.286***		68.707***	68.813***
	(47.436, 89.137)		(47.842, 89.572)	(48.044, 89.583)
45–49	71.943***		72.117***	71.720***
	(50.458, 93.428)		(50.617, 93.617)	(50.318, 93.122)
50–54	92.410***		92.501***	91.658***
	(70.226, 114.595)		(70.300, 114.701)	(69.560, 113.757)
55–59	117.808***		117.951***	116.385***
	(95.022, 140.594)		(95.149, 140.753)	(93.687, 139.083)
60–64	133.896***		133.966***	131.055***
	(110.124, 157.669)		(110.177, 157.756)	(107.374, 154.736)
65–69	129.427***		129.414***	127.524***
	(104.212, 154.643)		(104.181, 154.648)	(102.406, 152.641)
70–74	112.090***		112.337***	108.464***
	(84.163, 140.016)		(84.391, 140.283)	(80.645, 136.282)
75–79	92.148***		91.673***	86.858***
	(60.085, 124.212)		(59.586, 123.759)	(54.918, 118.798)

	AIM-R (specified model)	SLM	AIM	AIM-R-P3
80+	59.912*** (27.086, 92.738)		59.831*** (26.982, 92.680)	62.131*** (29.432, 94.829)
Age band × P3 group				
1				3.921 (-1.504, 9.346)
2				5.729* (0.137, 11.322)
3				4.057 (-1.561, 9.675)
4				0.395 (-4.890, 5.681)
5–9				-0.602 (-4.780, 3.576)
10–14				-4.173* (-8.338, -0.008)
15–19				6.425** (2.242, 10.609)
20–24				17.706*** (13.513, 21.899)
25–29				22.461*** (18.331, 26.592)
30–34				25.526*** (21.448, 29.604)
35–39				27.579*** (23.484, 31.674)
40–44				31.851*** (27.730, 35.971)
45–49				39.726*** (35.615, 43.837)

	AIM-R (specified model)	SLM	AIM	AIM-R-P3
50–54				46.348*** (42.282, 50.414)
55–59				52.584*** (48.511, 56.657)
60–64				60.728*** (56.667, 64.790)
65–69				66.946*** (62.837, 71.055)
70–74				79.943*** (75.777, 84.110)
75–79				96.177*** (91.856, 100.497)
80+				95.743*** (91.471, 100.014)
Constant	-85.616 (-380.718, 209.486)	435.236*** (415.833, 454.640)	-85.709 (-381.021, 209.603)	3.682 (-290.394, 297.757)
Observations	2,006,894	2,006,894	2,006,894	2,006,894
R^2	0.109	0.101	0.107	0.115
Adjusted R^2	0.108	0.101	0.107	0.115
Residual Std. Error	260.390 (df = 2006741)	261.431 (df = 2006862)	260.577 (df = 2006745)	259.377 (df = 2006721)
F Statistic	1,607.194*** (df = 152; 2006741)	7,299.430*** (df = 31; 2006862)	1,628.819*** (df = 148; 2006745)	1,522.883*** (df = 172; 2006721)
Note:	*p<0.05; **p<0.01; ***p<0.001			

Source: Sapere analysis based on PMS and NES data

Table 23 shows the predicted cost weights that arose from the coefficients of the AIM-R model in Table 22. A scan of this table helps determine which age and ethnic groups had similarly predicted costs, supporting grouping the age categories into those outlined in section 1.1.

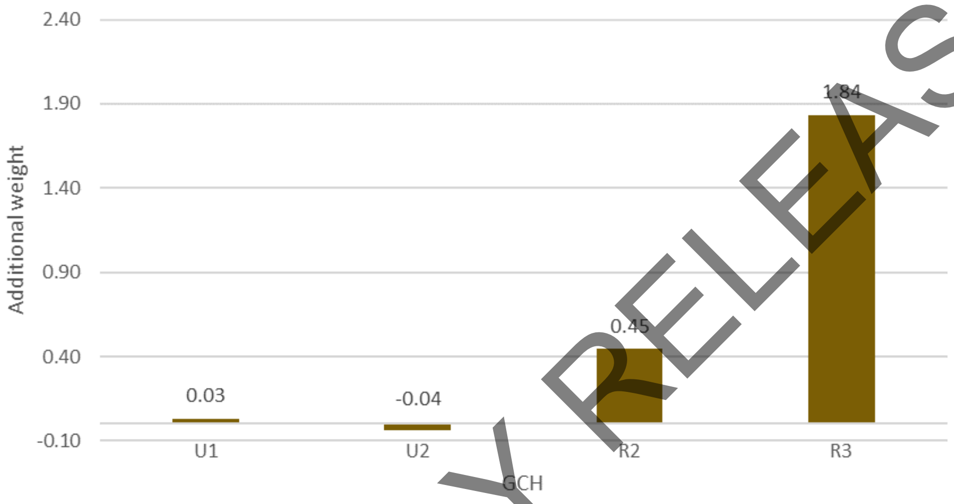
Table 23: Predicted capitation weights, assuming deprivation quintile 1, GCH R1, and P3 score < 0.2

		0	1	2	3	4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80+
Asian	M	7.4	5.9	3.5	3.7	3.5	1.8	1.6	1.0	1.3	1.2	1.3	1.5	1.7	1.9	2.2	2.5	3.0	3.3	3.7	4.3	4.9
	F	7.1	5.7	3.4	3.7	3.4	1.7	1.6	2.1	2.7	2.5	2.5	2.7	2.8	3.1	3.2	3.3	3.5	3.9	4.1	4.8	5.0
European	M	7.7	6.1	3.1	3.1	3.0	1.7	1.8	1.9	2.1	2.0	2.0	2.1	2.3	2.4	2.7	3.0	3.5	4.2	4.9	6.0	8.1
	F	7.5	5.9	3.1	3.1	2.9	1.7	1.8	3.0	3.5	3.3	3.2	3.3	3.4	3.6	3.7	3.8	4.1	4.8	5.4	6.5	8.2
Māori	M	6.5	4.9	2.7	2.4	2.6	1.4	1.5	1.7	1.9	2.2	2.3	2.6	2.9	3.3	3.8	4.3	4.9	5.4	6.3	7.3	8.4
	F	6.2	4.7	2.6	2.3	2.5	1.4	1.5	2.8	3.3	3.5	3.5	3.8	4.1	4.5	4.8	5.0	5.5	6.0	6.8	7.8	8.5
Pacific	M	7.4	5.6	3.6	3.3	3.4	2.4	2.0	1.9	2.3	2.6	3.0	3.3	3.9	4.3	5.0	6.0	6.6	6.9	7.4	7.9	7.9
	F	7.1	5.4	3.5	3.2	3.3	2.3	2.0	3.0	3.8	3.9	4.3	4.5	5.1	5.5	6.0	6.8	7.2	7.5	7.9	8.4	8.0
Other	M	7.2	5.4	3.5	3.5	3.4	1.7	1.6	1.6	1.7	1.7	1.8	2.0	2.3	2.6	2.8	3.3	3.6	3.9	4.8	5.7	6.5
	F	6.9	5.2	3.4	3.4	3.3	1.7	1.7	2.7	3.2	3.1	3.0	3.2	3.4	3.8	3.8	4.1	4.1	4.5	5.3	6.2	6.6

Our regression model estimate predicted no significant difference in cost weights between U1, U2 and R1 areas. Rurality did not play a significant role in the disparity of healthcare use for these three groups, all else held constant.

There were far more significant cost weights associated with those in more rural and isolated areas (R2 and R3), shown in Figure 40.

Figure 40: Predicted additional weight over R1, attributed to Geographical Classification of Health (GCH)



Appendix C R-script for analysis

Read in relevant packages

```
library(tidyverse)
library(magrittr)
library(dplyr)
library(lubridate)
library(DescTools)
library(ggplot2)
library(plotly)
library(readr)
library(readxl)
library(stringr)
```

Not all of these packages were required for the analysis; however, we recommend using functions from these packages to perform data quality and sense checks where necessary.

Data cleaning

We cleaned the data to the point where we have a large dataset containing:

- Appointments
- Facilities
- Encounters
- Patient Demographics
- Invoices
- Prescriptions
- Providers

Each of the datasets provided by PHOs may have inconsistencies between each other that need to be considered when cleaning. This includes removing duplicate rows, and ensuring that the key variables used in the analysis are consistently labelled with one another. Refer to Table 1 for the final values of variables used in the analysis.

In the first instance, we used the demographic information provided by the patient demographics dataset recorded by practices. If this was not available, we matched the NES characteristics using the encrypted NHI number and used these. We also used deprivation from the NES. P3 scores and GCH indices were also matched to the final dataset using the encrypted NHI number and domicile codes respectively. The final dataset should include the fields identified in Table 18. The steps to derive calculated columns will be outlined in the remainder of the appendix.

Some of the cleaning requirements we identified were:

Inconsistent labelling of gender

Different practices may label gender differently, some may use the initial of the gender (e.g., "M") while others may label it explicitly (e.g., "Male"). The NES labels genders as M, F, U and O. We identified the following labelling conventions, and relabelled these for consistency:

```
data  <- data %>%
  mutate(gender = ifelse(gender == "F", "Female", gender),
         gender = ifelse(gender == "M", "Male", gender),
         gender = ifelse(gender == "U" | is.na(gender) | gender == "" | gender == "Unspecified or Unknown" | gender
== "NULL" | gender == " ", "Unknown", gender),
         gender = ifelse(gender == "O" | gender == "Non-Binary" | gender == "Other Gender" | gender == "Transgender"
| gender == "Another Gender", "Other", gender))
```

Due to small numbers, we removed other and unknown genders from our weighting calculations.

Generating an age and ageBand variable

We calculated an age variable by taking the time difference between the appointment and the date of birth. Some practices provide appointments to the second, (e.g., "2023-01-01 13:00:00"), while others may only provide the date. The inconsistency will cause an error when calculating age. Therefore we created an appointDate variable that was consistent for all observations:

```
data <- data %>%
  mutate(appointDate = ymd(paste(substr(appoint,1,10))),
         dob = ymd(dob),
         age=(appointDate - dob)/365.25,
         age = ifelse(is.na(age), 2023-year(dob), age)) #If there is no appointment date attached
```

We then created the ageBand variable:

```
data <- data %>%
    mutate(ageBand = case_when(
    age == 0 ~ "0",
    age == 1 ~ "1",
    age >= 2 & age <= 4 ~ "2-4",
    age >= 5 & age <= 14 ~ "5-14",
    age >= 15 & age <= 24 ~ "15-24",
    age >= 25 & age <= 34 ~ "25-34",
    age >= 35 & age <= 49 ~ "35-49",
    age >= 50 & age <= 64 ~ "50-64",
    age >= 65 & age <= 69 ~ "65-69",
    age >= 70 & age <= 74 ~ "70-74",
    age >= 75 & age <= 79 ~ "75-79",
    age >= 80 ~ "80+",
    TRUE ~ NA_character
  )
```

Determining ethnicity

Using the ethnic codes, we determined ethnicity using the following mapping:

ethnic1	ethnicDescription	ethnicity	ethnicity in disaggregated analysis
10	European not further defined	European and Asian	European
11	NZ European	European and Asian	European
12	Other European	European and Asian	European
21	NZ Maori	Maori, Pacific and Other ethnicities	Maori
30	Pacific Island not further defined	Maori, Pacific and Other ethnicities	Pacific
31	Samoan	Maori, Pacific and Other ethnicities	Pacific
32	Cook Island Maori	Maori, Pacific and Other ethnicities	Pacific
33	Tongan	Maori, Pacific and Other ethnicities	Pacific
34	Niuean	Maori, Pacific and Other ethnicities	Pacific

ethnic1	ethnicDescription	ethnicity	ethnicity in disaggregated analysis
35	Tokelauan	Maori, Pacific and Other ethnicities	Pacific
36	Fijian	Maori, Pacific and Other ethnicities	Pacific
37	Other Pacific Island	Maori, Pacific and Other ethnicities	Pacific
40	Asian not further defined	European and Asian	Asian
41	Southeast Asian	European and Asian	Asian
42	Chinese	European and Asian	Asian
43	Indian	European and Asian	Asian
44	Other Asian	European and Asian	Asian
51	Middle Eastern	Maori, Pacific and Other ethnicities	Other
52	Latin American / Hispanic	Maori, Pacific and Other ethnicities	Other
53	African	Maori, Pacific and Other ethnicities	Other
61	Other ethnicity	Maori, Pacific and Other ethnicities	Other
94	Don't know	Removed from analysis	Removed from analysis
95	Refused to answer	Removed from analysis	Removed from analysis
97	Response unidentifiable	Removed from analysis	Removed from analysis
99	Not stated	Removed from analysis	Removed from analysis

Identifying dummy patients

```
data <- data %>%
  group_by(phoName, facilityId, patientId) %>%
  mutate(numAppoint = n()) %>%
  ungroup() %>%
  mutate(appointmentDecile = ntile(numAppoint, 10)) %>%
  mutate(dummyPat = ifelse(patientId == "TEMPPAT" | patientId == "TMPPATI" | numAppoint >= 400 | (appointmentDecile
== 10 & is.na(nhiEncrypted)), 1, 0))
```

It is important to keep dummy patients in the analysis, as they may be used in included non-contact templates. However, they should be excluded from the final output, which will occur naturally when filtering to non-empty encrypted NHI numbers.

Assigning template types, reassigning provider types, and defining excluded services

```
fteData <- data %>%
  mutate(templateType=case_when(tolower(providerName) %like any% c("%triage%") ~ "Services",
                              tolower(providerName) %like any% c("%acute nurse%",
                                                                  "%clinical nurse%",
                                                                  "%admin nurse%",
                                                                  "%practice nurse%",
                                                                  "%phone nurse%",
                                                                  "%script nurse%",
                                                                  "%prescription nurse%",
                                                                  "%rx nurse%",
                                                                  "%registrar%",
                                                                  "%student%") ~ "Generic",
                              tolower(providerName) %like any% c("admin%") ~ "Admin",
                              tolower(providerName) %like any% c("%phone%",
                                                                  "%triage%",
                                                                  "flu %",
                                                                  "% imms",
                                                                  "imms %",
                                                                  "%immunisation%",
                                                                  "%vaccine%",
                                                                  "%diabetes%",
                                                                  "%respiratory%",
                                                                  "%menz%",
                                                                  "%hpv%",
                                                                  "%b4%",
                                                                  "%cervical%",
                                                                  "%smear%",
                                                                  "%cx%",
                                                                  "%blood%",
                                                                  "%ltc%",
                                                                  "%queue%",
                                                                  "%immigration%",
                                                                  "%aclasta%",
                                                                  "%acute%",
```

```
                                                                "%car%") ~ "Services",
                    tolower(providerName) %like any% c("%medical%","%practice") ~ "Services",
                    tolower(providerName) %like any% c("% room%",
                                                                "%clinic 1%",
                                                                "%clinic 2%",
                                                                "%clinic 3%",
                                                                "%clinic 4%",
                                                                "%clinic 5%",
                                                                "%clinic 6%",
                                                                "%outreach%",
                                                                "%rx%",
                                                                "%script%",
                                                                "%prescription%",
                                                                "%medical%",
                                                                "%practice",
                                                                "%doctors%",
                                                                "%txt%",
                                                                "%sms%",
                                                                "%messages%") ~ "Admin",
                    tolower(providerName) %like any% c("%clinic%") ~ "Services",
                    TRUE ~ "Provider"),
        providerType=case_when(templateType=="Admin" ~ "Other",
                  tolower(providerName) %like any% c("dr%","%doctor%","%gp","%medical student%","%registrar%")
~ "GP",
                  tolower(providerName) %like any% c("%nurse practitioner%","np %","% np") ~ "NP",
                  tolower(providerName) %like any% c("%nurse%", "nelson east family medical centre") ~
"Nurse",
                  templateType %in% c("Generic","Services") ~ "Nurse",
                  TRUE ~ providerType),
        excludedService=case_when(tolower(providerName) %like any% c("%after hours%","%a&m%","%accident &
medical%","%accident and medical%") ~ 1L,
                              TRUE~ 0L))
```

Note that if the providerName field has any special characters, the above script will fail to execute. An error message will identify the relevant row number with special characters, and manual adjustment to the providerName for that row will need to be done to address this.

Defining patient contact and included template types

Clinical contact time was defined wherever there is evidence that the patient was in consult, i.e., if there is an arrival date, consult date, invoice date, or if the provider wrote notes or a prescription for the patient.

```
fteData <- fteData %>%
  mutate(appoint=ymd_hms(appoint)) %>%
  mutate(firstContact=ymd_hms(firstContact)) %>%
  mutate(patientContact=case_when(dummyPat==0 &
                                    (!is.na(arrived) |
                                       !is.na(consult) |
                                       !is.na(invoice) |
                                       encounter==1 |
                                       prescription == 1) ~ 1L, TRUE ~ 0L)) %>%
  mutate(include=case_when(templateType == "Admin" ~ 0L,
                           templateType == "Services" & patientContact==0 ~ 0L,
                           excludedService==1 ~ 0L, TRUE ~ 1L))
```

Excluding provider time if there are no patient contacts in a day

```
#The following creates the daily contact FTE per provider
sessions <- fteData %>%
  group_by(phoName, facilityId, providerType, providerId, appoint) %>%
  summarise(dailyContact = sum(ifelse(include==1,patientContact,0L))) %>%
  ungroup()

#The following joins the daily contact sessions per provider, back to the FTE dataset
fteData <- fteData %>%
  left_join(sessions) %>%
  mutate(include=ifelse(ifelse(is.na(dailyContact),0L,dailyContact)==0 & providerType=="GP",0L,include))
```

Assigning FTE and costs to different provider types

```
#FTE per patient per month
fteData <- fteData %>%
  mutate(annualWorkDays = ifelse(providerType == "Nurse", 220, 209),
        mutate(annualFTE=duration/(60*8*annualWorkDays))
```

This is a good point to save the dataset, as it will be extremely time-consuming to rerun everything to this point.

Summarising data at the patient–provider level

```
fteSummary <- fteData %>%
  filter(include == 1 & providerType %in% c("GP", "Nurse Practitioner", "Nurse")) %>%
  group_by(phoName, facilityId, practiceName, providerType, providerId,
           patientId, patientContact, ageBand, ethnicity,
           gender, deprivationQuint, p3score, gch, nhiEncrypted) %>%
  summarise(duration=sum(duration, na.rm=TRUE),
            annualFTE=sum(annualFTE, na.rm=TRUE),
            contacts=sum(patientContact, rm.na=TRUE)) %>%
  ungroup()
```

Calculating PHO scaling factors based on RCNZGP "Your Work Counts" survey data

```
# Generate providerData, remove providers that did not work the full year
providerData <- fteData %>%
  filter(providerType == "GP") %>%
  group_by(phoName, facilityId, providerId) %>%
  mutate(tenure = max(appoint) - min(appoint)) %>%
  filter(tenure >= 30000000) %>%
  ungroup()
```

```
#Remove outliers, providers who work more than 16 hours a day average
providerCheck <- providerData %>%
  group_by(phoName, facilityId, providerId) %>%
  summarise(sumDuration = sum(duration)) %>%
  mutate(remove = ifelse(sumDuration > 249600, 1, 0)) %>%
  ungroup()

providerData <- providerData %>%
  left_join(providerCheck, by=c("phoName", "facilityId", "providerId")) %>%
  filter(remove == 0) %>%
  select(-c("remove"))

## Remove contact consults with more than 3 hour duration
providerData <- providerData %>%
  filter((duration <= 180 & patientContact == 1) | patientContact == 0)


#Generate non-contact to contact time ratios
noncontactMean <- providerData %>%
  filter(patientContact == 0, !is.na(providerId)) %>%
  group_by(phoName) %>%
  summarise(NC = sum(duration)) %>%
  ungroup()

contactMean <- providerData %>%
  filter(patientContact == 1, !is.na(providerId)) %>%
  group_by(phoName) %>%
  summarise(C = sum(duration)) %>%
  ungroup()

collegeSurvey <- contactMean %>%
  left_join(noncontactMean, by=c("phoName")) %>%
  filter(C > 0, NC > 0) %>%
  mutate(ratio = NC/C,
         scalingFactor = 0.555983/ratio) #0.555983 is based on the weighted average of the non-contact to contact
time ratio in the RNZCGP survey data provided to Sapere
```

Scaling up FTE to factor in non-contact time

```
#Identify contact and non-contact time
timeSummary <- fteSummary %>%
  left_join(collegeSurvey, by=c("phoName")) %>%
  mutate(annualFTE = ifelse(patientContact == 0 & providerType == "GP", annualFTE * scalingFactor, annualFTE)) %>%
  group_by(phoName, facilityId, providerType) %>%
  summarise(totalContactFTE=sum(if_else(patientContact==1,annualFTE, 0,0), na.rm=TRUE),
            totalNonContactFTE=sum(if_else(patientContact==0,annualFTE,0,0), na.rm=TRUE),
            totalFTE=sum(annualFTE, na.rm=TRUE)) %>%
  ungroup()

#Scale contact FTE up by non-contact time
fteOutput <- fteSummary %>%
  inner_join(timeSummary, by=c("phoName"="phoName", "facilityId"="facilityId", "providerType"="providerType")) %>%
  mutate(nonContactScalingFactor =  ifelse(totalContactFTE != 0, totalNonContactFTE/totalContactFTE,
totalNonContactFTE))   %>%
  mutate(patientFTE=annualFTE*(1+nonContactScalingFactor)) %>%
  rename("contactFte"="annualFTE") %>%
  select(-c("totalContactFTE", "totalNonContactFTE", "totalFTE", "patientContact")) %>%
  ungroup()
```

Scaling up Nurse FTE using results from practice survey

We derived scaling factors using results from our practice survey based on the size of the practice. The scaling factors could be different under a different analysis; however, we apply the scaling factors across the sample population:

```
# Attach scaling factors based on total FTE of practice
providerScales <- fteOutput %>%
  group_by(phoName, facilityId, providerType) %>%
  summarise(FTE = sum(patientFTE)) %>%
  mutate(scalingFactor = 1,
         scalingFactor = ifelse(providerType == "Nurse" & FTE <= 2, 1.9128, scalingFactor),
```

```
        scalingFactor = ifelse(providerType == "Nurse" & FTE > 2, 1.5126, scalingFactor))
     ungroup()


# Join scaling factors back to fteOutput and assign costs
providerScales <- providerScales %>%
  select(phoName, facilityId, providerType, scalingFactor)

fteOutput <- fteOutput %>%
  left_join(providerScales) %>%
  mutate(newFTE = patientFTE * scalingFactor))
```

Assigning costs to patients

```
fteOutput <- fteOutput %>%
  mutate(salary = 262080,
         salary = ifelse(providerType == "Nurse", 106739, salary),
         salary = ifelse(providerType == "Nurse Practitioner", 151079, salary),
         cost = newFTE * salary)
```

Removing patients with incomplete information

At this point, it is necessary to remove patients with incomplete information from the dataset. Those without an observed age, gender, ethnicity, or socio-economic deprivation quintile are removed. There are a large number of missing P3 scores, and a somewhat large number of missing GCH indices – however, these can be retained as these are imputed and inferred respectively using the other characteristics of the patient.

Imputing missing P3 scores

We use the mice package in R to impute missing P3 scores using a method called Propensity Mean Matching (PMM).

```
# Aggregate FTE output to the individual level
indFteOutput <- fteOutput %>%
  filter(!is.na(nhiEncrypted), nhiEncrypted != "") %>%
```

```
  group_by(nhiEncrypted) %>%
  mutate(patientFTE = sum(newFTE),
         cost = sum(cost)) %>%
  slice(1) %>%
  ungroup()

# Create a subset of data that includes the variables needed for imputation
# Here it is necessary to include only the variables used to predict imputed P3 scores
data_for_imputation <- indFteOut put %>%
  select(p3score, ageBand, ethnicity, deprivationQuint, gender)

# Perform multiple imputation
imputed_data <- mice(data_for_imputation, method = 'pmm', m = 5, seed = 500)

# Extract the completed datasets
completed_data <- complete(imputed_data, action = "long", include = TRUE)


# Combine the completed data with the original dataset
indFteOutput <- indFteOutput %>%
  mutate(p3Score = if_else(is.na(p3score),
                           completed_data$p3score[completed_data$.imp == 1],
                           p3score))
```

Creating P3 groups

```
indFteOutput <- indFteOutput %>%
  mutate(p3Group = ntile(p3Score, 3))

# See where cuts are
check <- indFteOutput %>%
  group_by(p3Group) %>%
  summarise(max = max(p3Score),
            min = min(p3Score))
```

We check where the tercile cuts are, and clean this to be more round for presentation purposes:

```
indFteOutput <- indFteOutput %>%
  mutate(p3Group = 1,
         p3Group = ifelse(p3Score >= 0.2, 2, p3Group),
         p3Group = ifelse(p3Score > 0.6, 3, p3Group))
```

Inferring missing GCH using the practice mode

```
# Function to calculate the mode
  mode_function <- function(x) {
    ux <- unique(x)
    ux[which.max(tabulate(match(x, ux)))]
  }


 # Calculate the mode for each phoName and facilityId combination
  modes <- indFteOutput %>%
    filter(!is.na(gch)) %>%
    group_by(phoName, facilityId) %>%
    summarise(gch_mode = mode_function(gch), .groups = 'drop')

 # Join the modes back to the original dataframe
  indFteOutput <- indFteOutput %>%
    left_join(modes, by = c("phoName", "facilityId")) %>%
    mutate(gch = if_else(is.na(gch), gch_mode, gch)) %>%
    select(-gch_mode)  # Remove the temporary gch_mode column
```

Regression models for producing weights

SLM Model

```
SLM <- lm(cost ~ ageBand + gender + ethnicity + deprivationQuint + p3Group, data = indFteOutput)
```

AIM Model

```
AIM <- lm(cost ~ ageBand*gender + ageBand*ethnicity + deprivationQuint + p3Group, data = indFteOutput)
```

AIM-R Model

```
AIM_R <- lm(cost ~ ageBand*gender + ageBand*ethnicity + deprivationQuint + p3Group + gch, data = indFteOutput)
```

AIM-R-P3

```
AIM_R_P3 <- lm(cost ~ ageBand*gender + ageBand*ethnicity + deprivationQuint + ageBand*p3Group + gch, data =
indFteOutput)
```

AGIM

```
AGIM <- lm(cost ~ ageBand*gender + deprivationQuint + ageBand*p3Group + gch, data = indFteOutput)
```

We use the Stargazer package in R to report estimated coefficients:

```
stargazer(AIM_R, SLM, AIM, AIM_R_P3,
          type = "html",
          ci = TRUE,
          star.cutoffs = c(0.05, 0.01, 0.001),
          title = "Regression Results",
          out = "multiple_regression_results.html")
```

Generating weights from the specified model

Generate a dataset with unique combinations of variables

```
# Generate dataset containing every unique combination of variables
gender_levels <- unique(indFteOutput$gender)
deprivationQuint_levels <- unique(indFteOutput$deprivationQuint)
```

```
gch_levels <- unique(indFteOutput$gch)
p3Group_levels <- unique(indFteOutput$p3Group)
ethnicity_levels <- unique(indFteOutput$ethnicity)
ageBand_levels <- unique(indFteOutput$ageBand)

# Generate all possible combinations
combinations <- expand.grid(gender = gender_levels,
                            deprivationQuint = deprivationQuint_levels,
                            gch = gch_levels,
                            p3Group = p3Group_levels,
                            ethnicity = ethnicity_levels,
                            ageBand = ageBand_levels)
```

Predict costs with the specified model and the combinations dataset

```
combinations$predicted_cost <- predict(AIM_R, newdata = combinations) #Replace "AIM_R" with the specified model of
choice, if required.
```

Calculate weights baselined to Neo

```
calc <- combinations %>%
  mutate(scalingFactor = mean(predicted_cost),
         weight = predicted_cost / scalingFactor)


capitationWeights <- calc %>%
  select(gender, ageBand, ethnicity, deprivationQuint, p3Group, gch, weight) %>%
  mutate(ageBand = paste0(ageBand, " years")) #This step is done to avoid improperly formatted outputs when
exporting to CSV
```

About Sapere

Sapere is one of the largest expert consulting firms in Australasia, and a leader in the provision of independent economic, forensic accounting and public policy services. We provide independent expert testimony, strategic advisory services, data analytics and other advice to Australasia's private sector corporate clients, major law firms, government agencies, and regulatory bodies.

'Sapere' comes from Latin (to be wise) and the phrase 'sapere aude' (dare to be wise). The phrase is associated with German philosopher Immanuel Kant, who promoted the use of reason as a tool of thought; an approach that underpins all Sapere's practice groups.

We build and maintain effective relationships as demonstrated by the volume of repeat work. Many of our experts have held leadership and senior management positions and are experienced in navigating complex relationships in government, industry, and academic settings.

We adopt a collaborative approach to our work and routinely partner with specialist firms in other fields, such as social research, IT design and architecture, and survey design. This enables us to deliver a comprehensive product and to ensure value for money.

For more information, please contact:

David Moore

Phone: +64 4 915 7590

Mobile: +64 21 518 002

Email: dmoore@thinkSapere.com

Wellington	Auckland	Sydney	Melbourne	Canberra	Perth
Level 9 1 Willeston Street PO Box 587 Wellington 6140	Level 20 151 Queen Street PO Box 2475 Shortland Street Auckland 1140	Level 18 135 King Street Sydney NSW 2000	Level 5 171 Collins Street Melbourne VIC 3000	GPO Box 252 Canberra City ACT 2601	PO Box 1210 Booragoon WA 6954
P +64 4 915 7590	P +64 9 909 5810	P +61 2 9234 0200	P +61 3 9005 1454	P +61 2 6100 6363	P+61 8 6186 1410

www.thinkSapere.com

independence, integrity and objectivity