Australian Government
Department of Health and Aged Care

General Practitioner (GP) Supply and Demand Model - Methodology Paper

August 2024

Contents

Introduction

This document provides the detailed methodology for the Supply and Demand Model for the General Practitioner (GP) Workforce.

The GP Supply and Demand Model (the GP model) is a health workforce model that projects the supply and demand of Australia's primary health care GPs over a 25-year period. The GP model has been designed to simulate complex policy scenarios and analyse the geo-spatial GP workforce implications.

Methodology overview

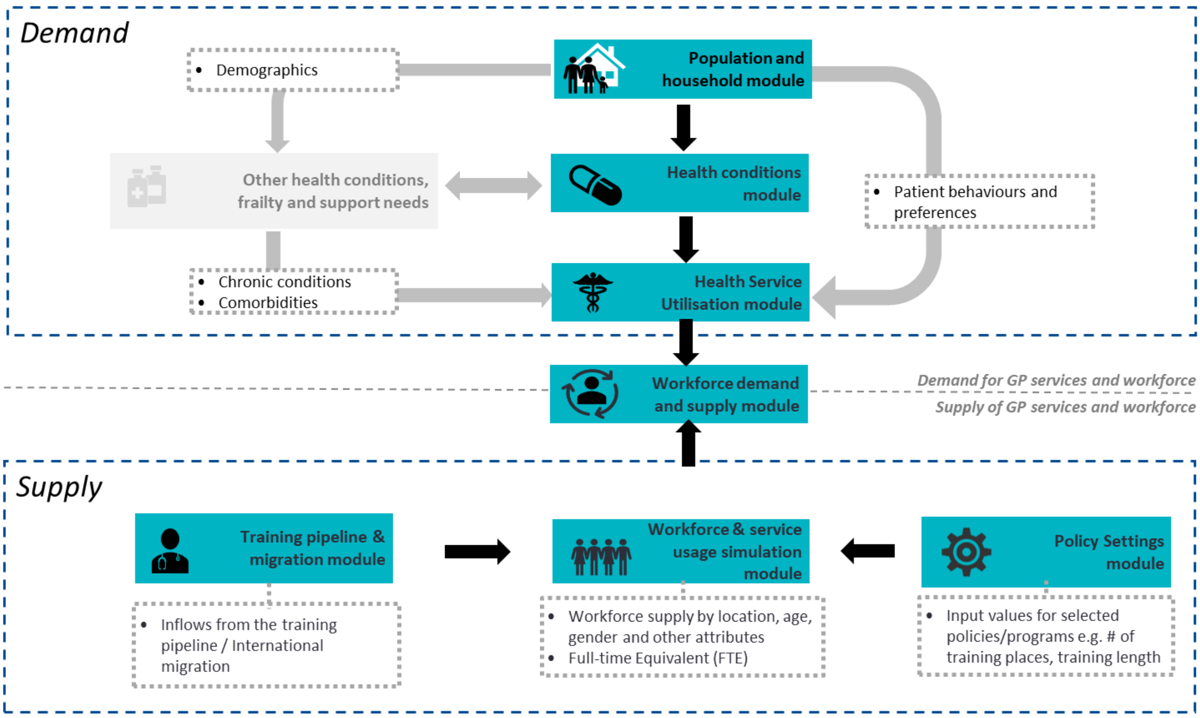
The following sections describe at a high-level the approach used to structure the GP Supply and Demand modelling. Further detail of the methodology, data, and assumptions are outlined in the subsequent sections.

To enable detailed scenario modelling of the GP workforce, the Department selected a micro-simulation modelling approach for both supply and demand. For supply, the starting stock for the simulation model is the GP workforce population in 2023, as observed in the MBS data. For demand, the starting stock for the simulation a 10% sample of the Australian population generated by the Population and household module.

Underpinning both the supply and demand data is the Department's existing GP Full-time Equivalent methodology as published on the Department Health Workforce Data website and in the Health Demand and Supply Utilisation Patterns Planning (HeaDS UPP) Tool.

The microsimulation model has been developed at the SA2 geographical level to allow results to be presented at higher level geographies – GP Catchment, PHN, MMM & State.

The modelling is organised into the following modules:



Demand Modules

Estimates of GP demand in terms of service volumes and full-time equivalent (FTE) are derived using three separate modules. Each model module runs independently and build upon results from the preceding module.

Population and Household Module

The population and household module creates a synthetic starting population representing Australian individuals, grouped into households. The module then generates the demographic variables for this population in the starting period, and models changes to the demographic variables in each projection period.

Population growth is estimated based on the trend in the ABS series B projections, combined with the ABS Household and Family Projections to capture age, gender and household types by SA2.

Demographic variables include age, gender, socio-economic indexes of area (SEIFA), Aboriginal and Torres Strait Islander status, culturally and linguistically diverse (CALD) status, labour force status, and participation in the National Disability Insurance Scheme (NDIS). Population growth dynamics are modelled within the microsimulation including mortality, births, and migration, and this module also includes transitions in households.

Health Conditions Module

The health conditions module models the prevalence of five key chronic conditions:

1. Asthma
2. Cardiovascular disease (CVD)
3. Diabetes
4. Mental health
5. Overweight/obesity.

These were agreed through consultation within the Department and selected as they are likely to result in an increased demand for primary health care services and is informative for health workforce planning.

The health conditions module builds upon the population and household module and simulates the prevalence of each condition at the individual level.

Health Service Utilisation module

The health service utilisation module estimates GP demand in terms of services, GP FTE and services per capita. The module produces demand estimates from the bottom-up, based on the patients age, gender and co-morbidity profile as produced by the preceding two modules.

The module uses a combination of GP MBS data and ABS ERP data to get an accurate picture of an individual's GP service utilisation. The module produces estimates based on the average service utilisation per cohort – these assumptions can be varied to produce demand scenarios.

Supply Modules

Estimates of GP supply both in terms of head count and full-time equivalent (FTE) are derived using two modules.

Most of the supply simulation happens in the main Workforce Simulation module with the additional module producing training inflow and policy input assumptions for the supply baseline estimates.

Workforce and Service simulation module

The workforce and service usage simulation module provides the main microsimulation engine that calculates GP supply over the next 25 years. The GP workforce is simulated using a combination of historically observed statistical distributions, policy setting assumptions and training inflow assumptions.

The microsimulation engine begins with the base year starting stock of GPs and simulates new entrants into and exits from the workforce each year. Each individual GP within the simulation is then progressed throughout their career based on their individual characteristics – such as age, gender and GP type.

Training pipeline and migration module

The training and pipeline module estimates training inflow numbers for the supply baseline. The module combines expert judgement with projection techniques to determine the baseline inflow assumptions for each training program. Expert judgement is used in conjunction with the historic inflow trends and changes to the workforce programs/ GP fellowship pathways.

Model Assumptions

#	Caveat/Limitation	Description and implications
1	Starting initial equilibrium assumptions	Initial equilibrium where FTE demand is met by FTE supply in the starting stock is assumed. This is achieved by calibrating national FTE demand to meet national FTE supply in the starting stock year of the microsimulation. Equilibrium here can be interpreted as the baseline for future projections in supply and demand. Unmet demand is accounted for in addition to the baseline.
2	Patient flows not captured	Supply and Demand modules are built separately and from different geographical perspectives. Demand Modules are built from the population's perspective where FTE demand is calculated based on population need, in their usual place of residence. Supply modules are built from a service location perspective where FTE supply is calculated based on service location. This must be considered when interpreting results at smaller geographies in which patients travel between different geographies to access GP services.
3	Differing implied growth rate assumptions	Population Growth rate assumptions are gathered from the ABS Population Projection Series B and play a fundamental role throughout the demand modules to calculate projected FTE demand. In contrast, supply modules are largely dependent on historical data and certain inflow assumptions made to calculate FTE supply. The different nature of implied growth rates will lead to a disparate relationship between supply and demand projections, which will be most present in geographies where historical profiles and projected demographic profiles are most contrasted.
4	Inflow assumptions	Inflow numbers are assumed based on expert consultation and vary between training programs. Inflow assumptions are based on historical trends. The percentage of AGPT places filled by training region is assumed to be static over the projection period based on the last 3 years of historical data. Assumptions have been made regarding the 3GA NONVR programs ceasing and flowing into PFP.
5	Health Risk Factor assumptions	Population risk factor variables required for the Health Condition module are simulated in the underlying synthetic population data in-line with the ABS NHS 2017-18 data. These have been adjusted based on the ABS NHS 2022 data.

		This data is sourced by age, gender and demographics from the ABS CURF data and is assumed to remain constant over the projection period.
6	Technological change	Technological improvements in the projection period that may affect workforce FTE in providing primary care is not considered.
7	Obesity omitted from HSU calculations	The Health Service Utilisation (HSU) module infers patients' chronic conditions based on their MBS item usage. Since the MBS does not have any specific obesity management items the current methodology doesn't factor in obesity/overweight prevalence.
8	Service per capita growth assumptions	Service per capita growth is calculated endogenous to the model in-line with the changes in needs relating to health condition prevalence rates. No exogenous service per capita growth factor is assumed as a model input.
9	Service duration growth assumptions	Service duration is assumed to be static across the projection period.
10	Average FTE per GP	There has been a decline in average FTE per GP in the recent years, the supply projections assume that the average FTE will remain at the 2023 level.

Detailed Module Methodology

Detailed supply methodology and inputs

Training Pipeline and Migration Module

The Training Pipeline and Migration Module calculates the sampling distributions of all individual agents' (an agent is a GP in this simulation model) transition through each year in the Workforce Simulation Module. Data is mainly estimated from sample proportions observed in MBS.

Inputs

#	Source	Description and use in model
1	GP Stats (Sourced from MBS dataset)	This provides unit records of General Practitioners including age, gender, primary SA2, FTE, GP type, place of initial medical qualification, 3GA programs, bond and moratorium status.
2	AGPT Dataset	This provides unit records of GP trainees currently not included in GP Stats because they are not currently billing Medicare. This data is added to the GP Stats data, so all registrars are included in the workforce headcount.

Methodology

New Agent Distributions

Calculates the sampling distribution of new agent attributes including Gender, Age, and place of initial medical qualification. Historical MBS data of new agents is used to determine new agent demographic probabilities and distributions. Sampling of these attributes is done sequentially in the following order:

1. New agents' Gender is sampled from a Binary distribution based on which training program they're being generated for.
2. New agents' place of initial medical qualification (AMG, IMG, FGAMS) is sampled from a Multinomial distribution based on training program and gender.
3. New agents' Age is sampled from a Gamma distribution based on their training program and gender.

Associated training attributes, such as whether the agent is in a moratorium, or a bonded program are then sampled from a Binary distribution based on training program, age, gender, and place of initial medical qualification.

Retention Distributions

Calculates the probability that a GP will exit the workforce on a given year of the simulation. This probability is modelled from the probability that a GP who is in MBS one year will be in MBS the following year using age, gender, place of initial medical qualification, GP type, and primary SA2 as covariates. This probability is modelled from unit-record MBS data using a Binomial model. To account for the relationship between GP age and the retention probability, a Gaussian Additive Model (GAM) is utilised.

Re-Entry Distributions

For GPs who are sampled to exit the workforce, we must then decide how long until they return, if at all. To do this, we use unit-record MBS data to classify all workforce exits as either temporary (1 to 3 years) or permanent (4 or more years). We then model this with a Multinomial model using age, gender, place of initial medical qualification, and GP type as covariates.

Full Time Equivalent (FTE) Output

Calculates the distribution used in sampling FTE when:

- A new GP is introduced,
- A GP re-enters the workforce having exited; or
- A GP changes their primary SA2 of occupation.

The FTE distribution is derived from comparable unit-record data in MBS. It's modelled by a Gamma regression with age, gender, place of initial medical qualification, and GP type as covariates. An SA2-level spatial effect is also used to account for locational differences.

Full Time Equivalent (FTE) Delta

Once a new FTE has been sampled, we modify its value over time by applying a multiplier referred to as the FTE Delta. The FTE delta captures the average year-on-year variation in a GP's FTE by age, gender, place of initial medical qualification, and GP type.

Geospatial Transition Rate

Models the event that a GP changes primary SA2. This component is derived from unit-record MBS data where we have modelled the event that a GP has a different primary SA2 in one year compared to their primary SA2 in the previous year. We model this event with a Binomial distribution using age, gender, place of initial medical qualification, and GP type as covariates. A spatial SA2-level change is also incorporated in the model to capture variations in this probability spatially.

Geospatial Allocations

Calculates the geographical SA2-level distribution of new and transitioning GPs by age, gender, place of initial medical qualification, GP type, and region restriction status (region Restricted Agents are identified in sampling as those GPs who are subject to bond or moratorium). This is done in a three-stage process where we first sample the GP's State, then MMM, then SA2. While the high-level State and MMM sampling uses age, gender, place of initial medical qualification, GP type and region restricted status as covariates, the low-level SA2 distribution is the same for all GPs except for AGPTs. In addition, for transitioning GPs we also use their previous State and MMM as covariates.

Non-Primary Allocations Sampling

Models the distribution of Non-Primary Places of Practice for GPs given their Primary Place of Practice. Historical MBS data within the sampling period is used to create Primary SA2 profiles constructed by modelling for the:

- Number of GP Practices individual GPs provide primary care in.
- FTE weighted distributions amongst the number of practices.
- Weighted Non-Primary SA2 locations.

Assumptions

#	Assumption	Detail	Implication
1	New Agents Grouped Sampling	PEP/FSP, IP/RGTS, RVTS AMDS, MDRAP/PFP and other 3GA programs are grouped due to sample size and the changing nature of training programs.	New Agents attributes are grouped and sampled in three training pipelines: Fully Comparable Vocationally Registered Professionals (VRs), Australian General Practitioner Training (AGPT) and Non-AGPT Trainees.
2	Geographical Allocations Grouped Sampling	PEP/FSP, IP/RGTS, RVTS AMDS, MDRAP/PFP and other 3GA programs are grouped due to sample size and the changing nature of training programs. Region Restricted Agents are sampled in a group manner for all fellowship pathways/Non VR programs and place of initial medical qualification.	Due to sample size in historical data available for new GPs in certain training programs, new agents are treated as being either Fully Comparable VRs, AGPT and Non-AGPT. For the same reason, region-restricted moratorium and bonded agents are sampled an initial geographical allocation based on grouped sampling distributions.
3	Probability of Billing Medicare in Year 1 of training	This assumption has been calculated using historical AGPT & MBS data. The probability of billing Medicare in Year 1 of training varies by State/RTO.	This assumption determines the number of new AGPT agents that will be generated each year by geography.

Workforce and Service Simulation Module

The workforce and service simulation module performs the main microsimulation. The model simulates individual agents at an annual timestep, updating agents' attributes based on historical distributions.

Inputs

#	Source	Description and use in model
1	2023 Starting Stock (sourced from MBS dataset)	This provides unit records of General Practitioners including age, gender, primary SA2, FTE, GP type, 3GA programs, bond and moratorium status.
2	AGPT dataset	This provides unit records of AGPT trainees currently not included in GP Stats because they are currently doing hospital terms and not yet billing Medicare. This data is added to the GP Stats data, so all registrars are included in the workforce headcount.
3	Sampling Data frames	Sampling data frames calculated in the previous module: • FTE sampling • Workforce retention/re-entry probabilities • FTE delta • Geospatial Transitions

Methodology

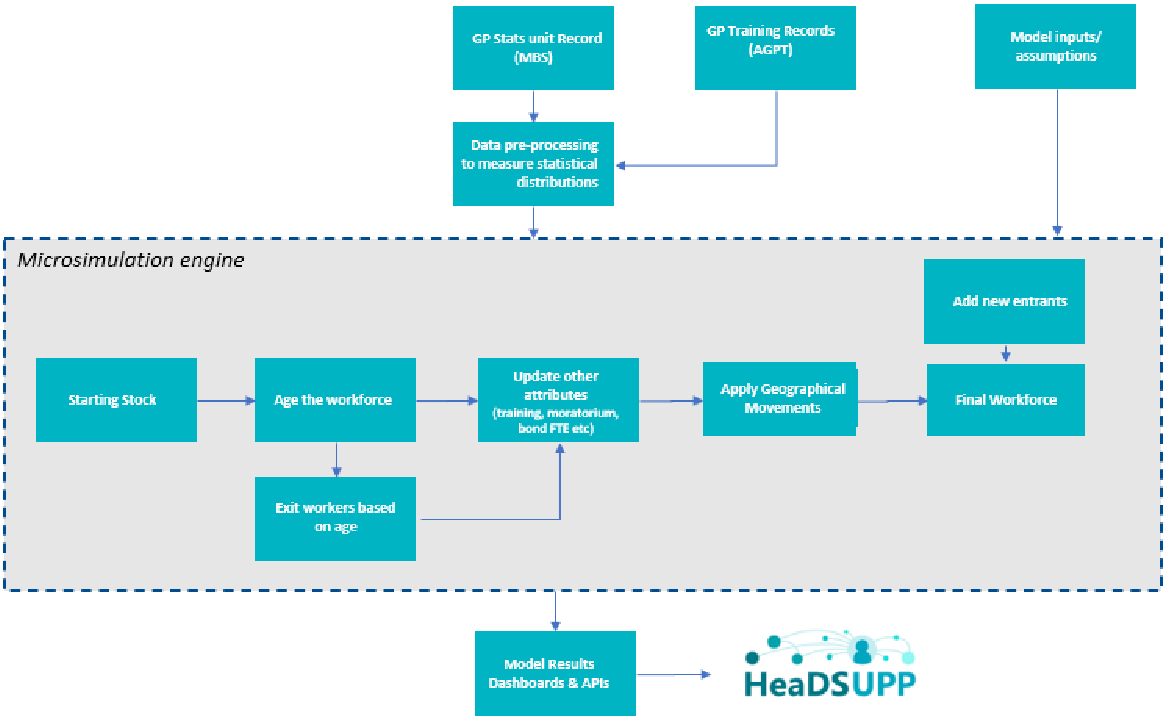
Each year the following steps are applied:

1. The workforce is aged, and some agents are exited based on their age, gender, place of initial medical qualification and GP type.
    a. Exits are sampled to determine if the exit is permanent or temporary based on age group and gender.
    b. Agents that temporarily exit will re-enter the workforce in a subsequent time period. The agent's records will be kept when they exit but their FTE drops to zero.
2. Moratorium status, bonded status and training length are reduced and once completed, attributes such as GP type and region restriction are updated (derived from bond and moratorium status once an exemption probability is applied).
3. Geographical movements are applied to agents based on historic interstate migration patterns broken down by age and region restriction status.
4. FTE is updated based on smoothed historical FTE year-on-year changes by age (FTE delta), unless an agent:
    a. Geographically Transitions to a different State or MMM.
    b. Transitions GP Type, through Fellowship or through a NONVR to GP trainee transition. In which case, a new FTE is sampled from the relevant Gamma

distribution for the agents age, gender and place of initial medical qualification.

5. Agents flagged for re-entry are brought back into the workforce based on a re-entry probability based on their age and gender. The agents FTE is sampled from the relevant Gamma distribution for the agents age, gender and place of initial medical qualification and becomes non-zero once they re-enter.
    a. Agents re-entering the workforce have their FTE adjusted to reflect the timing of their re-entry to the workforce.
6. New agents are generated synthetically in-line with the new trainee inflow numbers for each training program and new Non VR inflow numbers. These new agents have attributes sampled from the relevant distributions and allocated geographically according to their training and region restrictions. New agents are added to workforce either as:
    a. New trainees entering the training pipeline.
    b. Migrants joining the workforce from overseas (IMG subject to Moratorium region restriction)
    c. Re-entries to the workforce where the GP has been absent for more than four years.
7. Non-primary SA2s are allocated to each agent using a three-step sampling method derived by an agent's primary SA2 location and the historical profile of such location. The method involves allocating each agent:
    a. A number of non-primary SA2 locations in which they will provide primary care.
    b. The non-primary SA2 locations in which they will provide primary care.
    c. Weights that determine the distribution of FTE amongst the assigned locations.

The microsimulation approach is illustrated below:


GP Stats unit Record (MBS)
GP Training Records (AGPT)
Model inputs/ assumptions
Data pre-processing to measure statistical distributions
Microsimulation engine
Add new entrants
Starting Stock
Age the workforce
Update other attributes (training, moratorium, bond FTE etc)
Apply Geographical Movements
Final Workforce
Exit workers based on age
Model Results Dashboards & APIs
HeaDSUPP

Assumptions

#	Assumption	Detail	Implication
1	Region Restriction Exemption Probability	Agents that are subject to a moratorium have a region restriction exemption probability that is assumed to be 84%	This assumption means that most agents flagged as being subject to a moratorium will not be restricted to provide a majority of their total FTE throughout their moratorium in a DPA classified SA2.
2	Bond Start Probability	For agents who are in a bonded program, the probability that they begin their bond is assumed to be 5%	This assumption is static over time and means all agents will have the same probability of starting their bond.
3	Partially Comparable Percentage	The percentage of partially comparable agents is as follows: PEP/FSP, IP = 5.1% MDRAP/PFP, AMDS, Other 3GA = 5.13% RVTS = 5.1% The RVTS rate assumes that it is the same as IP rates due to data availability.	This percentage determines the amount of agents who receive partial recognition of previous training and thus complete the training program quicker.
4	Part Time percentage	The percentage of part - time agents is as follows: PEP/FSP = 19.6% IP = 55.8% MDRAP/PFP, AMDS, Other 3GA, RVTS = 55.8% The MDRAP/PFP, AMDS, Other 3GA and RVTS rates assumes that it is the same as IP rates due to data availability.	This percentage determines the amount of agents who complete the training program part-time and thus complete the training program slower.
5	Training Length	The training lengths are as follows: PEP/FSP = 3 years IP = 4 years MDRAP/FSP = 3 years AMDS = 3 years RVTS = 4 years	This determines the standard time taken to complete a training program which in term determines the number of GP trainees in the model at a given point in time.

		The AGPT training length is sampled for based on a historic distribution.	
6	Training Length – Partially Comparable	The partially comparable training lengths are as follows: PEP/FSP = 2 years Needs IP = 3 years MDRAP/PFP = 2 years AMDS = 2 years RVTS = 3 years	This determines the time taken to complete a training program given an agent is partially comparable which in term determines the number of GP trainees in the model at a given point in time.
7	Training Length – Part Time	The partially comparable training lengths are as follows: PEP/FSP = 6 years IP = 8 years MDRAP/PFP = 6 years AMDS = 6 years RVTS = 8 years	This determines the time taken to complete a training program part-time which in term determines the number of GP trainees in the model at a given point in time.
8	Geographic Concordances	The supply simulation is conducted at the SA2-level and subsequently aggregated to GP Catchment, PHN, MMM and DPA. The concordance from SA2s to other geographies has been calculated using the 2021 Census population by Mesh-Block as a weighting.	This assumption determines the GP Catchment, PHN, MMM and DPA an agent is in based on their primary SA2 At present, this may mean the distribution of FTE across GP Catchment, PHN and MMM may be slightly skewed. The impact is not expected to be drastic.
9	RGTS funded places	RGTS places are currently treated similar as IP places.	Inflow numbers based on historic inflow trends into IP and RGTS.

Detailed demand methodology and inputs

Population and Household Module

The population and household module creates a synthetic starting population representing Australian individuals, grouped into households. The module then generates the demographic variables for this population in the starting period, and models changes to the demographic variables in each projection period.

The synthetic population sample generated by this module is the key dataset underpinning all the demand modelling. It creates a unit record dataset that provides all the granular demographic and risk factors required for the health conditions and service utilisation module.

Inputs

#	Source	Description and use in model
1	Census of Population and Housing (2016)	Census microdata is a 1% sample of from the 2016 Census of population and housing. Census data is primarily used to generate demographic characteristics for individuals.
2	Population and household projections based off ABS data	Population and household projections developed by the Department based on ABS Series B population projections and the ABS Census distributions of household type.
3	Publicly available AIHW reports	Australia's mothers and babies data visualisations, which are publicly available web reports, were used to calibrate age gaps between mothers and their firstborn child in the synthetic starting population in developing household constraints.
4	Household, Income and Labour Dynamics in Australia (HILDA) survey	HILDA is a Department of Social Services (DSS) funded longitudinal study of Australian households and residents that commenced in 2001, providing data on income, the labour market and family dynamics. This data has been used to model employment transition variables.
5	National Health Survey (NHS) TableBuilder Pro (2017-18)	Health risk factors are added to the synthetic population data by extracting the SA2 level data from ABS TableBuilder. The health risk factor variables are required as inputs to the health condition prevalence models.
6	National Health Survey (NHS) (2022)	Population risk factor variables required for the Health Condition module are simulated in the underlying synthetic population data in-line with the ABS NHS 2017-18 data. These have been adjusted based on the summary NHS 2022 data published by ABS.

Methodology

The population and household module creates a synthetic starting population representing Australian individuals, grouped into households. The module then generates the demographic variables for this population in the starting period, and models changes to the demographic variables in each projection period.

The synthetic starting population is created separately for each of the SA2 regions in Australia. The creation of synthetic records for individuals and their allocation to households is performed through a technique called Iterative Proportional Updating (IPU). This produces a synthetic population that matches the ABS Series B population projections at a person-level (by age and gender) published by ABS on 23 Nov 2023., the ABS Census distributions of household types, and constraints such as age of a mother's age at first birth, age difference in couples and age distributions for household types.

Additional demographic variables are assigned to each synthetic individual in the starting period and projected forward in each projection period. The methodology used differs for each variable, and is outlined below:

1. The socio-economic indexes of area (SEIFA) decile of each synthetic individual is determined by their SA2. This variable is currently static through the projection period.
2. Aboriginal and Torres Strait Islander status and English proficiency variables are assigned together to each synthetic individual by age, gender and SA2 based on the distribution observed in the ABS 2016 Census of Population and Housing, preserving the relationship between these two variables. These variables are also static through the projection period.
3. The labour force status of each synthetic individual in the starting period is assigned by age, gender and SA2 based on the distribution observed in the ABS 2016 Census of Population and Housing. Changes in the labour force status in the projected years are modelled through the trends observed in the longitudinal HILDA data from 2013 to 2018, using demographic characteristics and employment history of individuals as predictors for an individual's labour force status in a given year.
4. Health risk factor variables are simulated based on ABS NHS 2017-18 data extracted from ABS Tablebuilder at the SA2 level. The variables are static and do not consider underlying trends in risk factors other than demographic changes.

Additional variables are also modelled and used in Aged Care modelling that also uses the synthetic population as input to policy simulations. These variables are:

5. The number of available carers within the household is determined based on the number of co-residents aged 15+, who are not full-time employed. This is then used to determine the likelihood of an individual to receive informal care from a co-resident within Module 4 of the Aged Care modelling.
6. An individual's participation in the National Disability Insurance Scheme (NDIS) is modelled based on the distribution of NDIS participants sourced from publicly available NDIA Participant data as at 30 June 2019. This is assigned by age, gender and SA2 for individuals with at least one support need in the starting population. This assumption around a minimum level of support needs for NDIS participants has been applied to account for the eligibility criteria into NDIS. New entrants into NDIS are not modelled in this iteration of the model, and a simplifying assumption has been applied such that individuals who are NDIS participants as at 30 June 2019 remain within the scheme for life. It is important to note that the model does not currently differentiate the likelihood of support needs and service use of NDIS participants to those of other individuals.

Population growth dynamics are modelled within the microsimulation including mortality, births, and migration. Individuals exit the population due to mortality, which is applied through frailty transitions that are calibrated to AGA Life tables 2015-17. Concurrently, births are modelled based on the ABS Series B population projections. Finally, migration of entire households (both immigration and emigration) is performed on a quarterly basis to match the population to the ABS Series B population projections for each SA2 region.

Household transitions are also modelled, where individuals exit existing households and form new single person households due to either relationship breakdown or children moving out of home. These transitions occur on a quarterly basis and are modelled using Census data provided by the Australian Bureau of Statistics. As household structures are modelled primarily as a determinant of availability for informal care, the transitions have focused on the break-up of households. Household formations have only been modelled to reflect new births; other reasons for household formation (such as through marriage) are not modelled at this stage.

Health Conditions Module

The health conditions module models the prevalence of five key chronic conditions:

1. Asthma
2. Cardiovascular disease
3. Diabetes
4. Mental health
5. Overweight/ obese.

These were agreed through consultation within the Department and selected as they are likely to result in an increased demand for primary health care services and is informative for health workforce planning.

Inputs

#	Source	Description and use in model
1	Synthetic population sample	This is used as the initial starting population for the microsimulation. This was a key output of the population and household module.
2	Australian Bureau of Statistics (ABS) National Health Survey (NHS) 2017/18 – Basic confidentialised unit record file (CURF)	This unit record dataset provides survey data at a household and individual level across Australia. It will be used to model the prevalence of chronic conditions using available demographic and health risk factors.
3	National Health Survey (NHS) (2022)	Population risk factor variables required for the Health Condition module are simulated in the underlying synthetic population data in-line with the ABS NHS 2017-18 data. These have been adjusted based on the summary NHS 2022 data published by ABS.
4	National Aboriginal and Torres Strait Islander Social Survey (NATSISS) Summary Table 16	This provides information on the prevalence of asthma, cardiovascular disease, diabetes and mental health by gender and Aboriginal and Torres Strait Islander status. This has been used to incorporate an Aboriginal and Torres Strait Islander adjustment.
5	National Aboriginal and Torres Strait Islander Health Survey, Table 4.3	This provides information on the prevalence of overweight/ obese by Age and Aboriginal and Torres Strait Islander status. This has been used to incorporate an Aboriginal and Torres Strait Islander adjustment.

Methodology

For the health conditions module, the priority was to consider modelling chronic conditions that are primarily treated in the primary care health system. As part of confirming the chronic condition scope, data discovery was performed to identify the potential variables in the ABS National Health Survey data for each chronic condition and understand how they differ from other data sources.

The definition of each chronic condition was agreed in conjunction with Departmental medical advisors and is outlined in the table below:

#	Chronic Condition	Definition
1	Asthma	Someone who has ever reported having asthma. This includes current, long-term and non-current asthma.
2	Cardiovascular Disease	Someone with any heart or circulatory conditions including any conditions which can be controlled with medication. It includes a variety of conditions such as: • Ischaemic heart disease (other ischaemic heart diseases, angina, heart attack) • Cerebrovascular diseases (stroke and other cerebrovascular diseases) • Heart failure • Oedema • Diseases of the arteries, arterioles and capillaries • Other heart disease This includes conditions that are both current and long-term but excludes conditions considered CVD risk factors such as low/high blood pressure, tachycardia, cardiac murmurs or hypertensive disease. These CVD risk factors are included as candidate predictors in the CVD regression modelling.
3	Diabetes	Someone has reported having Type 2 diabetes. This includes both current and ever diagnosed diabetes.
4	Mental Health	Someone who has organic mental problems, alcohol and drug problems, or mood (affective) disorders such as depression, anxiety related problems and other mental and behavioural problems. It includes: • Feeling anxious, nervous or tense • Depression • Feeling depressed • Anxiety disorders

		• Panic disorders/panic attacks • Phobic anxiety disorders • Post-traumatic stress disorder • Other problems of psychological development • Other mental or behavioural problems • Obsessive compulsive disorder • Behavioural, cognitive and emotional problems with usual onset in childhood/adolescence • Alcohol and drug problems • Other mood (affective) disorders
5	Overweight/Obese	Someone who has a BMI (Body Mass Index) score that translates to overweight, Obese class I, Obese class II, or Obese class III (i.e. BMI > 25).

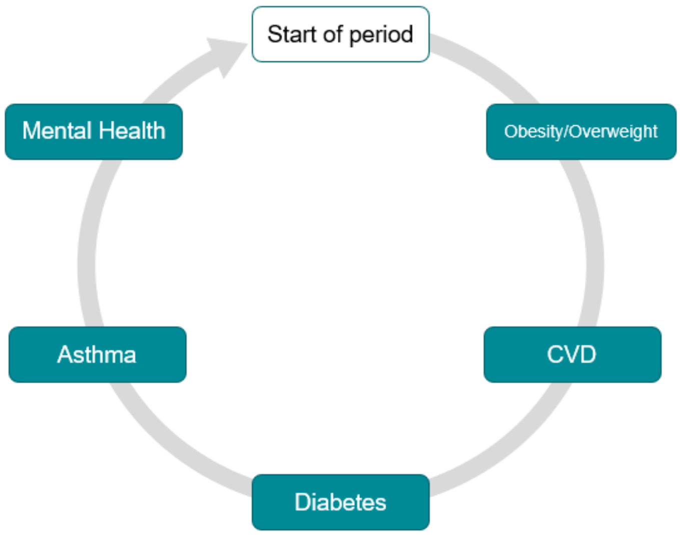
Prevalence models were developed for each chronic condition, accounting for co-morbidities, using regression techniques to produce transition propensity tables that define the likelihood of individuals having chronic conditions in the simulation period.

The output preparation steps are:

7. **Create a modelling dataset:** Conducted data preparation and veracity steps to build a modelling dataset based on the ABS National Health Survey 2017-18 data to contain the relevant chronic conditions and risk factors of interest. This included creating a training and testing dataset with a 70/30 split (training/validation data) for modelling purposes.
    1. **Model prevalence rates:** Iteratively fit logistic regression models for each chronic condition to the demographic variables to identify a model fit that is appropriate. Stepwise logistic regression models were developed (at the 5% level) and selected based on improvements in model AIC.

These logistic regression models are weighted with individual level weights in order to ensure that the ABS National Health Survey *"estimates conform to independently estimated distributions of the population by age, sex and area of usual residence. Calibration to benchmarks helps to compensate for over- or under-enumeration of particular categories of persons and households, which may occur due to the random nature of sampling or non-response".* The sum of all individual weights is the count of persons in the Australian population. To apply the individual weights to the regression, the inverse is taken.

For each projection period, the chronic condition prevalence models are simulated in a specified order. An illustrative summary of the order that the key variables are projected is provided in the diagram below:



2. **Create propensity tables:** Generate a propensity table for each chronic condition using the previously generated coefficient estimates for each combination of risk factors.
3. **Create Aboriginal and Torres Strait Islander status adjustment tables:** Adjustment factors for Aboriginal and Torres Strait Islander status are produced based on the National Aboriginal and Torres Strait Islander Health Survey aggregate statistics gender and age factors. Each prevalence model is only applied one adjustment factor, either based on age or gender. The relativities between different risk groups are applied to the base rates from Step 3.
4. **Run adjusted rates to produce starting prevalence rates:** The adjusted chronic condition prevalence rates are produced by taking the propensities based on all demographic characteristics (Step 3) and applying the relevant Aboriginal and Torres Strait Islander adjustment factors. Gender based Aboriginal and Torres Strait Islander adjustment factors are applied for the asthma, cardiovascular disease, diabetes and overweight/ obese models. Age-based rates are applied for mental health because gender rates weren't available for non-Aboriginal and Torres Strait Islanders.

   These adjusted rates are then calibrated (scaled up or down) by a calibration factor, which aims to ensure that the weighted average prevalence rates (weighted by the number of people in the synthetic population in each risk cohort) aligns to the propensities based on all demographic characteristics.

   Instances where cohorts have a small number of people in them can result in adjusted rates exceeding a value of 1. This only affects a small proportion of cohorts and these adjusted rates have been capped at a value of 1.
5. **Apply to the microsimulation**: The starting prevalence rates (Step 5) are applied to the synthetic population for each quarter, by simulating a random variable between 0 and 1 for each person in the population and comparing this to the determined probability for the person's risk factors. If this exceeds their defined chronic condition probability based on their demographic factors and co-morbidities, the individual is assigned with the chronic condition during that period.

Health Condition model predictors

Variable Type	Variable	Obesity Overweight	Diabetes	CVD	Asthma	Mental Health
Demographic	Age Band	✓	✓	✓	✓	✓
	Gender	✓	✓	✓	✓	✓
	Indigenous Status	✓	✓	✓	✓	✓
	SEIFA	✓	✓	✓	✓	✓
	CALD	×	×	×	✓	✓
Risk Factor	Tobacco Use	✓	×	×	×	×
	Alcohol Consumption	✓	✓	✓	×	×
	Diabetes Family History	✓	✓	✓	×	×
	Exercise/Physical activity	✓	×	×	×	×
	CVD Risk Factors	✓	✓	✓	×	×
Health Condition (comorbidity)	Obesity/Overweight	–	✓	✓	✓	✓
	Diabetes	–	–	✓	✓	✓
	CVD	–	–	–	✓	✓
Model Weight	NHS survey weight	✓	✓	✓	✓	✓

Table legend:

✓ *Factor included as a predictor*

× *Not included as predictor due to insignificance at a 5% level or violation of collinearity assumptions*

– *not available due to model ordering*

Assumptions

#	Assumption	Detail	Implication
1	Health Risk Factors assumptions are static	Health risk factors are added to the synthetic population data by extracting the SA2 level data from ABS TableBuilder from the NHS 2017-18 data. These have been adjusted based on the ABS NHS 2022 data. The health risk factor variables are required as inputs to the health condition prevalence models	The health condition prevalence projections may not reflect changes in population trends
2	Aboriginal and Torres Strait Islander adjustment	The NHS 2017-18 CURF data does not include the individuals Aboriginal and Torres Strait Islander status so could not be included as candidate variables for the logistic regression models. These models have been adjusted post-hoc based on additional ABS survey data.	Due to lack of identification in the NHS CURF data there may be Aboriginal and Torres Strait Islander individuals included in the model training data. The impact of this is unclear but expected to be minor

Health Service Utilisation Module

The Health Service Utilisation module estimates GP demand in terms of services, GP FTE and services per capita. The module produces demand estimates from the bottom-up based on the patients age, gender and co-morbidity profile as produced by the preceding two modules.

The module uses a combination of GP MBS data and ABS ERP data to get an accurate picture of an individual's GP service utilisation. The module produces estimates based on the average service utilisation per cohort (age, gender & co-morbidity profile), applied at the individual level.

The module produces two estimates of demand:

1. **Baseline demand** – demand projections under the initial equilibrium assumption where supply is assumed to meet demand in the base year.
2. **Unmet demand -** a level of demand for GP services that are not fully satisfied. It is the difference between what we expect the population to require, and what we have observed the population utilising.

   Unmet demand is calculated by applying the average GP service utilisation levels to each age, gender, and chronic condition cohort, and then adjusted based on the expected proportion of the population accessing GP services per year by age and gender.

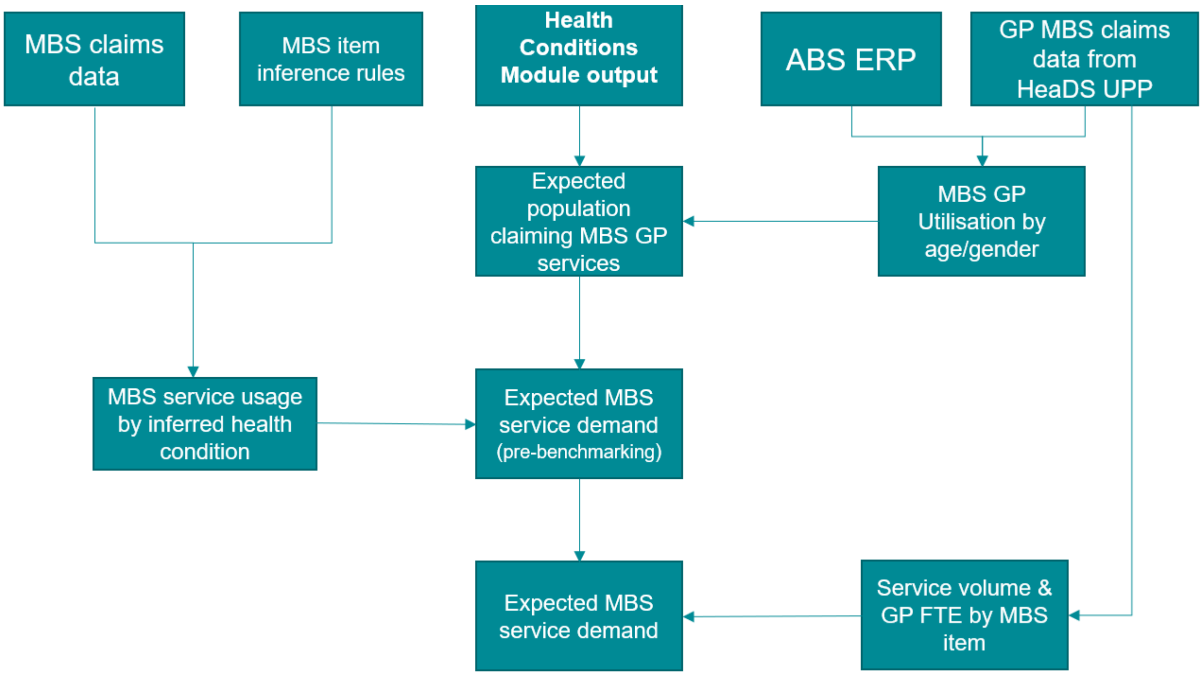
Inputs

#	Source	Description and use in model
1	Health Conditions module output	Health conditions module output disaggregated by: • Calendar Year • Age Band • Gender • State/SA2 • Health Condition
2	MBS data	MBS service usage data averaged over 2021-2023 are disaggregated by: • Age Band • Gender • Inferred Health Condition These data are further disaggregated to capture average service levels of the top 10 GP MBS items claimed by service volume. These top 10 GP items cover approximately 85 per cent of all GP services. Disaggregating the GP FTE/Service estimates by item type allows the model to capture different levels of complexity and effort required by cohorts with different health needs.

3	ABS ERP data	ABS ERP population data is compared to MBS patient data by age and gender to calculate the percentage of each cohort utilising GP services.
4	MBS Item Categorisation	Health conditions have been inferred based on the individual's usage of certain MBS items.

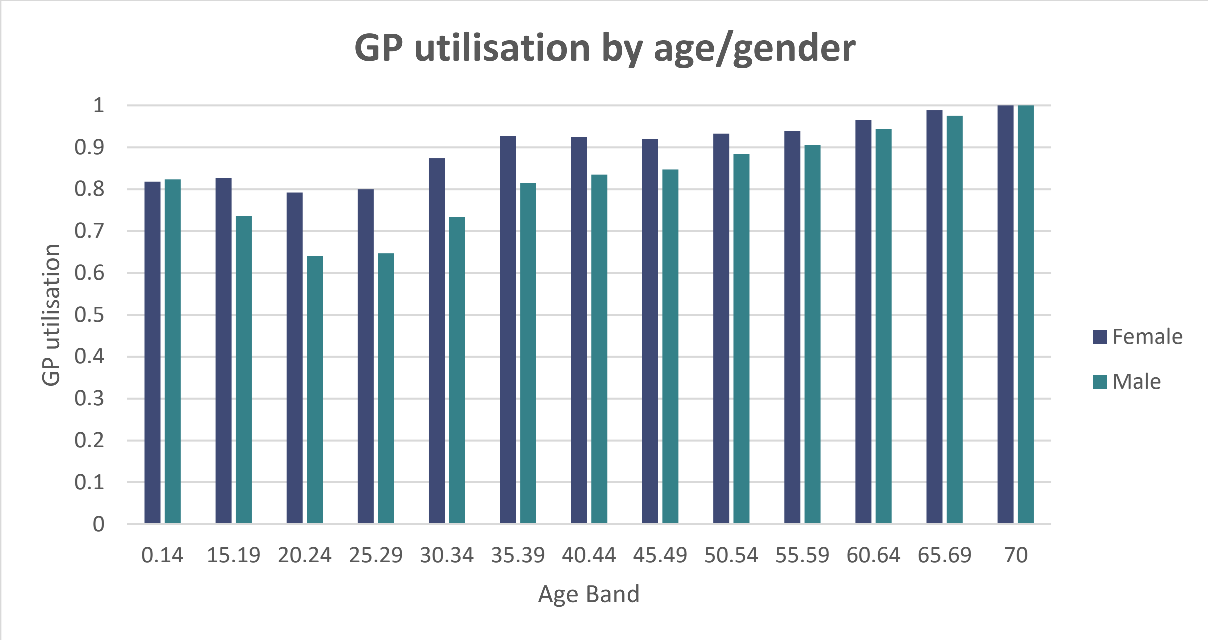
Methodology

The HSU module processing steps are outlined in the diagram below:



1. **Aggregate Health Conditions module output**: Take the Health Conditions Module output containing prevalence estimates at the SA2 level for each age, gender and co-morbidity profile.
2. **Prepare MBS service utilisation data:** Calculate the average GP service utilisation profile for each combination of age, gender and co-morbidity cohort:
   a. MBS GP items are categorised by inferred chronic condition. This is used to infer chronic conditions to individuals based on their service usage patterns in the base year.
   b. The top 10 MBS GP items by volume are modelled explicitly, with the remaining low volume items aggregated together.
3. **Calculate the expected population interacting with MBS GP services**: The overall proportion of the population claiming GP services is published annually by the Department in the GP Stats report ('% of the population that were patients'). This compares the count of unique patient IDs in the MBS claims data to the ABS ERP data for the corresponding period.

The same calculation was expanded to disaggregate the data by age and gender – results shown below:



4. **Apply the MBS service usage data to the health conditions module output:** Build bottom-up estimates of GP demand by:
    a. Weight each population cohort by the MBS/ERP data to derive the expected population using GP services
    b. Multiply the average service usage to the weighted population to derive the expected GP FTE/service demand (Unmet demand)

Calculation formula:

GP Service Volume = Population in cohort * % cohort accessing GP Services * Average Services

GP FTE = GP Service Volume in cohort * average GP FTE per item

5. **Apply initial equilibrium assumption to produce Baseline Demand projections:** apply final calibrations factors by geography to benchmark the GP demand to historical levels.

Assumptions

#	Assumption	Detail	Implication
1	Inferred health conditions based on MBS items	Health conditions are inferred to individuals in the MBS data based on their claiming activity. The inferred prevalence rates of each in-scope condition were validated against the ABS NHS prevalence rate by age and gender.	For most in-scope conditions the presence of MBS claiming activity in the base year was sufficient to replicate the expected prevalence rate with the MBS inferred rates To improve the identification of individuals with CVD the identification window was extended to any CVD related MBS item claimed in the last 5 years. To ensure that the inferred Type 2 diabetes prevalence rate matched expectations all gestational diabetes related items delivered outside of the GP setting were excluded. Several proposed future enhancements are outlined to address potential improvements in the inference methodology.
2	GP FTE/duration per item static across projection period	The observed GP FTE per item observed in the base year is assumed to remain static across the projection period.	The model will not capture changes in GP behaviours relating to consultation duration for certain items.

Glossary

Term	Definition
FTE	Full Time Equivalent
GP Type	There are three types of GPs in Australia, Vocationally Registered (VR), Non-Vocationally Registered (Non-VR) and GP Trainees. 1. VR GPs are fully qualified GPs that can perform all services and bill Medicare at the fully rate, whilst non-VR GPs provide the same services and level of care but are only allowed to bill Medicare at a discounted rate. 2. Non-VR GPs are encouraged to complete a training program and become a VR GP with many programs incentivising this transition. 3. GP Trainees are still undertaking training and are not fully qualified, however provide some services to patients.
Place of initial medical qualification	The place of initial medical qualification indicates where the individual completed medical school: 1. Australia Medical Graduate (AMG) – these individuals graduated from an Australian medical school. 2. International Medical Graduate (IMG) – these individuals graduated from an international medical school that is not accredited to Australian requirements. 3. Foreign graduates of accredited medical schools (FGAMS) – these individuals are international trainees who have graduated from a medical school that is accredited to Australian requirements.
AGPT	Australian General Practitioner Training
PEP	Practice Experience Program
FSP	Fellowship Support Program
IP	Independent Pathway
RVTS	Remote Vocational Training Scheme
RGTS	Rural General Training Scheme
MDRAP	More Doctors for Rural Australia Program
AMDS	Approved Medical Deputizing Service
3GA programs	Section 3GA programs are programs specified in the Regulations, for the purposes of s 3GA of the Act. 3GA Programs allow non-vocationally recognised doctors to work towards their qualification on a workforce or training program.
Moratorium	International medical graduates (IMGs) must work in a priority area for a minimum of 10 years from medical registration as required under 19AB of the Health Insurance Act 1973.

Bonded return of service	Australian Medical Graduates that were in a Commonwealth Supported Place (CSP) in a medical course at an Australian university are required to do a return of service by working in an eligible regional, rural and remote area for a specified period of time after the completion of their medical course.
DPA	The Distribution Priority Area (DPA) classification identifies locations in Australia with a shortage of general practitioner (GP) services. International Medical Graduates (IMGs) must work in a DPA to be eligible to access Medicare.

www.hwd.health.gov.au

All information in this publication is correct as at August 2024