

RESEARCH

Open Access



Measuring unequal distribution of pandemic severity across census years, variants of concern and interventions

Quang Dang Nguyen^{1†}, Sheryl L. Chang^{1,2*†}, Christina M. Jamerlan¹ and Mikhail Prokopenko^{1,2}

Abstract

Background The COVID-19 pandemic stressed public health systems worldwide due to emergence of several highly transmissible variants of concern. Diverse and complex intervention policies deployed over the last years have shown varied effectiveness in controlling the pandemic. However, a systematic analysis and modelling of the combined effects of different viral lineages and complex intervention policies remains a challenge due to the lack of suitable measures of pandemic inequality and nonlinear effects.

Methods Using large-scale agent-based modelling and a high-resolution computational simulation matching census-based demographics of Australia, we carried out a systematic comparative analysis of several COVID-19 pandemic scenarios. The scenarios covered two most recent Australian census years (2016 and 2021), three variants of concern (ancestral, Delta and Omicron), and five representative intervention policies. We introduced pandemic Lorenz curves measuring an unequal distribution of the pandemic severity across local areas. We also quantified pandemic bimodality, distinguishing between urban and regional waves, and measured bifurcations in the effectiveness of interventions.

Results We quantified nonlinear effects of population heterogeneity on the pandemic severity, highlighting that (i) the population growth amplifies pandemic peaks, (ii) the changes in population size amplify the peak incidence more than the changes in density, and (iii) the pandemic severity is distributed unequally across local areas. We also examined and delineated the effects of urbanisation on the incidence bimodality, distinguishing between urban and regional pandemic waves. Finally, we quantified and examined the impact of school closures, complemented by partial interventions, and identified the conditions when inclusion of school closures may decisively control the transmission.

Conclusions Public health response to long-lasting pandemics must be frequently reviewed and adapted to demographic changes. To control recurrent waves, mass-vaccination rollouts need to be complemented by partial NPIs. Healthcare and vaccination resources need to be prioritised towards the localities and regions with high population growth and/or high density.

Keywords Agent-based modelling, COVID-19, SARS-CoV-2, Pandemic inequality, Urbanisation effects

[†]Quang Dang Nguyen and Sheryl L. Chang have contributed equally to this work.

*Correspondence:

Sheryl L. Chang

sheryl.chang@sydney.edu.au

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

On 30 January 2020, the COVID-19 was recognised by the World Health Organisation (WHO) as a public health emergency of international concern: the WHO's highest level of alert. On 11 March 2020, this was followed by the WHO declaring the outbreak a pandemic [1]. On 5 May 2023, that is, 170 weeks since announcing the global health emergency, the WHO declared an end to the emergency, while continuing to refer to COVID-19 as a pandemic [2]. Over this time, the COVID-19 pandemic has had a profound impact, causing significant loss of life, reducing life expectancy [3, 4], seriously challenging healthcare systems [5], and adversely affecting socio-economic activity worldwide [6]. By mid-June 2023, the pandemic had caused almost 800 million confirmed cases and 7 million confirmed deaths [7].

Over the course of pandemic, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which causes COVID-19 has mutated from its ancestral strain into a number of lineages and sub-lineages, varying in terms of infectivity and virulence. Several of these variants have been designated by the WHO as variants of concern, including the highly transmissible lineages B.1.617.2 (Delta) [8, 9] and B.1.1.529 (Omicron) [10, 11]. In Australia, these variants have triggered significant new waves of the pandemic [12–15].

In response, the public health systems worldwide employed a diverse range of intervention policies. The initial reactions involved non-pharmaceutical interventions (NPIs) [16–18]. Typically, the NPI interventions combined various components, such as border closures and travel restrictions, case isolation, home quarantine, school closures, and comprehensive “stay-at-home” orders comprising social distancing. Once safe and effective vaccines became available, this was followed by mass vaccination campaigns around the globe [19–21]. Vaccination rollouts differed with respect to (a) population coverage, ranging from partial to nearly complete; (b) rollout schemes, e.g., preemptive, progressive, or boosting; as well as (c) different vaccine combinations, e.g., priority and general vaccines [21]. Each vaccine had efficacy variations with respect to (i) the susceptibility-reducing efficacy, (ii) the disease-preventing efficacy, and (iii) the transmission-limiting efficacy [19, 21]. In addition, vaccine effectiveness was nonlinearly diminishing over time [22]. The practice of dealing with multiple variables, objectives and constraints confounded many potential effects of various intervention policies: for example, school closures have been found to contribute differently under different circumstances [18, 23].

The combined effects of the evolving viral lineages and complex intervention policies have been difficult to systematically analyse, model and predict. For example, the

persistence of the Omicron variant in Australia and the resulting recurrent waves were explained by a nuanced combination of the new transmissible sub-variants, the fluctuating adoption of NPIs, and the waning immunity from prior infections and vaccinations [15]. Importantly, such complex effects become sensitive to demographic variations in heterogeneous populations spanning different age groups, household sizes, socio-economic profiles and jurisdictions. In general, to study a pandemic which has lasted more than three years, one needs to account for demographic changes which play an increasingly salient role. This influence often remains concealed due to the lack of high-resolution data and presence of jurisdictional barriers, socio-political biases, and other factors [24, 25].

Here, we aim to examine some of these public health challenges and carry out a systematic simulation-based analysis of several COVID-19 pandemic scenarios, using Australia as a case study. We use two most recent census years (2016 and 2021) as the alternative demographic settings within which each pandemic scenario is simulated. Our comparative analysis contrasts three variants of concern which made an impact in Australia: the ancestral strain, the Delta and the Omicron lineages. For every scenario, five representative intervention policies are compared, ranging from (1) baseline (i.e., no interventions), to (2) partial NPIs without vaccination, (3) partial preemptive vaccination without NPIs, (4) mixed intervention with both partial NPIs and partial vaccination, and (5) partial lockdown including school closures.

In order to compare 30 possible scenarios ($2 \times 3 \times 5$), we apply an agent-based model (ABM) which simulates an artificial population generated using the high-resolution Australian census data. The ABM has been previously calibrated and validated for several pandemic stages in Australia during the last four years [13, 15, 18, 21, 26].

This study identifies and explains, in context of different variants and policies, several coupled nonlinear effects of the population growth and heterogeneity on the pandemic severity. In particular, we study how pandemic peaks may be amplified by distributed changes in the population size or the changes in population density.

Importantly, the study introduces a novel measure of pandemic inequality—*pandemic Lorenz curves*—and demonstrates that the pandemic severity may be distributed unequally across local areas. For example, while the pandemic inequality may reduce when the population or the disease transmissibility grow, it may increase with more stringent interventions or in non-urban areas. We also measure *pandemic biomodality*, which characterises formation of distinct urban and regional pandemic waves. In addition, we measure *bifurcations in the*

effectiveness of interventions, e.g., school closures, relative to variants of concern.

Methods

Agent-based modelling COVID-19 pandemic with Australian census

We simulated several scenarios of the COVID-19 pandemic in Australia using a well-established agent-based model previously validated for several pandemic waves and variants of concern [13, 15, 18, 21]. Our model has two fundamental components: (i) a simulated Australian population generated to represent key demographic features of the Australian census data, and (ii) a stochastic agent-based model for disease transmission and control, detailed in Appendix B and Appendix C, respectively.

Our model comprises stochastically generated anonymous agents that represent the population of Australia: about 23.4 million using 2016 census, and 25.4 million using 2021 census. The population is partitioned into Statistical Areas (SAs) at different resolutions, e.g., SA2 level represents suburbs. The surrogate Australian population is constructed using a number of high-resolution datasets, including demographic data from the Australian census, international air traffic reports from the Bureau of Infrastructure and Transport Research Economics (BITRE), and educational registration records from the Australian Curriculum, and Assessment and Reporting Authority (ACARA). The generated population matches the population characteristics in terms of age, gender, household composition, student enrollment, workforce mobility, and international travel, as detailed in Appendix B.

Disease transmission follows a discrete-time simulation, updating states of each agent over time. Following initial infections “seeded” around international airports, the transmission is probabilistically simulated by considering agent interactions across multiple social layers (mixing contexts), given different contact and transmission rates within both residential and work/study contexts.

If an agent is exposed to the disease in one of their mixing contexts, it goes through several health states following the natural history disease model: Susceptible, Latent, Infectious (asymptomatic or symptomatic), and Removed (recovered or deceased). For an agent i , let us denote the set G_i consisting of all mixing contexts in which this agent interacts (e.g., workplace or school, grade and class in daytime cycles; and household, household cluster, neighbourhood and community in nighttime cycles). At time cycle n , the probability for an susceptible agent i becoming infected across context $g \in G_i$ is determined as follows:

$$p_i^g(n) = 1 - \prod_{j \in A_g \setminus \{i\}} (1 - p_{j \rightarrow i}^g(n)) \tag{1}$$

where $A_g \setminus \{i\}$ represents the set of agents in the context $g \in G_i$ excluding agent i , and $p_{j \rightarrow i}^g(n)$ denotes the instantaneous probability that an infectious agent j who shares the context g with susceptible agent i , transmits the infection to agent i . The transmission probability $p_{j \rightarrow i}^g$ is determined by epidemiological characteristics and variant-specific natural history of the disease (see Appendix C, and section “Variants of concern”).

Then, the infection probability for agent i across all mixing contexts is calculated as follows:

$$\begin{aligned} p_i(n) &= 1 - \prod_{g \in G_i(n)} (1 - p_i^g(n)) \\ &= 1 - \prod_{g \in G_i(n)} \prod_{j \in A_g \setminus \{i\}} (1 - p_{j \rightarrow i}^g(n)) \end{aligned} \tag{2}$$

Various intervention policies may reduce this probability (see section “Intervention policies”, Appendix “Non-pharmaceutical interventions” and “Vaccination”).

Simulated pandemic scenarios

Variants of concern

Our model has been calibrated to match key COVID-19 characteristics across three variants of concern: ancestral (i.e., the strain initially detected in Wuhan, which was prevalent in Australia in 2020) [18], Delta (i.e., B.1.617.2 variant, prevalent in Australia in 2021) [13], and Omicron (i.e., B.1.1.529 variant, prevalent in Australia during 2022) [15]. Over the last four years, these variants have not only evolved towards higher infectivity (i.e., higher basic reproductive number, R_0), but have also exhibited distinct characteristics in the disease natural history. In this study, we performed a lateral comparison of these three variants, each following a different natural disease history (as shown in Fig. 1), combined with the intervention policies described in section “Intervention policies”, and using the data from two most recent census years (2016 and 2021). See Appendix C for a detailed parameterisation of the relevant epidemiological characteristics.

Intervention policies

NPIs considered in our model include case isolation (CI), home quarantine (HQ), social distancing (SD), and school closures (SC), each affecting different agents based on their health states (infected or susceptible), age groups (school-aged or not), and household compositions (if there is an infected household member).

In simulating pandemic scenarios, we assumed a moderate level of preemptive vaccination coverage of the

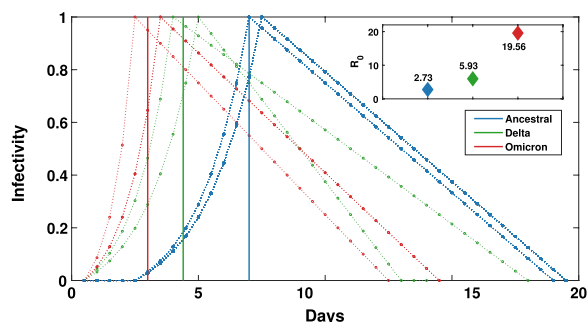


Fig. 1 Model of the natural history of three COVID-19 variants: ancestral (blue), Delta (green), and Omicron (red). The illustrated profiles are sampled from 2 random agents. Each profile rises exponentially until reaching the infectivity peak, followed by a linear decrease until full recovery. Vertical lines mark the mean incubation period for the three considered variants (ancestral: blue, Delta: green, and Omicron: red), with the means following a log normal distribution. The mean incubation period and recovery period for each of the variants are reported in Appendix Table 6. The inset shows R_0 of the three considered variants

population (50%), accounting for the combined effects of a relatively high vaccination coverage in Australia [27] and low diminishing vaccine efficacy [28]. In line with prior studies [13, 15, 21], the vaccination scheme distributes two types of vaccines (priority and general), each with the vaccine efficacy defined in terms of reducing susceptibility, preventing symptoms of the disease, and limiting further transmission, as described in Appendix section "Vaccination".

We simulated five specific policies, each with a different combination of preemptive vaccination coverage and NPIs. These policies cover a wide range of scenarios, starting from the "live-as-usual" intervention-free

scenario to the lockdown-like scenario with strong restrictions limiting population mobility and social interactions, as shown in Fig. 2. Specifically, these policies can be summarised as follows:

- Policy 1—baseline: no NPIs and no preemptive vaccination coverage, representing a scenario without any interventions.
- Policy 2—partial NPIs: some NPIs implemented (CI, HQ and SD at 70% compliance level), with no preemptive vaccination coverage. This represents a pandemic intervention scenario feasible without vaccines.
- Policy 3—partial vaccination: 50% preemptive vaccination of the population before a pandemic wave. This represents a scenario developing in a population with partially acquired immunity, but without any restrictions on social interactions during the pandemic.
- Policy 4—mixed intervention: some NPIs implemented during a pandemic wave (CI, HQ and SD at 70% compliance level) assuming that 50% of the population has been preemptively vaccinated prior to the pandemic wave. This represents a scenario with partial acquired immunity in the population, followed by further restrictions on social interactions during the wave.
- Policy 5—partial lockdown: all NPIs implemented (CI, HQ, SD at 70% compliance level, and SC in addition) and 50% preemptive vaccination of the population. This scenario represents a lockdown of the partially immunised population using strong restrictions on social interactions.

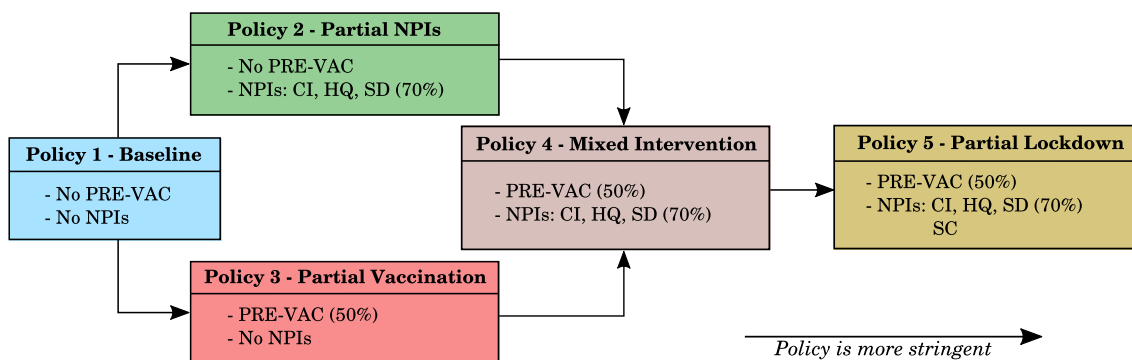


Fig. 2 Five simulated intervention policy scenarios. PRE-VAC: preemptive vaccination prior to the pandemic. NPIs: non-pharmaceutical interventions. Policies are considered to be more stringent moving from left to right. The macro- and micro-parameters for NPI-related policies are summarised in Appendix Table 7. Parameters relating to the vaccination coverage and vaccine efficacy are summarised in Appendix Table 8

Lorenz curves: measuring unequal distribution of pandemic severity

Different communities may experience impacts of interventions in significantly different ways, and these complex effects may not uniformly aggregate into the national pandemic dynamics (e.g., nationwide incidence and cumulative incidence). To examine distribution of the overall pandemic severity across different local areas, we introduce a novel technique based on Lorenz curves.

Lorenz curves, initially proposed to evaluate the degree of inequality in wealth distribution in society [29], have since been applied in many other domains, such as economics [30, 31], underpinning the well-known Gini index which measures income/wealth distribution across a population [32], and biology [33, 34]. In this study, we proposed and constructed the pandemic Lorenz curves that capture inequality in the distribution of cumulative incidence at the SA2 level, and compared their shapes across the considered scenarios. The pandemic Lorenz curve dissects the nationwide cumulative incidence at the SA2 level, tracing it across all SA2 areas and assessing their relative contribution to the pandemic severity.

Figure 3 shows a simplified example to demonstrate possible shapes of pandemic Lorenz curve where the x-axis represents the cumulative fraction of SA2 residential population, ranked by their local attack rate (cumulative incidence over the SA2 residential population), and the y-axis represents the fraction of cumulative incidence at the ‘global’ national level, contributed by the corresponding fraction of SA2 residential population.

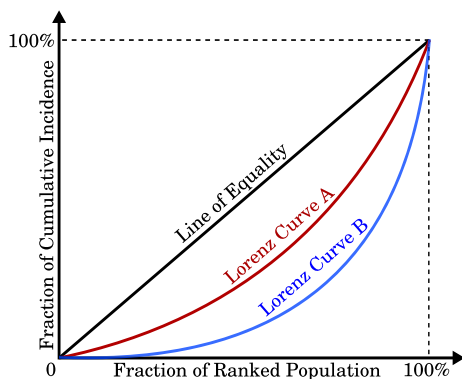


Fig. 3 Pandemic Lorenz curves measuring inequality in distribution of the pandemic severity. The black line represents the line of equality where each SA2 contributes equally to the cumulative incidence. A curve closer to the line of equality (i.e., Lorenz Curve A, shown in red) indicates that the contributions of SA2 residential areas towards the aggregate cumulative incidence in response to a specific intervention policy A are more equally distributed than the contributions of these areas under policy B which are traced by the curve shaped further away from the line of equality (i.e., Lorenz Curve B, shown in blue)

Appendix D provides a detailed explanation of the Lorenz curves introduced in our study to measure the unequal distribution of pandemic severity.

Results

We present our results in three parts matching our key objectives, firstly examining the effects of population heterogeneity on pandemic severity across two census years (section "Effects of population heterogeneity on pandemic severity"), then exploring pandemic spread, under the considered policies, in terms of urbanisation (section "Effects of urbanisation on pandemic spread"), and finally examining varying effects of a specific intervention policy—school closures—across the three considered variants of concern (section "Effects of school closures across variants of concern").

Effects of population heterogeneity on pandemic severity

We related the population heterogeneity with the pandemic severity observed in simulated scenarios across different intervention policies. Here, we measured the pandemic severity as the normalised incidence per million (unless specified otherwise), computed as the ratio between the detected incidence cases to the total population (in millions) for the considered census year. We assessed the population heterogeneity in terms of population increase at the ‘global’ national level (section "Population growth amplifies pandemic peaks") and at the ‘local’ SA2 level (sections "Changes in population size amplify incidence peak more than changes in density" and "Pandemic severity is distributed unequally across local (SA2) areas"). Appendix A provides more information about the population structure captured by the Australian Bureau of Statistics (ABS).

Population growth amplifies pandemic peaks

Our results show that the population growth, i.e., the 8.6% population increase between 2016 and 2021, produced a nonlinear response effect on the pandemic severity. If the pandemic severity was proportional to the population growth, the relative change in incidence between the two years could be expected to be zero (i.e., flat line in Fig. 4, bottom row). However, comparison of the simulated pandemic scenarios between 2016 and 2021 produces a relative change in normalised incidence which non-trivially diverges from a flat profile (Fig. 4, bottom row), especially in scenarios with less stringent policies (e.g., policies 1, 2 and 3). It is clear that the divergence is positive (i.e., a higher normalised incidence in 2021, indicating an amplified nonlinear response to population growth) around the incidence peak, followed by a negative oscillation (i.e., a lower normalised incidence in 2021, suggesting a negative response to population

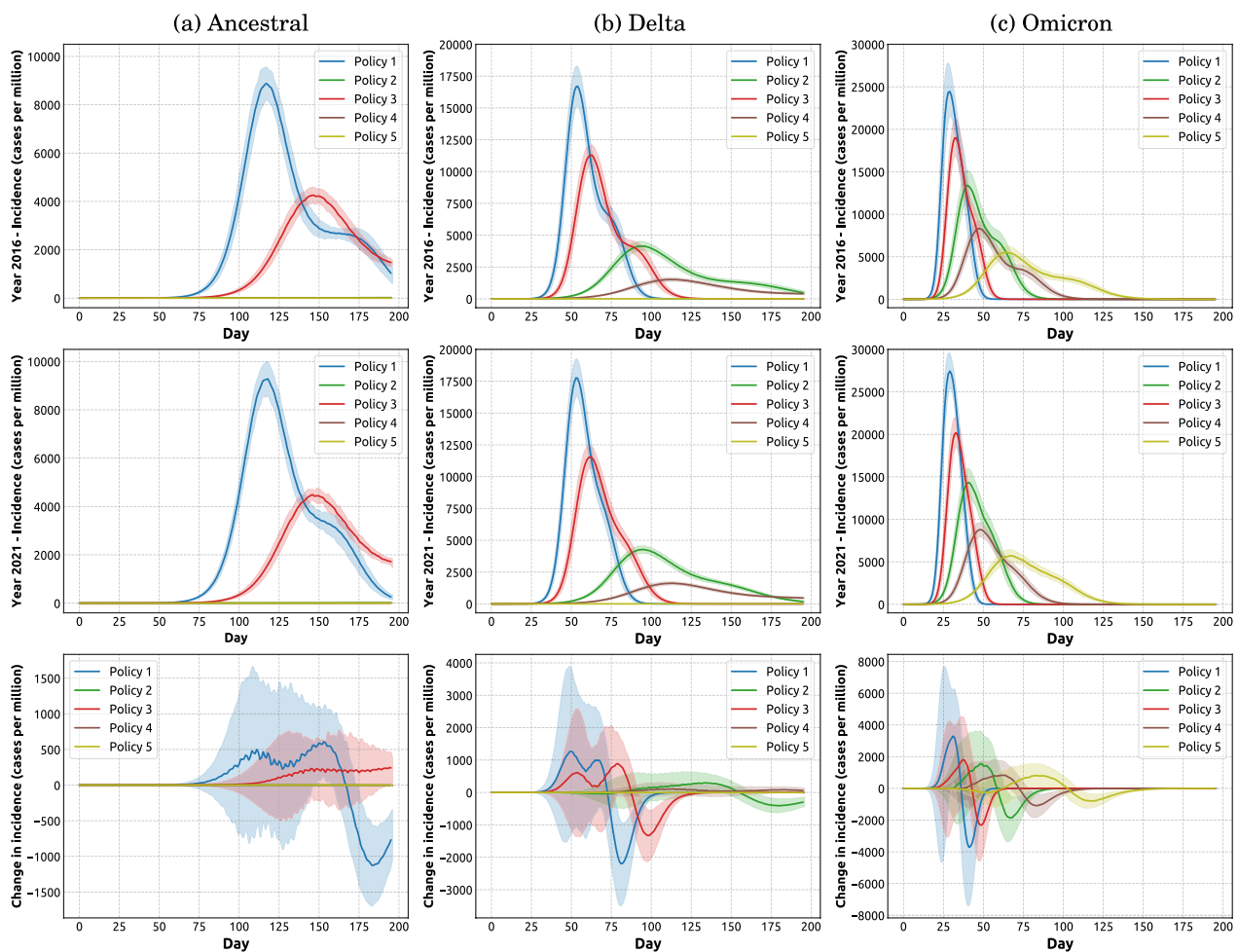


Fig. 4 Impact of different intervention policies on pandemic severity for three considered variants simulated for two census years (top row: 2016; middle row: 2021; bottom row: relative change between years). Each column compares the impact of five intervention policies for one variant of concern: **a** ancestral; **b** Delta; **c** Omicron. See Fig. 2 for a detailed description of the considered intervention policies. Coloured shaded areas around solid lines show standard deviation. Each profile corresponds to one intervention policy and is computed as the average over 100 runs

growth) after the peak. Figure 5 (top row) directly contrasts the pandemic profiles across two census years.

This nonlinear response pattern (an early amplification compensated by a late negative oscillation) is manifested for the baseline scenarios across all variants of concern, as well as the scenarios with less stringent policies for the Omicron variant, as shown in Fig. 4 (bottom row). In general, the amplification effect is more notable in the scenarios associated with more transmissible variants (Fig. 5). Specifically, for the Omicron variant, the incidence peak for Policy 1 (baseline) simulated for 2021 census data (Fig. 5a, red profiles, solid and dashed lines), is 12.21% higher than its counterpart produced for 2016 census. This nonlinearity arises due to the population distribution which non-uniformly affects the pandemic severity, amplifying the peak incidence (see the following subsection).

Changes in population size amplify incidence peak more than changes in density

To investigate possible causes that have contributed to the nonlinear response effects observed at national level, we considered the demographic changes which occurred between the two census years at SA2 level. The higher resolution offered at this level enabled us to trace how the local areas contribute to the disproportionate response in relative incidence between the two census years, 2016 and 2021 (see Appendix A for the demographic statistics and structure captured by census data). Specifically, we examined changes in the incidence peaks across 2147 SA2 areas which were registered in both 2016 and 2021 census years. In doing so, we used the baseline scenario implementing Policy 1 (no interventions) and correlated the peak incidence with specific demographic changes: the residential population size, the population density,

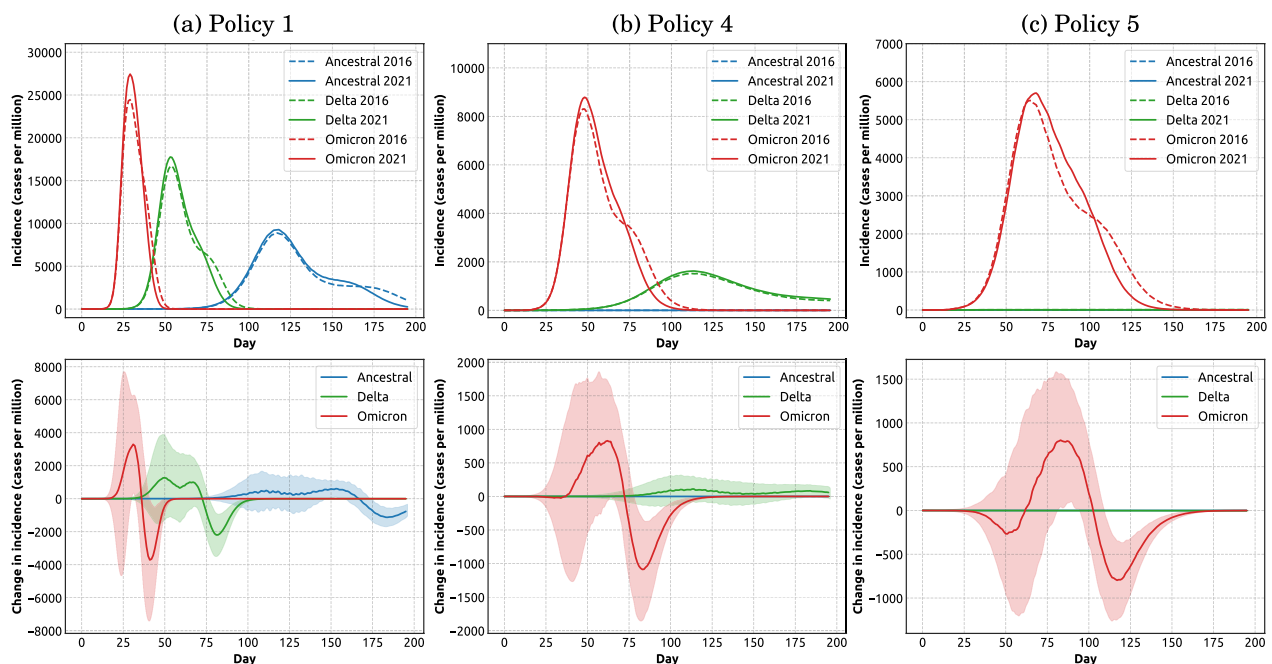


Fig. 5 A comparison of pandemic severity for different policies across three considered variants (ancestral: blue; Delta: green; Omicron: red) and two census years (solid line: 2021; dashed line: 2016). The severity of each variant is measured by cases per million (top row). The change in incidence (bottom row) is calculated as the difference of incidence cases per million between two census years. Each column compares the impact of three variants for one intervention policy: **a** Policy 1; **b** Policy 4; and **c** Policy 5. See Fig. 2 for a detailed description of the considered intervention policies. Coloured shaded areas around solid lines (in bottom row) show standard deviation. Each profile (solid and dashed lines) corresponds to one intervention policy and is computed as the average over 100 runs

and the populations residing in households of different sizes, all measured at SA2 level.

We found that there is a strong positive correlation between the changes of peak incidence and the changes in the “usual residential” population. This finding is supported by high correlation coefficients across all variants ($0.64 \leq r \leq 0.91$, as shown in Fig. 6). That is, an SA2 with a greater net population influx in the five-year period between 2016 and 2021 is highly likely to have a higher spike in the incidence peak. The impact of population increase is further amplified for highly transmissible variants, resulting in a greater peak incidence difference shown by the steeper slope observed for the Omicron variant (Fig. 6, in red). In addition to the correlation with the usual residential population, we also found a weaker positive correlation ($0.34 \leq r \leq 0.45$) between the changes in peak incidence and the changes in population density, suggesting that SA2 areas which develop a higher population density may also have a higher incidence peak across all variants (Appendix Fig. 20).

We also found that the identified nonlinear response patterns cannot be attributed to differences in the average household size between census years. Although the

SA2 areas with a higher average household size experience slightly worse pandemic effects across all considered variants (Appendix Fig. 21), there is no clear correlation between the difference in average household size and the peak incidence difference between 2016 and 2021, as summarised in Appendix Table 11 for Pearson correlation coefficient, and shown in Fig. 22, across the considered variants.

To further investigate a potential impact of the household size on the nonlinear pandemic response, we partitioned each SA2 population into two sub-populations: (i) residents of large households (i.e., with at least five household members), and (ii) residents of small households (i.e., with up to four household members). Irrespective of the household size, we observed a strong correlation between each sub-population and the peak incidence. This is shown in Fig. 23 for large households, and Fig. 25 for small households. In particular, the sub-population residing in small households shows higher correlations for all considered variants across two census years, with high Pearson coefficients ranging between 0.95 and 0.99 (Appendix Table 12).

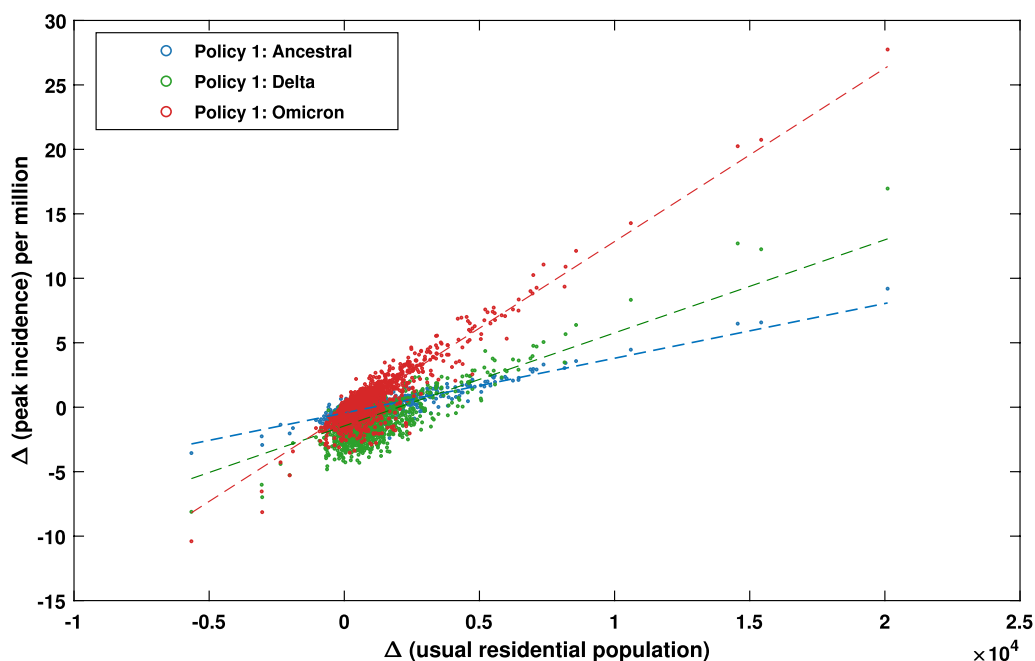


Fig. 6 Positive correlation between the usual residential population difference and the peak incidence difference between 2016 and 2021 at SA2 resolution for three considered variants: Ancestral (blue), Delta (green), and Omicron (red). Dashed lines represent linear fitting for each of the profiles (see Appendix Table 10 for statistical analysis). Data points corresponding to each SA2 are computed as the average over 100 runs. Total number of overlapping SA2 between 2016 and 2021 census years: 2147. Pearson correlation coefficients: $r_{Ancestral} = 0.7717$, $r_{Delta} = 0.6447$, $r_{Omicron} = 0.9002$

However, when considering the correlation between the peak incidence difference between 2016 and 2021 and the difference in a partitioned sub-population—large-household (Fig. 24) or small-household (Fig. 26)—we found that, for the Delta variant, the peak incidence difference is stronger correlated with difference in the large-household sub-population ($r = 0.611$) than difference in the small-household sub-population ($r = 0.522$). Appendix Table 12 summarises the comparative analysis.

Pandemic severity is distributed unequally across local (SA2) areas

Using Lorenz curves, we examined how the local (SA2) areas are impacted by different intervention policies across two demographic settings representing two census years. Figure 7a and b shows that for the baseline scenario without any interventions (Policy 1), the pandemic effects are distributed equally across the areas, for all considered variants simulated for both census years (Appendix Fig. 27 provides a different layout of the same results). Comparing across two census years, we note that Lorenz curves produced for 2021 census are closer to the line of equality (i.e., the diagonal line) for the less transmissible variants (i.e., ancestral and Delta). This suggests that for these two variants, the SA2 areas contribute

to the aggregate attack rate at the national level more equally in 2021 scenarios than in their 2016 counterparts.

For the ancestral and Delta variants, more stringent intervention policies led to a more diverging contribution pattern (i.e., Lorenz curve shaped further away from the line of equality). For example, for the Delta variant, using 2016 census data, we found that the increase in the fraction of cumulative incidence from 20 to 80% under Policy 1 (Fig. 7a top row, blue profile) is attained by an equal increase (i.e., 20 to 80%) in the population ranked by their local SA2 attack rate. However, the same increase in the incidence fraction (i.e., 20 to 80%) under Policy 4 (Fig. 7a top row, purple profile), was attained by a smaller set of SA2s comprising only approximately 40% to 85% of the population ranked by their local attack rate. This finding indicates that the pandemic severity is distributed more unequally under more stringent interventions (including NPIs and vaccinations). Specifically, SA2 areas with a higher local attack rate (i.e., ranked higher on x-axis) account for a higher fraction of the cumulative incidence.

We also note that the unequal distribution of pandemic severity diminishes for the highly transmissible Omicron variant (shown in Fig. 7c) where all simulated intervention policies failed to adequately slow down the transmission. This resulted in all SA2 areas contributing equally to

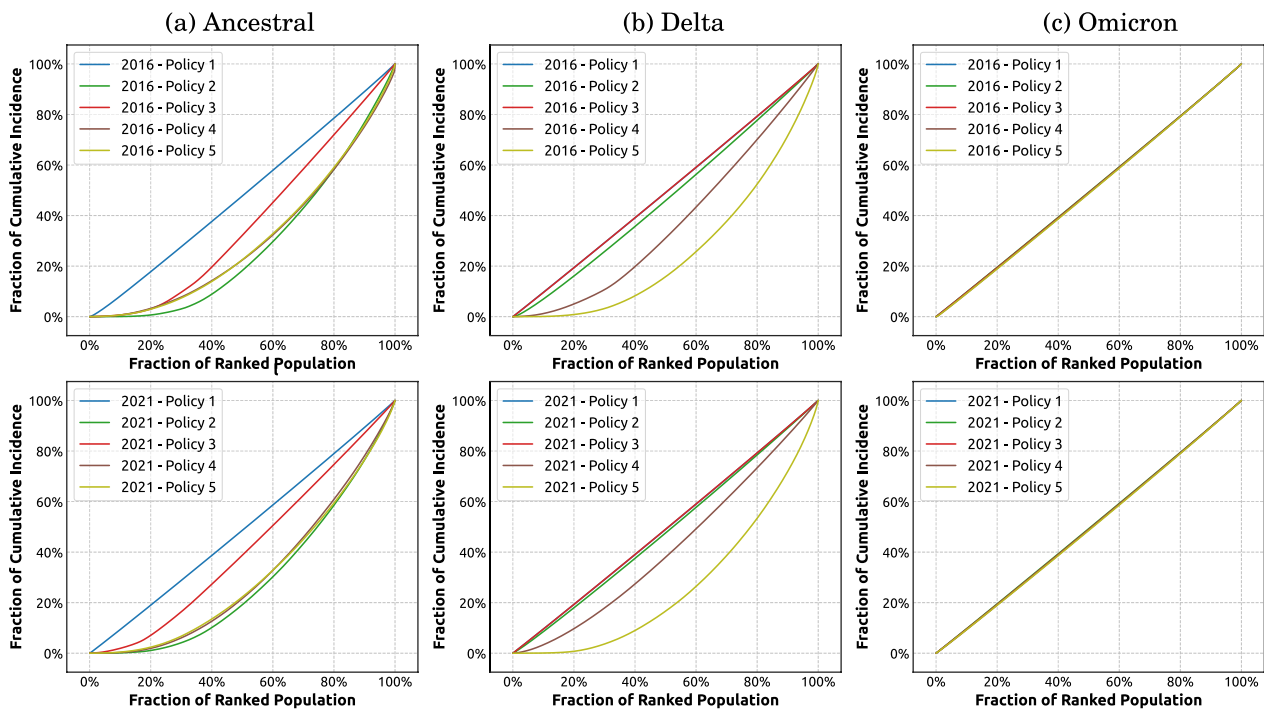


Fig. 7 Pandemic Lorenz curves measuring distribution of pandemic effects across SA2 areas for considered variants, years and policies. Each column compares the impact of five intervention policies for one variant: **a** ancestral; **b** Delta; **c** Omicron. Top row: 2016; bottom row: 2021. Refer to Fig. 2 for a detailed description of the considered intervention policies. Each profile corresponds to one intervention policy and is computed as the average over 100 runs

the national-level cumulative incidence, with all Lorenz curves overlapping with the line of equality.

Effects of urbanisation on pandemic spread

The bimodal pandemic dynamics is characterised by the incidence peak (attributable to predominantly urban spread), followed by an inflexion point which shapes around a smaller secondary wave (attributable to mostly regional and rural areas) [35]. Typically, the first wave triggered by the international air traffic is rapidly shaped

in urban populations concentrated near international airports. In contrast, the pandemic progression into non-urban regions (i.e., areas further away from international airports) is significantly slower. The confluence of these factors resulting in bimodal dynamics in Australia was detected and described in context of simulating the pandemic influenza using 2011 and 2016 census data [35]. Our simulations of the COVID-19 pandemic scenarios in this study also produced bimodal dynamics, especially for the ancestral and Delta variants using 2016 census

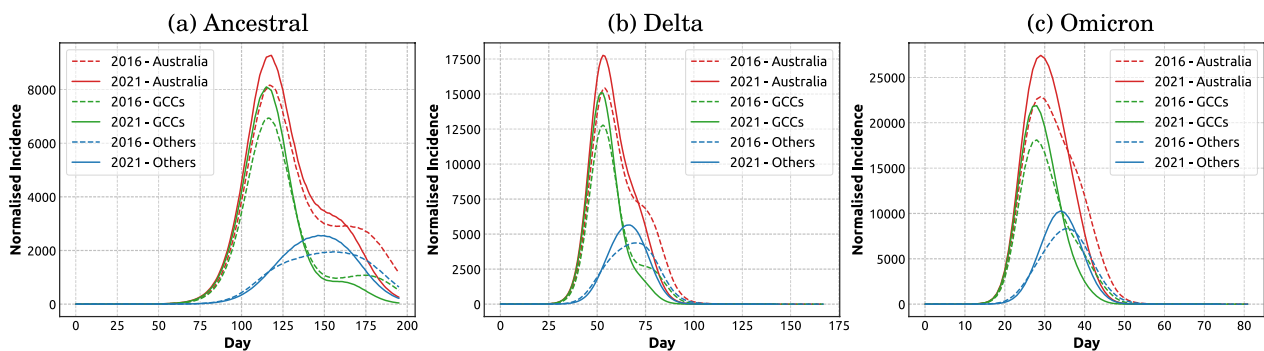


Fig. 8 Comparison of pandemic waves in Greater Capital Cities (GCCs) and all other areas. Each column compares the baseline scenario (Policy 1) for 2016 and 2021 census data, for a variant of concern: **a** ancestral; **b** Delta; **c** Omicron. Each profile is computed as the average over 100 runs

data (Fig. 8). However, the bimodality is less prominent in scenarios using 2021 census data, indicating a shift of the pandemic dynamics between urban and regional regions.

This observation is also supported by the pandemic dynamics examined at the SA2 level. Appendix Fig. 19 shows that in comparison to the simulated results for 2016 census, the time gap between the first and second pandemic waves (defined in terms of the number of SA2 areas that peaked on a given day) for 2021 census has become shorter at the SA2 level. These findings indicate that there are more intricate structural demographic changes in the population which occurred between 2016 and 2021, beyond a uniform population growth.

To explain this phenomenon, we assessed the pandemic progression during the baseline scenario (Policy 1), by tracing pandemic waves in urban and non-urban SA2 areas (see Appendix Table 13 for the population statistics). Figure 8 compares the Greater Capital Cities (GCCs) against other areas (i.e., all remaining SA2s). We note that the two distinct pandemic progression profiles (urban and non-urban) are mostly separated by the initial conditions: the national incidence peak is largely attributed to urban SA2 areas while the inflexion point observable at a later stage of the pandemic is caused by a secondary wave emerging in non-urban SA2 areas.

For the baseline scenarios traced for two census years, our results showed that the bimodal dynamics is diminishing from 2016 to 2021 data. This can be explained by two factors: (i) a higher incidence peak occurring in both GCCs and other urban areas, and (ii) an accelerated pandemic progression into non-urban areas (or at least, into areas outside of GCCs). This effect was observed for

scenarios across all variants of concern, with more transmissible variants showing a greater increase in the incidence peak. This observation suggests a faster pandemic spread for 2021 scenarios, indicating that the urbanisation increased over the five-year period, captured by the census statistics (see Appendix Section A), reducing the bimodal dynamics.

We also observed that the bimodality is weakened for the Omicron variant. This is explained by a reduced difference in the peak incidence timing between GCCs and non-GCCs. In other words, the two waves, urban (primary) and non-urban (secondary), tend to merge into a primary significant wave with a single incidence peak. This can be attributed to the high transmissibility of the Omicron variant which suppresses the impact of population heterogeneity. When the transmission was adequately slowed down (due to interventions, for example, implementing Policy 4 and Policy 5), the two pandemic waves became more separable, leading to notable bimodality (Appendix Fig. 28).

It is also worth noting that we do not consider re-infections in these scenarios. In other words, a higher incidence peak occurring earlier in the simulation corresponds to a reduction of the susceptible population, thus exhausting the susceptible population sooner and consequently weakening bimodality.

Effects of school closures across variants of concern

Finally, we examined the effects of school closures across three variants of concern by comparing pandemic scenarios between Policy 4 (Mixed intervention) and Policy 5 (Partial lockdown). Our results suggested that the effectiveness of school closures varies significantly for

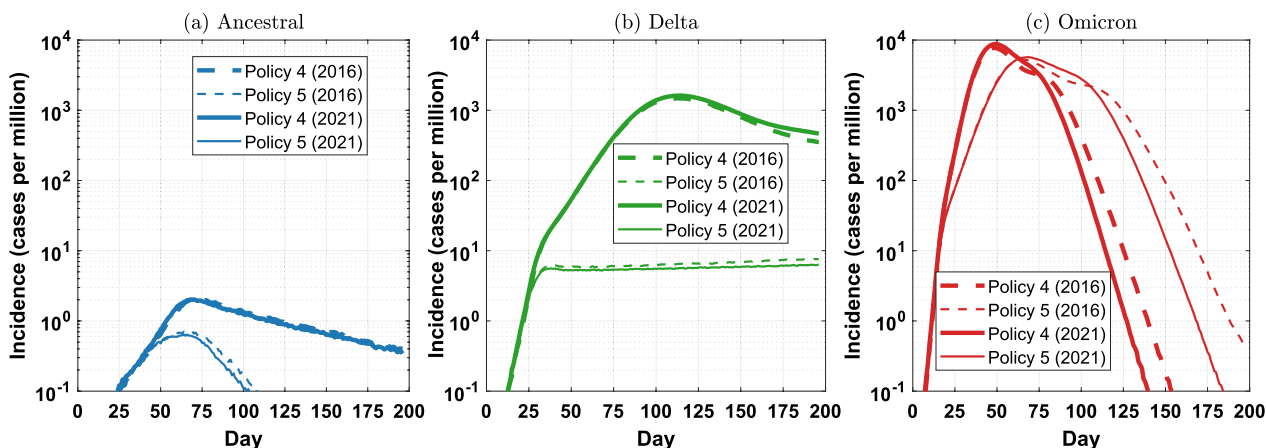


Fig. 9 Effects of school closures combined with NPIs for three considered variants (log scale): **a** ancestral; **b** Delta; **c** Omicron. School closures effectively control the spread of Delta variant, producing a sharp difference in incidence. Such a bifurcation is not observed for the ancestral and Omicron variants. Each profile corresponds to one intervention policy and is computed as the average over 100 runs. Appendix Fig. 31 shows these plots on linear scale

different variants. The effects of school closures were most prominent for the Delta variant with a two-order reduction in peak incidence (from over 1000 cases per million to under 10 cases per million), resulting in a sharp difference shown in Fig. 9b. Such a bifurcation was observed only in scenarios for the Delta variant and was not detected for variants with either lower R_0 (i.e., ancestral variant, Fig. 9a) or a higher R_0 (i.e., the Omicron variant, Fig. 9c).

For the ancestral variant, Policy 4 was sufficiently effective in containing the spread with new cases kept at a very low level. Although school closures could further reduce the peak incidence, the reduction (from around 2 cases per million to around 0.6 cases per million) would be marginal. This could be observed for both 2016 and 2021 census years (Fig. 9a).

For the Omicron variant, school closures delay the incidence peak by approximately 25 days with a sizable reduction from nearly 9000 cases per million to under 6000 cases per million (Fig. 9c). This reduction, however, would still be insufficient to curb the spread. This finding suggests that school closures could slow down the spread of the Omicron variant to some extent but would be inadequate for suppressing incidence, due to the extremely high R_0 of the Omicron variant.

Overall, these observations suggest that when coupled with NPIs and partial vaccination, school closures can be a highly effective policy for pandemic suppression of variants comparable in transmissibility with the Delta variant. These benefits, however, may not eventuate for either significantly less or significantly more transmissible variants of concern.

Discussion

In this study, we carried out a systematic comparison of pandemic scenarios across census years, variants of concern and intervention policies. The considered scenarios combined five different intervention policies with three dominant variants impacting Australia between 2020 and 2023 (ancestral, Delta, and Omicron), using a surrogate population generated based on 2016 and 2021 Australia census data. Our simulation results revealed that the population growth and heterogeneity nonlinearly affect pandemic dynamics, amplifying the peak incidence for all variants of concern. These nonlinear response effects highlight a complex interplay between the demographic and pandemic characteristics, which may amplify pandemic impacts on continuously growing populations worldwide.

Firstly, we focused on the Australian population which grew by 8.6% between 2016 and 2021 (net gain of nearly 2 million people). However, the population growth has been distributed non-uniformly across the country,

affecting both local residential population size and population density. Such unevenly distributed demographic changes directly affect pandemic progression, and our results demonstrated that the estimation of pandemic severity cannot be accomplished by a simple scaling of the outcomes obtained for previous datasets, such as 2016 census data. This strongly suggests that pandemic models and public health policies need to be frequently reviewed and adapted to changes in the population heterogeneity, especially when dealing with long-lasting pandemics similar to COVID-19.

We showed that the effectiveness of intervention policies differed across variants of concern (Figs. 4 and 5). Coupled with the observed amplification effects of demographic changes (Fig. 6), these findings call for specific interventions aimed to reduce the amplification effect, e.g., prioritising (1) the SA2 areas with a higher population growth, and (2) the SA2 areas with a higher density increase. Growth in large-household sub-populations may deserve additional attention, especially when dealing with specific variants, e.g., variants comparable in their epidemiological profile with the Delta variant.

At the same time, some policy-related findings were independent of the considered variants or demographics. In particular, the scenarios considered in this study strongly suggested that a partial preemptive vaccination rollout with 50% coverage and limited vaccine efficacy (Policy 3) would likely only be effective in a combination with partial NPIs, comprising a mixed intervention (Policy 4). Applied on its own, Policy 3 was unable to control the pandemic spread across all considered scenarios, as shown in Fig. 4. Thus, the ongoing and future immunity boosting vaccination rollouts need to aim at high population coverage comparable with original mass-vaccination campaigns, while still being complemented by partial NPIs. Similarly, partial NPIs (Policy 2) cannot provide a principled solution on its own. While managing to control pandemic scenarios for the less transmissible ancestral variant, partial NPIs did not succeed in preventing sizable incidence peaks in scenarios for the Delta and Omicron variants, but only delayed these peaks (Fig. 4). In general, such delays may be useful in rolling out booster vaccinations, reinforcing the point that a mixed intervention which combines partial NPIs and partial vaccination (such as Policy 4) may provide an adequate intervention.

Our study has also highlighted the “pandemic inequality”, with certain SA2 areas contributing to the nationwide cumulative incidence stronger than others. We found that the pandemic inequality reduces when the population or the disease transmissibility grows (Fig. 7). However, the opposite tendency—increasing pandemic inequality—was observed with more stringent

interventions (Appendix Fig. 27) or in non-urban areas (Appendix Fig. 30). While this inequality was lesser in scenarios for the highly transmissible Omicron variant, the findings still suggest that a resource prioritisation scheme is needed, with the interventions targeting the localities and regions which have been experiencing a high population growth and/or developing a high density.

Secondly, we quantified the pandemic effects of urbanisation, distinguishing between the first wave affecting Greater Capital Cities and a delayed second wave developing in regional and rural areas (Fig. 8). This bimodality indicates that, to be effective, the intervention efforts need to adapt during a pandemic progression, with the initial focus on metropolitan centres followed by a shift to non-urban areas, in anticipation of the corresponding peaks. Such geo-spatial redistribution of healthcare and vaccination resources needs to account for the transmissibility of dominant variants of concern, as more transmissible variants accelerate the pandemic, bring the urban and non-urban waves closer in time, and shorten the time between their peaks.

Finally, we evaluated the role of school closures in suppressing the pandemic transmission caused by different variants of concern in a partial lockdown-like scenario in combination with other interventions. Crucially, we demonstrated that the effects of school closures are highly dependent on the dominant variant, with the more decisive effect observed only for the Delta variant (given the range of other policy-defining parameters, i.e., the vaccination coverage of 50% and the NPI compliance with 70% SD), as illustrated in Fig. 9. This also reinforces the suggestions that policy makers should not assume that interventions will have the same effect across different variants. In addition, this highlights the possibility that some interventions can compensate others in specific circumstances: for example, when the preemptive vaccination or booster uptake is lower or slower than anticipated, school closures and stricter NPIs may be required to compensate for the lack of immunity.

In summary, the study highlighted the need for geospatially and demographically tailored, proactive and agile interventions, in contrast to general-purpose, reactive and rigid policies.

Limitations and future work

This study of pandemic severity did not include considerations of (a) socio-economic factors, and (b) disease burden in terms of hospitalisations, ICU occupancy and mortality. As demonstrated in our previous studies [13, 15, 26], these components can be included within an ABM study but substantially increase its scope.

Our simulations ran over a period of 196 days, without considering re-infections. Given the considered simulation horizon, this limitation has a minor effect discussed in our study of recurrent waves [15].

Our focus on the three dominant variants of concern rather than on their numerous sub-lineages (which may co-circulate) allowed us to distill some of the salient public health lessons. We believe these lessons would remain relevant across other sub-variants, including co-circulating ones.

We re-iterate that our aim was to compare key pandemic scenarios rather than replicate the 2021 incidence of COVID-19 in Australia. Thus, we intentionally did not use the air traffic data between 2020 to 2021 due to the severe disruption of international travel caused by the travel restrictions during the COVID-19 pandemic at the time.

Our ABM includes a substantial number of parameters, which have been calibrated to different variants or estimated using available epidemiological evidence. As more data become available, the parameter ranges may change and some estimates and findings may be refined. At the same time, a comprehensive sensitivity analysis provides strong evidence that the model and its outcomes are robust to parameter changes.

Finally, we did not model the differences in vaccine efficacy across variants of concern (including ancestral, Delta, and Omicron variants). This should not impact the outcomes over the considered simulation horizon. Nevertheless, our model will be extended in near future, addressing these limitations (re-infections and multiple co-circulating variants with different immunity profiles).

Conclusion

In pursuing our objectives, we solved several methodological challenges, extending the range of applicability for agent-based pandemic modelling. Firstly, we incorporated the ABS census data for 2021, thus accounting for the most recent demographic information for Australia, in terms of the population structure, age distribution, household composition, and commuting flow patterns. This “upgrade” is important because previous similar studies used the data from the Australian census of 2016, scaling the modelling outcomes by approximately 10% to account for the larger population. Our results showed that the demographic changes over the five-year period contribute to the pandemic outcomes in more subtle nonlinear ways that often cannot be captured by a uniform scaling. To study these nuanced contributions, we employed Lorenz curves characterising an unequal distribution of pandemic effects.

Secondly, we addressed a well-known inconsistency between low-resolution and aggregated high-resolution census data brought about in 2016 by the ABS anonymity policy compliance system [36, 37]. In order to reduce this mismatch, we reconstructed a surrogate high-resolution 2021 commuter topology (Section "Surrogate domestic commuting network" of Appendix). This allowed us to examine nuanced effects of the pandemic scenarios on urban and regional areas, and measure pandemic bimodality.

Overall, the extended ABM, coupled with the reconstruction techniques, offered a versatile approach to model comparative scenarios with multiple variants of concern, simulated across different demographic settings (census years) and for distinct intervention policies. A combination of census-based ABM and pandemic Lorenz curves provided a unique high-resolution method to not only simulate different pandemic scenarios across varying demographics and variants, but also evaluate the unequally distributed effects of feasible intervention policies. This, in particular, allowed us to emphasise the divergent role of school closures as a complementary NPI—with respect to the disease transmissibility, exemplifying a bifurcation in the effectiveness of school closures.

In summary, the presented results illustrate how comparative analysis measuring distribution of the pandemic severity across different dimensions can help in improving public health preparedness and response to future pandemics. In particular, the study highlights that rigorous pandemic modelling can provide insights into the impact of complex demographic factors on the spread of infectious diseases over medium- to long-term.

Appendix A: Population data

The population data used in our model is drawn from 2016 and 2021 Australian Census data published by the Australian Bureau of Statistics (ABS). Australian Census comprises a large number of hierarchical data fields categorised by geographical, demographic, and socio-economic parameters [40]. In our model, we use the demographic fields in Australian census (extracted utilising ABS TableBuilder [41]) to generate an artificial population as a “digital twin” of the Australian population, aiming to capture its main characteristics with high fidelity. Here, we describe the demographic census data structures and examine salient features of the Australian population mobility and the international air traffic.



Fig. 10 Snapshot of the Australian population: locations and geographical areas. Capital city of each state is annotated with their usual residential population (P) and geographical area (A) in 2016 (in red) and in 2021 (in blue), respectively. The state names are abbreviated as follows: Western Australia (WA), Northern Territory (NT), South Australia (SA), Queensland (QLD), New South Wales (NSW), Victoria (VIC) and Tasmania (TAS). Note that changes in the size of geographical areas are due to boundary changes

Census data structure

The 2021 census reports that the Australian population reached 25.4 million, following an 8.6% increase (2,020,896 million people) since 2016 census counts. There are eight states and territories in Australia, each with a capital city. Figure 10 shows the geographical representation of the states and their corresponding capital cities which comprise about two thirds of Australian population.

Each state is further divided into different levels of statistical areas (without gaps or overlaps) following a nested hierarchical framework of geographic levels [40]. Each succeeding level is a breakdown of the previous level into smaller areas defined as follows: States and Territories (ST/T), Statistical Area Level 4 (SA4), Statistical Area Level 3 (SA3), Statistical Area Level 2 (SA2), Statistical Area Level 1 (SA1), and Mesh Blocks (MB). Table 1 summarises the number of statistical areas at each level and the corresponding usual residential population range. Note that 2021 census reported more statistical areas across all levels due to changes in the partitioning approach. Due to merging and splitting of areas, the new partitioning approach resulted in a net increase of 7.5% for the number of SA1 areas and 7% increase for the number of SA2 areas.

In addition to the residential population, a wide range of other demographic features are also reported for each level. The following features are incorporated into the artificial population generation: age, gender, household composition, and commuting patterns. See Appendix section B for more details regarding how these features are incorporated in our model.

In order to reduce the risk of individual- or household-level identification, all data reported in the census are anonymised via several perturbation methods including, but not limited to, the removal of aggregated microdata falling under a set threshold. It is important to note that the ABS data perturbation policies may differ across datasets and may change between census years.

Table 1 The hierarchical levels in Australian census structure

	2016 census	2021 census	population range (approx.)
MB	358,122	368,286	N.A
SA1	57,523	61,845	200–800
SA2	2310	2473	3000–25,000
SA3	358	359	30,000–130,000
SA4	107	108	100,000–500,000

The population for each level refers to the usual residential population, excluding visitors

Mobility: travel to work

The short-distance commuting patterns are reported via the travel-to-work (TTW) dataset, detailing the accumulated number of employed individuals commuting between their place of usual residence (UR) and a place of work (POW). We use this dataset to construct a commuter mobility network (detailed in section "Surrogate domestic commuting network"). Commuter counts are accumulated on several statistical levels:

- UR: reported at SA1, SA2, SA3, SA4 level
- POW: reported at DZN, SA2, SA3, SA4 level
- number of employed commuters (either full-time or part-time)

Another level (DZN, destination zone) is used to describe the POW population at a resolution comparable to SA1 but following a different set of partition rules. Both SA1 and DZN accumulate to attain exact partitions at SA2 level.

The perturbation protocol that the ABS adopted for privacy protection removes some edges between UR and POW (that is, UR ↔ POW edges). This removal is done if the commuter counts are below a certain threshold. In our previous work [37], we showed that the edge removal causes large discrepancies in the edge counts between different statistical area levels. When generating the artificial population, we address this problem by minimising the discrepancies between different SA levels, improving consistency of the population representation across the levels (see section "Travel to work").

Urban structure and international air traffic

Australia is a highly urbanised society with heterogeneously distributed population as annotated, see Fig. 10. The eight capital cities, despite their rather small geographical area size, account for more than two thirds of the Australian residential population, while the remaining land (mostly regional areas) comprises only one third of the population.

The capital cities and their surrounding areas are defined by the ABS as the Greater Capital Cities (GCC) including the urban areas (city centres) and the areas in some proximity, to account for the residents who work, shop, and socialise within the city. We examine the urban structure, particularly GCCs, for two reasons.

Firstly, the population densities of GCCs are considerably higher compared to other areas of Australia, allowing for a greater social interaction from both residents and working commuters. This increased likelihood of mixing aggravates the spread of infectious diseases [42–44]. In

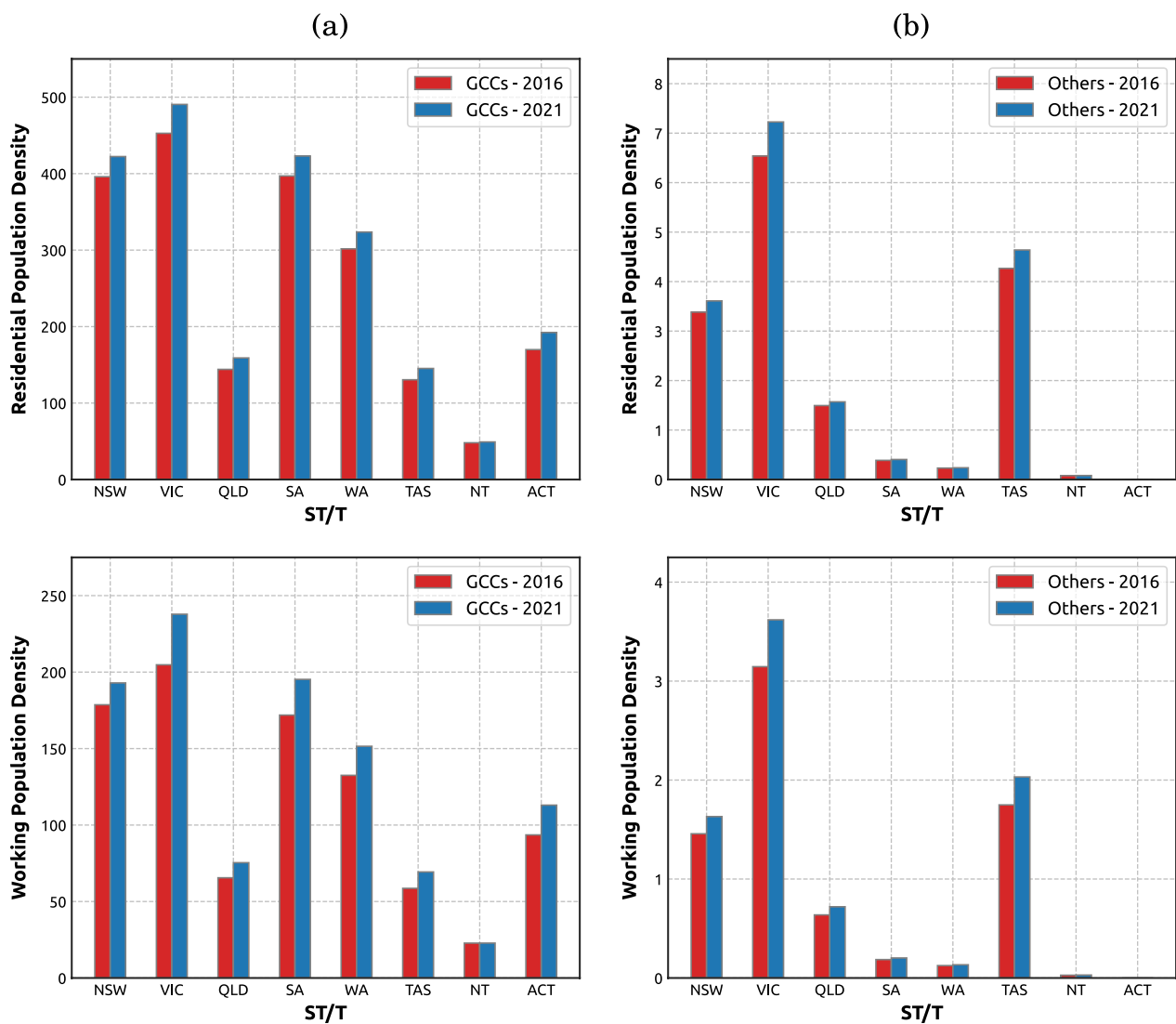


Fig. 11 Comparison of residential population density and working population density in different states and territories (ST/T) of Australia partitioned by GCCs (a), and other areas (b). Note that the population density scale in (b) is more than 50 times lower than that in (a). The Australian Capital Territory (ACT) predominantly comprises urban areas and thus does not show a comparable density for non-GCC areas

addition, both residential and working population densities continued to increase between 2016 and 2021 for all GCCs despite the brief disruption of migration and movement during the COVID-19 pandemic, as shown in Fig. 11. Furthermore, the urbanisation level in GCCs greatly varies across the country. Figures 11a and 12 illustrate a stark contrast between the population-dense GCCs (Sydney, NSW GCC; and Melbourne, VIC GCC) and cities with noticeably lower population density, e.g., Hobart (TAS GCC). This large variability also suggests a high heterogeneity within the GCC population.

Secondly, the local and international air traffic is concentrated in GCCs, increasing the potential of disease introduction by long-range travel (i.e., international sources) [45], as well as its subsequent spread. Due to Australia’s geographic isolation, the international airports located in GCCs take the vast majority of international passenger inflow. We note that airports of several non-GCC cities, such as Newcastle (NTL) and Sunshine Coast (MCY), may also take a relatively small number of international incoming passengers, as shown in Fig. 13. The air traffic data is reported by the Australian Bureau

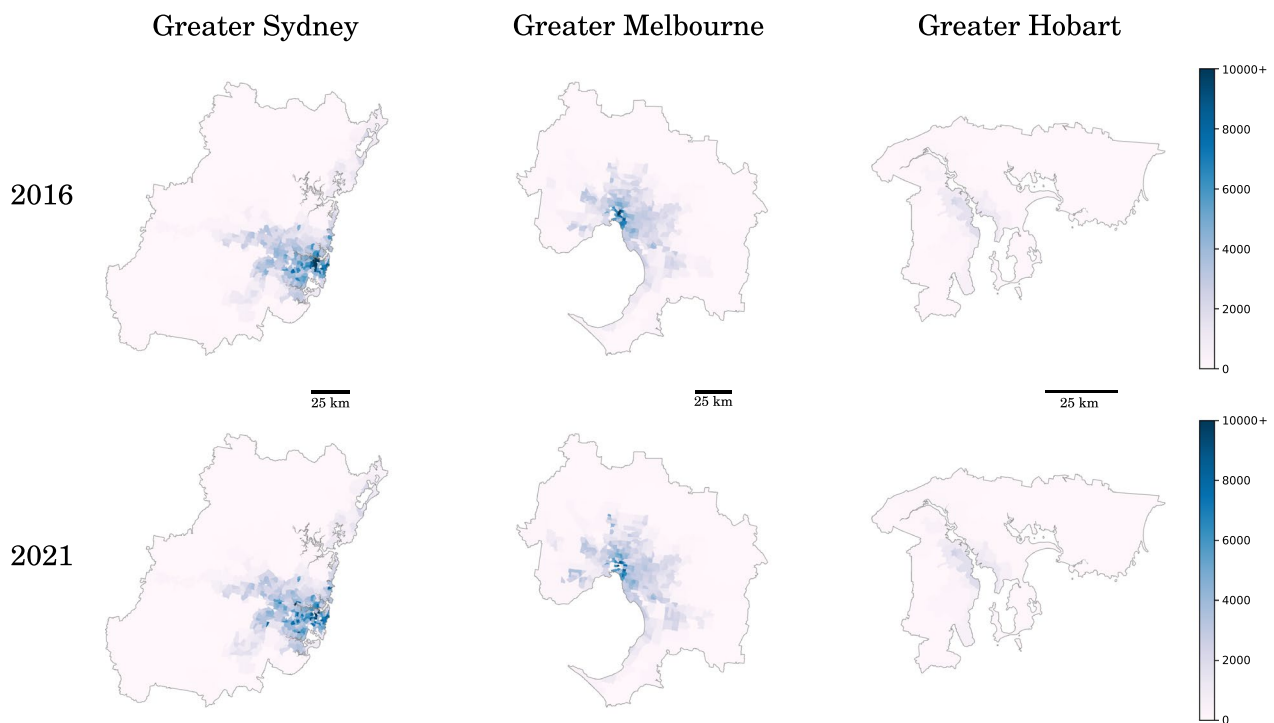


Fig. 12 High variation of the population density across three Greater Capital Cities of Australia: Greater Sydney (left), Greater Melbourne (middle), and Greater Hobart (right). Population-dense areas are highlighted in darker colours, with intensity defined by the colour bar on the right. Note that boundaries in these figures are shown at SA2 resolution

of Infrastructure, Transport, and Regional Economics (BITRE) [46], and we use this data to scale the number of initial infections in proportion to the international air traffic inflow, following earlier studies [18, 47]. The actual traffic data in 2021 were severely impacted due to travel restrictions imposed during the COVID pandemic. To account for this impact, in this study we use the BITRE air traffic data between 2003 and 2019 (inclusive) in estimating the projected air traffic volume in 2021. See section "International air traffic" below for more details.

Appendix B: Generation of artificial population

This section describes the construction of a surrogate Australian population using high-resolution datasets, including the ABS demographic fields from the Australian census, international air traffic reports from the BITRE, and educational registration records from ACARA. These datasets provide high fidelity details encompassing age, gender, household composition, student enrollment, workforce mobility, and international travel. Using these datasets, we generate a representative artificial population capturing unique demographic

characteristics and mobility characterising the Australian population.

This section is structured as follows. Section "Agent-based generation" explains the generation of individual agents and their matching demographic characteristics. Section "Surrogate domestic commuting network" details our approach to reconstructing the commuting network which included travel to school and travel to work depending on agents' age groups. Section "International air traffic" elaborates on the estimation of international air traffic in 2021 using traffic records prior to the COVID-19 pandemic. The user guide of our open-source software provides more technical details [39].

Agent-based generation

In this section, we describe agent generation procedure. Initially, we generate households and then assign each virtual agent with a household determined by the household composition in their residential SA1 areas. This is followed by assignment of demographic characteristics, including age and gender. The distribution of these agents in each SA1 matches the number reported in census

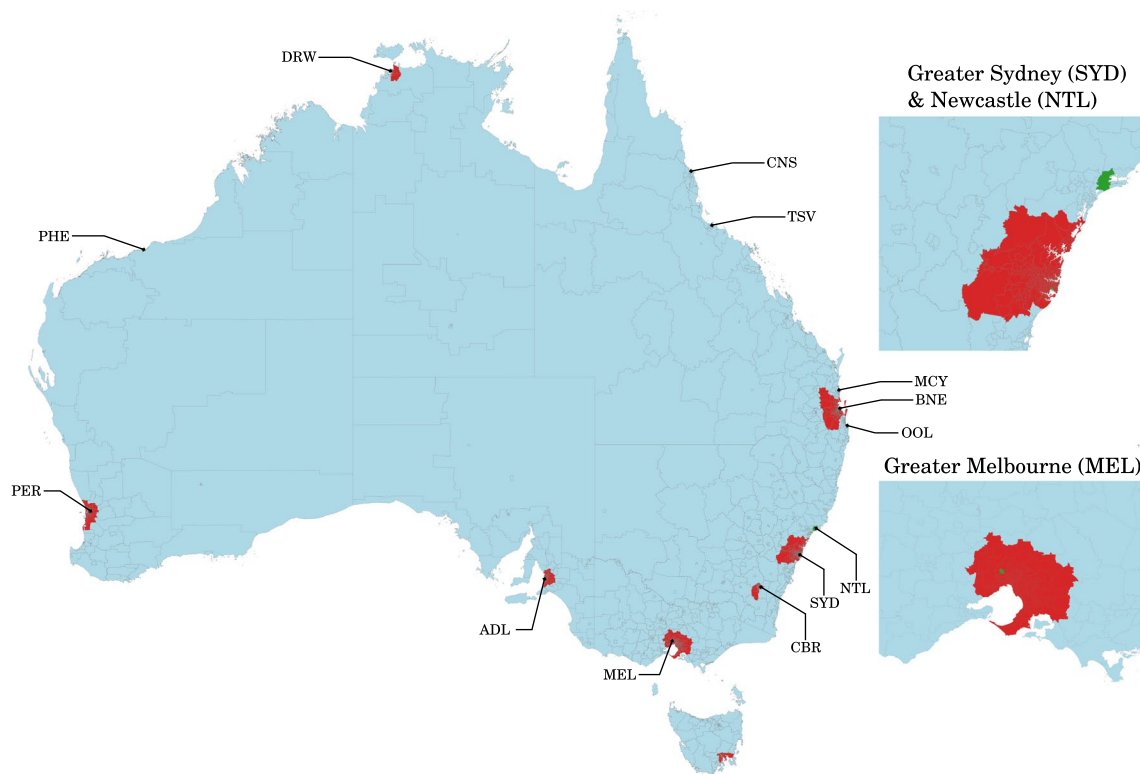


Fig. 13 Geographic representation of Greater Capital Cities (in red), international airports (having international air traffic in 2019, in green), and other regions (in blue) across Australia. The boundaries are delineated based on the 2021 SA2-resolution map of Australia. For visual clarity, some Australian islands are not shown. Hobart is not annotated as there were no incoming international passengers to Hobart during the considered period (i.e., in 2019). The 3-letter code names for the annotated international airports are described in Table 4

reports. We then generate agents across all SA1 areas to construct an Australia-wide population. In our population generation process, we excluded all non-geographic regions (e.g., “Migratory/offshore/shipping” and “No usual address”) as these regions have no residential population with specific addresses.

Household composition

Each agent is initially assigned to a household. The distribution of household size in each SA1 area matches the census data on household composition (i.e., the count of families in terms of the number of dependent children and/or non-family members, excluding overseas visitors), corresponding to the following five categories:

- single adults,
- couples without children (i.e., two adults),
- couples with one to four children,
- single-parent families with one adult and one to five children,
- non-family groups with two to six adults

A cumulative distribution function (CDF) of household composition is built for each SA1 area.

SA1 population, age, gender, and residency.

Within the household, attributes of each agent (age, gender, and residency) are assigned to match the census records capturing the demographic diversity at the SA1 resolution. Specifically, for each SA1, the ABS summarises the number of individuals residing in this region, and the number of male/female residents categorised by five age groups as follows:

- early childhood (0–4 years old),
- school age (5–18 years old),
- young adult (19–29 years old),
- middle age (30–64 years old),
- elderly (65 years old and above).

CDFs of age groups are produced for male adults (aged 19 and older), female adults (aged 19 and older), male children (aged 18 and younger), and female children

(aged 18 and younger). Other CDFs of children genders and adult genders are also created to support the sampling processes requiring specification on agents' gender (see following section).

Generation of simulated population

While each agent is assigned with demographic and household attributes as detailed in earlier sections, the total residential population and household count in each SA1 area are constrained by the aggregate number per SA1 reported by census. We assign the attributes as following:

- Step 1: For a given SA1 area, if the number of generated agents is less than the number of residential population reported by census, sample a household based on the CDF of households constructed from the ABS data.
- Step 2: Within each of the generated household, assign agents with genders and age groups (male/female adults and/or children) based on the corresponding CDFs.

For example, if the household instance sampled from the census data is a couple comprising two adults of opposing gender, with one child, the generated household will consist of three household members:

- an agent representing the father in this household, generated with its age sampled from the male adult CDF of age groups,
- an agent representing the mother in this household, generated with its age sampled from the female adult CDF of age groups,
- an agent representing the child in this household, generated with its gender sampled from the CDF of

genders for children, and its age sampled from the CDF of age groups for children corresponding to the sampled gender.

This process continues until all SA1 areas are considered and the aggregate artificial populations matches the population statistics reported by census. During this process, every three consecutively generated households form a household cluster, representing a residential context with close contacts around the agents' residency.

With the population generated at SA1 level, we follow the partitioning method that the ABS uses to merge SA1 populations into SA2 populations. We note that no agents are generated during this process as all SA1 areas are included within SA2 areas. We assume that SA2 areas are spatially distributed and the geographic distance between any pair of SA2 areas is calculated as the distance between the centroids of two SA2 areas defined by the ABS-provided shapefiles representing geographical boundaries. This spatial distribution plays a crucial role in generating school placements for children and working group placements for adults.

As a result, the generated aggregate artificial population is a "digital twin" of the Australian population, matching key census statistics, including the number of SA2 and SA1 areas, households, household clusters, and total agents. These statistics are summarised in Table 2. Note that in addition to residential contexts, a number of other mixing contexts is also reported in Table 2. DZNs and working groups (i.e., work-related environments) are allocated for adults, while schools, grades, and classes (i.e., education-related environments) are allocated for children. The allocation of these contexts is used in generating the commuting network, detailed in section "Surrogate domestic commuting network".

Table 2 Summary of demographic attributes and mixing contexts of the surrogate population generated from 2016 and 2021 Census

Number of	Census 2016 based	Census 2021 based	Increase (%)
SA2s (SLAs)	2310	2454	6.23
SA1s (CDs)	57,523	61,811	7.45
DZNs	9136	9307	1.87
Working Groups	470,608	530,988	12.83
Schools	9,463	9605	1.50
Grades	56,778	57,630	1.50
Classes	174,564	189,078	8.31
Households	9,656,841	10,802,052	11.86
Household Clusters	2,435,274	2,723,287	11.83
Population	23,406,541	25,428,029	8.64

The 4th column shows the relative increase for each attribute over the five-year period. The population expansion results in an increase across all residential and mixing contexts

Surrogate domestic commuting network

In addition to the residential mixing contexts (see section "Agent-based generation"), our model also considers interactions across spatially distributed agents which occur in workplaces and schools. The mobility patterns describing commuting to places of work and education are represented by a realistic commuting network constructed using the census commuting data and student registration records from ACARA [48]. In this network, we allocate population flows specifically capturing the student and worker flows between mutually exclusive sets, e.g., between the usual residential areas (SA1 areas) and the places of work (DZNs). In this section, we describe the construction of the student and worker flows in the commuting network in two parts: travel to school, applicable for agents from 5 to 18 years of age; and travel to work, applicable for agents over 18 years of age.

Travel to school

In our surrogate population, nearly 18% of the population are school-aged children (from 5 to 18 years old), with the vast majority of them (approximately 95%) going to school on weekdays during day time and interacting with other school-attending children and teachers. These at-school interactions are enabled by integrating "student flows" into the commuting network. Each flow is set between the student residence (which has been previously assigned during the generation of simulated agents; see section "Agent-based generation") and the destination zones (i.e., schools).

Using the school enrollment numbers and locations reported by ACARA, the school locations are pseudo-deterministically chosen within destination zones, while accounting for school capacity and geographic proximity to areas with a sufficient number of children and teachers.

The allocation of schools is based on the following assumptions. The probability of students attending a school is set in proportion to the geographic distance between the school and their residence, within the school's catchment zone. This process emulates the tendency of students to initially prioritise schools in the closest proximity. If schools nearby are unavailable, students are assumed to expand their search to nearby options within the maximum travel distance (150 km). This ensures that the vast majority of students are assigned to schools by the end of the generation process.

Within each school, students are further organised into smaller groups representing grades and classes, where classes represent the primary units of regular interaction for students with the highest frequency of contacts, followed by grades. The contact rates in the school-environment are summarised in Table 5.

Upon completion of the student-to-school assignments, we estimate the number of teachers working at a school based on the enrolment numbers of the school, the student-to-staff ratio (2:17) (approximated from historical ABS data), and the estimated number of three teachers per class. We then randomly sample from the ensemble of adults working in the school's DZN to match the estimated number of teachers. The remaining workforce within the DZN is then assigned to non-school-related working groups detailed in the following section.

Travel to work

We then construct a travel-to-work network for the remaining working population, capturing daily frequent contacts at the workplace where each edge represents a worker flow between agents' usual-residential areas (UR, previously assigned as detailed in section "Agent-based generation") and the corresponding places of work (POW). Within the working population, each agent is also assigned to a working group with a maximum of 20 people at work to simulate a realistic working environment where a small group of individuals frequently interact on a daily basis. This assignment process matches the key census attributes reported in travel-to-work (TTW) data [41] (i.e., the census table for employment, income, and education representing the number of commuters between UR and POW).

The work-related census reports partition the population into multiple statistical area (SA) levels and Destination Zones (DZN), in ascending resolution. We note that DZN (i.e., the highest resolution for POW) does not align precisely with SA1 partitions (i.e., the highest resolution for UR). We use DZN and SA1 as two separate partitions simulated in different time cycles, daytime and nighttime, respectively (see Appendix section C below for more details). Each agent residing in a SA1 is randomly assigned to work in a DZN and the total number of registered commuters between the SA1 (considered as UR) and the DZN (considered as POW) must match the corresponding commuting volume reported by census. This process uses the high-resolution worker flow data between SA1 and DZN to construct a high-fidelity TTW network between all URs and POWs. However, the implementation of privacy-protecting policies adopted by the ABS heavily perturbed the data, resulting a sizable discrepancy between high-resolution data (e.g., SA1–DZN) to lower-resolution data (e.g., SA2–SA2). We consider and address this discrepancy in detail, as described below.

Inconsistencies in ABS data for travel to work The census data contains inconsistencies between UR and POW data due to the adoption of new privacy protection measures, implemented within the ABS anonymity policy

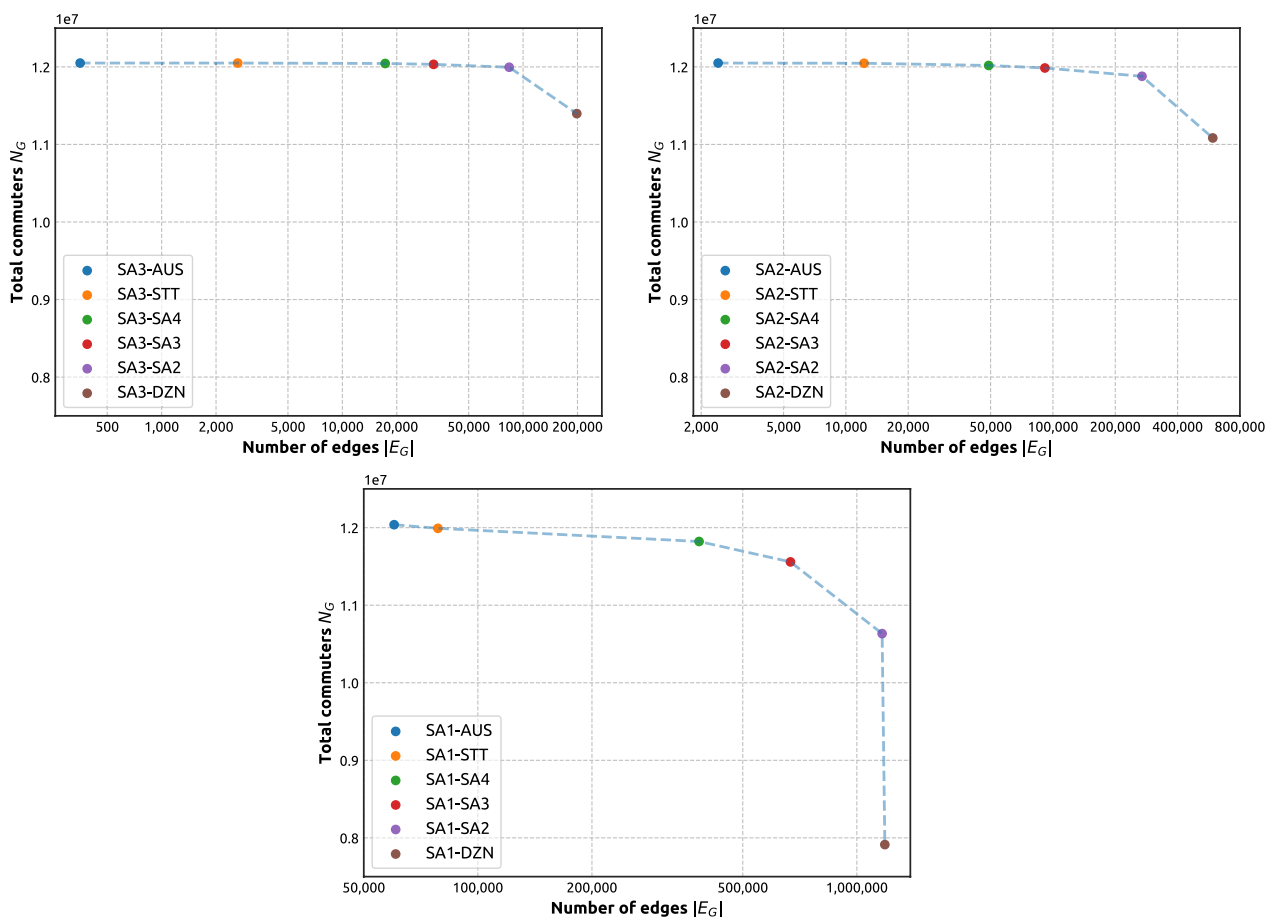


Fig. 14 Visualisation of the discrepancies at different resolutions observed in 2021 census commuting data. Each subplot traces combinations between UR and POW for all possible resolutions. Across the UR ↔ POW flows, considered at comparable or higher resolutions (i.e., SA3–DZN, SA2–DZN, and SA1–DZN), the edge numbers are noticeably lower towards the right hand side on x-axis, with a reduced number of commuters (y-axis)

compliance system over different census years [36, 37]. Such inconsistencies produce a mismatch in the commuting flows (between UR and POW) reported at different SA partitions. These inconsistencies therefore require a reconstruction of some commuting flows between UR and POW, added to match the aggregate totals.

Specifically, the total number of commuters aggregated at a higher resolution, e.g., between SA1 (UR) and DZN (POW), is usually less than its counterpart obtained from a lower resolution, e.g., between SA2 (UR) and SA2 (POW), as shown in Fig. 14). This results from various ABS privacy protection measures aimed to avoid re-identification from high-resolution commuting data, particularly when commuting volume is low. As a result, the working mobility in 2016 census shows a 34% difference between the aggregated and true commuters [37]. Here, we report a similar observation for 2021 census,

with a difference of a 25.6% between the SA1–DZN and SA1–SA2 commuting datasets, and a 34.27% difference between the SA1–DZN and the total number of reported commuters in Australia (see Table 3).

Prior work resolving inconsistencies in 2016 census In our prior work [37], we proposed an algorithm that uses a statistical re-sampling process to generate and add new edges to the SA1–DZN commuting network. This considerably improved the consistency of commuting flow across different resolutions. This algorithm used the commuting data from the older 2011 census as a baseline (existing prior to the implementation of privacy protection protocols), to supplement the commuting data in different resolutions in 2016 census. The main steps of the algorithm are summarised as follows:

Table 3 Summary of working flows between the usual residential (UR) and the place-of-work (POW), reported at different resolutions of statistical areas defined by census

UR-POW Resolutions	Number of edges	Number of commuters	Reduction in registered commuters
SA1-AUS	60,126	12,038,052	Reference
SA1-STT	78,461	11,991,363	0.39%
SA1-SA4	383,454	11,821,831	1.80%
SA1-SA3	668,132	11,558,487	3.98%
SA1-SA2	1,165,986	10,633,702	11.67%
SA1-DZN	1,185,396	7,913,093	34.27%
SA2-AUS	2415	12,049,439	Reference
SA2-STT	12,242	12,046,956	0.02%
SA2-SA4	48,929	12,018,402	0.26%
SA2-SA3	91,426	11,985,868	0.53%
SA2-SA2	269,261	11,878,866	1.42%
SA2-DZN	591,375	11,083,493	8.02%
SA3-AUS	354	12,049,397	Reference
SA3-STT	2637	12,049,398	0.0%
SA3-SA4	17,257	12,043,898	0.05%
SA3-SA3	31,971	12,032,775	0.14%
SA3-SA2	83,778	11,995,534	0.45%
SA3-DZN	198,299	11,397,507	5.41%

At each resolution, the number of edges specifies the number of destinations (connections between these regions), while the number of commuters represents the aggregated number of commuters in all edges at this resolution. The last column shows the discrepancy (a reduction measured in percentage) between the specified resolution and the reference resolution. Note that census data are pre-processed, with all non-geographic regions (e.g., "Migratory/offshore/shipping" and "No usual address") excluded

- Step 1: Calculate the historical (2011) distribution $P(w|N_{SA1})$ of the number of commuters between UR (SA1) and POW (DZN) (i.e., edge weights of the edges (SA1, DZN) in the TTW network, denoted by w), given a working population size in SA1 level (denoted by N_{SA1}).
- Step 2: Sample additional edges for each UR unit (SA1) based on $P(w|N_{SA1})$. The list of partial candidate edges, $L = \{(SA1_1, w_1), \dots, (SA1_n, w_n)\}$, is generated. The total sampled edge weights for an SA1 must be bounded by the inconsistencies between SA2 and SA1 levels (i.e., the difference between the number of commuters residing in an SA2 and the aggregated number of commuters residing in the corresponding decomposed SA1s).

- Step 3: Assign workplaces (DZN) for the newly sampled edges, i.e., to complete the list L : $L = \{(SA1_1, DZN_1, w_1), \dots, (SA1_n, DZN_n, w_n)\}$. These assignments are constrained by the number of remaining unassigned commuters in each DZN and in the SA2-SA2 commuting network when compared with the SA1-DZN network.
- Step 4: Eliminate reclaimed edges (i.e., the SA1-DZN edges that the ABS already declared) and concatenate the remaining generated edges with the existing edges in 2016 census.

An optimised approach to address the inconsistencies in 2021 census We optimised the algorithm originally presented in [37] and applied the optimisation to construct the travel-to-work network using 2021 census data. The optimised algorithm reduces the discrepancy in SA1-DZN travel-to-work data by using the lower resolution commuting data as a reference point, since the privacy protocols mostly affect the higher resolution commuting data. For example, there is only a marginal reduction (1.42%) in the number of commuters observed in the SA2-SA2 network, compared to a significantly more sizeable 34.27% reduction observed in the higher-resolution SA1-DZN network, as summarised in Table 3.

The optimised re-sampling algorithm reconstructing the surrogate TTW network is detailed in Algorithm 1, with the changes to the original algorithm [37] highlighted in blue. Similar to the original approach, in the optimised algorithm, a number of unassigned working-age agents in an $SA1_i$ region is selected at random and assigned to a random working group in a DZN_i region. The assignment of the residential area and the destination zone is constrained by the edge weight w_i of the corresponding edge $(SA1_i, DZN_i, w_i)$ in the surrogate network. We introduced several additional constraints to check the eligibility of newly generated edges (step (3) of Algorithm 1), before adding them to the edge-list of the surrogate network. This modification ensures that the sampled artificial edges are not duplicates of existing edges and improves the quality of the final surrogate edge-list. This however had a cost of significantly lengthening the processing time relative to the original algorithm [37]. To speed up the computation, we adjusted the sampling process by sampling a uniformly random number of edges ($rand(N)$) in the existing edge-list for each DZN at a time. This method significantly reduced the computational cost.

Algorithm 1 Optimised algorithm generating the 2021 travel-to-work (TTW) surrogate network

Input:

- [A] 2016 – List of surrogate UR-POW (SA1-DZN) TTW edges (reconstructed by [37]),
- [B] 2016 – List of working population size in SA1 resolution,
- [C] 2021 – List of working population size in SA1 resolution,
- [D] 2021 – List of UR-POW (SA1-DZN) TTW edges (from ABS census data),
- [E] 2021 – List of POW-UR (DZN-SA2) TTW edges (from ABS census data),
- [F] 2021 – List of numbers of remaining unassigned commuters (not in D) for each UR-POW (SA2-SA2) TTW edge,
- [G] 2021 – List of numbers of remaining unassigned working adults (not in D) residing in particular SA1s,
- [I] 2021 – List of numbers of remaining unassigned working adults (not in D) working in particular DZNs,
- [K] 2021 – DZN-SA2 correspondence list,
- [L] 2021 – SA1-SA2 correspondence list,
- [N] Maximum number of edges allowed to be selected in step 3.g.

Output:

- [O] 2021 - List of additional UR-POW (SA1-DZN) edges supplementing D to represent reconstructed TTW network

Procedure:

(1) Calculate the historical (2016) distribution $P(w|N_{SA1})$ from [A,B]:

w : An edge weight in the SA1-DZN travel-to-work network, representing the number of commuters living in a SA1 and working in a DZN.

N_{SA1} : Total number of commuters living in a SA1, or working population size of a SA1. N_{SA1} is grouped into bins, used in the construction of $P(w|N_{SA1})$.

An illustrative example of the distribution $P(w|N_{SA1})$ is shown below. Each column represents a distribution of edge weight given a bin value of N_{SA1} .

	N_{SA1}		
w	[0,9]	[10,19]	...
1	0.01	0.08	...
2	0.02	0.05	...
3	0.10	0.01	...
...

(2) Construct lists of additional partial edge candidates ($SA1, w$) for each SA1 based on $P(w|N_{SA1})$ and [C,G]:

For each $SA1_i$ listed in G:

a. From C: determine its bin size N_{SA1_i} .

b. From G: determine the number of working adults residing in $SA1_i$, but not assigned a

DZN: $R(SA1_i)$.

c. Determine $P(w|N_{SA1_i})$ from step 1.

d. Construct $L(SA1_i) = \left\{ w_i \mid \begin{matrix} w_i \sim P(w|N_{SA1_i}) \\ \sum_{w_i \in L(SA1_i)} w_i \leq R(SA1_i) \end{matrix} \right\}$

(3) Construct additional complete edges ($SA1_i, DZN_i, w_i$) based on [E,I,K,L] and $L(SA1_i)$:

Repeat until convergence. For each DZN_i in I:

a. From K: search for $SA2_{POW}$ that DZN_i locates.

b. From E: construct $L_{SA2-UR} = \{SA2_i | (DZN_i, SA2_i) \in E\}$.

c. From [L, L_{SA2-UR}]: construct

$$L_{SA1-UR} = \left\{ SA1_i \mid \begin{matrix} SA1_i \in SA2_i, \text{ where } SA2_i \in L_{SA2-UR} \\ (SA1_i, DZN_i) \notin D \cup O \end{matrix} \right\}$$

d. From [I,O]: determine the remaining unassigned working adults (not in $D \cup O$) in DZN_i :

$R(DZN_i)$.

e. From [F,O]: determine the remaining unassigned commuters (not in $D \cup O$) in the UR-

POW ($SA2_{UR,i} - SA2_{POW}$) TTW edge: $R(SA2_{UR,i}, SA2_{POW})$.

f. Construct $L_f = \left\{ (SA1_i, DZN_i, w_i) \mid \begin{matrix} SA1_i \in L_{SA1-UR} \\ w_i \in L(SA1_i) : \text{sampling without replacement} \\ w_i \leq R(DZN_i) \\ w_i \leq R(SA2_{UR,i}, SA2_{POW}), \text{ where } SA2_{UR,i} \in L_{SA2-UR} \end{matrix} \right\}$

g. Sample for rand(N) candidate edges without replacement: $(SA1_i, DZN_i, w_i)$ from L_f . Add sampled edge(s) to O.

(4) Concatenate O and D to reconstruct the SA1-DZN TTW network.

Comparison of the network reconstruction results Our optimised algorithm significantly improved the mapping between SA2–SA2 and SA1–DZN networks from 69.77% (between the census SA2–SA2 and SA1–DZN networks) to 92.87% (between our reconstructed SA2–SA2 and census SA2–SA2 networks), thus accounting for missing edges. It provided further improvement in comparison to the prior algorithm [37], which reconstructed 90.66% of the mapping between census SA2–SA2 and SA2–SA2 networks, as shown in Fig. 16a). The re-sampling process reduces the number of unassigned commuters in the SA2–SA2 TTW network from 3,469,391 to 835,534, as shown in Fig. 16b).

The optimised algorithm reconstructs the SA2–SA2 surrogate network with a smaller mean squared error (MSE) and a higher correlation with the true commuter flow (i.e., census SA2–SA2 network), as shown in Fig. 16c and d).

The resultant commuting patterns generated from 2021 census are shown in Fig. 15 with the Greater Capital Cities (GCC) annotated within each state.

We note that although our solution effectively rectified most of the discrepancies between SA1–DZN and SA2–SA2 networks using the re-sampling procedure, it cannot provide a complete mapping between the SA2–SA2 network and the lower resolution networks because the algorithm does not introduce new edges (i.e., new commuting flows) beyond the edges already present in the census SA2–SA2 network (see Table 3 and Fig. 14b). However, the proposed re-sampling algorithm is able to produce a surrogate travel-to-work network that adequately represents the census data by supplementing the vast majority of missing edges, as shown in Figs. 15 and 16.

International air traffic

In our model, the initial infection cases are seeded proportionally to the international air traffic inflow. The scenarios for 2016 use the BITRE air traffic data for this year without any modifications. In contrast, to simulate the pandemic scenarios for 2021, we estimate the

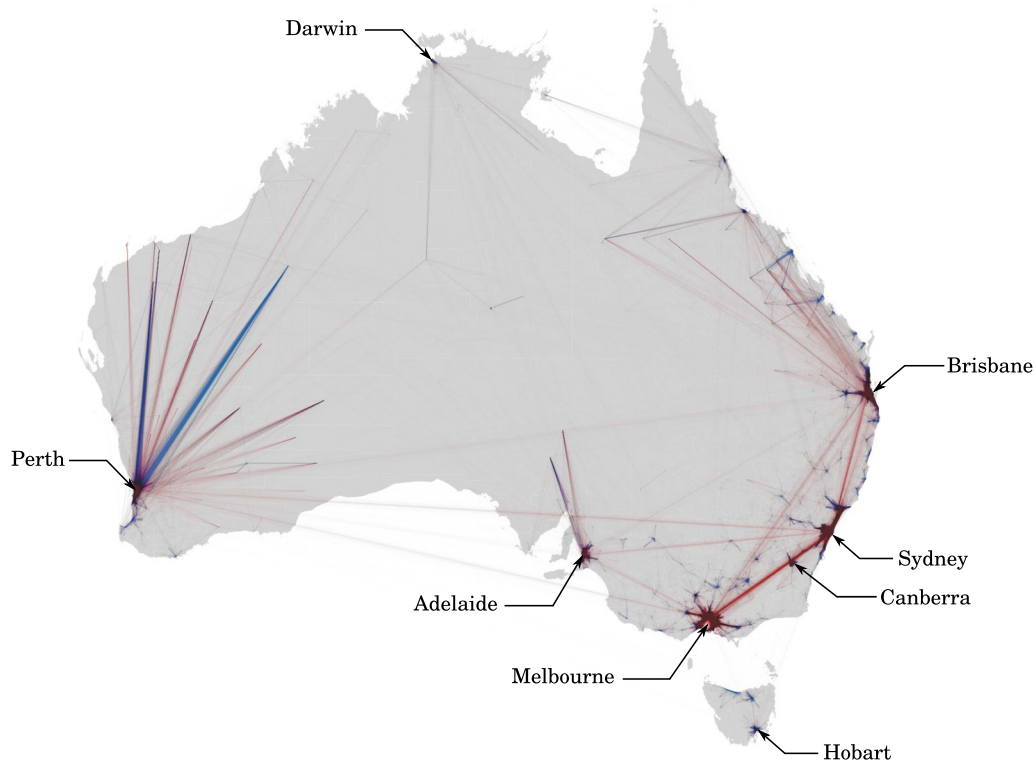


Fig. 15 Visualisation of commuting patterns between usual residences (URs) and places of work (POWs), using 2021 census data. Each commuting flow is represented by an edge connecting the centroids of the corresponding UR and POW. Red lines represent the commuting edges between URs (in SA1 level) and POWs (in DZN level) directly taken from 2021 census, while blue lines depict the reconstructed edges created by our algorithm to match the aggregate totals

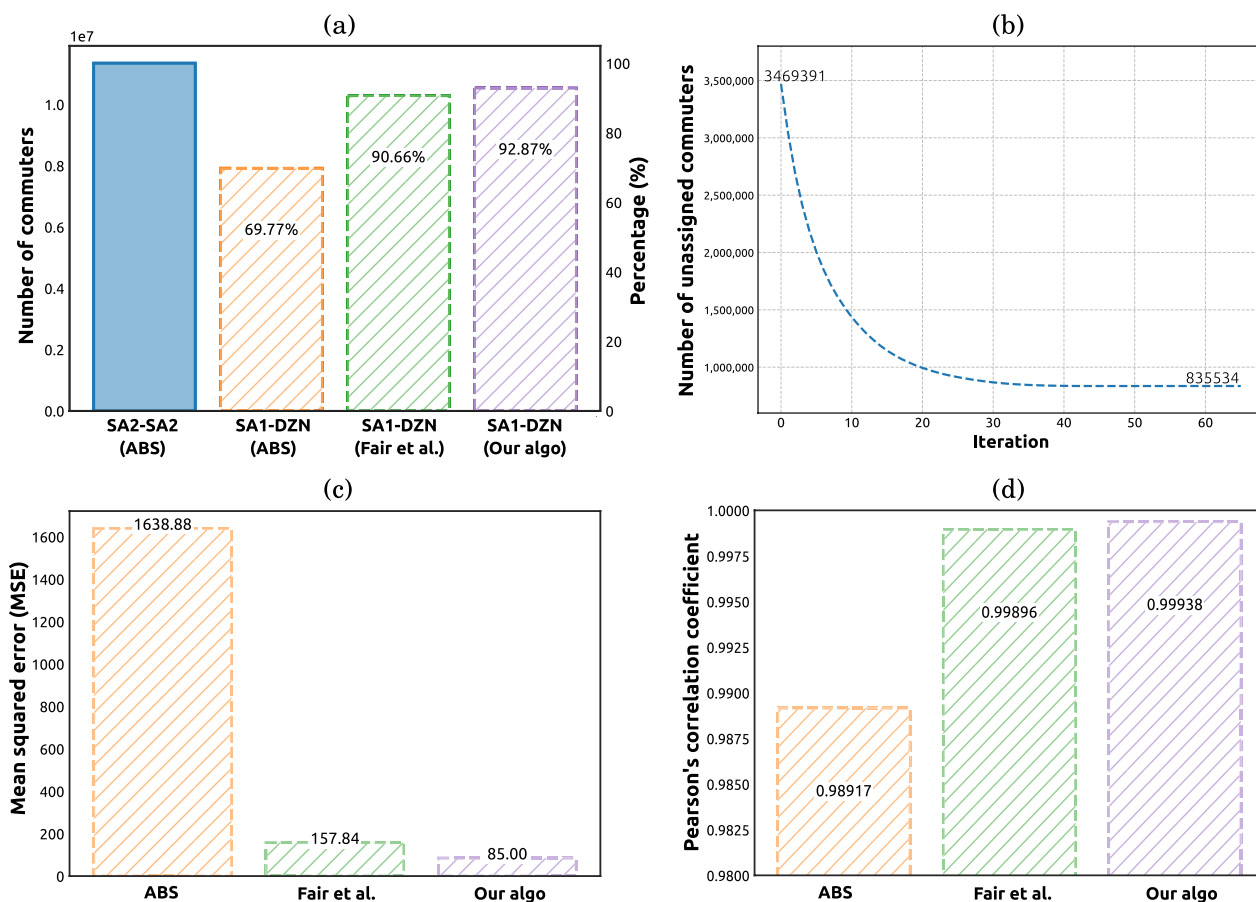


Fig. 16 Results of network reconstruction using the 2021 TTW data. **a** Number of commuters between the reference TTW SA2–SA2 network (from ABS Data), the TTW SA1–DZN network (from ABS Data), the TTW SA1–DZN network (refined by the original algorithm of Fair et al. [37], adapted for census 2021), and the TTW SA1–DZN network (refined by the optimised algorithm). **b** Number of unassigned commuters, counted in SA2–SA2 network but not in SA1–DZN network, over refining iterations. **c** Mean squared error between the SA2–SA2 TTW network from ABS and the reconstructed SA2–SA2 TTW network from the refined SA1–DZN networks. **d** Pearson's correlation coefficient between the SA2–SA2 TTW network from ABS and the reconstructed SA2–SA2 TTW network from the refined SA1–DZN networks

international air travel for this year based on the BITRE air traffic data between 2003 and 2019. We intentionally do not use the air traffic data between 2020 to 2021 due to the severe disruption of international travel due to the travel restrictions during the COVID-19 pandemic.

We approximated the daily average number of incoming international passengers (IIP) at the international airports across Australia 2021 by utilising the international air traffic data up to 2019 [46] which reports the total number of IIP for a given year (reported by June) at various international airports across the country. We then extrapolated the number of IIP in this period by using a weighted multimodal linear regression model. We placed a higher weight ($\gamma = 0.8$) on more recent data points and

constructed the linear regression model which produced the most optimised estimates with small weighted losses. We note that while larger international airports (e.g., Sydney, Melbourne, Brisbane, and Adelaide, shown in Fig. 17) have a sustained inflow of IIP since 2003, small airports (e.g., Darwin) may not have a continuous IIP until the more recent years. In such cases, we could not produce an accurate weighted linear model due to the lack of data; instead, we estimated the IIP in 2021 by averaging the IIP in available years. A comprehensive summary of the estimation methods employed for each international airport is provided in Table 4. The linear regression model for eligible airports is shown in Fig. 17.

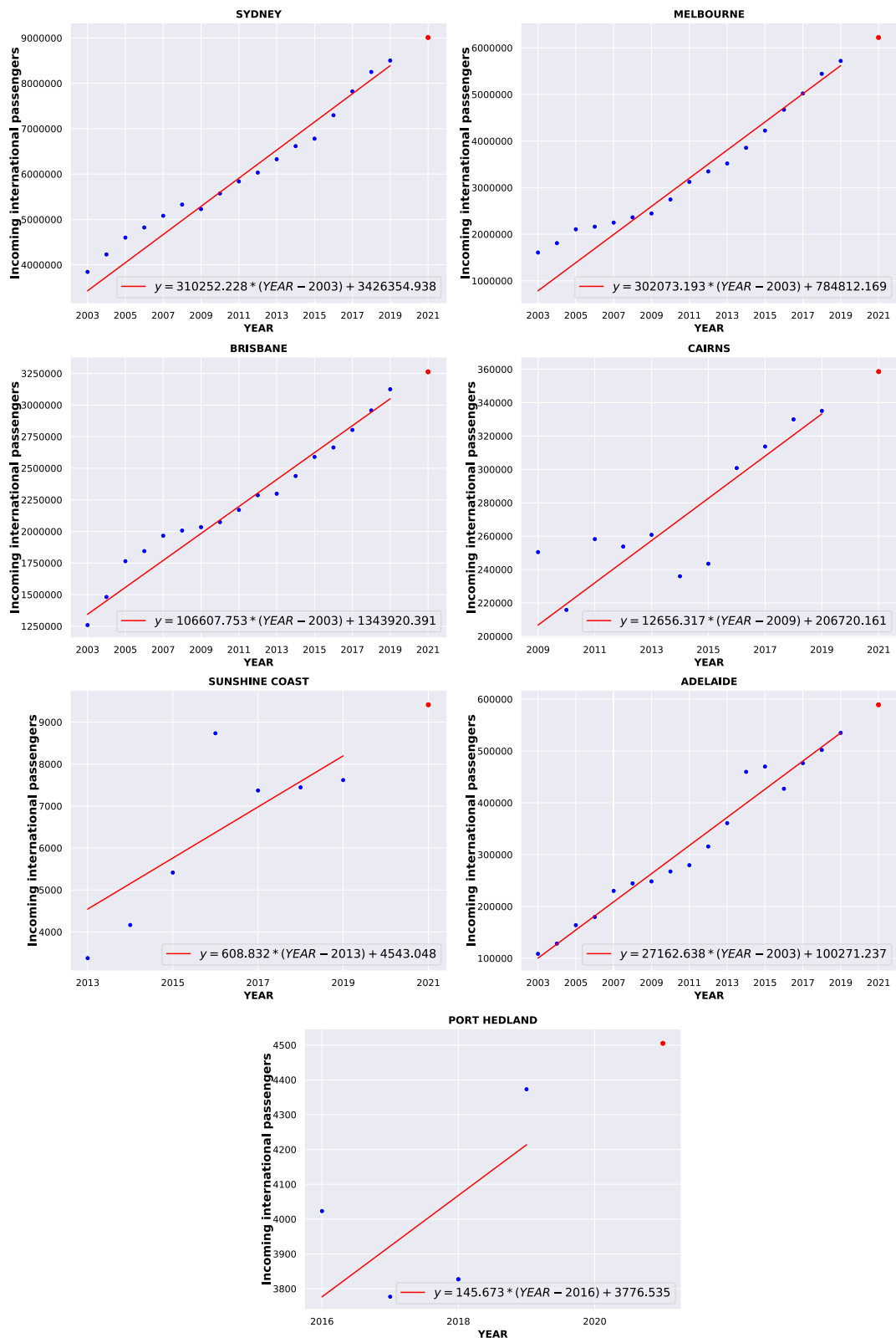


Fig. 17 Weighted linear regression models for the considered airports across Australia. The incoming international passengers (IIP) numbers in previous years up to 2019 are shown in blue; the projected IIP numbers in 2021 are shown in red. Solid red line represents the fitted linear regression model

Table 4 Estimated number of incoming international passengers (IIP) in the international airports across in Australia in 2021, projected from the pre-pandemic BITRE data

Airport	Code	State	Location (SA2)	Estimated IIP	Estimation method
Sydney	SYD	NSW	117011325	9,010,895	Weighted Linear Regression
Newcastle	NTL	NSW	106031125	3493	Same as 2019
Melbourne	MEL	VIC	210051248	6,222,130	Weighted Linear Regression
Brisbane	BNE	QLD	302031036	3,262,860	Weighted Linear Regression
Cairns	CNS	QLD	306011140	358,596	Weighted Linear Regression
Gold Coast	OOL	QLD	309021231	534,733	Averaging data from 2017, 2018, and 2019
Sunshine Coast	MCY	QLD	316031426	9414	Weighted Linear Regression
Townsville	TSV	QLD	318021475	1031	Same as 2019
Adelaide	ADL	SA	404031104	589,199	Weighted Linear Regression
Perth	PER	WA	506021121	2,206,807	Averaging data from 2017, 2018, and 2019
Port Hedland	PHE	WA	510021269	4505	Weighted Linear Regression
Darwin	DRW	NT	701011001	128,585	Averaging data from 2017, 2018, and 2019
Canberra	CBR	ACT	801031114	39,374	Averaging data from 2017, 2018, and 2019

The estimated IIP is used for the initial infection seeding (see section "Results")

Appendix C: Transmission and control model

Transmission model

The transmission of SARS-CoV-2 in our Agent-based Model of Transmission and Control of the COVID-19 pandemic in Australia (AMTraC-19) is simulated using agent interactions in the generated surrogate population. The model operates in discrete time steps, with each step representing a 12-hour cycle (either daytime cycle or nighttime cycle). In daytime cycle, interactions occur within the working/studying groups, i.e., in the work and education contexts; in nighttime cycle, interactions occur in the residential mixing contexts within communities (SA2 wide), neighbourhoods (SA1 wide), household clusters, and households. We also account for differences between weekdays and weekends. A weekday (Monday to Friday) consists of a daytime cycle and a nighttime cycle, while a weekend consists of two nighttime cycles (i.e., no work-related activities).

The initial infections are seeded as imported infections dependent on the air traffic using international passenger data published by the Bureau of Infrastructure and Transport Research Economics (BITRE). At each time step, until border restrictions are imposed, new infections are introduced within a 50-km radius of major international airports, in proportion to the international travel influx.

A simulation scenario begins with an initial distribution of infections determined by a binomial distribution with the probability proportional to the average

number of incoming passengers at the international airports (see section "Urban structure and international air traffic"). These infected agents are then assigned to a randomly selected residential area within some distance to an international airport. This seeding process takes place at the start of each simulation day and continues daily until the international border closures are triggered by a defined cumulative incidence threshold.

To re-iterate, at time cycle n , the probability for an susceptible agent i becoming infected across context $g \in G_i$ is determined as follows (see Eq. 1):

$$p_i^g(n) = 1 - \prod_{j \in A_g \setminus \{i\}} (1 - p_{j \rightarrow i}^g(n)) \tag{3}$$

where $A_g \setminus \{i\}$ represents the set of agents in the context $g \in G_i$ excluding agent i , and $p_{j \rightarrow i}^g(n)$ denotes the instantaneous probability that an infectious agent j , who shares the context g with susceptible agent i , transmits the infection to agent i :

$$p_{j \rightarrow i}^g(n) = \kappa f(n - n_j|j) q_{j \rightarrow i}^g \tag{4}$$

A global transmission scalar, denoted by κ , is used to calibrate the reproductive number R_0 . The time cycle n_j indicates the moment when agent j becomes infected, while the function $f(n - n_j|j)$ represents the natural history of the disease, reflecting the infectivity of agent j over time. If agent j is not infected, $n < n_j$ and $f(n - n_j|j) = 0$. If

Table 5 Daily transmission probabilities $q_{j \rightarrow i}^g$ from infected agent j to susceptible agent i for different mixing contexts and interaction types

Mixing context	Type of interaction	Daily transmission probability ($q_{j \rightarrow i}^g$)
Household (size 2)	Any to child (0–18)	0.09335
	Any to adult (19+)	0.02420
Household (size 3)	Any to child (0–18)	0.05847
	Any to adult (19+)	0.01495
Household (size 4)	Any to child (0–18)	0.04176
	Any to adult (19+)	0.01061
Household (size 5)	Any to child (0–18)	0.03211
	Any to adult (19+)	0.00813
Household (size 6)	Any to child (0–18)	0.02588
	Any to adult (19+)	0.00653
Household Cluster	Child (0–18) to child (0–18)	0.00400
	Child (0–18) to adult (19+)	0.00400
	Adult (19+) to child (0–18)	0.00400
	Adult (19+) to adult (19+)	0.00400
Working Group	Adult (19+) to adult (19+)	0.00400
School	Child (0–18) to child (0–18)	0.00029
Grade	Child (0–18) to child (0–18)	0.00158
Class	Child (0–18) to child (0–18)	0.00865
Neighborhood	Any to child (0–4)	0.035×10^{-5}
	Any to child (5–18)	1.044×10^{-5}
	Any to adult (19–64)	2.784×10^{-5}
	Any to adult (65+)	5.568×10^{-5}
	Community	Any to child (0–4)
Community	Any to child (5–18)	2.608×10^{-6}
	Any to adult (19–64)	6.960×10^{-6}
	Any to adult (65+)	13.92×10^{-6}

Numbers in brackets show age groups

agent j is infected, $n \geq n_j$ and $f(n - n_j | j) \geq 0$. The infectivity increases exponentially since the latent period ends (the latent period can be set to zero for some variants), until it reaches its peak at $f(n - n_j | j) = 1.0$. Subsequently, during the recovery period, the infectivity decreases linearly to zero, leading to the agent transitioning to the Removed state. The changes in infectivity

of an infected agent are plotted in the main manuscript, see Fig. 1. The upper bounds of the daily probabilities of transmission from agent j to agent i , denoted by $q_{j \rightarrow i}^g$, which depends on age and mixing contexts. Table 5 summarises the daily transmission probabilities for age-dependent interactions in different mixing contexts. These probabilities have been derived and validated in prior studies [13, 15, 18, 21], and despite being relatively small, need to be specified with a sufficient precision, given a large number of interacting agents.

The infection probability for agent i across all mixing contexts is calculated as shown in the main text, see Eq. 2:

$$p_i(n) = 1 - \prod_{g \in G_i(n)} \prod_{j \in A_g \setminus \{i\}} (1 - p_{j \rightarrow i}^g(n)) \tag{5}$$

At the end of each time cycle, a Bernoulli sampling process based on the probability $p_i(n)$ is used to decide whether a susceptible agent i acquires the infection.

An infected agent can be either symptomatic or asymptomatic. The probability of symptomatic illness is determined with respect to the infection probability $p_i(n)$, by incorporating a scaling factor σ that represents the fraction of symptomatic cases among the total cases:

$$p_i^d(n) = \sigma(i) p_i(n) \tag{6}$$

where $\sigma(i)$ is defined based on the age of agent i using a piece-wise approach: for adults (age > 18), $\sigma_a = 0.67$; and for children (age ≤ 18), $\sigma_c = 0.268$, following previous studies [13, 15, 18, 21]. Asymptomatic agents in the model have a lower infectivity compared to their symptomatic counterparts, governed by a scaling factor of α_{asympt} ($0 \leq \alpha_{asympt} \leq 1$).

In reality, not all infection cases are detected, particularly when case detection relies on voluntary self-reporting. To capture this feature, the model employs two variables adjusting the extent of case detection as the probability of detecting symptomatic cases (π_{symp}) and the probability of detecting asymptomatic cases (π_{asympt}).

In this study, we model the transmission of three COVID-19 variants of concern in a wide of range of scenarios in combination with different intervention policies. Some epidemiological parameters are set differently across the three variants, while others are kept constant for all variants. A summary of parameterisation is provided in Table 6.

Table 6 Model parameterisation for the three considered variants of concern (i.e., ancestral, Delta, and Omicron) adopted by AMTraC-19

Model parameters	Ancestral	Delta	Omicron	Note
κ	2.75	5.3	23	Global transmission scalar
T_{inc} , mean	5 ($\mu=1.609 \sigma=0.00001$)	4.4 ($\mu=1.396 \sigma=0.413$)	3 ($\mu=1.013 \sigma=0.413$)	Incubation period (log-normal)
T_{rec} , mean and range	12 [12,12]	10.5 [7, 14]	9 [7,11]	Recovery period, mean and range (uniform)
T_{lat} , fixed	2	0	0	Latent period
α_{asympt}	0.3			Asymptomatic transmission scalar
σ_a	0.67			Probability of symptoms (age > 18)
σ_c	0.268			Probability of symptoms (age ≤ 18)
π_{symp}	0.1			Daily case detection probability (symptomatic)
π_{asympt}	0.01			Daily case detection probability (asymptomatic)
R_0 , mean and 95% CI	2.77 [2.73,2.83]	5.97 [5.93, 6.00]	19.56 [19.12, 19.65]	Basic reproductive number
T_{gen} , mean and 95% CI	7.62 [7.53, 7.70]	6.88 [6.81, 6.94]	5.42 [5.38, 5.44]	Generation/serial interval

The last two rows show the corresponding basic reproductive number (R_0) and generation/serial interval (T_{gen}), calibrated in our previous studies [13, 15, 18, 21]

Non-pharmaceutical interventions

The model incorporates various non-pharmaceutical interventions (NPIs) aimed to mitigate SARS-CoV-2 transmission, including case isolation (CI), home quarantine (HQ), school closures (SC), and social distancing (SD). Each NPI is characterised by: (i) macro-distancing level, setting a fraction of the population that adopts, or complies with, the intervention, and (ii) micro-distancing parameters which quantify the adjusted (usually reduced) interaction strengths between the agents who have adopted the NPI and other agents in the shared mixing groups. The infection probability $p_i(n)$ for individuals who have adopted an NPI is adjusted as follows:

$$p_i(n) = 1 - \prod_{g \in G_i(n)} \left[1 - F_g(i) \left(1 - \prod_{j \in A_g \setminus \{i\}} (1 - F_g(j) p_{j \rightarrow i}^g(n)) \right) \right] \tag{7}$$

where $F_g(j)$ is the strength of the interaction between agent j and other agents in the mixing context g . For NPI-adopting agents j , the interaction strength is modified: $F_g(j) \neq 1$. For non-adopting agents j , the interaction strength is unchanged: $F_g(j) = 1$.

While CI and HQ are activated from the beginning of the simulation and last throughout the entire simulation period, SD and SC are triggered only if the cumulative incidence nationwide exceeds certain threshold. Thus, in general, an NPI is activated once the cumulative incidence exceeds a defined, NPI-dependent, threshold. Agents are assigned to adopt NPIs according to a Bernoulli process. In addition, CI only applies to symptomatic cases or the detected infections of asymptomatic individuals. Agents can adopt more than one NPIs when multiple NPIs are active, but their interaction strengths are set to the value of $F_g(j)$ following a descending

Table 7 The macro-distancing parameters (population fractions) and micro-distancing (interaction strengths) for the considered NPIs

Intervention	Macro-distancing (population fractions)			Micro-distancing (interaction strengths)			
	Compliance level	Duration T (days)	Threshold (cases)	Household	Community	Workplace\ School	Duration t (days)
CI	0.7	196	0	1.0	0.25	0.25	$D(i)$
HQ	0.5	196	0	2.0	0.25	0.25	7
$SC^{s/t}$	1.0	110	100	1.0	0.5	0	110
SC^P	0.25	110	100	1.0	0.5	0	110
SD	0.7	196	400	1.0	0.25	0.1	196

The micro-duration of CI is limited by the disease progression in the affected agent i , $D(i)$. Compliance levels for SC for students/teachers and parents can be different, as indicated by distinct settings for $SC^{s/t}$ and SC^P , respectively

priority assignment order: CI, HQ, SD, and SC. In this study, we assume static adoption/compliance-with NPI for CI, HQ, SC, and SD (i.e., adoption/compliance level remains constant throughout the simulation). The NPI parameterisation, encompassing both macro-distancing and micro-distancing levels, is presented in Table 7.

Vaccination

In this study, we consider several scenarios with a partial vaccination preemptively rolled out before a pandemic wave (Policies 3, 4, and 5 as shown in Fig. 2 in the main manuscript). In these scenarios, the vaccination coverage is set to 50% of the population prior to the simulation start, corresponding to 11.7 million agents using 2016 census, or 12.712 million agents using 2021 census. The vaccines are distributed to different age groups as follows: 8.33% of vaccinated agents are aged 18 years or younger ($age \leq 18$), 83.34% are aged between 18 and 65 years ($18 < age < 65$), and 8.33% are aged 65 years or older ($age \geq 65$). In our model, the vaccination roll-out contains two types of vaccines: the “priority” vaccine with a higher clinical efficacy (Priority, $VE^c = 0.7$), representing BNT162b2 (Pfizer/BioNTech) and Moderna; and the “general” vaccine with a medium efficacy (General, $VE^c = 0.5$), representing ChAdOx1 nCoV-19 (Oxford/AstraZeneca). These values align with clinical observations following booster vaccination against the Omicron variant [49]. For simplicity, we use the clinical vaccine efficacy against Omicron as a reference point for all considered variants, aiming to investigate the impact of imperfect vaccines in a comparative way.

We further decompose VE^c into two efficacy components: the susceptibility-reducing efficacy (VE^s) and the disease-preventing efficacy (VE^d), following [13, 15, 21]:

$$VE^c = VE^d + VE^s - VE^s \times VE^d \tag{8}$$

with the efficacy components set as specified in Table 8.

We also consider the transmission-limiting efficacy (VE^t) set at $VE^t = 0.4$, observed for the considered vaccines against the Omicron variant [50]. We carried out the sensitivity analysis in terms of vaccine efficacy components in prior studies [13, 21], testing a range of VE^t and VE^c values and showing robustness of the model to changes in these efficacy components. The model parameterisation with respect to vaccination is provided in Table 8.

The transmission probability of infecting a susceptible agent i , accounting for both vaccination and NPIs, is derived as follows:

$$p_i(n) = 1 - \prod_{g \in G_i(n)} [1 - (1 - VE_i^s)F_g(i) \left(1 - \prod_{j \in A_g \setminus \{i\}} (1 - (1 - VE_j^t)F_g(j) p_{j \rightarrow i}^g(n)) \right)] \tag{9}$$

where for all vaccinated agents j : $VE_j^t = VE^t$, $VE_j^s = VE^s$ and $VE_j^d = VE^d$, and for all unvaccinated agents j : $VE_j^t = VE_j^s = VE_j^d = 0$. The values of VE^s , VE^d , and VE^t are selected according to the vaccine type (priority or general) allocated to the agent.

Table 8 Simulation parameters for the preemptive vaccine rollout simulated within the ABM

Parameter	Value (census 2016)	Value (census 2021)	Reference
Priority vaccine coverage	5.850M	6.356M	25% of total population
General vaccine coverage	5.850M	6.356M	25% of total population
Vaccine allocation [$age \leq 18$]	0.975M	1.059M	8.33% of vaccinated population
Vaccine allocation [$18 < age < 65$]	9.750M	10.594M	83.34% of vaccinated population
Vaccine allocation [$age \geq 65$]	0.975M	1.059M	8.33% of vaccinated population
Priority, VE^c	0.7		[49]
Priority, VE^s	0.452		Derived
Priority, VE^d	0.452		Derived
Priority, VE^i	0.4		[50]
General, VE^c	0.5		[49]
General, VE^s	0.293		Derived
General, VE^d	0.293		Derived
General, VE^i	0.4		[50]

The probability of an agent experiencing illness (symptomatic) is further influenced by the disease-preventing efficacy, (VE_i^d) , as follows: $p_i^d(n) = (1 - VE_i^d) \sigma_{a|c} p_i(n)$, given the fractions of symptomatic adults and children, denoted by σ_a and σ_c , respectively.

Sensitivity analysis

We have carried out an extensive sensitivity analysis for a wide range of model parameters and multiple variants of concern, including different profiles of the disease natural history [13, 18, 26], the global transmission scalar [15, 18, 26], the asymptomatic infectivity [15, 18, 26], the fractions of symptomatic adults and children [13, 15, 18, 26], the NPI compliance levels (e.g., SD compliance [13, 15, 18]), the SD intervention threshold [26], the interaction strength in different mixing contexts (at home, work, or community [18]), and the vaccination coverage and vaccine efficacy components [18, 21]. These studies demonstrated

robustness of the model to parameter changes within acceptable ranges supported by epidemiological evidence.

Appendix D: Lorenz curves

The pandemic Lorenz curves measure unequal contributions of SA2 areas towards the aggregate nationwide pandemic severity (introduced in section "Lorenz curves: measuring unequal distribution of pandemic severity" of the main manuscript) and are constructed as follows:

1. Record the national cumulative incidence at the end of simulation (196 days), $CI_{AUS,T=196}$.
2. Compute local attack rate, AR_{SA2} , as the ratio between the cumulative incidence and the total population, both taken at SA2 level, as $AR_{SA2} = \frac{CI_{SA2,T=196}}{UR_{SA2}}$, where $CI_{SA2,T=196}$ is the local cumulative incidence at the end of the simulation (196 days) and UR_{SA2} is the usual residential population.
3. Rank SA2 areas by their local attack rate in ascending order. Using this rank, add one SA2 at a time to accumulate the population fraction reaching the national total, from 0% to 100%, until all SA2 areas are considered. This forms the x-axis.
4. Following the rank of SA2 areas in terms of their local attack rate, determined in the previous step, add one SA2 at a time to compute the fraction of the national cumulative incidence for the considered SA2 areas, from 0% to 100%, until all SA2 areas are considered. This forms the y-axis.

Table 9 Lorenz curves setup: demographic and epidemic characteristics of a simplified example

SA2	CI	UR	AR	AR _{SA2} rank
SA _x	12	40	12/40 = 0.3	1
SA _y	15	24	15/24 = 0.625	3
SA _z	18	36	18/36 = 0.5	2
National total	45	100	45/100 = 0.45	NA

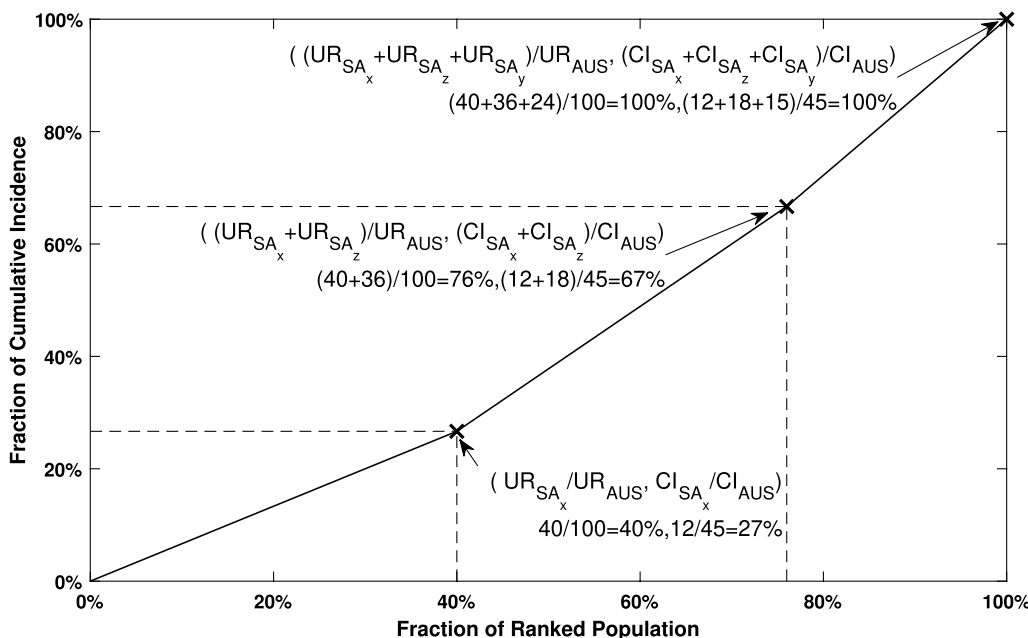


Fig. 18 Constructing pandemic Lorenz curve for the simplified example, using the data presented in Table 9

- Connect all data points to form a pandemic Lorenz curve.

Let us consider a simplified example of the national population, comprising the populations of three SA2 areas, namely SA_x , SA_y , and SA_z , with demographic and epidemic characteristics specified in Table 9.

Using Table 9, we can compute the corresponding pandemic Lorenz curve, shown in Fig. 18:

In this example, the three considered SA2 areas contribute unequally towards the global attack rate, with SA_x having the smallest gradient and thus deviating furthest from the line of equality (i.e., diagonal line). When using

the census data, we follow this process for all SA2 areas across the country: 2,310 SA2 areas in 2016 and 2,454 SA2 areas in 2021.

Appendix E: Supporting results

In this study, we systematically compared various COVID-19 pandemic scenarios across different census years, intervention strategies, and variants of concern. This section complements results presented in section "Results" in the main manuscript and follows the same structure, providing supporting results in terms of the effects of population heterogeneity on pandemic

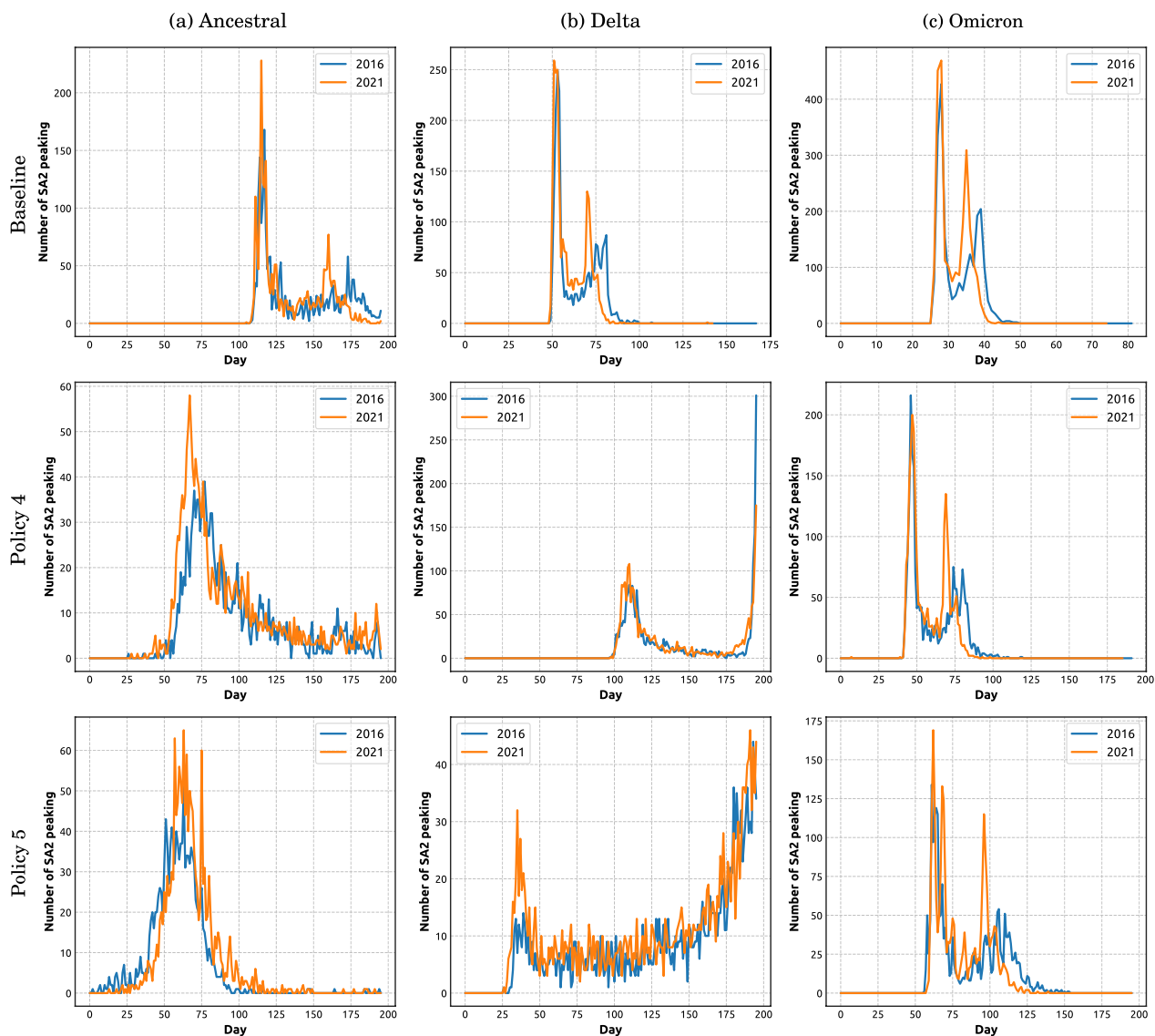


Fig. 19 Number of SA2 areas exhibiting an incidence peak in the simulated time period, across three considered policies: Policy 1, Policy 4, and Policy 5, and three variants: **a** ancestral; **b** Delta; **c** Omicron. Each plot is averaged over 100 runs

severity (Section E.1), the impact of urbanisation on the spread of the virus (Section "Effects of urbanisation on pandemic spread"), and the effects of school closures across variants of concern (Section "Effects of school closures across variants of concern").

Effects of population heterogeneity on pandemic severity

Population growth amplifies pandemic peaks

Figure 19 shows that in comparison to 2016 results (solid blue line), there are more SA2 areas in 2021 that peak around similar times across all variants and policies (solid orange line). In cases where two peaks are observed (e.g., Fig. 19c), in 2021, the time difference between the two peaks is shortened, explaining the weakened bimodal dynamics (Figs. 5 and 8 in main manuscript) as the two incidence peaks tend to merge into a single but wider incidence wave.

Changes in population size amplify incidence peak more than changes in density Fig. 20 shows a weaker yet significant correlation between the usual residential population density difference and the peak incidence difference, computed between 2016 and 2021 at the SA2 resolution for three considered variants. The statistical analysis of the linear fits, shown in Figs. 20 and 6, is summarised in Table 10 below.

Table 10 Statistical analysis of the linear fits shown in Fig. 9 in main manuscript (effects of population size difference) and Appendix Fig. 20 (effects of population density difference), in terms of the square of the correlation (R^2) and the root mean squared error (RMSE)

Variant of concern	Usual residential population		Population density	
	R^2	RMSE	R^2	RMSE
Ancestral	0.607	0.296	0.206	0.393
Delta	0.432	0.694	0.15	0.788
Omicron	0.81	0.561	0.25	1.02

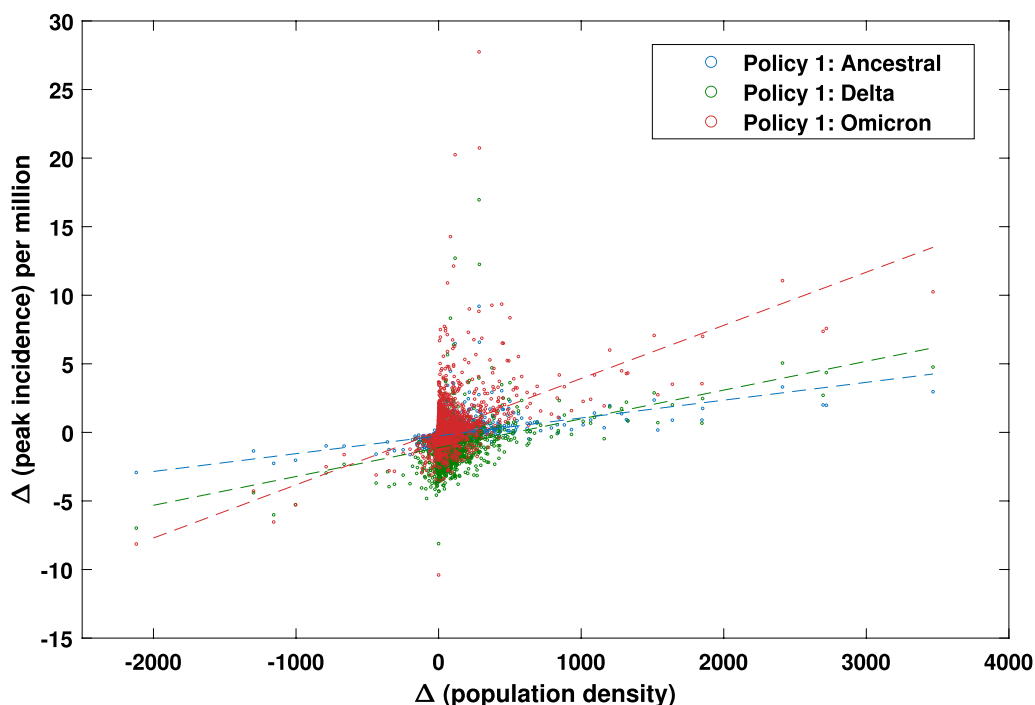


Fig. 20 Correlation between the usual residential population density difference and the peak incidence difference, computed between 2016 and 2021 at the SA2 resolution for the three considered variants: ancestral (blue), Delta (green), and Omicron (red). Data points corresponding to each SA2 area are derived as the averages over 100 runs. Dashed lines represent linear fit for each of the profiles. Total number of overlapping SA2 areas between 2016 and 2021 census years: 2,147. Pearson correlation coefficients: $r_{ancestral} = 0.4150$, $r_{Delta} = 0.3426$, $r_{Omicron} = 0.4509$. These are lower correlation coefficients than the ones obtained for the population size difference, as reported in main manuscript, Fig. 6

We performed further analyses investigating the impact of large households on pandemic severity at SA2 level by looking at (i) the average household size, (ii) the population within an SA2 area residing in households of two different sizes: large households (i.e., at least five household members), and small households (i.e., up to four household members). Note that this partition is based on the number of total household members, regardless of the household composition.

Fig. 21 shows some correlation between the average household size and the peak incidence for three considered variants in two census years at the SA2 resolution ($0.23 < r < 0.34$). However, there is no clear correlation between the average household size difference and

Table 11 Pearson correlation coefficients (i) between the peak incidence and the average household size, as shown in Appendix Fig. 21 (effect of the average household size in 2016 and 2021), and (ii) between the peak incidence and the difference in average household size between 2016 and 2021, as shown in Appendix Fig. 22 (effects of the difference in average household size)

Variant of concern	Average household size		Δ (average household size)
	2016	2021	
Ancestral	0.330	0.244	0.078
Delta	0.333	0.234	0.163
Omicron	0.327	0.218	0.031

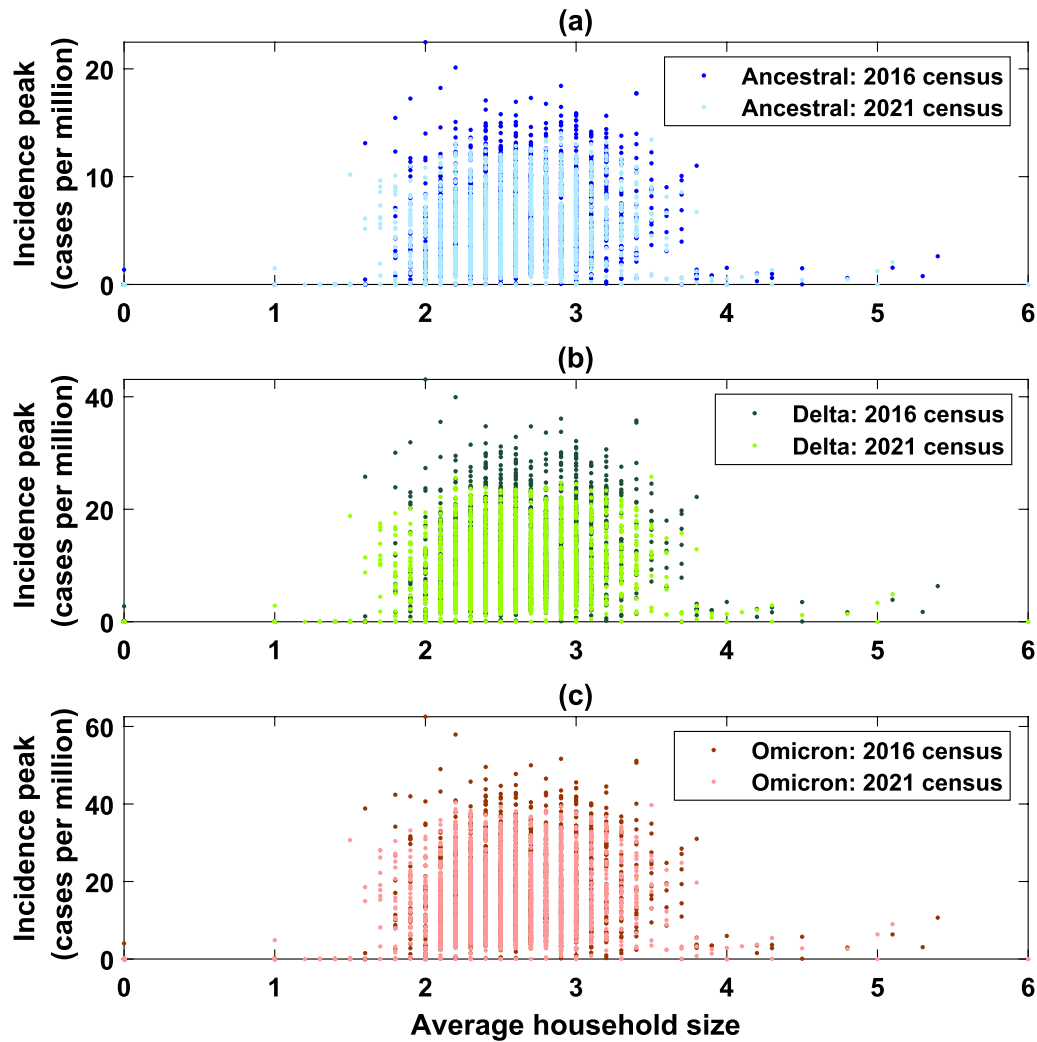


Fig. 21 Correlation between the average household size and the peak incidence between 2016 and 2021 at SA2 resolution for three considered variants: **a** ancestral, **b** Delta, and **c** Omicron. Data points corresponding to each SA2 are computed as the average over 100 runs. Total number of overlapping SA2 areas between 2016 and 2021 census years: 2147. Refer to table 11 for Pearson correlation coefficients

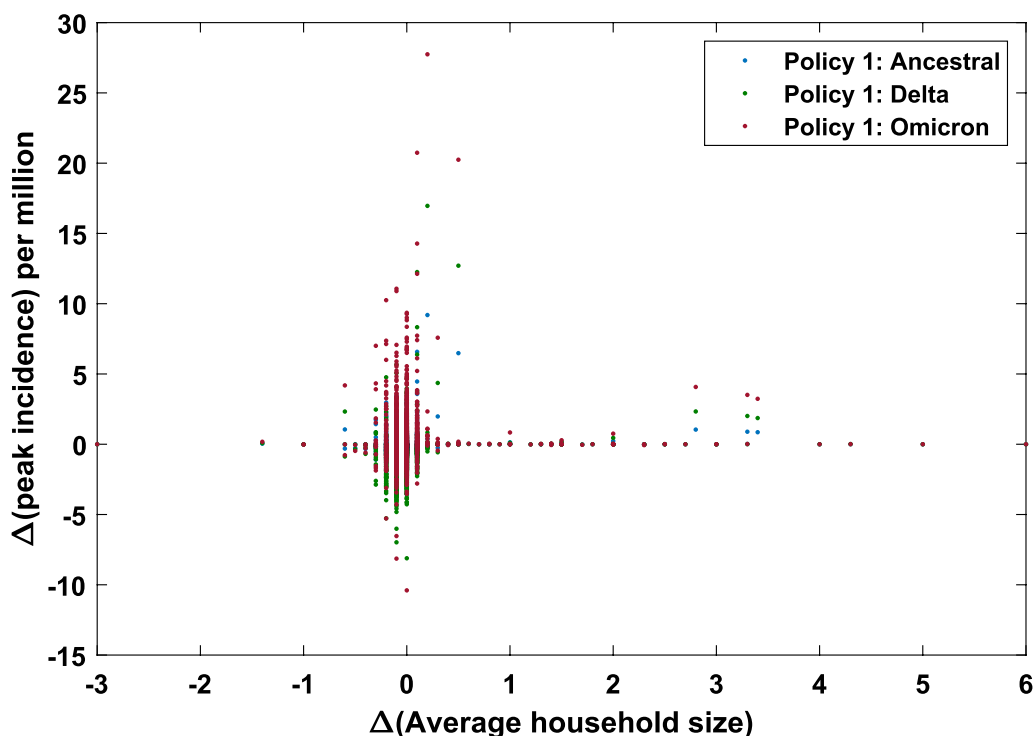


Fig. 22 There is no clear correlation between the difference in average household size and the peak incidence difference, computed between 2016 and 2021 at SA2 resolution for three considered variants: ancestral (blue), Delta (green), and Omicron (red)

the peak incidence difference between 2016 and 2021 ($0.03 < r < 0.17$). The correlation coefficients are summarised in Table 11 below.

It is evident that there is a strong correlation between a sub-population and the peak incidence regardless of the household size (Figs. 23, 25 and Table 12), with the sub-population residing in small households showing a higher correlation in each census year.

The correlation between the peak incidence difference and the sub-population difference (Fig. 24 for large households, and Fig. 26 for small households) is slightly weaker, see Table 12. Interestingly, for the Delta variant, the correlation between the peak incidence difference and the difference in large-household sub-population is higher ($r = 0.611$) than its counterpart for small households ($r = 0.522$).

Pandemic severity is distributed unequally across local (SA2) areas In this section, we present pandemic Lorenz curves directly comparing variants of concern for different policies and census years (Fig. 27).

Effects of urbanisation on pandemic spread

Over the last five years, the Australian population has increased in both urban and non-urban (i.e., regional and rural) areas, leading to redefinition of SA2 boundaries (see Table 13). There are several non-GCC cities in Australia with international airports (see Fig. 13 in Appendix, shown in green). The inclusion of the SA2 areas surrounding these cities into analysis of the urban areas does not affect the observations reported in section "Effects of urbanisation on pandemic spread" of main manuscript. In this

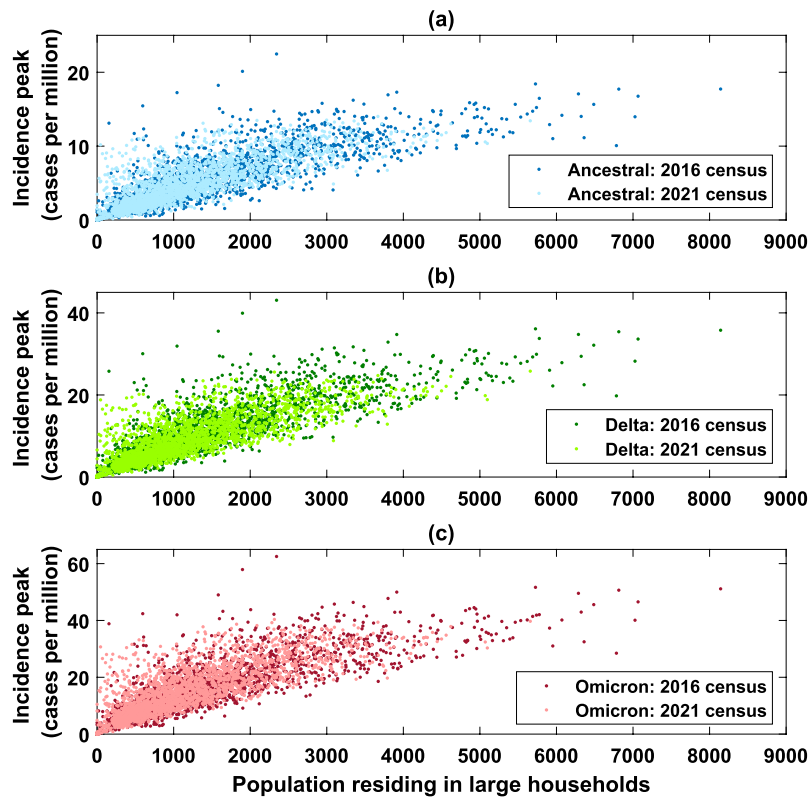


Fig. 23 Strong positive correlation between the sub-population residing in large households in an SA2 area and the peak incidence, determined at SA2 resolution across two census years (2016 and 2021) and three considered variants: **a** ancestral, **b** Delta, and **c** Omicron. Data points corresponding to each SA2 are computed as the average over 100 runs. Pearson correlation coefficients r are summarised in Table 12

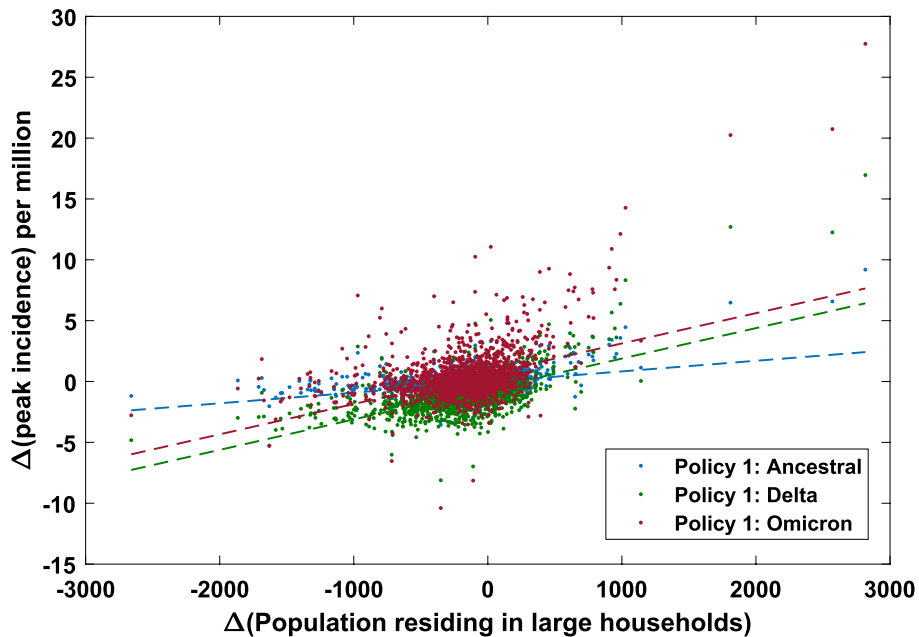


Fig. 24 Strong correlation between the difference in sub-population residing in large households and the peak incidence difference between 2016 and 2021 at SA2 resolution for three considered variants: ancestral (blue), Delta (green), and Omicron (red). Dashed lines represent linear regression for each of the profiles. Data points corresponding to each SA2 are computed as the average over 100 runs. Pearson correlation coefficients r are summarised in Table 12

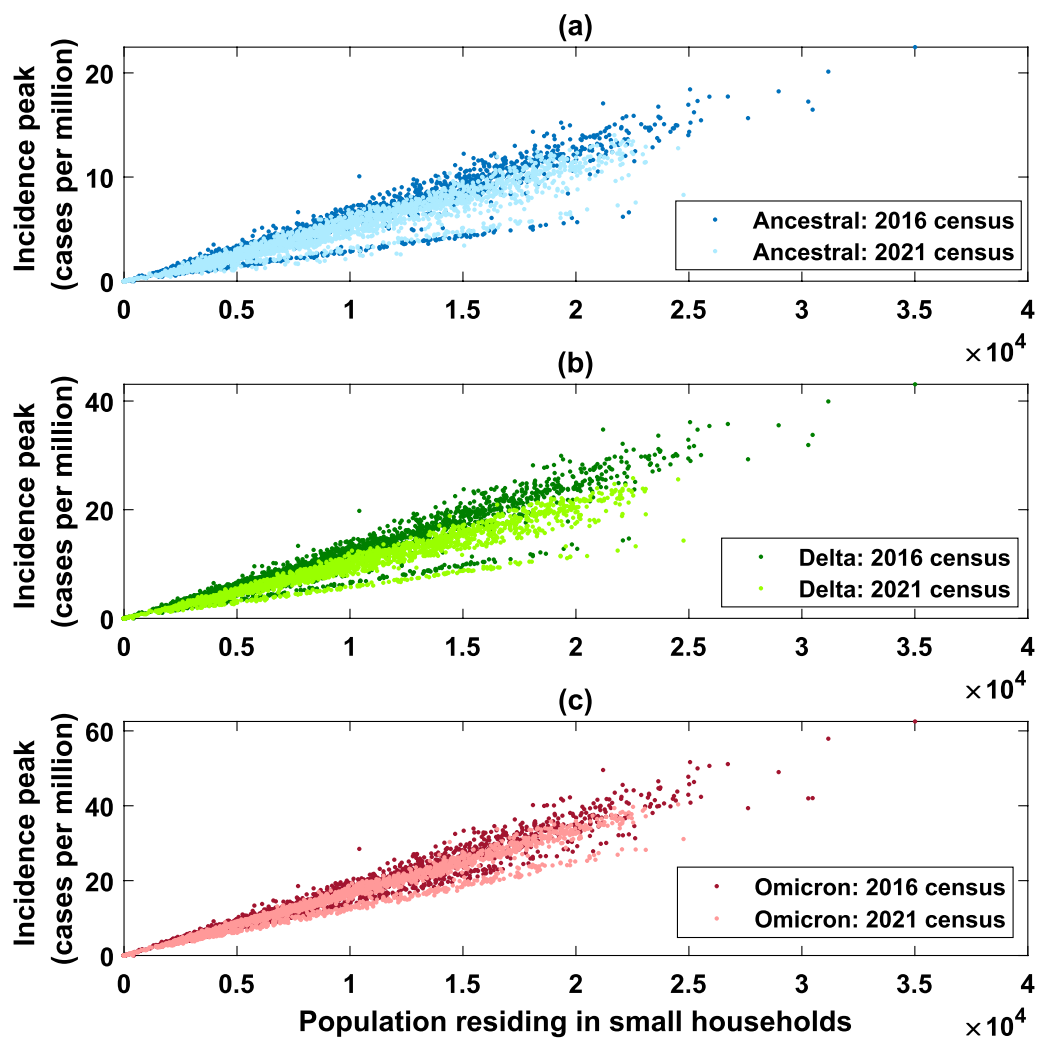


Fig. 25 Strong positive correlation between the sub-population residing in small households in an SA2 area and the peak incidence, determined at SA2 resolution across two census years (2016 and 2021) and three considered variants: **a** ancestral, **b** Delta, and **c** Omicron. Data points corresponding to each SA2 are computed as the average over 100 runs. Pearson correlation coefficients r are summarised in Table 12

Table 12 Pearson correlation coefficients between sub-populations residing in households of different sizes and the peak incidence, as well as between differences in these quantities

Variant of concern	Large households			Small households		
	2016	2021	$\Delta(\text{large households})$	2016	2021	$\Delta(\text{small households})$
Ancestral	0.841	0.808	0.415	0.958	0.959	0.712
Delta	0.845	0.506	0.611	0.966	0.964	0.522
Omicron	0.839	0.801	0.437	0.982	0.979	0.844

(i) For large households, compare with Appendix Fig. 23 (effects of sub-populations residing in large households in 2016 and 2021) and Fig. 24 (effect of the difference in this sub-population). (ii) For small households, compare with Appendix Fig. 25 (effects of sub-population residing in small households in 2016 and 2021) and Fig. 26 (effect of the difference in this sub-population)

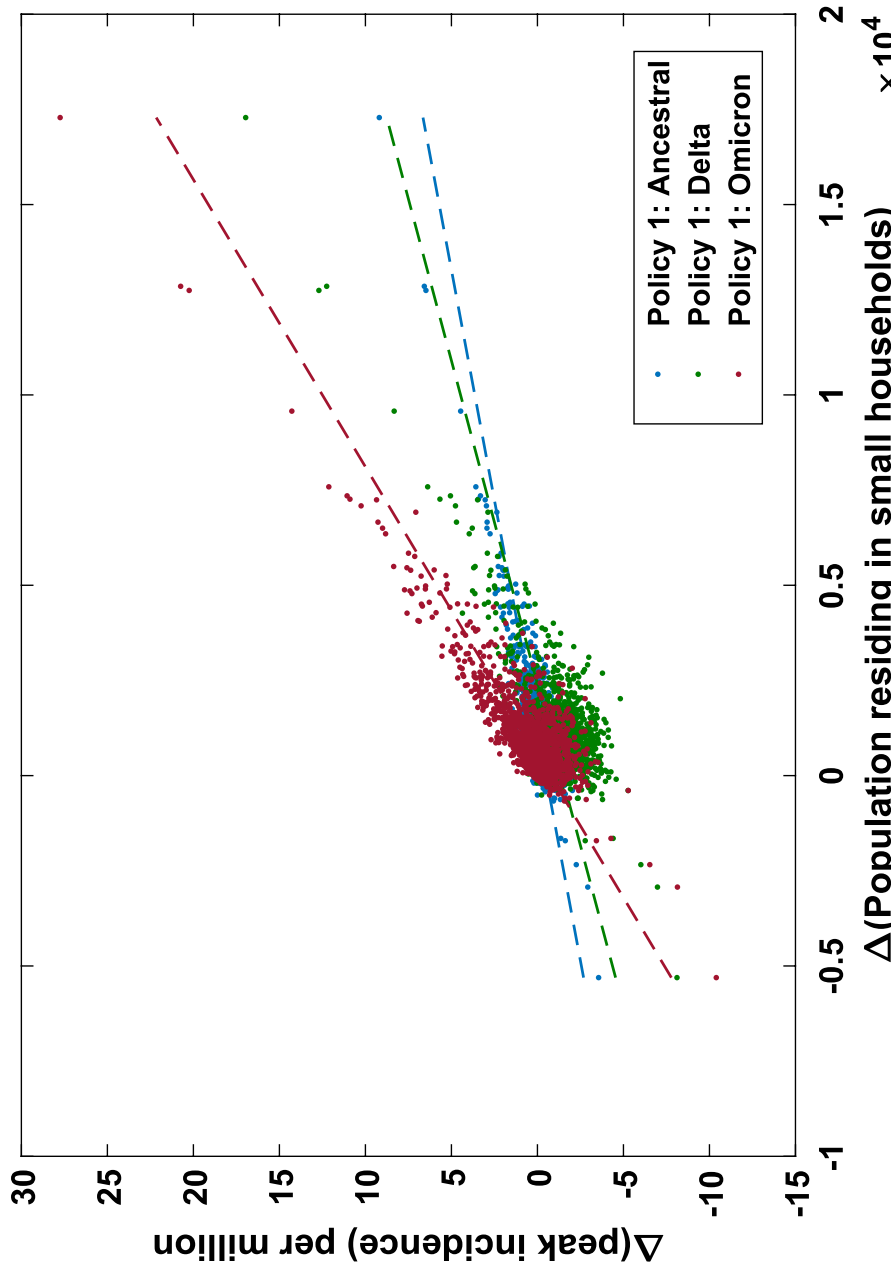


Fig. 26 Strong correlation between the difference in sub-population residing in small households and the peak incidence difference between 2016 and 2021 at SA2 resolution for three considered variants: ancestral (blue), Delta (green), and Omicron (red). Dashed lines represent linear regression for each of the profiles. Data points corresponding to each SA2 are computed as the average over 100 runs. Pearson correlation coefficients r are summarised in Table 12

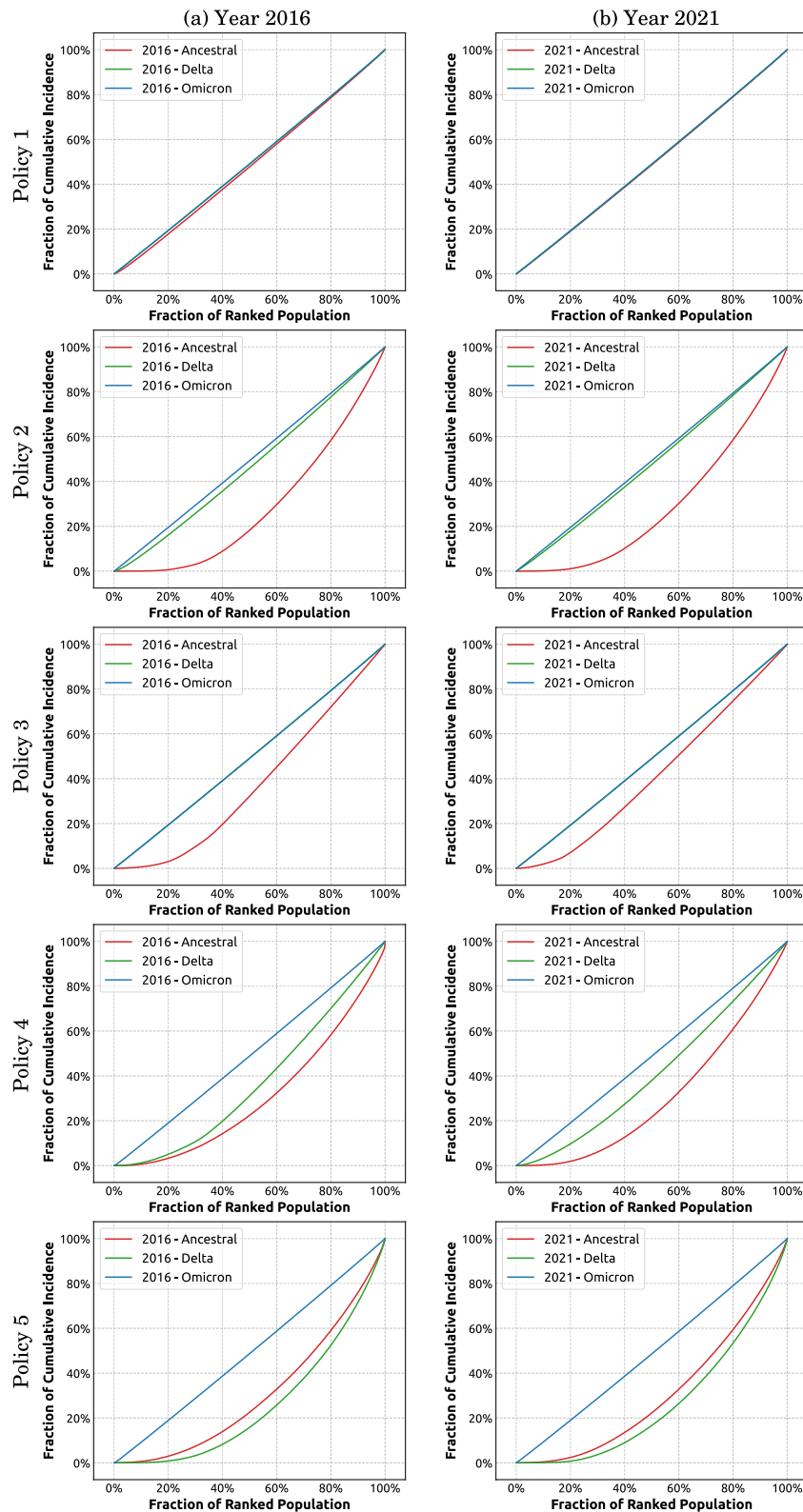


Fig. 27 Pandemic Lorenz curves for the considered policies across the three variants and two census years. Each row compares the impact of a policy using 2016 census (left) and 2021 census (right) for three variants. Refer to Figure 2 in the main manuscript for a detailed description of the considered intervention policies. Each profile corresponds to one intervention policy and is computed as the average over 100 runs

Table 13 Population distribution of 2016 and 2021 census years, following the urban and non-urban partition explained in section "Effects of urbanisation on pandemic spread" in main manuscript, for Greater Capital Cities (GCCs) and cities with international airports (APs)

	GCCs / non-urban areas		GCCs & APs / non-urban areas		Total
2016 census					
Population	15,612,130	/ 7,794,205	17,359,223	/ 6,047,112	23,406,335
SA2 areas	1,319	/ 991	1,498	/ 812	2,310
2021 census					
Population	17,059,063	/ 8,368,966	18,973,019	/ 6,455,010	25,428,029
SA2 areas	1,460	/ 994	1,650	/ 804	2,454

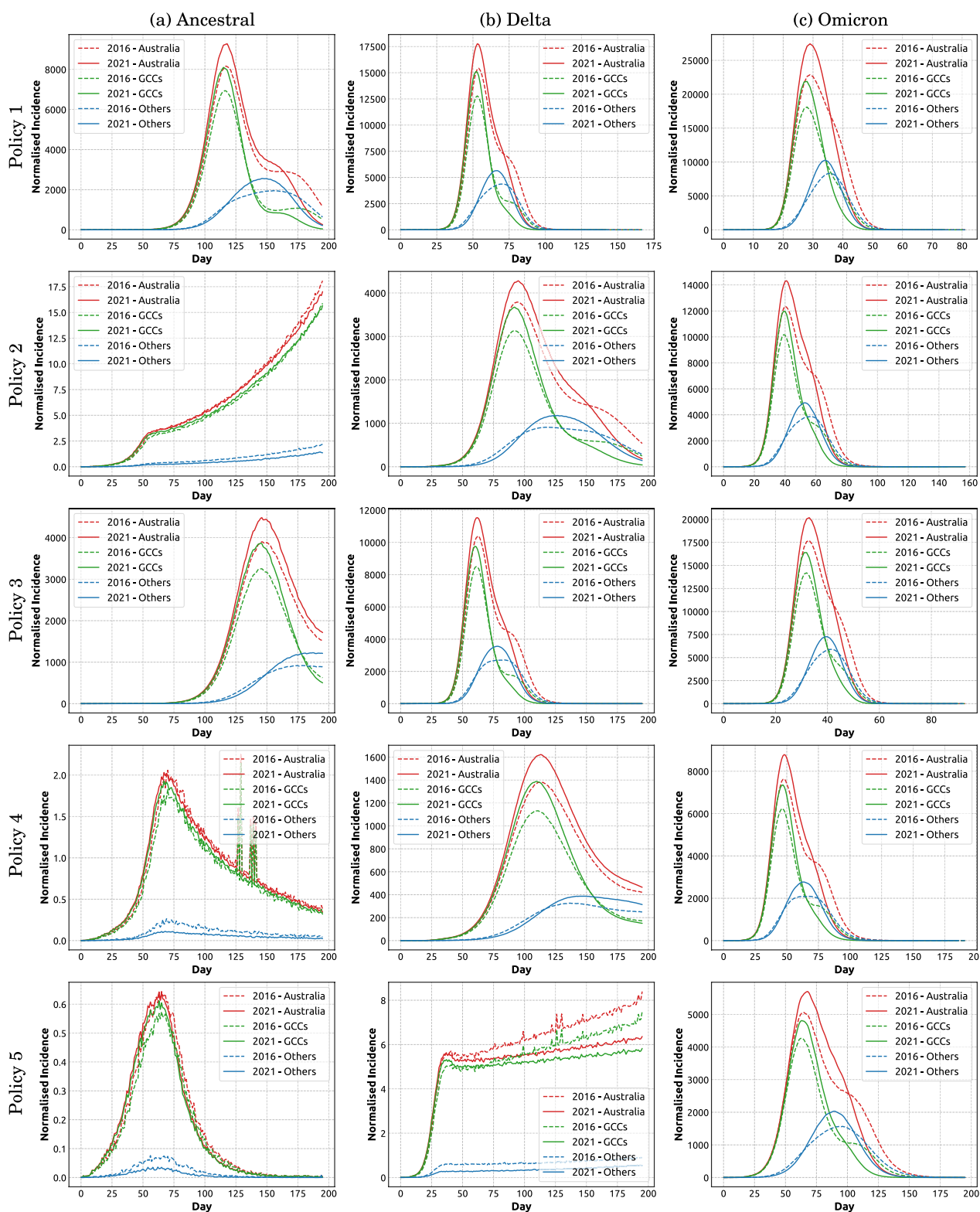


Fig. 28 GCCs versus other non-urban areas. Effects of urbanisation on pandemic dynamics for the considered policies, across three variants, using 2016 and 2021 census years. We partition SA2 areas in Australia as Greater Capital Cities (GCCs), and other non-urban areas. Each column compares the impact of five intervention policies for a variant of concern: **a** ancestral; **b** Delta; **c** Omicron. Each plot is averaged over 100 runs

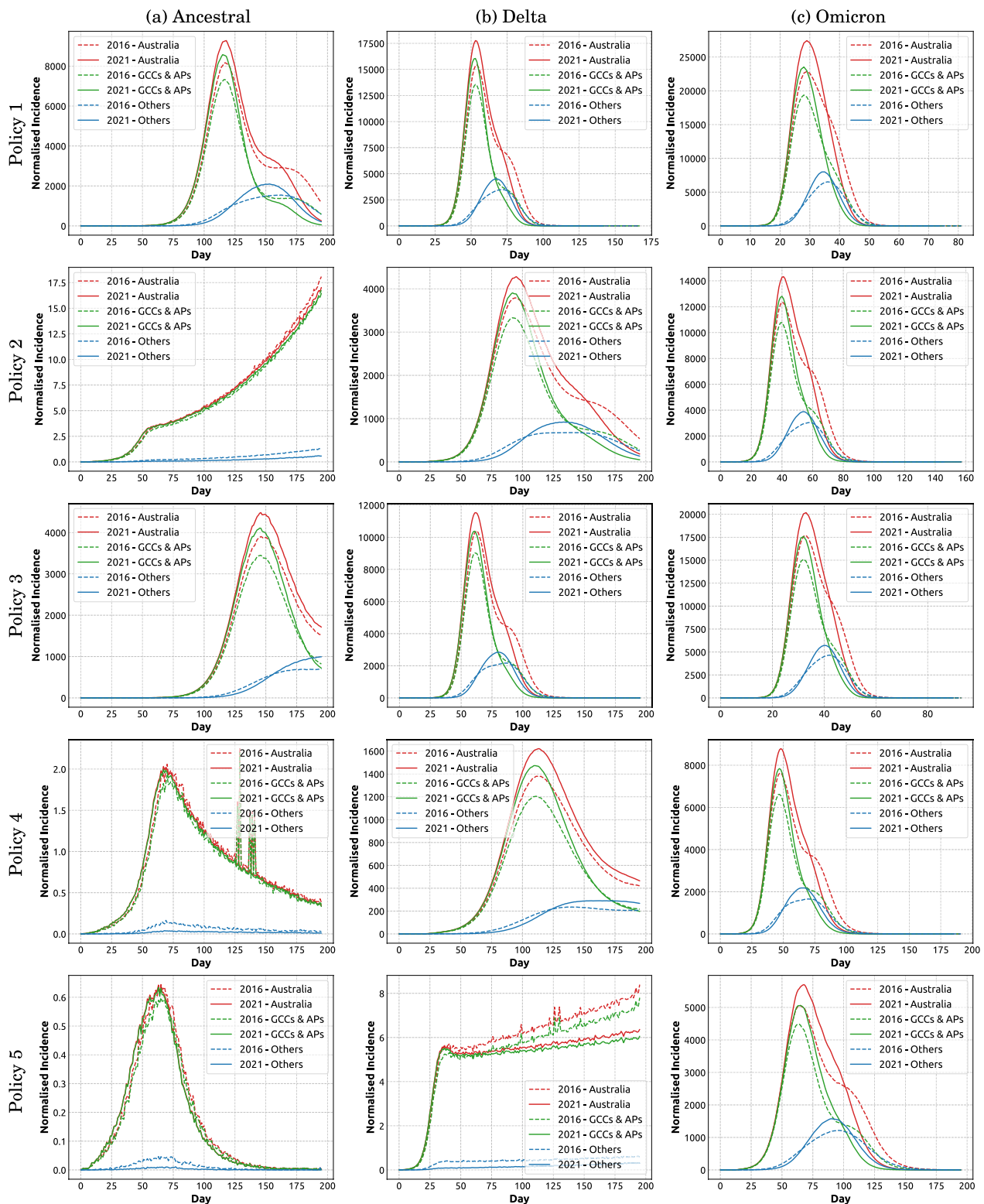


Fig. 29 GCCs and APs versus other non-urban areas. Effects of urbanisation on pandemic dynamics for considered policies, across three variants using 2016 and 2021 census years. We partition SA2 areas in Australia as Greater Capital Cities (GCCs) and those close to international airports (APs), and other non-urban areas. Each column compares the impact of five intervention policies for a variant of concern: **a** ancestral, **b** Delta, **c** Omicron. Each plot is averaged over 100 runs

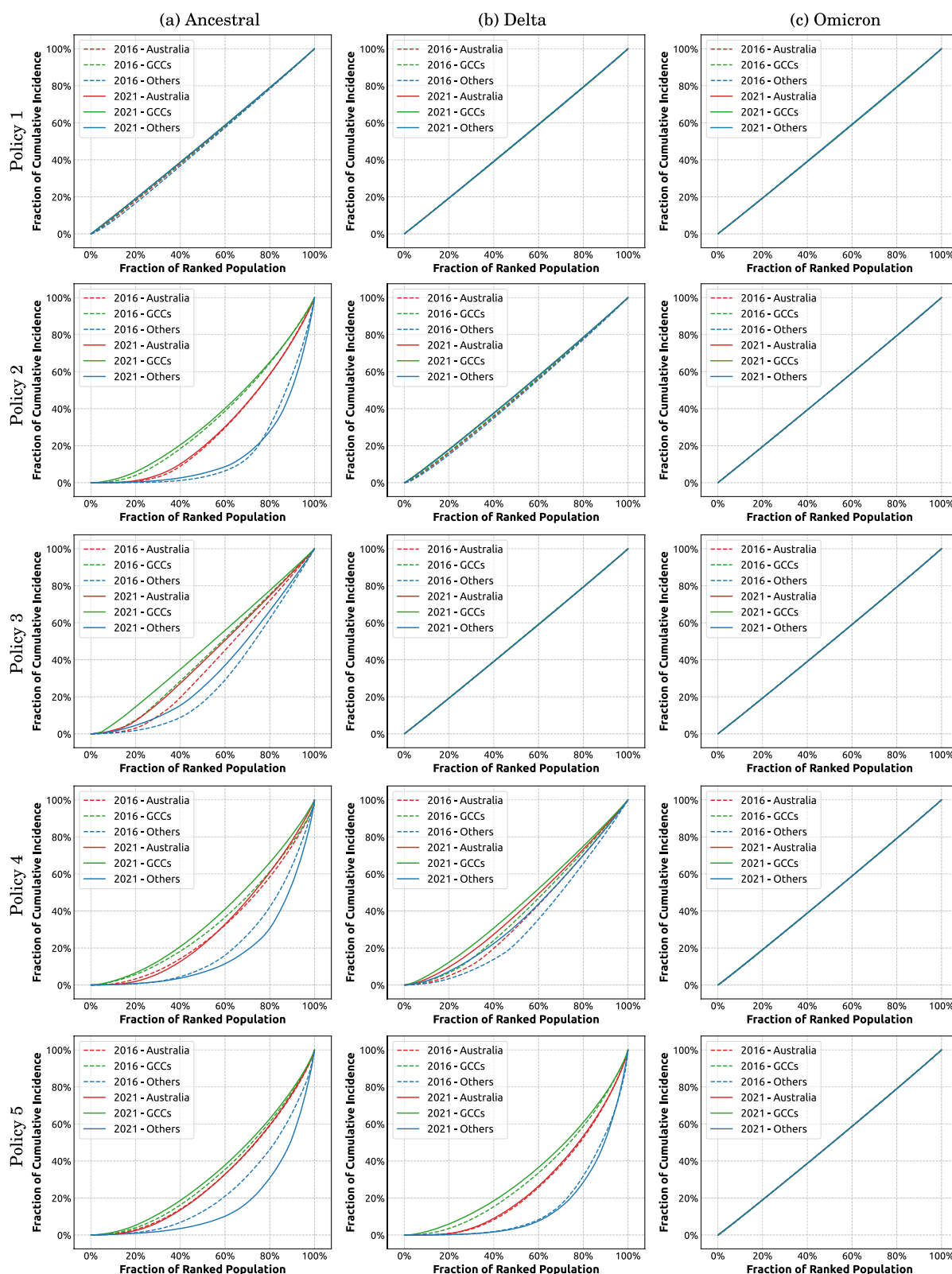


Fig. 30 Pandemic Lorenz curves for considered variants across policies and census years. Each column compares the impact of five intervention policies for a variant of concern for GCCs and non-GCCs: **a** ancestral; **b** Delta; **c** Omicron. Each profile corresponds to one intervention policy for Australia (red), GCCs (green), or other non-urban areas (blue), and is computed as the average over 100 runs

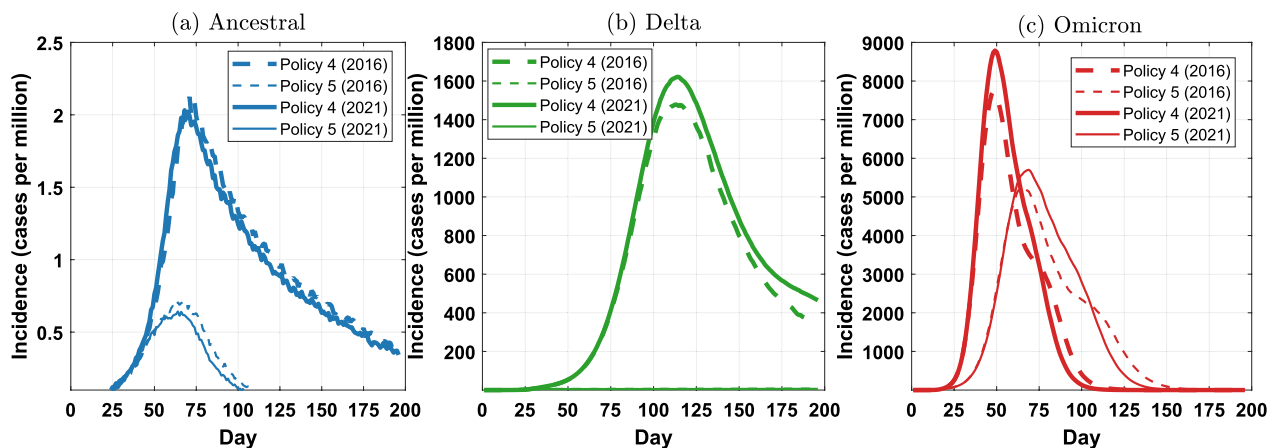


Fig. 31 Effects of school closures combined with NPIs for three considered variants (on linear scale). **a** ancestral; **b** Delta; **c** Omicron. Figure 9 in main manuscript shows these plots on log scale. Each profile corresponds to one intervention policy and is computed as the average over 100 runs

section, we compare the effects of growing urbanisation on pandemic dynamics for Greater Capital Cities (GCC) considered without and with cities having international airports (AP), as shown by Figs. 28 and 29, respectively. We also examine pandemic Lorenz curves in context of this urbanisation (Fig. 30).

Effects of school closures across variants of concern

Figure 31 shows the effects of school closures combined with NPIs for three considered variants on linear scale.

Acknowledgements

This work was supported by the Australian Research Council grant DP220101688 (MP, SLC and CMJ) and the University of Sydney's Digital Science Initiative (DSI) Research Pilot Project funding scheme (SLC, MP and QDN). Additionally, SLC is supported by the University of Sydney Infectious Diseases Institute Seed Grant 220182. CMJ is supported by the University of Sydney Faculty of Engineering Research Stipend Scholarship (ERSS). We are grateful to Oliver Cliff, Cameron Zachreson and Nathan Harding for contributing to earlier versions of the open-source code of AMTraC-19. We also thank Vitali Sintchenko and Tania Sorrell for their insights on epidemiological and public health aspects of this study, as well as Tim C. Germann and Sara Del Valle for helpful discussions regarding pandemic agent-based modelling. We extend our appreciation to Aranya Ghose for his assistance in data curation facilitating generation of the artificial 2021 population. Finally, we thankfully acknowledge the use of high-performance computing cluster, Artemis, provided by the Sydney Informatics Hub at the University of Sydney.

Author contributions

MP, QDN, and SLC designed the computational experiments. QDN implemented and tested the software code and supporting algorithms. QDN and CMJ performed data curation and validation. QDN and SLC carried out the computational experiments, processed simulation data, and prepared figures. MP supervised the study. All authors drafted the original article, contributed to editing of the article, and approved the final manuscript.

Availability of data and materials

We used anonymised data from the 2016 and 2021 Australian Census obtained from the Australian Bureau of Statistics (ABS), the Australian Curriculum and Assessment and Reporting Authority (ACARA), and the Bureau of Infrastructure and Transport Research Economics (BITRE). These datasets are accessible publicly, except the travel-to-work data and household composition data which can be obtained from the ABS on request. Simulation and post-processing data are available at Zenodo [38]. The source code of AMTraC-19 is also available at Zenodo [39].

Declarations

competing interests

The authors declare no competing interests.

Author details

¹Centre for Complex Systems, Faculty of Engineering, The University of Sydney, Camperdown, NSW, Australia. ²Sydney Institute for Infectious Diseases, The University of Sydney, Westmead, NSW, Australia.

Received: 27 June 2023 Accepted: 18 October 2023

Published online: 29 October 2023

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020. Accessed on 21 March 2020. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>.
2. Rigby J, Satija B. WHO declares end to COVID global health emergency. Reuters (5 May 2023). Accessed on 06 May 2023. <https://www.reuters.com/business/healthcare-pharmaceuticals/covid-is-no-longer-global-health-emergency-who-2023-05-05/>.
3. Andrasfay T, Goldman N. Reductions in 2020 US life expectancy due to COVID-19 and the disproportionate impact on the Black and Latino populations. *Proceed Natl Academy Sci.* 2021;118(5): e2014746118.
4. Aburto JM, Schöley J, Kashnitsky I, Zhang L, Rahal C, Missov TI, Mills MC, Dowd JB, Kashyap R. Quantifying impacts of the COVID-19 pandemic through life-expectancy losses: a population-level study of 29 countries. *Int J Epidemiol.* 2022;51(1):63–74.
5. Miller IF, Becker AD, Grenfell BT, Metcalf CJE. Disease and healthcare burden of COVID-19 in the United States. *Nat Med.* 2020;26(8):1212–7.

6. Del Rio-Chanona RM, Mealy P, Pichler A, Lafond F, Farmer JD. Supply and demand shocks in the COVID-19 pandemic: an industry and occupation perspective. *Oxford Rev Econ Policy* 2020;36(Supplement_1):94–137 <https://doi.org/10.1093/oxrep/graa033>.
7. Mathieu E, Ritchie H, Rodés-Guirao L, Appel C, Giattino C, Hasell J, Macdonald B, Dattani S, Beltekian D, Ortiz-Ospina E, Roser M. Coronavirus pandemic (COVID-19). *Our World in Data*. 2020.
8. Campbell F, Archer B, Laurenson-Schafer H, Jinnai Y, Konings F, Batra N, Pavlin B, Vandemaële K, Van Kerkhove MD, Jombart T, et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Eurosurveillance*. 2021;26(24):2100509.
9. Milne GJ, Carrivick J, Whyatt D. Mitigating the SARS-CoV-2 Delta disease burden in Australia by non-pharmaceutical interventions and vaccinating children: a modelling analysis. *BMC Med*. 2022;20(1):1–13.
10. Karim SSA, Karim QA. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. *The Lancet*. 2021;398(10317):2126–8.
11. Duong BV, Larpruenrudee P, Fang T, Hossain SI, Saha SC, Gu Y, Islam MS. Is the SARS CoV-2 Omicron variant deadlier and more transmissible than Delta variant? *Int J Environ Res Public Health*. 2022;19(8):4586.
12. COVID-19 in Australia. (2021). Accessed on 15 November 2022. <https://www.covid19data.com.au/>.
13. Chang SL, Cliff OM, Zachreson C, Prokopenko M. Simulating Transmission Scenarios of the Delta Variant of SARS-CoV-2 in Australia. *Frontiers in Public Health*. 2022; 10.
14. Porter AF, Sherry N, Andersson P, Johnson SA, Duchene S, Howden BP. New rules for genomics-informed COVID-19 responses—Lessons learned from the first waves of the Omicron variant in Australia. *PLoS Genet*. 2022;18(10):1010415.
15. Chang SL, Nguyen QD, Martiniuk A, Sintchenko V, Sorrell TC, Prokopenko M. Persistence of the Omicron variant of SARS-CoV-2 in Australia: the impact of fluctuating social distancing. *PLOS Global Public Health*. 2023;3(4):0001427. <https://doi.org/10.1371/journal.pgph.0001427>.
16. Walker PG, Whittaker C, Watson OJ, Baguelin M, Winskill P, Hamlet A, Djafaara BA, Cucunubá Z, Olivera Mesa D, Green W, et al. The impact of COVID-19 and strategies for mitigation and suppression in low-and middle-income countries. *Science*. 2020;369(6502):413–22.
17. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Whittaker C, Zhu H, Berah T, Eaton JW, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*. 2020;584(7820):257–61.
18. Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia. *Nat Commun*. 2020;11(1):5710. <https://doi.org/10.1038/s41467-020-19393-6>.
19. Moore S, Hill EM, Tildesley MJ, Dyson L, Keeling MJ. Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study. *Lancet Infect Dis*. 2021;21(6):793–802.
20. Harris RJ, Hall JA, Zaidi A, Andrews NJ, Dunbar JK, Dabrera G. Impact of vaccination on household transmission of SARS-CoV-2 in England. *N Engl J Med*. 2021;385:759–60.
21. Zachreson C, Chang SL, Cliff OM, Prokopenko M. How will mass-vaccination change COVID-19 lockdown requirements in Australia? *The Lancet Regional Health—Western Pacific*. 2021; 14:100224. <https://doi.org/10.1016/j.lanwpc.2021.100224>
22. Szanyi J, Wilson T, Scott N, Blakely T. A log-odds system for waning and boosting of COVID-19 vaccine effectiveness. *Vaccine*. 2022;40(28):3821–4.
23. Viner RM, Russell SJ, Croker H, Packer J, Ward J, Stansfield C, Mytton O, Bonell C, Booy R. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. *Lancet Child Adolesc Health*. 2020;4(5):397–404.
24. Kontis V, Bennett JE, Rashid T, Parks RM, Pearson-Stuttard J, Guillot M, Asaria P, Zhou B, Battagliani M, Corsetti G, et al. Magnitude, demographics and dynamics of the effect of the first wave of the COVID-19 pandemic on all-cause mortality in 21 industrialized countries. *Nat Med*. 2020;26(12):1919–28.
25. Volk AA, Brazil KJ, Franklin-Luther P, Dane AV, Vaillancourt T. The influence of demographics and personality on COVID-19 coping in young adults. *Personal Individ Differ*. 2021;168: 110398.
26. Nguyen QD, Prokopenko M. A general framework for optimising cost-effectiveness of pandemic response under partial intervention measures. *Sci Rep*. 2022;12(1):19482. <https://doi.org/10.1038/s41598-022-23668-x>.
27. Australian Government, Department of Health and Aged Care: COVID-19 vaccination rollout update. Accessed on 15 March 2023. (2023). <https://www.health.gov.au/resources/collections/covid-19-vaccination-rollout-update>.
28. Feikin DR, Higdon MM, Abu-Raddad LJ, Andrews N, Araos R, Goldberg Y, et al. Duration of effectiveness of vaccines against SARS-CoV-2 infection and COVID-19 disease: results of a systematic review and meta-regression. *Lancet*. 2022; 399(10328):924–944 [https://doi.org/10.1016/S0140-6736\(22\)00152-0](https://doi.org/10.1016/S0140-6736(22)00152-0).
29. Lorenz MO. Methods of measuring the concentration of wealth. *Publ Am Stat Assoc*. 1905;9(70):209. <https://doi.org/10.2307/2276207>.
30. World Bank: Applications of Lorenz curves in economic analysis. Accessed on 01 December 2022. 1975. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/526831468183558343/Applications-of-Lorenz-curves-in-economic-analysis>.
31. Summary Indicators of Income and Wealth Distribution | Australian Bureau of Statistics (2022). Accessed on 09 May 2023. <https://www.abs.gov.au/statistics/detailed-methodology-information/concepts-sources-methods/survey-income-and-housing-user-guide-australia/2019-20/summary-indicators-income-and-wealth-distribution>.
32. Ceriani L, Verme P. The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. *J Econ Inequal*. 2012;10:421–43.
33. Wittebolle L, Marzorati M, Clement L, Balloi A, Daffonchio D, Heylen K, De Vos P, Verstraete W, Boon N. Initial community evenness favours functionality under selective stress. *Nature*. 2009;458(7238):623–6. <https://doi.org/10.1038/nature07840>.
34. Damgaard C, Weiner J. Describing inequality in plant size or fecundity. *Ecology*. 2000;81(4):1139–42. [https://doi.org/10.1890/0012-9658\(2000\)081\[1139:DIIPSO\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[1139:DIIPSO]2.0.CO;2).
35. Zachreson C, Fair KM, Cliff OM, Harding N, Piraveenan M, Prokopenko M. Urbanization affects peak timing, prevalence, and bimodality of influenza pandemics in Australia: results of a census-calibrated model. *Sci Adv*. 2018;4(12):5294. <https://doi.org/10.1126/sciadv.aau5294>.
36. Australian Bureau of Statistics: Data confidentiality guide. Accessed on 09 May 2023. 2021. <https://www.abs.gov.au/about/data-services/data-confidentiality-guide>
37. Fair KM, Zachreson C, Prokopenko M. Creating a surrogate commuter network from Australian Bureau of Statistics census data. *Sci Data*. 2019;6(1):150. <https://doi.org/10.1038/s41597-019-0137-z>.
38. Nguyen QD, Chang SL, Jamerlan CM, Prokopenko M. AMTraC-19 (v8.0) Dataset: Systematic Comparison of the COVID-19 Pandemic Scenarios. <https://doi.org/10.5281/zenodo.8067859>
39. Chang SL, Nguyen QD, Zachreson C, Cliff OM, Prokopenko M. AMTraC-19 Source Code: Agent-based Model of Transmission and Control of the COVID-19 pandemic in Australia. Zenodo. <https://doi.org/10.5281/zenodo.5778217>.
40. Australian Bureau of Statistics. Accessed on 01 November 2022. <https://www.abs.gov.au/>.
41. Australian Bureau of Statistics: Census of Population and Housing. Accessed on 01 December 2022. <https://tablebuilder.abs.gov.au/>.
42. Dye C. Health and urban living. *Science*. 2008;319(5864):766–9. <https://doi.org/10.1126/science.1150198>.
43. Eubank S, Guclu H, Anil Kumar VS, Marathe MV, Srinivasan A, Toroczkai Z, Wang N. Modelling disease outbreaks in realistic urban social networks. *Nature*. 2004;429(6988):180–4. <https://doi.org/10.1038/nature02541>.
44. Yashima K, Sasaki A. Epidemic process over the commute network in a metropolitan area. *PLoS ONE*. 2014; 9(6). <https://doi.org/10.1371/journal.pone.0098518>.
45. Brueckner JK. Airline traffic and urban economic development. *Urban Stud*. 2003;40(8):1455–69. <https://doi.org/10.1080/0042098032000094388>.
46. Bureau of infrastructure and transport research economics: Airport traffic data. Accessed on 01 December 2022. 2022. <https://data.gov.au/dataset/ds-dga-cc5d888f-5850-47f3-815d-08289b22f5a8/details>.
47. Cliff OM, Harding N, Piraveenan M, Erten EY, Gambhir M, Prokopenko M. Investigating spatiotemporal dynamics and synchrony of influenza epidemics in Australia: an agent-based modelling approach. *Simul Model Pract Theory*. 2018; 87, 412–431 <https://doi.org/10.1016/j.simpat.2018.07.005>.

48. Australian Curriculum, Assessment and Authority (ACARA): School profiles and locations. Accessed on 01 December 2022. <https://acara.edu.au/contact-us/acara-data-access>.
49. Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, Gallagher E, Gower C, Kall M, Groves N, O'Connell A-M, Simons D, Blomquist PB, Zaidi A, Nash S, Iwani Binti Abdul Aziz N, Thelwall S, Dabrera G, Myers R, Amirthalingam G, Gharbia S, Barrett JC, Elson R, Ladhani SN, Ferguson N, Zambon M, Campbell CNJ, Brown K, Hopkins S, Chand M, Ramsay M, Lopez Bernal J. COVID-19 vaccine effectiveness against the Omicron (B.1.1.529) variant. *N Engl J Med.* 2022; 386(16), 1532–1546
50. Buchan SA, Chung H, Brown KA, Austin PC, Fell DB, Gubbay JB, Nasreen S, Schwartz KL, Sundaram ME, Tadrous M, Wilson K, Wilson SE, Kwong JC. Estimated effectiveness of COVID-19 vaccines against Omicron or delta symptomatic infection and severe outcomes. *JAMA Netw Open.* 2022;5(9):2232760–2232760.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

