

Forecasting the burden of type 2 diabetes in Singapore using a demographic epidemiological model of Singapore

Thao P Phan,¹ Leontine Alkema,^{1,2} E Shyong Tai,^{1,3,4} Kristin H X Tan,¹ Qian Yang,¹ Wei-Yen Lim,^{1,5} Yik Ying Teo,^{1,2,6} Ching-Yu Cheng,^{1,7,8,9} Xu Wang,¹ Tien Yin Wong,^{7,8} Kee Seng Chia,¹ Alex R Cook^{1,2,10,11}

To cite: Phan TP, Alkema L, Tai ES, *et al.* Forecasting the burden of type 2 diabetes in Singapore using a demographic epidemiological model of Singapore. *BMJ Open Diabetes Research and Care* 2014;**2**:e000012. doi:10.1136/bmjdr-2013-000012

Received 15 December 2013
Revised 20 May 2014
Accepted 24 May 2014

ABSTRACT

Objective: Singapore is a microcosm of Asia as a whole, and its rapidly ageing, increasingly sedentary population heralds the chronic health problems other Asian countries are starting to face and will likely face in the decades ahead. Forecasting the changing burden of chronic diseases such as type 2 diabetes in Singapore is vital to plan the resources needed and motivate preventive efforts.

Methods: This paper describes an individual-level simulation model that uses evidence synthesis from multiple data streams—national statistics, national health surveys, and four cohort studies, and known risk factors—aging, obesity, ethnicity, and genetics—to forecast the prevalence of type 2 diabetes in Singapore. This comprises submodels for mortality, fertility, migration, body mass index trajectories, genetics, and workforce participation, parameterized using Markov chain Monte Carlo methods, and permits forecasts by ethnicity and employment status.

Results: We forecast that the obesity prevalence will quadruple from 4.3% in 1990 to 15.9% in 2050, while the prevalence of type 2 diabetes (diagnosed and undiagnosed) among Singapore adults aged 18–69 will double from 7.3% in 1990 to 15% in 2050, that ethnic Indians and Malays will bear a disproportionate burden compared with the Chinese majority, and that the number of patients with diabetes in the workforce will grow markedly.

Conclusions: If the recent rise in obesity prevalence continues, the lifetime risk of type 2 diabetes in Singapore will be one in two by 2050 with concomitant implications for greater healthcare expenditure, productivity losses, and the targeting of health promotion programmes.

Key messages

- Asians in general, and Singaporeans in particular, are increasingly at risk of diseases such as type 2 diabetes that are associated with modern, high calorie, sedentary lifestyles.
- Forecasts of the prevalence of diseases such as type 2 diabetes require models that quantify and predict the changing dynamics of the drivers of the epidemic, including population age structure and evolving obesity levels.
- Using evidence synthesis, Bayesian inference, and individual-based modeling, we have developed forecasts of type 2 diabetes prevalence and incidence to 2050 in different segments of the population of Singapore, with ethnic minorities bearing a disproportionate burden, a marked rise in the incidence of type 2 diabetes in the workforce, and a rise in the lifetime risk to one in two.
- *It's tough to make predictions, especially about the future.*—Yogi Berra

energy-dense foods and reduced physical activity. Examples abound: rates of T2DM are 1.3–4.8 times higher among American Asians and Pacific Islanders than in Europeans living in the Americas,^{3–6} 2–3 times higher among American Japanese than in Japanese living in Japan,^{7–9} and 1.9–6 times higher among South Asians versus Europeans living in Europe.^{10 11} India (51 million) and China (43 million) already have more people with type 2 diabetes than the USA (27 million),¹² but as lifestyles and diets in rapidly developing Asia become increasingly urbanized, it therefore must be expected that the burden of T2DM will continue to grow in the most populous continent.

Singapore is a microcosm of Asia. Three broad ethnicities, corresponding to the three major population centers in Asia, are represented in the city-state: East Asians, in the Chinese majority, South East Asians, via the Malay, and South Asians of mostly Indian



CrossMark

For numbered affiliations see end of article.

Correspondence to

Dr Alex R Cook;
alex.richard.cook@gmail.com

INTRODUCTION

Type 2 diabetes mellitus (T2DM) looms large over Asia. Asians, especially South Asians, are predisposed toward T2DM to a greater extent than ethnic Europeans.^{1 2} At even greater risk are ethnic Asians living in Europe or the Americas, where this predisposition is accentuated by the adoption of modern, urban lifestyles rich in processed,

and Sri Lankan descent. Over the past few decades, these groups have been exposed to significant changes in lifestyle, diet, and other environmental influences that are typical of a high-income society, changes that are reflected in the doubling of the prevalence of T2DM from 5% in the 1980s¹³ to 11% in 2010.¹⁴ Rapidly ageing, increasingly sedentary, Singapore pre-sages the problems other Asian countries will face in the decades ahead.

Since T2DM is one of many competing public health issues that will accompany the ageing of Singapore, as in Asia, it is vital to be able to forecast the future burden of T2DM to facilitate rational planning of public health campaigns. To predict involves positing a model that encapsulates epidemiological and medical subject-area expertise on the main drivers of T2DM at the individual and population levels. Rigorous subsequent parameterization of the model ensures its relevance to the population to which it is applied. The degree of complexity of the model depends on the objective of the analysis and the data available: neither too simplistic, lest it fail in extrapolation to scenarios it was not validated for; nor more complex than is needed to meet those objectives. Methods used in other settings to forecast the evolving burden of chronic diseases include microsimulation models positing assumptions on future obesity and physical activity trends,¹⁵ extrapolating linear regressions of the prevalence of overweight and adjusting geographic distribution using deprivation indices,¹⁶ forecasting the changing demography of a country with or without increasing incidence,^{17–20} and modeling body mass index (BMI) and its impact on development of T2DM and related complications.^{21–22}

One of the most challenging issues in developing a model for a future public health phenomenon is that the health of a population is never in a stable equilibrium. Although the observed rise in T2DM prevalence from 8.6% (95% CI 7.7% to 9.6%) in 1992 to 11.3% (95% CI 10.3% to 12.3%) in 2010 in Singapore can be attributed to ageing, as the age-specific prevalence has remained relatively static since the 1990s,¹⁴ it would be misleading to forecast the future prevalence of T2DM by applying the historic age-specific prevalence of T2DM to a projected age distribution at some future time point—for the age-specific prevalence of *obesity*, and overweight, another important risk factor of T2DM,²³ have risen substantially in most demographic segments over this time period.²⁴ This rise foreshadows an increase in the age-specific prevalence of T2DM, as the increasingly obese young of today become the increasingly diabetic old of tomorrow. Predictions must hence incorporate ageing and secular trends in obesity, reflecting changes in diet and physical activity, as otherwise they may severely understate the future burden. Furthermore, evidence of a genetic contribution to T2DM from familial aggregation (the risk of T2DM increases twofold to fivefold for individuals having a family history,^{25–26} while heritability of T2DM—the proportion of phenotypic variance

attributed to genetic factors—has been estimated at approximately 26% in a Danish population-based twin study²⁷) suggests that the effect of genetics should also be incorporated.

This paper describes a demographic, epidemiological model of Singapore and its use in forecasting the total prevalence of T2DM (diagnosed and undiagnosed) to 2050. The model is an individual-based model which represents each resident in the city-state, past (from 1990) and future (to 2050), thereby facilitating the incorporation of obesity trends, both secular and over an individual's life span. The model incorporates demographic processes including the mass migration Singapore has experienced over the past two decades, submodels for the evolution of each individual's yearly BMI and genetic risk of T2DM, and a T2DM onset submodel, and data from national statistics, nationally representative cross-sectional surveys, longitudinal studies, molecular epidemiological cohort studies, and the literature, analyzed using Bayesian statistical methods.

METHODS

The model contains submodels as depicted in [figure 1](#) and summarized below. Mathematical details are provided in the online supplementary methods.

Demographic model

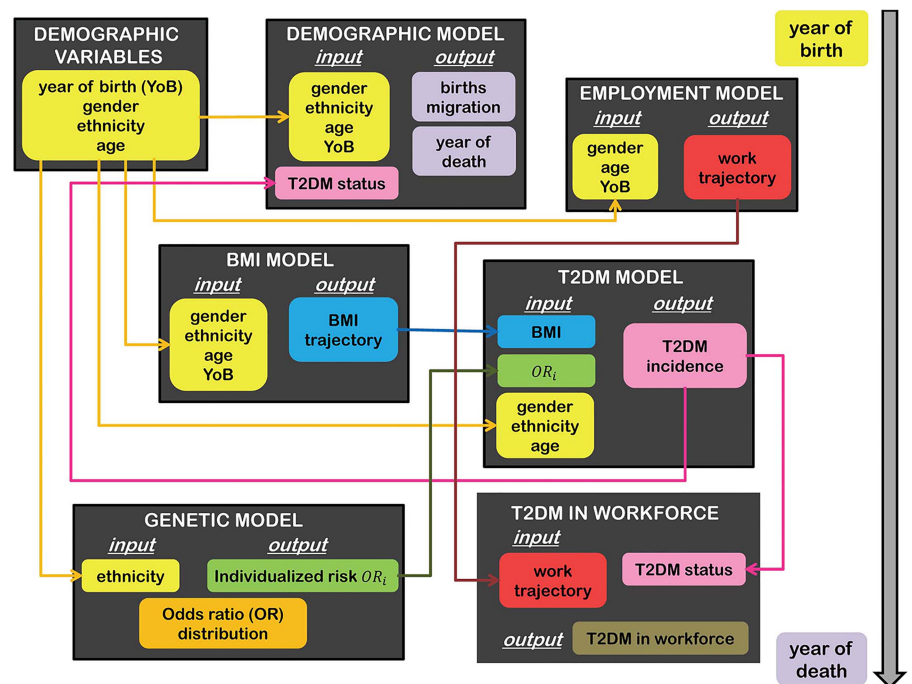
The model is incremented in units of 1 year and tracks the resident population of Singapore from 1990 to 2050. Individuals die according to mortality rates that vary by age, year of birth, gender, and for the three main ethnicities of Singapore—Chinese, Indian, Malay—and a fourth category aggregating others (mostly of mixed ethnicity; other South East Asians; and Europeans²⁸).

The mortality rate is parameterized as a smooth spline function stratified by gender with proportional hazards for other effects, including T2DM status.

Fertility rates differ for each age, year of birth, and ethnicity, with ethnicity assumed to be inherited maternally. The fertility rate is modeled as a Gaussian function with parameters that are functions of demographic factors.

Migration (outward and, especially, inward: the population grew from 3 million in 1990 to 5.1 million in 2010) is represented by a baseline migrant age profile curve, a spline curve stratified by ethnicity and gender, with a random effect applied to each year to reflect the economic situation and government policy changes. The parameters of these models are estimated from national statistics released by the Singapore Department of Statistics, in particular the 1990, 2000, and 2010 censuses of population, the annual yearbook of statistics, which conveyed information on the size, age structure, gender, and ethnic composition over time, and life tables by gender and fertility and mortality rates. These rates and statistics (except the censuses and life tables) had for the most part a resolution only to 5-year age bands—coarser for older ages—and no information on

Figure 1 Overview of model structure. Boxes represent submodels; arrows indicate direction of information flow between submodels. BMI, body mass index; T2DM, type 2 diabetes mellitus.



migration. Crude birth rates and death rates by ethnicity during the period 1990–2010 were obtained from the Report on the Registration of Births and Deaths.

A three-state Markov model describes how the resident population moves between work, unemployment, and out of the workforce, conservatively assuming no correlation between T2DM and workforce participation, and no changes in retirement ages. In this, weekly transition probabilities between states vary by age, calendar year, and gender, and are estimated from the annual Report on the Labour Force in Singapore, which provides data on the resident population by 5-year age groups, gender, and workforce status, and unemployed resident population by 10-year age groups and duration of unemployment (in weeks), and the General Household Survey in 1995 and 2005, which provide data by 5-year age groups, gender, and workforce status. We obtained estimates of weekly age-dependent transition probabilities via Markov chain Monte Carlo (MCMC) methods.^{29 30}

BMI model and data

We developed an individual, hierarchical model of BMI trajectories over an adult life span, stratified by gender and ethnicity. In this, an individual's BMI over time is described by Gaussian fluctuations around a sequence of connected lines, with joints at age 35, 55, and 75. Each individual has a different starting BMI (at age 18) and three BMI gradients, which are assumed to come from a multivariate normal distribution whose hyperparametric mean and covariance are common to all individuals of that gender and ethnic group. These hyperparameters are estimated using (1) longitudinal data from the Singapore Prospective Study programme (SP2), which contains BMI measurements and T2DM

status at two of three time points (the 1992 or 1998 National Health Survey (NHS) and a follow-up visit around 2005), and (2) aggregate data from the 2004 and 2010 NHSs on the proportions of four BMI categories (underweight, normal weight, overweight, and obese) within age bands (18–29, 30–39, 40–49, 50–59, 60–69) and gender/ethnicity groups.

Genetic risk factor model and data

From a combined list of 44 single-nucleotide polymorphisms (SNPs) previously reported to be associated with T2DM in an Asian population,^{31–33} association analysis between the SNPs and T2DM was performed using additive logistic regression on SNPTEST software³⁴ in three cohort studies—the Singapore Chinese Eye Study (SCES, with 302 people with type 2 diabetes and 1089 without), the Singapore Malay Eye Study (SiMES, with 794 patients with type 2 diabetes and 1420 non-diabetes), and the Singapore Indian Eye Study (SINDI, with 977 people with type 2 diabetes and 1169 without). Fourteen SNPs were collectively selected for the p value threshold of 0.05 in at least one cohort study representing one major ethnic group in Singapore (see online supplementary table S1). To account for heterogeneous genetic risks within ethnic groups, the joint distribution of 14 SNPs associated with T2DM in each of the three main ethnicities of Singapore was determined from the corresponding cohort study. Assuming representativeness of the cohorts and no gender bias in the distributions of the associated risk alleles, the frequencies of all 16 384 allele combinations of 14 SNPs were determined within these groups. A point estimate of the odds ratio (OR) for each SNP from a meta-analysis³⁵ was combined with these frequencies to determine the distribution of

ORs for T2DM for each ethnicity. As the distribution of ORs conditional on ethnicity was approximately log normal, we derived its mean and SD by weighing log ORs with associated allele frequencies. To prevent double counting the effect of ethnicity on T2DM incidence and genetic risks, we standardized the ORs such that the weighted mean OR within each ethnic group was 1 (see online supplementary figure S1). In the simulation model, for people belonging to the three main ethnicities, individuals' genetic risks, which were modeled to be conditionally independent of their BMI trajectories given ethnicity, were selected randomly from the appropriate distribution of ORs. For people belonging to other ethnic groups, the distribution of ORs of the Chinese majority was used.

Type 2 diabetes onset model and data

Using the same longitudinal data as in the BMI model, we generated a single putative BMI trajectory that matches the observed data for each individual using importance sampling. This was then used together with age, gender, and ethnic group within a logistic model for T2DM incidence. The cumulative probability of developing T2DM between the two observation times was derived by summation and used to generate the likelihood function, which permitted estimation using MCMC. In the simulation model, the probability of progressing from a non-diabetic state to T2DM was calculated annually conditional on the individual's demographics and BMI and genetic risk, with the effects assumed to operate multiplicatively in the ORs.

All participants provided written informed consent.

Sensitivity analysis

We also developed a model in which BMI and genetics were excluded as risk factors and the risk of getting T2DM was a function of age, ethnicity, and gender only. This model is described in the online supplementary methods.

Software

All statistical analyses were performed in R V3.0.0³⁶ or JAGS V3.1.0^{37 38} using customized scripts that took around 24 h to run on a desktop for each model and demographic group. All graphics were created using the grid package.³⁹ Simulations were run in C++ with individuals represented as objects, linked to their mothers, with attached static and dynamic variables. The simulation was initialized using the demographic structure described in the 1990 census, with individuals added to the population when their mothers gave birth or when they immigrated to Singapore. Multiple runs using different random number seeds and parameters, drawn from the posterior distribution to account for parametric uncertainty, were used to build up a Monte Carlo sample, with each simulated population queried to output characteristics, such as the number of people with type 2 diabetes within any age range at any time.

The C++ code was compiled using the GCC compiler, and runs, covering the time horizon 1990–2050 and around 6.25 million individuals, took an average of 3 min for one whole run.

RESULTS

Incidence of type 2 diabetes

Incidence rates were estimated and projected from the fitted model by extracting new, potentially undiagnosed cases of T2DM among various demographic segments. Crude incidence rates, past and future, are tabulated in [table 1](#) by gender and ethnicity. Incidence rates are expected to double over the period 1990–2050 for all the demographic groupings considered. Among the Chinese, the incidence is expected to rise from 5 (95% prediction interval 4–5) per 1000 woman-years to 9 (7–10), or 6 (5–6) per 1000 man-years to 12 (10–13), over these six decades. For Malays, the rise is steeper (7 (6–8) to 14 (13–16) among women or to 17 (15–18) in men), and for Indians, steeper still, with an annual incidence of 17 (16–19) to 19 (17–21) per 1000-person years by 2050, from 8 (7–10) to 10 (9–12) in the 1990s.

Prevalence of type 2 diabetes

The total prevalence of T2DM (diagnosed and undiagnosed) among Singapore adults (age 18–69) is projected to rise from 7.3% (6.8–8%) in 1990 to 15% (13.8–16.2%) in 2050 ([figure 3B](#)). Modeled past and projected future age-specific prevalence rates are depicted in [figure 2](#). The prevalence was generally markedly higher in Indians and Malays than Chinese Singaporeans, with Malays and Indians having a risk profile roughly the same as a Chinese 10 years their senior. Although prevalence among female Chinese sexagenarians is projected to stay relatively constant, rates are expected to grow substantially in other groups: by 2050, we expect 35% (31–39%) of Chinese men aged 60–69 having T2DM, and around half of the Malays and Indians of that age group ([figure 2](#)). A moderate risk in prevalence among young adults is forecast (see online supplementary table S2).

Age and overweight

The projected rise in the total prevalence of T2DM in Singapore is driven by two factors: the modeled ageing and fattening of the population. The age pyramid ([figure 3D–G](#)) is predicted to become increasingly top heavy, with the proportion of the population under age 20 falling from 25.2% (2010) to 15.9% (2050), and the proportion over the age of 60 soaring from 13.3% to 31.9% over the same time frame. The effect of this rise in the prevalence of the main risk factor (advanced years) is compounded by a dramatic rise in obesity and overweight levels. The fraction of the population that is obese is predicted to quadruple from 4.3% in 1990 to 15.9% by 2050, while those overweight are projected to expand in number from 24.6% in 1990 to 38.6% by 2050 ([figure 3A](#)). This projected increase in BMI at the

Table 1 Modeled and forecast crude type 2 diabetes incidence rates, per 1000 person-years

Period	Chinese		Indian		Malay		Total	
	Female	Male	Female	Male	Female	Male	Female	Male
1990–2000	5 (4–5)	6 (5–6)	8 (7–10)	10 (9–12)	7 (6–8)	7 (6–8)	5 (5–6)	6 (5–7)
2000–2010	5 (5–6)	7 (6–8)	11 (9–12)	13 (11–15)	8 (7–10)	8 (7–10)	6 (6–7)	8 (7–8)
2010–2020	6 (5–7)	8 (7–9)	11 (10–13)	13 (11–15)	10 (9–12)	11 (9–12)	7 (6–8)	9 (8–10)
2020–2030	7 (6–8)	9 (8–11)	14 (12–15)	15 (13–17)	12 (11–13)	13 (11–14)	8 (7–9)	10 (9–12)
2030–2040	8 (7–9)	11 (9–12)	16 (14–18)	17 (15–20)	13 (12–15)	15 (13–16)	9 (8–10)	12 (11–13)
2040–2050	9 (7–10)	12 (10–13)	17 (16–19)	19 (17–21)	14 (13–16)	17 (15–18)	10 (9–11)	13 (12–14)

Numbers in parentheses are 95% prediction intervals.

population level can be attributed to all subgroups (see online supplementary figure S3–S7). The forecast rise in obesity levels is most stark for Malays and Indians (hitting 40% among Malay women aged over 40), but the large Chinese majority is also expected to see a rise in obesity levels of around 10 percentage points (see online supplementary table S3).

The confluence of these factors will, if the projections hold true, lead to a rise in the number of those in the workforce living with T2DM, a proxy for the impact of T2DM on productivity and corporate health insurance plans, from 97 600 (89 800–106 100) in 1990 to 321 600 (293 000–353 700) by 2050 (figure 3C). The type 2 diabetic population is predicted to increase from 358 500 (333 900–386 100) in 2010 to 673 200 (624 700–727 400) in 2030 and to 909 300 (839 700–986 900) in 2050.

Model validation

Demographic structure

The model is seeded with the 1990 census. It reproduces the 2010 census very accurately (figure 3E), save for a slight underprediction of the number of women aged 25–40, which we attribute to migration.

BMI trajectories

The distribution of each pair of BMI observations for the SP2 participants agrees well with the posterior predictive distribution of trajectories within each ethnic group and gender demographic segment (see online supplementary figure S2).

Prevalence of type 2 diabetes

The modeled overall proportion of patients with type 2 diabetes closely corresponds to results of the NHSs, except for the outlying 2004 survey (figure 3B). It is not known why the 2004 NHS is so discrepant from the other NHSs. Prevalence of T2DM within age, gender, and ethnic groups is similar between the model and data (figure 2), though the small sample sizes on stratification lead to unstable empirical estimates with broad uncertainty intervals, so the concordance is not perfect.

Sensitivity analysis

We also developed a simpler model for T2DM that does not take into account BMI changes and genetic effects, with population ageing being the main factor contributing to the increase in prevalence of T2DM in this model. Consequently, projected T2DM prevalence among Singaporean adults aged 18–69 by 2050 for the simpler model is 11.8% (11.2–12.6%), lower than in the full model (15%) even though the overall historic prevalence of the two models is quite close to each other (7.1% for the reduced model and 7.3% for the full model in 1990; see online supplementary figure S9). The reduced model for T2DM assumes that the effect of age on T2DM risk is constant over time. As a result, the lifetime risk for Singaporean adults aged 18–69 for this model does not change much over time, from 38.9%

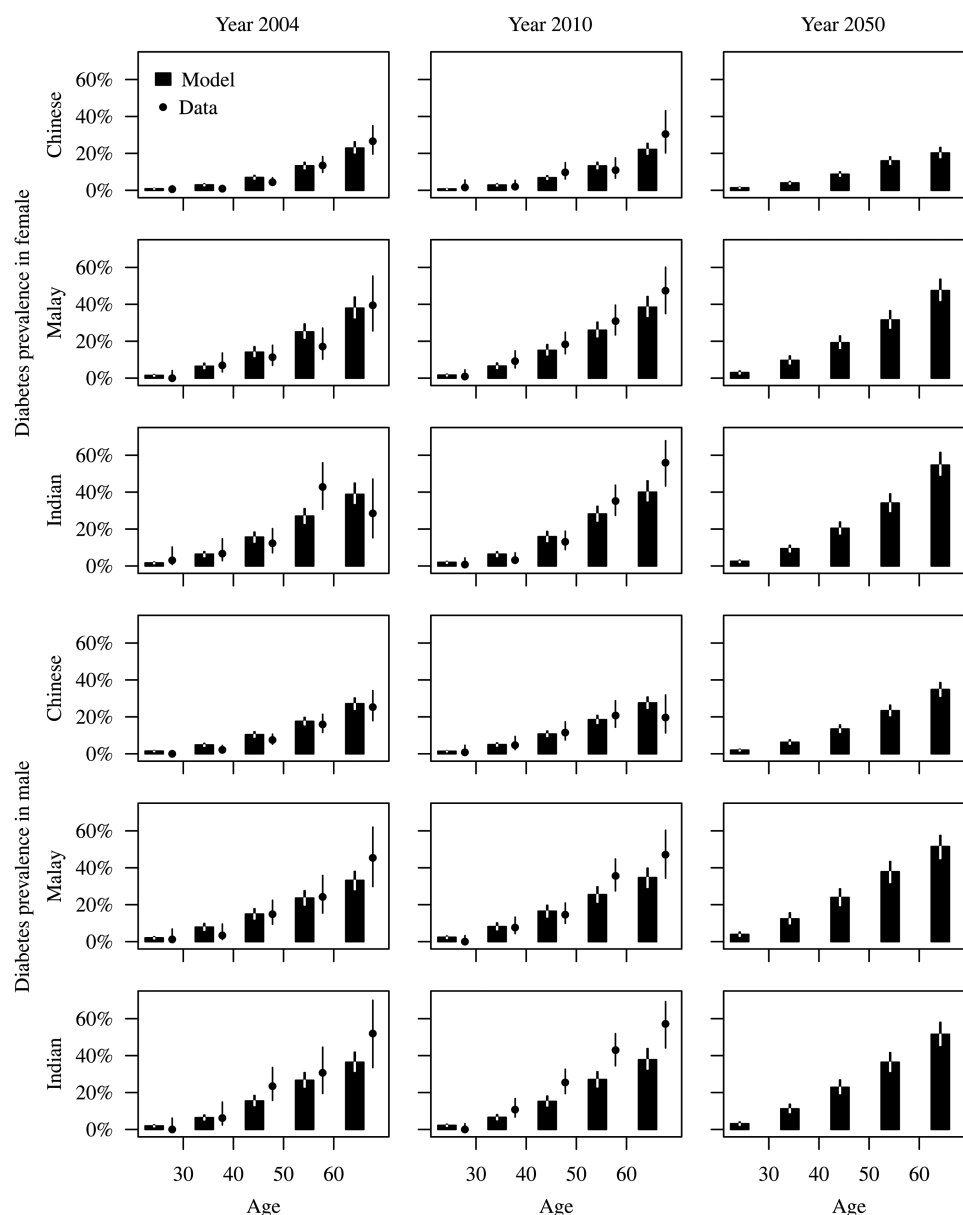


Figure 2 Age-specific, gender-specific, and ethnicity-specific prevalence estimates and forecasts of (diagnosed and undiagnosed) type 2 diabetes. Model forecasts are presented as bars with 95% prediction intervals. Data are indicated by dots with 95% empirical CIs.

(36.3–41.9%) in 1990 to 39.2% (36.9–42.5%) in 2050. In the full model, lifetime risk for T2DM for Singaporean adults aged 18–69 is projected to rise from 34.5% (31.9–38.2%) in 1990 to 43.8% (40.8–47.5%) in 2050 as the increasing BMI trend is accounted for. An interesting observation is the gender difference in projected lifetime risk of T2DM in the two models (see online supplementary table S4). For the reduced model, women would have a marginally higher lifetime risk than men (39.9% vs 38.4% in 2050). In the full model, however, women are forecast to have a lower lifetime risk than men (lifetime risks of T2DM by 2050 are 37% in women and 51% in men). This is due to Chinese women, the largest group of women in Singapore, not experiencing the same rise in overweight as did men and other

women, so that simple forecasts based on current age prevalence would substantially underestimate future prevalence of T2DM in all groups other than Chinese women.

DISCUSSION

Modeling provides a way to explore what-if scenarios quickly and cost effectively. In this paper, we use modeling to answer the question: If the recent rise in obesity levels in Singapore were maintained, what would the effect on the prevalence of T2DM be one generation from now? The answer is worrying: a rise in the overall prevalence from 1 in 13 to around 1 in 6 working age adults, a lifetime risk of around 1 in 2, and an increasing

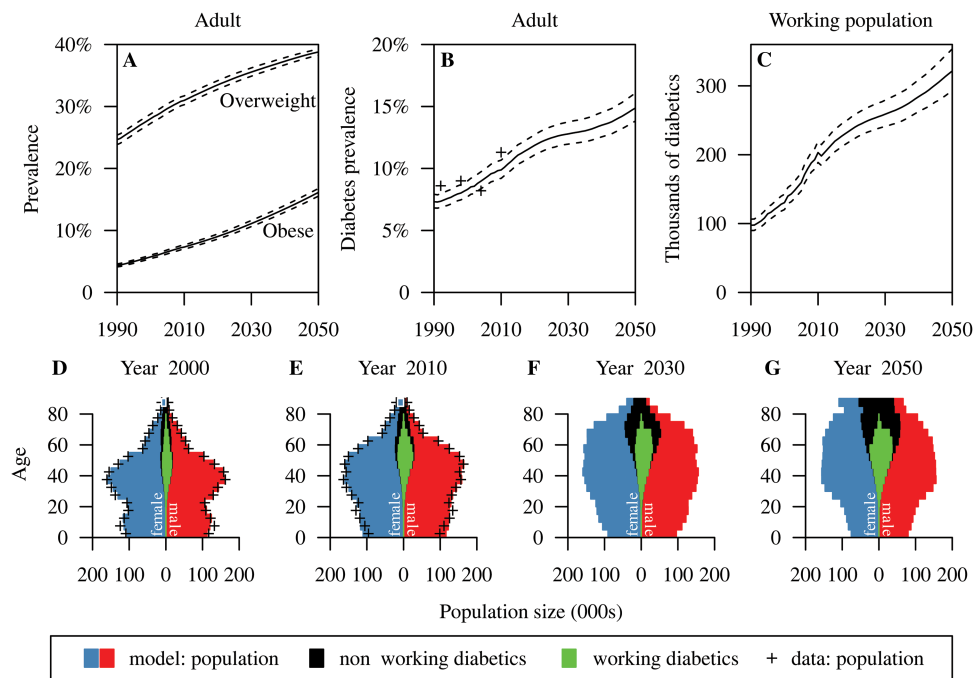


Figure 3 Obesity and type 2 diabetes forecasts. Top: forecast prevalence of obesity and overweight in adults (A), forecast prevalence of type 2 diabetes among working age adults (B) and number of patients with type 2 diabetes in the workforce (C). Means and 95% prediction intervals are plotted. For prevalence, point estimates from the National Health Surveys are overlaid. Bottom (D–G): modeled age pyramids with patients with type 2 diabetes and diabetic workers overlaid. Red and blue bars indicate women and men, respectively; black bars indicate patients with type 2 diabetes (not in the workforce) of both genders; and green bars indicate working diabetics. The + symbol indicates data from the censuses of 2000 and 2010.

burden of T2DM in the workplace. T2DM has been estimated to reduce a worker's productivity by around a third in the USA,⁴⁰ due to disability, premature mortality, early retirement, and absenteeism, in that order,⁴¹ while in Canada, those with T2DM were found to be between 150% and thrice as likely not to be in the labor force, and to have an income approximately 25% lower than non-diabetics.⁴² Employers in Singapore will have to decide whether they should take responsibility for preventive action, such as screening or weight loss programmes, to mitigate future losses.

Singapore is an ideal test bed for public health research in Asia. Not only does its Chinese, Indian, and Malay population make it a miniature of Asia as a whole, but other countries in Asia are likely to look increasingly like Singapore, as they become increasingly developed, urbanized, sedentary, and aged. The current prevalence of T2DM in populations comparable genetically and culturally, but at an earlier stage of development, is markedly lower (in mainland Chinese sexagenarians 19%⁴³ vs 25% in Singapore Chinese, in elderly Malays in Malaysia 21%⁴⁴ vs 37%), boding ill for the future elsewhere in Asia.

All modeling studies make some degree of simplifying assumptions. In this study, the risk of developing T2DM is determined by demographic factors, a secular trend, genetics, and current BMI, as a proxy for overweight and general ill health. The model averages over other factors that have a role include epigenetics,⁴⁵ physical

activity,⁴⁶ diet,⁴⁷ family history,⁴⁸ socioeconomic status, and pregnancy.⁴⁹ The formulation as an individual-level model allows observed variability between individuals to be characterized, along with risk factors that vary dynamically over lifetimes. The genetic risk model includes 14 SNPs that are significantly associated with T2DM in the Singapore population and assumes that these have an effect independent of BMI, as none of the 14 SNPs has been reported to be associated with overweight/obesity in the Singapore setting.⁵⁰ Future work should verify this assumption and might incorporate the effect of genetic factors on BMI, for which twin studies indicate that estimated BMI heritability is 47–90%.⁵¹ The primary data source on weight that was available to us was aggregate statistics from the NHS, which on stratification led to small demographic segments with substantial sampling variability. The model we used pools information between age groupings and from cohort studies, which we believe yields more reliable estimates. Even more reliable estimates of change might result from mining medical records, which would permit relaxation of the distributional assumptions used herein.⁵² The data used to parameterize the models of BMI and T2DM were from cohort studies based on nationally representative samples of adults, overcoming the common difficulty in generalizing from cohort studies to the general population, though this means that BMI trajectories in childhood and adolescence were not modeled. With data on childhood obesity, the prevalence of which has risen

globally in recent decades,⁵³ we would have been able to capture any recent changes in this critical age period. Future work should address this data paucity. The model assumes no interaction between overweight/T2DM and workforce participation except interactions mediated by demographics, and research is needed in the Singapore setting to elucidate whether any additional interactions are present. The T2DM model does not incorporate prediabetes, an intermediate state of T2DM when blood glucose level shows abnormalities—impaired fasting glucose or impaired glucose tolerance—but does not exceed the threshold determining T2DM. Introducing a prediabetic state into the T2DM model would stratify the non-diabetic population into low-risk and high-risk groups, enhancing the capability of the model for possible intervention evaluation. This would, however, require additional information on reversion rates from a prediabetic state to a normal state or on progression rates from a prediabetic state to T2DM in the Singapore context.

The long-term goal of this modeling project is to bring together three models: the present one, which projects the *prevalence* of T2DM in different subpopulations, a model of *outcomes*—from more complications from macrovascular diseases (eg, cardiovascular disease) and microvascular diseases (eg, kidney, nerve, and eye diseases), to healthcare expenditure and workplace absenteeism—and a model of *interventions*, such as healthy eating or active lifestyle programmes. Taken together, these would allow the effectiveness, and cost effectiveness, of different health promotion interventions to be assessed *in silico* to enhance the evidence base of public health decision making by determining how much of a reduction to levels of overweight and obesity would be needed to substantially reduce the burden of T2DM, and how much can realistically be achieved by health promotion campaigns.

Author affiliations

¹Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore

²Department of Statistics and Applied Probability, Faculty of Science, National University of Singapore, Singapore

³Division of Endocrinology, National University Hospital and National University Health System, Singapore

⁴Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore

⁵Ministry of Health, Singapore

⁶Life Sciences Institute, National University of Singapore, Singapore

⁷Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore

⁸Singapore Eye Research Institute, Singapore

⁹Center for Quantitative Medicine, Duke-NUS Graduate Medical School, Singapore

¹⁰Program in Health Services and Systems Research, Duke-NUS Graduate Medical School, Singapore

¹¹Yale-NUS College, Singapore

Contributors TPP developed and programmed the model, performed statistical analysis, and wrote the paper. LA developed the model, provided demographic expertise, and wrote the paper. EST provided endocrinological expertise and wrote the paper. KHXT performed statistical analysis and wrote

the paper. QY performed statistical analysis and wrote the paper. WYL provided epidemiological expertise and wrote the paper. YYT, CYC, XW, and TYW contributed to genetic analysis and wrote the paper. KSC conceived the study, provided public health expertise, and wrote the paper. ARC conceived the study, supervised the study, developed the model, and wrote the paper. He is also the guarantor of this paper.

Funding This work was supported by the Population Health Metrics and Analytics project, funded by the Ministry of Health, Singapore. ARC is also grateful for funding from the NUS Initiative to Improve Health in Asia. Grant number NIHA-2011-1-004.

Competing interests None.

Ethics approval The SP2, SCES, SINDI, and SIMES studies were approved by the Singapore Health Services Centralised Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

1. Chan JC, Malik V, Jia W, *et al.* Diabetes in Asia: epidemiology, risk factors, and pathophysiology. *JAMA* 2009;301:2129–40.
2. Ramachandran A, Snehalatha C, Shetty AS, *et al.* Trends in prevalence of diabetes in Asian countries. *World J Diabetes* 2012;3:110–17.
3. Gupta LS, Wu CC, Young S, *et al.* Prevalence of diabetes in New York City, 2002–2008: comparing foreign-born South Asians and other Asians with U.S.-born whites, blacks, and Hispanics. *Diabetes Care* 2011;34:1791–3.
4. McNeely MJ, Boyko EJ. Type 2 diabetes prevalence in Asian Americans: results of a national health survey. *Diabetes Care* 2004;27:66–9.
5. Mohanty SA, Woolhandler S, Himmelstein DU, *et al.* Diabetes and cardiovascular disease among Asian Indians in the United States. *J Gen Intern Med* 2005;20:474–8.
6. Lee JW, Brancati FL, Yeh HC. Trends in the prevalence of type 2 diabetes in Asians versus whites: results from the United States National Health Interview Survey, 1997–2008. *Diabetes Care* 2011;34:353–7.
7. Fujimoto WY, Leonetti DL, Kinyoun JL, *et al.* Prevalence of diabetes mellitus and impaired glucose tolerance among second-generation Japanese-American men. *Diabetes* 1987;36:721–9.
8. Nakanishi S, Okubo M, Yoneda M, *et al.* A comparison between Japanese-Americans living in Hawaii and Los Angeles and native Japanese: the impact of lifestyle westernization on diabetes mellitus. *Biomed Pharmacother* 2004;58:571–7.
9. Franco LJ. Diabetes in Japanese-Brazilians—influence of the acculturation process. *Diabetes Res Clin Pract* 1996;34(Suppl): S51–7.
10. Mather HM, Keen H. The Southall Diabetes Survey: prevalence of known diabetes in Asians and Europeans. *BMJ (Clin Res Ed)* 1985;291:1081–4.
11. Jenum AK, Holme I, Graff-Iversen S, *et al.* Ethnicity and sex are strong determinants of diabetes in an urban Western society: implications for prevention. *Diabetologia* 2005;48:435–9.
12. Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Res Clin Pract* 2010;87:4–14.
13. Thai AC, Yeo PP, Lun KC, *et al.* Diabetes mellitus and its chronic complications in Singapore: an increasing healthcare problem. *Ann Acad Med Singapore* 1990;19:517–23.
14. Ministry of Health, Singapore, Epidemiology and Disease Control Division. *National Health Survey 2010*. Singapore: Government of Singapore, 2011.
15. Shi L, van Meijgaard J, Fielding J. Forecasting diabetes prevalence in California: a microsimulation. *Prev Chronic Dis* 2011;8:A80.
16. Holman N, Forouhi NG, Goyder E, *et al.* The Association of Public Health Observatories (APHO) diabetes prevalence model: estimates

- of total diabetes prevalence for England, 2010-2030. *Diabet Med* 2011;28:575-82.
17. Waldeyer R, Brinks R, Rathmann W, *et al.* Projection of the burden of type 2 diabetes mellitus in Germany: a demographic modelling approach to estimate the direct medical excess costs from 2010 to 2040. *Diabet Med* 2013;30:999-1008.
 18. Imperatore G, Boyle JP, Thompson TJ, *et al.* Projections of type 1 and type 2 diabetes burden in the U.S. population aged <20 years through 2050: dynamic modeling of incidence, mortality, and population growth. *Diabetes Care* 2012;35:2515-20.
 19. Boyle JP, Thompson TJ, Gregg EW, *et al.* Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. *Popul Health Metr* 2010;8:29.
 20. Magliano DJ, Shaw JE, Shortreed SM, *et al.* Lifetime risk and projected population prevalence of diabetes. *Diabetologia* 2008;51:2179-86.
 21. Huang ES, Basu A, O'Grady M, *et al.* Projecting the future diabetes population size and related costs for the U.S. *Diabetes Care* 2009;32:2225-9.
 22. Begg S, Vos T, Goss J, *et al.* An alternative approach to projecting health expenditure in Australia. *Aust Health Rev* 2008;32:148-55.
 23. Astrup A, Finer N. Redefining type 2 diabetes: 'diabesity' or 'obesity dependent diabetes mellitus'? *Obes Rev* 2000;1:57-9.
 24. Low S, Chin MC, Deurenberg-Yap M. Review on epidemic of obesity. *Ann Acad Med Singapore* 2009;38:57-9.
 25. Pierce M, Keen H, Bradley C. Risk of diabetes in offspring of parents with non-insulin-dependent diabetes. *Diabet Med* 1995;12:6-13.
 26. Valdez R, Yoon PW, Liu T, *et al.* Family history and prevalence of diabetes in the US population. The 6-year results from the National Health and Nutrition Examination Survey (1999-2004). *Diabetes Care* 2007;30:2517-22.
 27. Poulsen P, Kyvik KO, Vaag A, *et al.* Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance—a population-based twin study. *Diabetologia* 1999;42:139-45.
 28. Rocha ZL. Multiplicity within singularity: racial categorization and recognizing 'mixed race' in Singapore. *J Curr Southeast Asian Aff* 2011;30:95-131.
 29. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 1970;57:97-109.
 30. Metropolis N, Rosenbluth AW, Rosenbluth MN, *et al.* Equation of state calculations by fast computing machines. *J Chem Phys* 1953;21:1087.
 31. Sim X, Ong RT-H, Suo C, *et al.* Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *PLoS Genet* 2011;7:e1001363.
 32. Kooner JS, Saleheen D, Sim X, *et al.* Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet* 2011;43:984-9.
 33. Cho YS, Chen C-H, Hu C, *et al.* Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* 2012;44:67-72.
 34. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* 2010;11:499-511.
 35. Wang X, Chua HX, Chen P, *et al.* Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Hum Mol Genet* 2013;22:2303-11.
 36. Team RDC. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria, 2013. <http://www.R-project.org/>
 37. Plummer M, ed. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the Third International Workshop on Distributed Statistical Computing*, 2003.
 38. Plummer M. JAGS version 3.1.0 user manual. 2011.
 39. Murrell P. *R graphics*. 2nd edn. The R series. London: Chapman and Hall/CRC, 2011.
 40. Ng YC, Jacobs P, Johnson JA. Productivity losses associated with diabetes in the US. *Diabetes Care* 2001;24:257-61.
 41. Vijan S, Hayward RA, Langa KM. The impact of diabetes on workforce participation: results from a national household sample. *Health Serv Res* 2004;39(6 Pt 1):1653-69.
 42. Kraut A, Walld R, Tate R, *et al.* Impact of diabetes on employment and income in Manitoba, Canada. *Diabetes Care* 2001;24:64-8.
 43. Yang W, Lu J, Weng J, *et al.* Prevalence of diabetes among men and women in China. *N Engl J Med* 2010;362:1090-101.
 44. Rampal S, Rampal L, Rahmat R, *et al.* Variation in the prevalence, awareness, and control of diabetes in a multiethnic population: a nationwide population study in Malaysia. *Asia Pac J Public Health* 2010;22:194-202.
 45. Gluckman PD, Hanson MA, Cooper C, *et al.* Effect of in utero and early-life conditions on adult health and disease. *N Engl J Med* 2008;359:61-73.
 46. Bassuk SS, Manson JE. Epidemiological evidence for the role of physical activity in reducing risk of type 2 diabetes and cardiovascular disease. *J Appl Physiol* 2005;99:1193-204.
 47. Hu FB. Globalization of diabetes: the role of diet, lifestyle, and genes. *Diabetes Care* 2011;34:1249-57.
 48. Harrison TA, Hindorf LA, Kim H, *et al.* Family history of diabetes as a potential public health tool. *Am J Prev Med* 2003;24:152-9.
 49. Kjos SL, Buchanan TA. Gestational diabetes mellitus. *N Engl J Med* 1999;341:1749-56.
 50. Dorajoo R, Blakemore A, Sim X, *et al.* Replication of 13 obesity loci among Singaporean Chinese, Malay and Asian-Indian populations. *Int J Obes (Lond)* 2012;36:159-63.
 51. Elks CE, Den Hoed M, Zhao JH, *et al.* Variability in the heritability of body mass index: a systematic review and meta-regression. *Front Endocrinol (Lausanne)* 2012;3:29.
 52. Penman AD, Johnson WD. The changing shape of the body mass index distribution curve in the population: implications for public health policy to reduce the prevalence of adult obesity. *Prev Chronic Dis* 2006;3:A74.
 53. Wang Y, Lobstein T. Worldwide trends in childhood overweight and obesity. *Int J Pediatr Obes* 2006;1:11-25.

SUPPLEMENTARY MATERIAL

This document provides additional detail on the model components used to forecast diabetes incidence and prevalence in Singapore, along with further results.

Supplementary Methods

This section describes the mathematical formulation of the subcomponents of the model in detail, along with the approach used for parameter estimation.

General notation

The model formulation described in the following sections makes use of the following common notation.

Indices	Subscripts or superscripts
i	individual index
x	age index
g	gender index ($g = 0$ for females and $g = 1$ for males)
r	race index (Chinese $r = 0$; Malay $r = 1$; Indian $r = 2$; Other $r = 3$)
t	time index ($t = 0$ for starting year 1990)

Demographic model

The demographic model has three subcomponents, describing mortality, fertility and migration.

Mortality

Using the Lee–Carter model [1] as a state-space model [2], we have the following form for the log-mortality of the g th gender, at age a and time t :

$$m_{x,t}^g = \alpha_x^g + \beta_x^g \kappa_t^g$$

where α_x^g is the (natural) log of the mortality rate for that group in 1990, β_x^g accounts for mortality changes over time, and varies by 5-year age groups, and κ_t^g is assumed to be a random walk with drift d_m^g :

$$\kappa_t^g = \kappa_{t-1}^g + d_m^g + \epsilon_{m,t}^g,$$

with $\epsilon_{m,t}^g$ a random error, assumed to be normally distributed with mean 0 and variance $\sigma_{\epsilon_{g,m}}^2$.

Race-specific mortality rates are assumed to be proportional to the baseline mortality rate, with proportional hazard ratios allowed to be different for three broad age ranges:

young (age 0–19), *adult* (age 20–59) and *old* age (age 60 and above). The log mortality rate for race r is then

$$m_{x,t}^{g,r} = m_{x,t}^g + \gamma_{A_x}^{g,r}$$

where A_x indicates which of the three age ranges x falls in.

Fertility

We use a Gaussian function to model the shape of the age-specific fertility rates, $f_{x,t}$:

$$f_{x,t} = \theta_2 \exp(\delta_t) \frac{1}{\sqrt{2\pi\sigma_f^2}} \exp\left(\frac{-[x - \mu_f + \theta_1(t-1)]^2}{2\sigma_f^2}\right).$$

We assume that the age of peak fertility, μ_f , increases linearly with time (i.e. a birth cohort effect) with a constant rate θ_1 , while a further secular trend δ_t of the overall fertility rate that follows a random walk with drift. Race-specific fertility rates are modelled as proportional to age-specific fertility rates of the total population, i.e.

$$f_{x,t}^r = f_{x,t} \exp(\tau_r).$$

Migration

To model the net-migration number for each year, we re-parameterise the simplified migration schedule model of Castro and Rogers [3] without the retirement peak:

$$M_{x,t} = (a_1 b_1 \exp(-a_1 x) + b_2 \exp(-a_2(x - \mu_2) - \exp(-\lambda_2(x - \mu_2))) + b_0) \rho_t.$$

Pre-labour force migration (i.e. children) is represented by a single negative exponential curve with descent parameter a_1 , while the labour force age migration is represented by a left-skewed unimodal curve with λ_2 as ascent parameter, a_2 as descent parameter and μ_2 controlling the age of peak migration. For identifiability, b_2 is fixed to 1, and ρ_t is time-varying scale parameter reflecting economic fluctuations. All the parameters are allowed to vary by gender and race. For forecasts, we fix ρ_t to match government “guidelines” for projected future population growth, as reported in the national press. It is worth noting that these reports imply a reduced influx of immigrants relative to past patterns.

Employment model

We propose a discrete-time Markov chain model for employment status of Singapore residents with three states: working (W), unemployed (U) and economically inactive (I). In the model, weekly transition probabilities depend on age and gender. The transition probability matrix of the model is:

$$\begin{matrix} & W_{x,t}^g & U_{x,t}^g & I_{x,t}^g \\ \begin{matrix} W_{x,t-1}^g \\ U_{x,t-1}^g \\ I_{x,t-1}^g \end{matrix} & \begin{pmatrix} 1 - u_x^g \cdot \exp(\omega_t) & u_x^g \cdot \exp(\omega_t) & 0 \\ w_x^g & 1 - w_x^g - v_x^g & v_x^g \\ s_x^g & 0 & 1 - s_x^g \end{pmatrix} \end{matrix}$$

with time parameter $\omega_t \sim N(0, \sigma_\omega^2)$.

We use a cubic spline function to model the age-specific transition probabilities with 5 knots at age 15, 30, 45, 60 and 70, for both genders separately. To obtain the distribution of durations unemployed in weeks, we extend the transition matrix from 3 states $(W_{x,t}^g, U_{x,t}^g, I_{x,t}^g)$ to 42 states $(W_{x,t}^g, U_{x,t}^{g,1}, \dots, U_{x,t}^{g,40}, I_{x,t}^g)$ where $U_{x,t}^{g,k}$ denotes the number of people age x , gender g , being unemployed for k weeks at time t , except $U_{x,t}^{g,40}$ denotes people being unemployed for 40 weeks or more.

Body mass index model

Model

The BMI trajectory of individual i (of gender g and race r) is modelled to follow an overall trend μ_i with an annual variability around the trend assumed to be independent and identically distributed Gaussian:

$$B_{i,x} | \mu_{i,x}, \sigma^2 \sim N(\mu_{i,x}, \sigma^2).$$

We developed a piecewise linear model for individuals' overall BMI trend $\mu_{i,x}$ with 4 chosen breakpoints at ages 18, 35, 55 and 75, for which that individual's mean BMI is specified parametrically. A natural cubic spline basis function S is used to interpolate intermediate means between those 4 breakpoints:

$$\mu_{i,x} = S(\mu_{i,18}, \mu_{i,35}, \mu_{i,55}, \mu_{i,75}; x).$$

The BMI values at the breakpoints are set to be:

$$\begin{aligned} \mu_{i,18} &= \varphi_{i,0} + \beta_{g,r} \cdot (y_i - 1950), \\ \mu_{i,35} &= \mu_{i,18} + \varphi_{i,1} \cdot (35 - 18), \\ \mu_{i,55} &= \mu_{i,35} + \varphi_{i,2} \cdot (55 - 35), \text{ and} \\ \mu_{i,75} &= \mu_{i,55} + \varphi_{i,3} \cdot (75 - 55), \end{aligned}$$

i.e. there is a secular trend based on year of birth, y_i , on the BMI at age 18 within any demographic segment of the population. Individualised BMI parameters governing BMI at different stages of adulthood ($\varphi_{i,0}$ to $\varphi_{i,3}$) are taken to have multivariate Gaussian distribution over each demographic segment of the population as a whole, with hyper-parameters estimated independently for each race and gender combination:

$$\begin{pmatrix} \varphi_{i,0} \\ \varphi_{i,1} \\ \varphi_{i,2} \\ \varphi_{i,3} \end{pmatrix} \sim \text{MVN} \left(\begin{pmatrix} \phi_0 \\ \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix}, \Sigma_{g,r} \right).$$

Hyperpriors

We use the following hyperpriors:

$$\begin{aligned} \begin{pmatrix} \phi_0 \\ \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix} &\sim \text{MVN} \left(\begin{pmatrix} 20 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{\Omega}^{-1} \right), \\ \Sigma^{-1} &\sim \text{Wish}(\mathbf{\Omega}, 5), \\ \beta^{g,r} &\sim U(-100, 100). \end{aligned}$$

Average BMI at age 18 for people born in 1950 was presumed to lie in the range 10–30, which we represented by mean 20 and variance 5^2 , hence the specific values in the hyperprior for ϕ_0 . As we expect variability in BMI from year to year to be small and of the order of ± 1 we assume standard Gaussian distribution for variability around individuals' BMI trends and set $\sigma = 0.5$. Prior correlations among ϕ_0 , ϕ_1 , ϕ_2 and ϕ_3 are set to be 0. Hence, we have $\mathbf{\Omega} = \text{Diag}(5^{-2}, 1, 1, 1)$. Face validity of these priors was established by presenting simulated trajectories to nutritional epidemiologists.

Parameter estimation

This hierarchical model is parameterised using longitudinal data from the Singapore Prospective Study Programme (SP2) [4] which contain two BMI measurements for each individual in the study. The first time point corresponds to the 1992 National Health Survey (NHS) or 1998 NHS and the second to a follow-up visit around 2005. The likelihood from these data is augmented by a contribution from aggregate data from the 2004 and 2010 NHSs on the numbers in four BMI categories (underweight, normal weight, overweight and obese) within age bands and gender/race groups, assuming no differences in proportions of age groups and sex ratios among the 3 ethnic groups in the survey sample. Using the total sample size and the marginal proportions in age groups, ethnic groups and genders, we estimate the number of people in each gender, ethnic group, age group and BMI category combination in the survey sample. For these data, we generate $N_{a,b}^{g,r}$ individuals with known gender g and race r , with age assumed uniformly distributed in age group a , simulated BMI values from the model above and evaluated via Monte Carlo the fraction within each BMI category b , a process similar to Approximate Bayesian Computation [5].

The full Bayesian hierarchical model for BMI was implemented using JAGS 3.1.0.[6-7] Data preparation and analysis was done in R and JAGS was called from R [8] using the package `rjags` [9]. For each demographic segment, we ran three parallel chains from randomly selected starting points, with a burn-in of 20 000, thinning at 10, and 100 000 iterations post-burnin. Convergence was assessed by visually assessing the trace and the posterior density plot of all parameters (**Supplementary Figure S10–S15**). The computing time, under a desktop Intel Core 4 CPU 2.83 GHz 7.6GB RAM, was about 10–12 hours for minority groups (Indian male/female and Malay male/female) and nearly 1 day for Chinese males and females for whom the sample size was larger.

Diabetes incidence model

Diabetes Mellitus data

The SP2 longitudinal data also provided information on diabetic status at the same two time points as BMI was measured. Although the classification of diabetes mellitus status in National Health Survey 1992 and 1998 was based on the result of 2 tests (fasting-glucose test and oral glucose tolerance test), the diagnosis of diabetes in the follow-up study in 2005 was based only on the result of the fasting glucose test, which has lower sensitivity. We therefore adjusted the estimates of the diabetes prevalence to account for the underestimated prevalence at the second time point, as described below.

Denote the true diabetes mellitus status of individual i at time point 1 (1992 or 1998) and time point 2 (around 2005) by Y_1^i and Y_2^i , respectively, where $Y_k^i = 1$ means that the individual i has diabetes at time point k and 0 otherwise. At time point 1, Y_1^i may only take the value of 0 or 1, and its value is assumed known as the oral glucose tolerance test is taken to be a gold standard. At time point 2, if an individual tested positive on the fasting glucose test, Y_2^i is known to be 1, but if negative on the fasting glucose test, his or her true diabetic status is assumed to be unknown, i.e. Y_2^i could be either 0 or 1. (Note that this applies only to individuals without diabetes at the first time point. Those with diabetes are assumed not to be cured and are excluded from this part of the model fitting.)

We assume that $Y_1^i \sim \text{Bern}(\pi_1^i)$ and $(Y_2^i | Y_1^i = 0) \sim \text{Bern}(\pi_2^i)$. Let us denote F_k^i as the value of fasting glucose test at time point k and T_k^i as the classification result based on fasting glucose test. Using the standard threshold for fasting glucose, we have:

$$\begin{aligned} T_k^i &= \begin{cases} 0 & \text{if } F_k \leq 7, \\ 1 & \text{otherwise.} \end{cases} \\ T_1^i &\sim \text{Bern}(p \cdot Y_1^i) \\ T_2^i &\sim \text{Bern}(p \cdot Y_2^i), \end{aligned}$$

i.e. where p is the sensitivity of fasting glucose at this threshold relative to the gold standard. The distribution for Y_2^i can therefore be estimated as the information from Y_1^i provides information on p .

We use logistic regression to fit the model:

$$\begin{aligned} \text{logit}(\pi_1^i) &= \alpha_1 + \beta_1 B_1^i + \beta_2 x_1^i + \beta_3 1[g^i = 1] + \beta_4 1[r^i = 1] + \beta_5 1[r^i = 2] + \beta_6 F_1^i \\ \text{logit}(\pi_2^i | Y_1^i = 0) &= \alpha_2 + \beta_1 B_2^i + \beta_2 x_2^i + \beta_3 1[g^i = 1] + \beta_4 1[r^i = 1] + \beta_5 1[r^i = 2] + \beta_6 F_2^i \end{aligned}$$

where B_i is the BMI value at time point i , x_i is age at that time point, $1[A]$ the indicator function equal to 1 if A is true and 0 otherwise.

The following non-informative priors were taken for this submodel:

$$\begin{aligned} \alpha_1 &\sim \text{U}(-100, 100), \\ \alpha_2 &\sim \text{U}(-100, 100), \\ \beta_k &\sim \text{U}(-5, 5), \text{ for } k = 1, \dots, 6, \\ p &\sim \text{U}(0, 1). \end{aligned}$$

The model was implemented in JAGS 3.1.0.[6-7] Data preparation and analysis was done in R and JAGS was called from R [8] using the package `rjags` [9]. 3 parallel chains were set to run at different initial conditions. After burn-in and thinning, 3 chains of 10 000 samples from each chain were kept. Convergence was assessed visually by the trace plot and the posterior density plot of all parameters (**Supplementary Figure S16**). The computing time on a normal desktop was about 12 hours.

BMI trajectory and importance sampling

To fit the logistic regression model for diabetes risk as a function of BMI and demographics requires knowing the BMI in each year between the 2 time points of the SP2 study, rather than the 2 observed BMI values for which we actually have data. We thus imputed a BMI trajectory for each individual using the fitted model for BMI using importance sampling.

First we generate a sample of BMI trajectories of size 2000 for each individual in the SP2 study, using their year of birth, race and gender, and the estimated BMI model using the following steps:

- The mean of BMI at age 18 was calculated based on year of birth of individual i ;
- 4 parameters for underlying piecewise linear model were generated using the estimated multivariate Gaussian distribution from the BMI model
- BMIs at 4 ages (18, 35, 55 and 75) were calculated from the piecewise linear model and the overall BMI trend of individual i was obtained by a cubic spline function interpolating those 4 points;
- a BMI trajectory was generated by simulating $B_{i,x}^{g,r} | \mu_{i,x}^{g,r}, \sigma^2 \sim N(\mu_{i,x}^{g,r}, \sigma^2)$.

Step 2–4 were repeated to reach the desired sample size. Then, for each individual, one of these 2000 samples is then selected with probability proportional to the likelihood of getting the 2 observed BMI values from the sampled trajectories. This was then used in subsequent analysis.

Incidence model

We denote the annual risk of developing diabetes for each non-diabetic individual i at age x by $p_{i,x}^D$. The model accounts the individual i 's age x , gender g and ethnic group r and BMI B at age x , and takes a logit link for this probability:

$$\text{logit}(p_{i,x}^D) = \theta + \alpha x + \beta B_i(x) + \gamma g + \delta_1 1[r = 1] + \delta_2 1[r = 2]$$

where δ_1 is the Malay race effect (relative to Chinese), δ_2 the Indian race effect (ditto) and $1[A]$ the indicator function equal to 1 if A is true and 0 otherwise.

We denote the observed diabetic status in 2 time points for individual i as $D_{1,i}$ and $D_{2,i}$ and age at the 2 time points as $x_{1,i}$ and $x_{2,i}$. For each individual, an accumulated

risk is calculated from age 18 until the first survey so that the likelihood contribution is derived from

$$D_{1,i} \sim \text{Bern} \left(1 - \prod_{x=10}^{x_{1,i}} [1 - p_{i,x}^D] \right).$$

An analogous likelihood contribution for those non-diabetic at the time of the first survey, for the accumulated risk from the first survey until follow-up is given by

$$(D_2|D_1 = 0) \sim \text{Bern} \left(1 - \prod_{x=x_1}^{x_{2,i}} [1 - p_{i,x}^D] \right).$$

Parameter estimation

The parameters of the diabetes model were estimated using the Metropolis-Hasting algorithm [10-11]. with multivariate normal proposal densities, a burn-in of length 10 000, and 3 independent chains of 50 000 iterations each, with every 5th iteration retained for analysis. Convergence was assessed by visually assessing the trace and the posterior density plot of all parameters (**Supplementary Figure S17**). It took about 1.5 hours to simulate BMI trajectories for these individuals and 6.5 hours to run the MCMC on a desktop Intel Core 4 CPU 2.83 GHz 7.6GB RAM.

Mortality rates for diabetics are taken to be proportional to that of the general population, multiplied by the estimated hazard ratios for all-cause mortality according to diabetes status and by ethnic groups for participants from the National Health Survey 1992 [12].

Sensitivity analysis

We developed an alternative diabetes model using only age, race and gender as risk factors. After stratifying by race and gender, the age (t_i) at disease onset is assumed to follow a Weibull distribution, $t_i \sim W(\alpha, \beta)$. For each race and gender combination, we have:

$$p_{[x_1, x_2]}^{g,r} = \frac{\sum_{k=x_1}^{x_2} P(t_i^{g,r} \leq k)}{x_2 - x_1 + 1}$$

where $p_{[x_1, x_2]}^{g,r}$ and $N_{[x_1, x_2]}^{g,r}$ are modeled diabetes prevalence and total number of people in age group $[x_1, x_2]$, respectively. The likelihood of diabetes prevalence is:

$$\hat{p}_{[x_1, x_2]}^{g,r} \sim N \left(p_{[x_1, x_2]}^{g,r}, \frac{p_{[x_1, x_2]}^{g,r} \cdot (1 - p_{[x_1, x_2]}^{g,r})}{N_{[x_1, x_2]}^{g,r}} \right)$$

where $\hat{p}_{[x_1, x_2]}^{g,r}$ is estimated diabetes prevalence of people of race r and gender g in age group $[x_1, x_2]$ obtained from 3 recent National Health Survey (1998, 2004 and 2010). We replaced $N_{[x_1, x_2]}^{g,r}$ with $\hat{N}_{[x_1, x_2]}^{g,r}$ as estimated sample size based on the total sample sizes and marginal proportions of gender, race and age group in survey samples.

The prior distributions for $\alpha^{g,r}$ and $\beta^{g,r}$ are taken to be improper uniform distributions on the positive part of the real line. Using the Metropolis Hasting algorithm,^{10–11} these parameters are updated by drawing a sample from a bivariate normal distribution centred on the current value with covariance matrix selected on pilot runs. **(Supplementary Figure S18)**

Supplementary Tables

Genetic risk model

Supplementary Table S1: Summary Genome-wide associated study (GWAS) results of the selected 14 SNPs. RAF: Risk allele frequency; OR: odds ratio; p: p-values are calculated from logistics regression using SNPTEST software with bold denoting significant association ($p < 0.05$).

SNP	Associated gene	Risk allele	SCES (Chinese)			SiMES (Malay)			SINDI (Indian)		
			RAF	OR	p	RAF	OR	p	RAF	OR	p
rs11642841	FTO	A	0.046	0.761	0.026	0.198	0.868	0.051	0.317	0.964	0.542
rs12779790	CDC123; CAMK1D	G	0.174	1.128	0.261	0.148	1.175	0.036	0.177	1.101	0.123
rs13266634	SLC30A8	C	0.532	0.627	6.35E-07	0.563	0.896	0.120	0.759	0.902	0.282
rs1353362	TSPAN8- LGR5	C	0.251	0.797	0.027	0.246	0.979	0.972	0.374	0.995	0.948
rs1387153	MTNR1B	T	0.463	1.257	0.013	0.425	0.974	0.794	0.369	0.999	0.881
rs1470579	IGF2BP2	C	0.236	1.247	0.031	0.331	0.988	0.859	0.456	1.125	0.226
rs231362	KCNQ1	G	0.888	1.487	0.011	0.841	1.032	0.531	0.737	1.029	0.676
rs2334499	HCCA2	T	0.813	0.969	0.783	0.546	1.096	0.144	0.312	0.857	0.038
rs243021	BCL11A	A	0.662	1.022	0.822	0.540	1.040	0.576	0.470	0.898	0.048
rs5015480	HHEX	C	0.186	0.941	0.603	0.256	0.863	0.063	0.369	0.858	0.006
rs5215	KCNJ11	C	0.351	0.856	0.095	0.387	0.861	0.023	0.350	0.983	0.686
rs7754840	CDKAL1	C	0.378	0.836	0.058	0.364	0.950	0.395	0.233	0.870	0.036
rs7903146	TCF7L2	T	0.021	1.344	0.297	0.038	1.329	0.2621	0.264	1.242	0.002
rs8042680	PRC1	A	0.997	0.601	0.633	0.965	0.612	0.003	0.762	0.952	0.960

Prevalence of type 2 diabetes

Supplementary Table S2: Forecast age, race and gender specific prevalence of type 2 diabetes (total: diagnosed and undiagnosed), per 100 population, to 2050. Numbers in parentheses are 95% prediction intervals.

Age	Total		Chinese		Indian		Malay	
	Female	Male	Female	Male	Female	Male	Female	Male
2010								
18-29	1 (1-1)	2 (1-2)	1 (1-1)	1 (1-2)	2 (1-3)	2 (2-3)	2 (1-2)	2 (2-3)
30-39	3 (3-4)	5 (5-6)	3 (2-3)	5 (4-6)	6 (5-8)	7 (6-8)	7 (5-8)	8 (7-11)
40-49	8 (8-9)	12 (11-13)	7 (6-8)	11 (10-12)	16 (13-19)	16 (14-19)	15 (13-18)	17 (14-20)
50-59	16 (14-18)	20 (18-22)	13 (12-15)	19 (16-21)	28 (24-32)	29 (25-35)	26 (23-31)	26 (22-30)
60-69	25 (22-28)	29 (26-32)	22 (19-25)	28 (25-31)	39 (34-45)	42 (37-48)	38 (33-44)	36 (31-41)
2030								
18-29	1 (1-2)	2 (2-2)	1 (1-1)	2 (1-2)	2 (2-3)	3 (2-4)	2 (2-3)	3 (2-4)
30-39	5 (4-5)	7 (6-8)	4 (3-4)	6 (5-7)	8 (7-10)	10 (8-13)	8 (6-10)	10 (8-13)
40-49	9 (8-10)	14 (12-16)	6 (5-7)	12 (10-14)	20 (17-24)	21 (18-26)	16 (14-20)	20 (16-24)
50-59	16 (14-18)	24 (21-26)	12 (11-14)	21 (19-23)	33 (28-38)	34 (29-39)	30 (25-34)	32 (27-37)
60-69	26 (24-29)	35 (31-38)	21 (19-25)	32 (28-35)	47 (42-53)	46 (41-53)	44 (39-50)	44 (38-51)
2050								
18-29	2 (1-2)	2 (2-3)	1 (1-2)	2 (2-2)	3 (2-3)	3 (2-4)	3 (2-4)	4 (3-5)
30-39	5 (5-6)	8 (7-9)	4 (3-5)	7 (5-8)	9 (8-11)	12 (10-15)	10 (8-12)	13 (10-16)
40-49	11 (10-12)	16 (14-18)	9 (8-10)	14 (12-16)	21 (17-25)	24 (21-29)	19 (16-23)	25 (20-29)
50-59	19 (17-21)	27 (24-30)	16 (14-18)	24 (21-27)	34 (29-39)	38 (34-44)	32 (27-36)	39 (33-44)
60-69	26 (24-29)	39 (35-43)	20 (17-23)	35 (31-39)	54 (48-60)	54 (48-61)	47 (41-53)	52 (46-58)

Age and Overweight

Supplementary Table S3: Estimated and forecast age, race and gender specific prevalence of obesity, per 100 population, to 2050. Numbers in parentheses are 95% prediction intervals.

Age	Total		Chinese		Indian		Malay	
	Female	Male	Female	Male	Female	Male	Female	Male
1990								
18-29	2 (2-3)	2 (2-3)	0 (0-0)	2 (1-2)	7 (5-9)	3 (2-5)	8 (6-11)	7 (5-9)
30-39	5 (4-5)	5 (4-5)	2 (2-2)	4 (3-4)	15 (12-17)	8 (7-11)	18 (15-21)	9 (8-11)
40-49	7 (6-7)	5 (5-6)	3 (3-4)	4 (4-4)	19 (17-22)	11 (10-14)	27 (24-29)	10 (9-12)
50-59	9 (8-9)	4 (4-5)	5 (4-5)	3 (2-3)	19 (17-21)	10 (8-12)	30 (28-33)	10 (9-11)
60-69	9 (9-10)	4 (4-5)	6 (5-7)	2 (2-2)	18 (15-21)	9 (7-11)	30 (27-33)	11 (9-13)
2010								
18-29	3 (3-4)	5 (4-6)	1 (1-1)	3 (3-4)	11 (8-14)	5 (3-8)	12 (9-14)	13 (10-16)
30-39	6 (5-6)	9 (8-9)	1 (1-2)	7 (6-7)	22 (19-25)	13 (11-16)	23 (20-26)	17 (15-20)
40-49	8 (8-9)	10 (9-10)	2 (2-3)	7 (6-8)	27 (25-30)	16 (14-19)	33 (30-35)	19 (17-21)
50-59	9 (9-10)	7 (7-8)	4 (3-4)	5 (5-6)	28 (25-30)	15 (13-17)	36 (34-38)	18 (17-20)
60-69	9 (8-9)	5 (5-6)	5 (4-5)	4 (3-4)	26 (22-29)	13 (10-16)	35 (32-38)	18 (15-20)
2030								
18-29	8 (7-9)	9 (8-10)	5 (4-6)	7 (5-8)	10 (8-13)	8 (5-12)	22 (18-26)	22 (18-27)
30-39	11 (10-12)	14 (13-15)	7 (7-8)	11 (10-12)	23 (20-26)	18 (15-22)	28 (25-32)	28 (25-32)
40-49	10 (9-10)	15 (14-16)	2 (2-2)	11 (10-13)	37 (34-41)	23 (20-26)	38 (35-41)	31 (29-34)
50-59	10 (10-11)	12 (12-13)	3 (2-3)	9 (8-10)	38 (35-41)	21 (18-24)	42 (40-45)	31 (28-33)
60-69	10 (10-11)	10 (9-11)	4 (3-4)	6 (6-7)	36 (32-40)	18 (15-22)	40 (36-45)	29 (25-32)
2050								
18-29	12 (11-13)	15 (13-16)	8 (7-10)	11 (10-13)	14 (11-18)	13 (9-17)	34 (29-40)	34 (29-40)
30-39	16 (15-17)	21 (20-23)	11 (10-12)	18 (16-19)	28 (24-32)	26 (22-30)	42 (39-46)	42 (39-46)
40-49	17 (16-18)	23 (22-24)	11 (10-12)	19 (17-20)	36 (33-40)	31 (27-35)	46 (44-49)	46 (44-49)
50-59	16 (15-17)	20 (19-21)	10 (9-10)	15 (14-16)	36 (33-39)	28 (25-32)	46 (43-48)	46 (43-48)
60-69	12 (11-13)	16 (14-17)	3 (2-3)	11 (9-12)	47 (42-52)	25 (22-30)	47 (43-51)	43 (39-47)

Sensitivity analysis

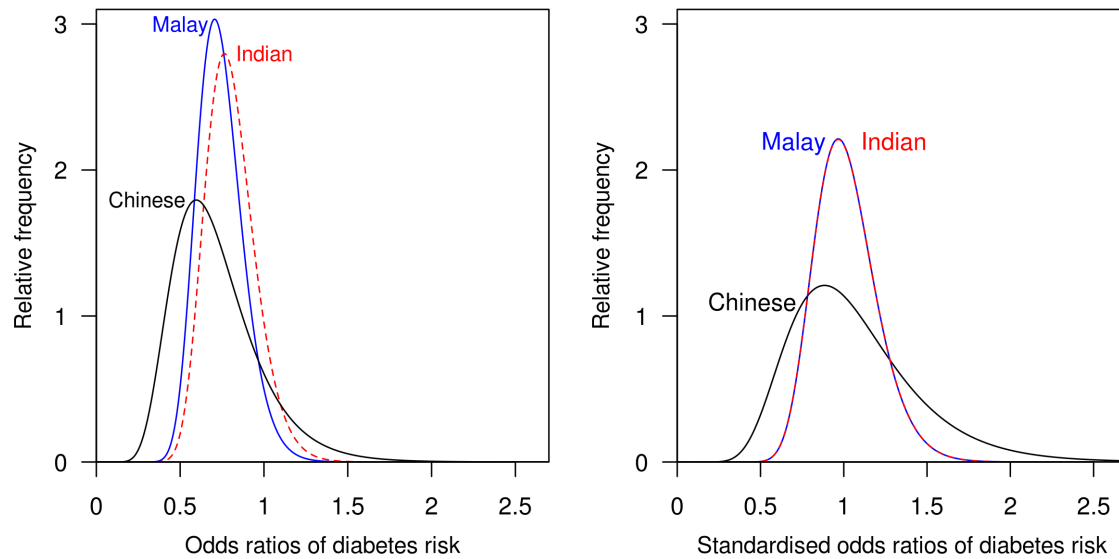
Supplementary Table S4: Estimated lifetime risk (percent) of developing type 2 diabetes for adults aged 18 and above over time, Singapore.

year	Model without BMI and genetic risk			Full model		
	Male (%)	Female (%)	Both (%)	Male (%)	Female (%)	Both (%)
1990	38.9 (34.8-43.8)	38.1 (35.7-44.6)	38.9 (36.3-41.9)	37 (33.4-41.2)	32.2 (28.6-36.3)	34.5 (31.9-38.2)
2000	37.8 (34.2-43.2)	39.5 (35.7-44.5)	38.9 (35.9-41.7)	38.5 (35.1-42.7)	32.4 (29-36.3)	35.4 (32.9-38.8)
2010	38.2 (34.2-44.3)	39.8 (35.8-45.3)	39 (36.2-42.6)	41.2 (37.5-45.6)	32.6 (29.4-35.9)	36.9 (34.1-40.3)
2020	38.9 (35-44.7)	39.9 (36.3-45.6)	39.6 (36.6-42.8)	43.9 (40.1-48.7)	33.6 (30.4-37)	38.7 (35.9-42.2)
2030	39.1 (35.2-44.5)	39.9 (36.5-45.7)	39.7 (37-42.8)	46.4 (41.9-50.9)	34.3 (30.9-38.2)	40.2 (37.2-43.5)
2040	38.7 (35-44.2)	39.8 (36.3-45.8)	39.5 (36.9-42.7)	48.2 (43.8-53.1)	35.3 (32-38.9)	41.6 (38.8-45.3)
2050	38.4 (34.7-44.2)	39.9 (36.8-45.5)	39.2 (36.9-42.5)	51 (46.1-55.9)	37 (33.6-40.9)	43.8 (40.8-47.5)

Supplementary Figures

Genetic effects on diabetes incidence

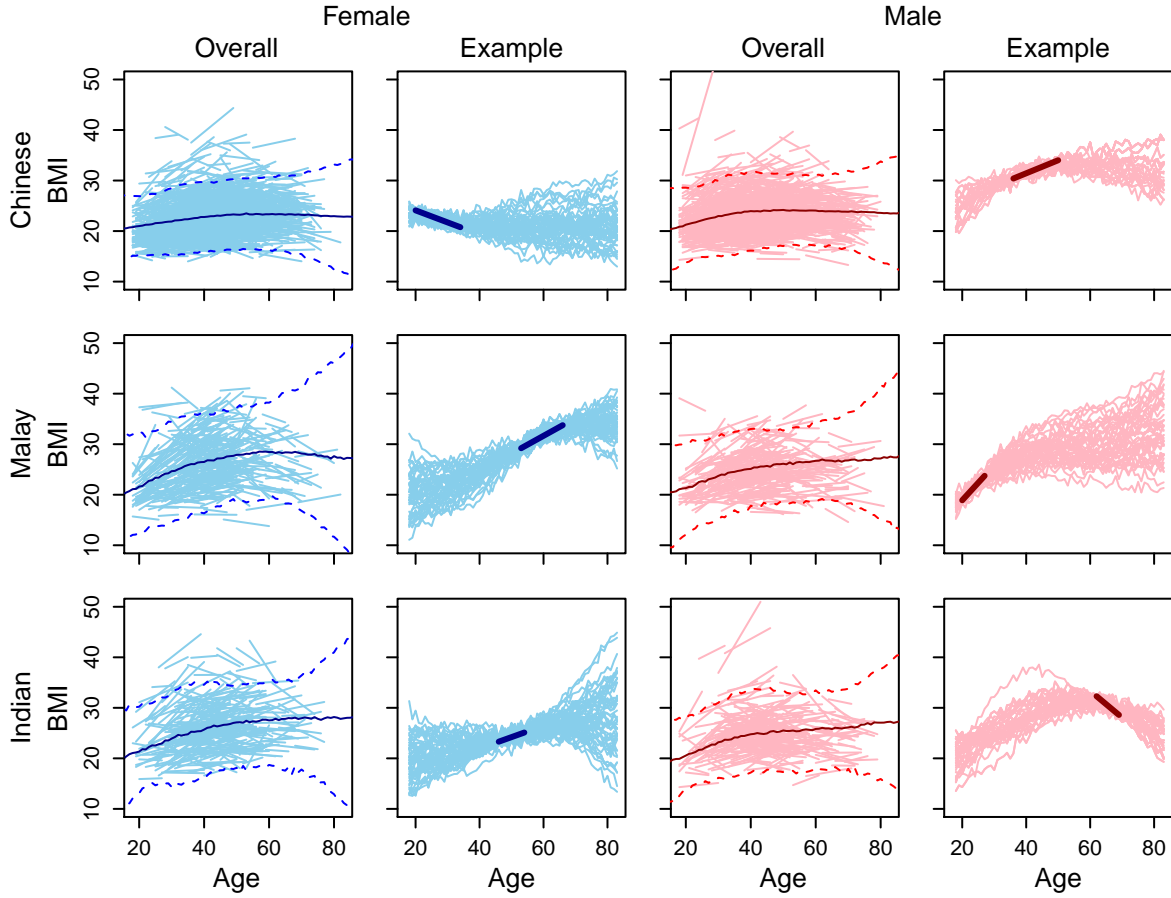
Supplementary Figure S1: Distribution of odds ratios of diabetes before and after standardisation.



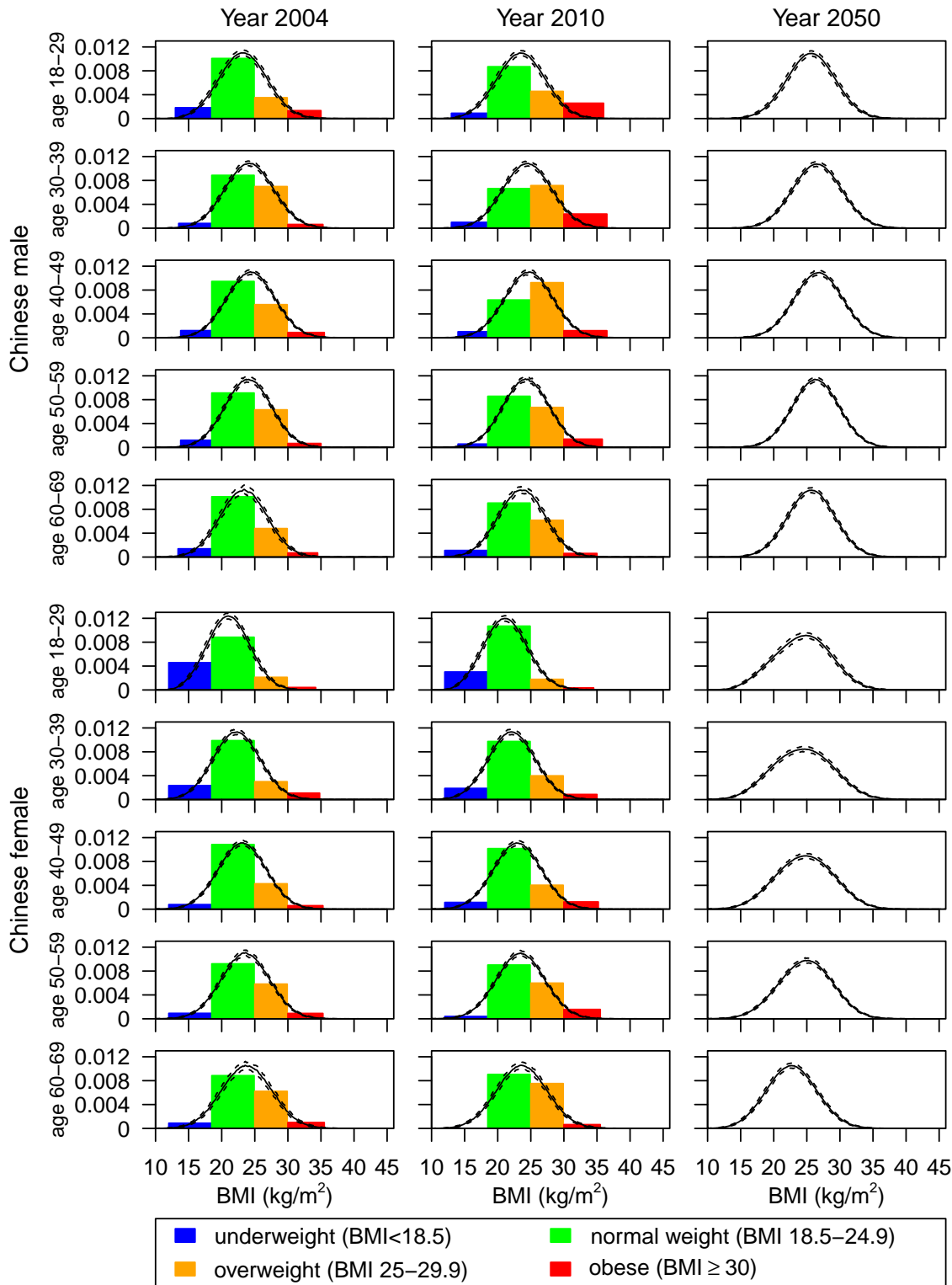
Body mass index model

In this section we provide more detailed results for the BMI model. **Supplementary Figure S2** provides the fitted model and examples of individual BMI trajectories for selected individuals. **Supplementary Figures S3–S8** compare the prevalence of different BMI categories in different age groups, gender and race with the National Health Survey results.

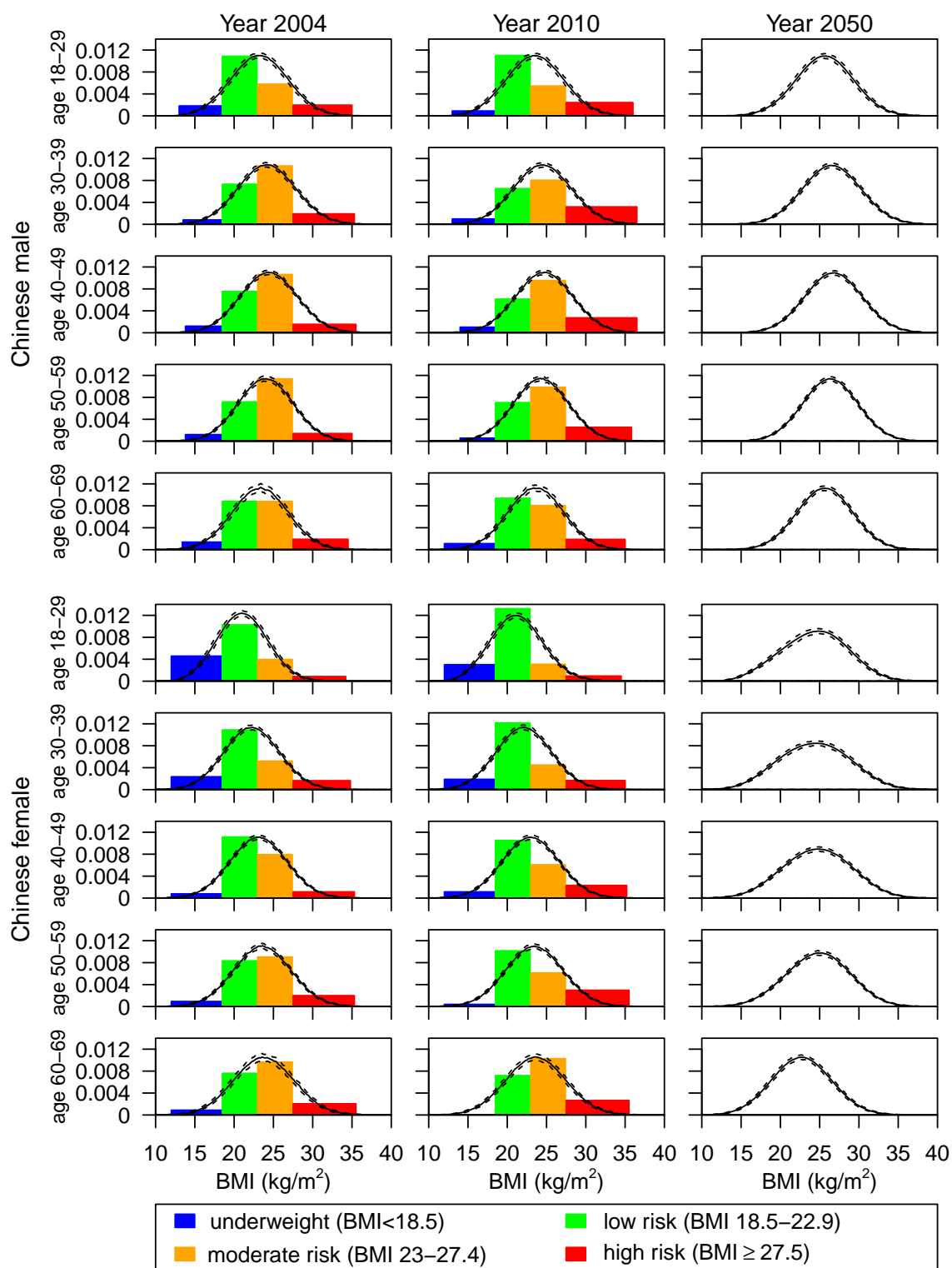
Supplementary Figure S2: Overall model and example of simulated BMI trajectories by race and gender. Overall BMI model: mean and 95% prediction interval are plotted with overlaid line segments representing individuals' BMI measurements at 2 time points. Example: individual data are indicated by dark lines and simulated BMI trajectories are indicated by light lines.



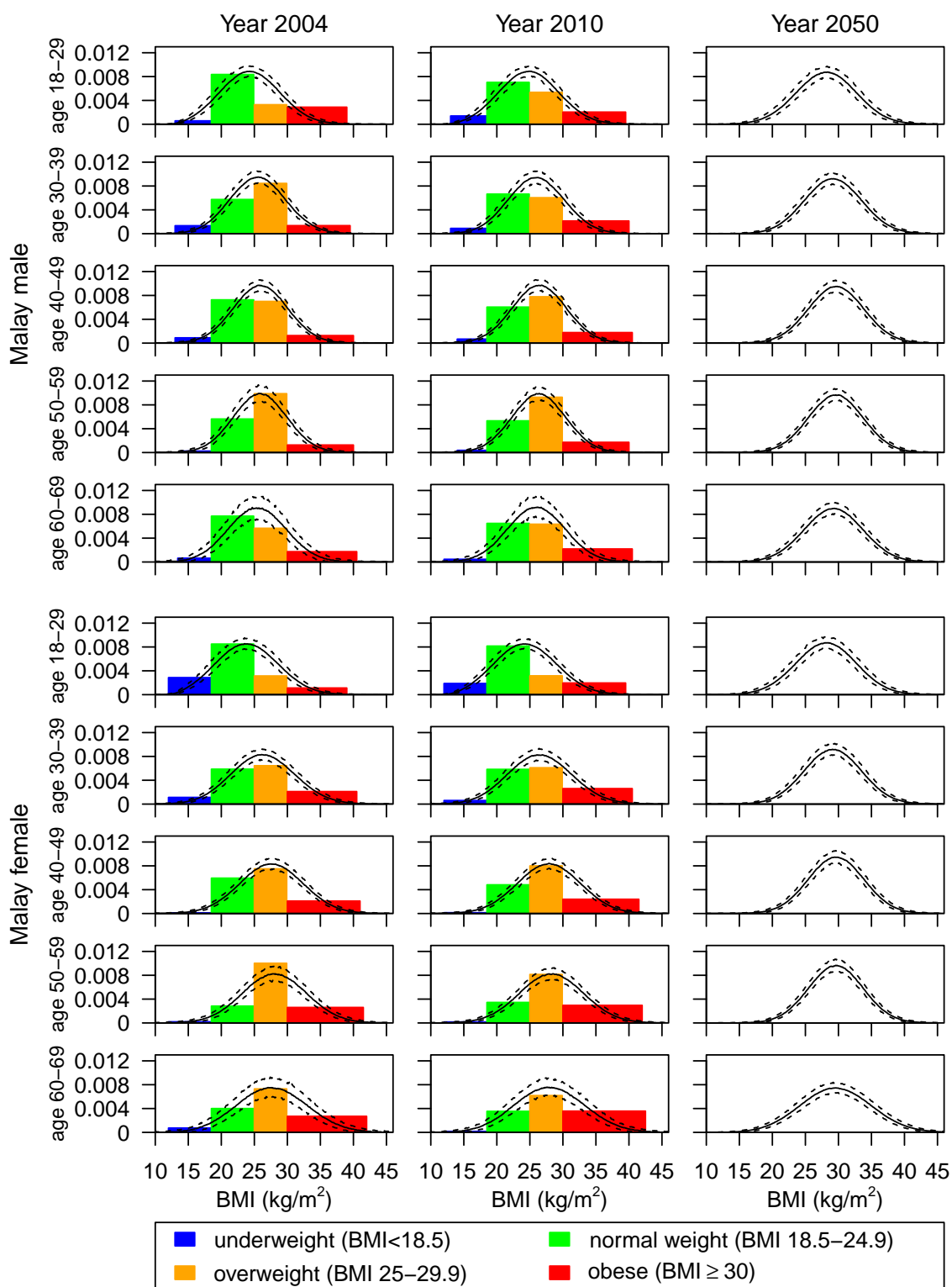
Supplementary Figure S3: Estimated and forecasted BMI distribution of Chinese by age groups and gender and aggregated statistics of 4 weight status. Mean and 95% prediction intervals are plotted. Data are presented as bars.



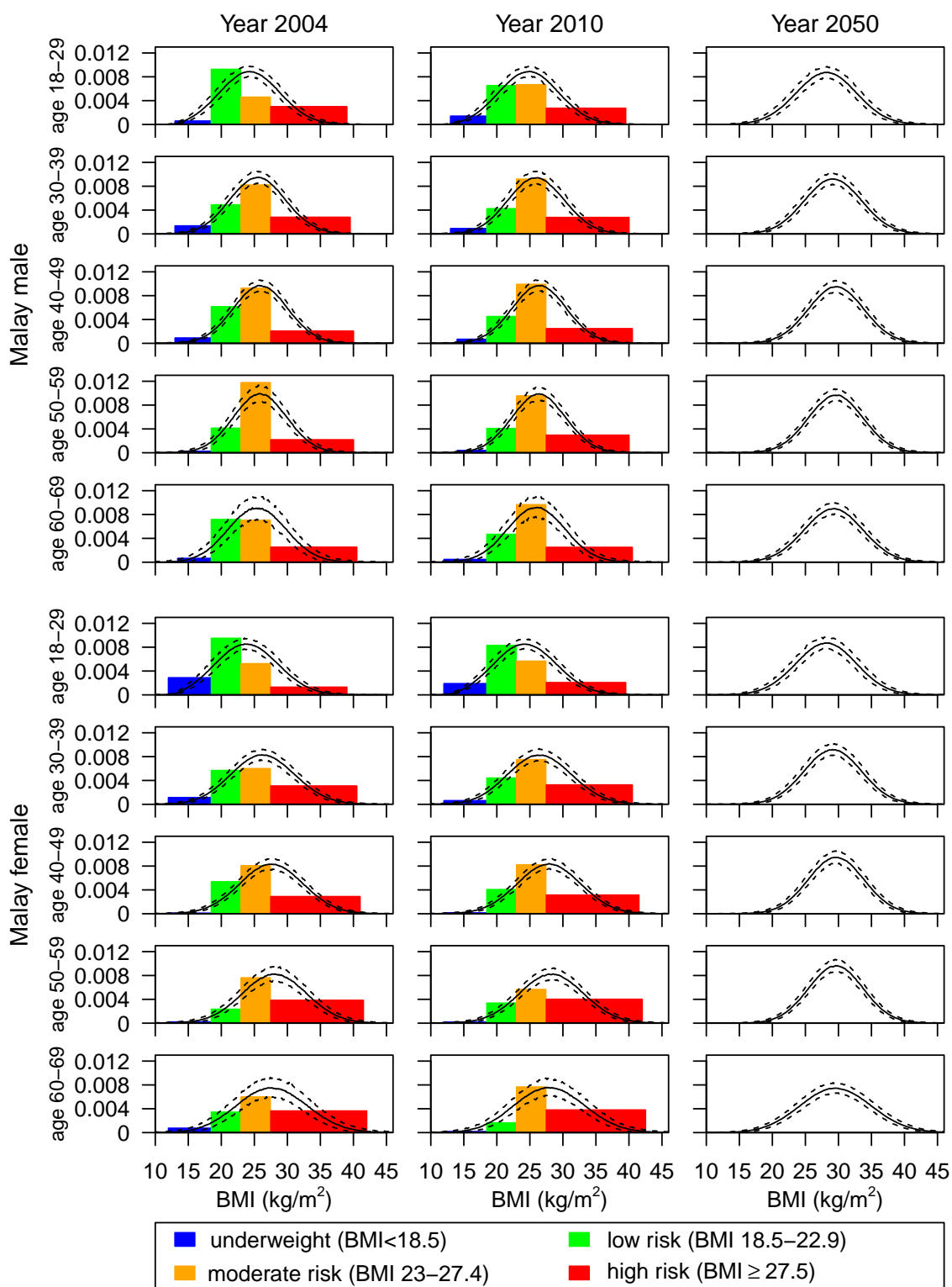
Supplementary Figure S4: Estimated and forecasted BMI distribution of Chinese by age groups and gender and aggregated statistics of 4 BMI risk category status. Mean and 95% prediction intervals are plotted. Data are presented as bars.



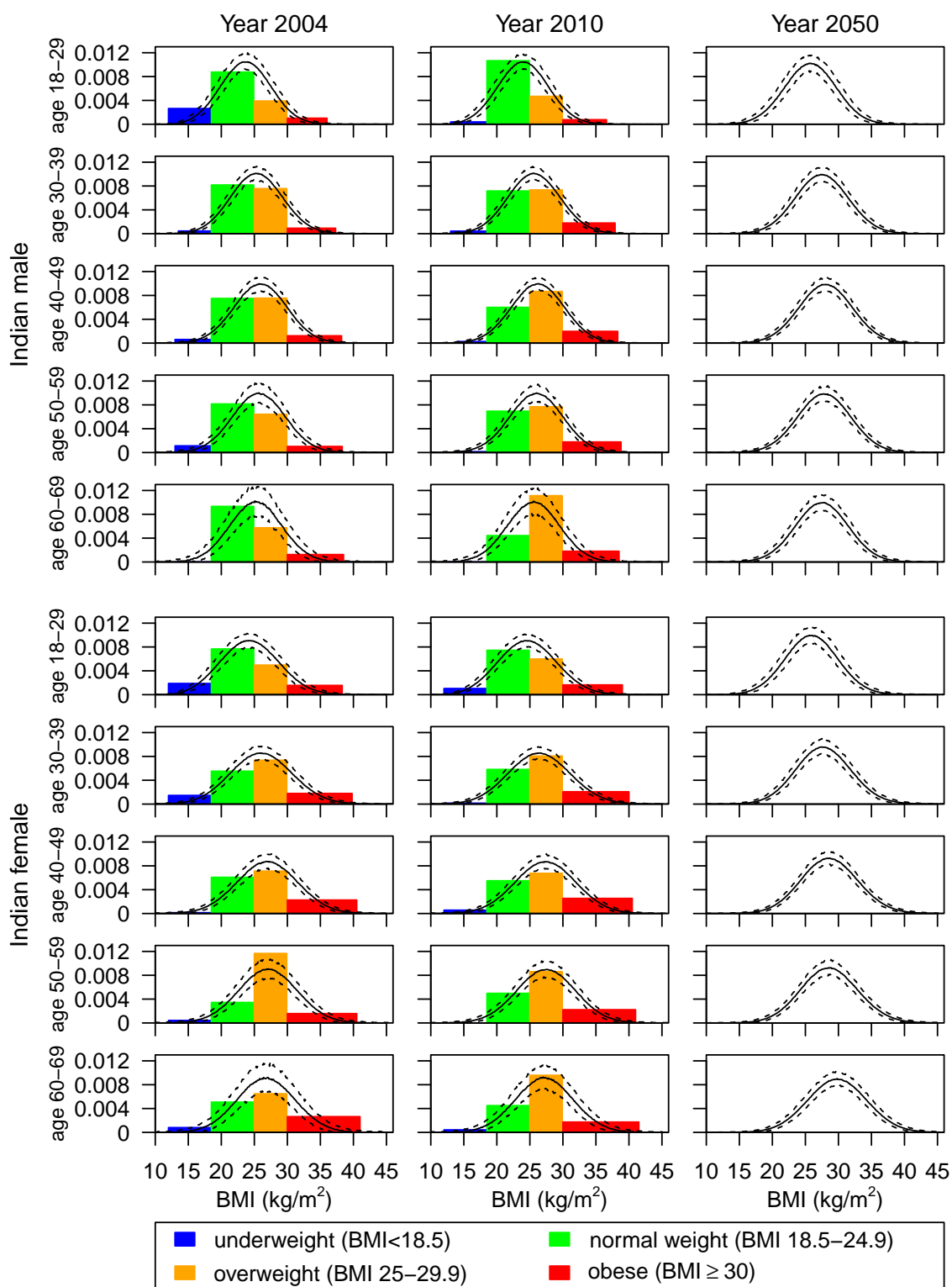
Supplementary Figure S5: Estimated and forecasted BMI distribution of Malay by age groups and gender and aggregated statistics of 4 weight status. Mean and 95% prediction intervals are plotted. Data are presented as bars.



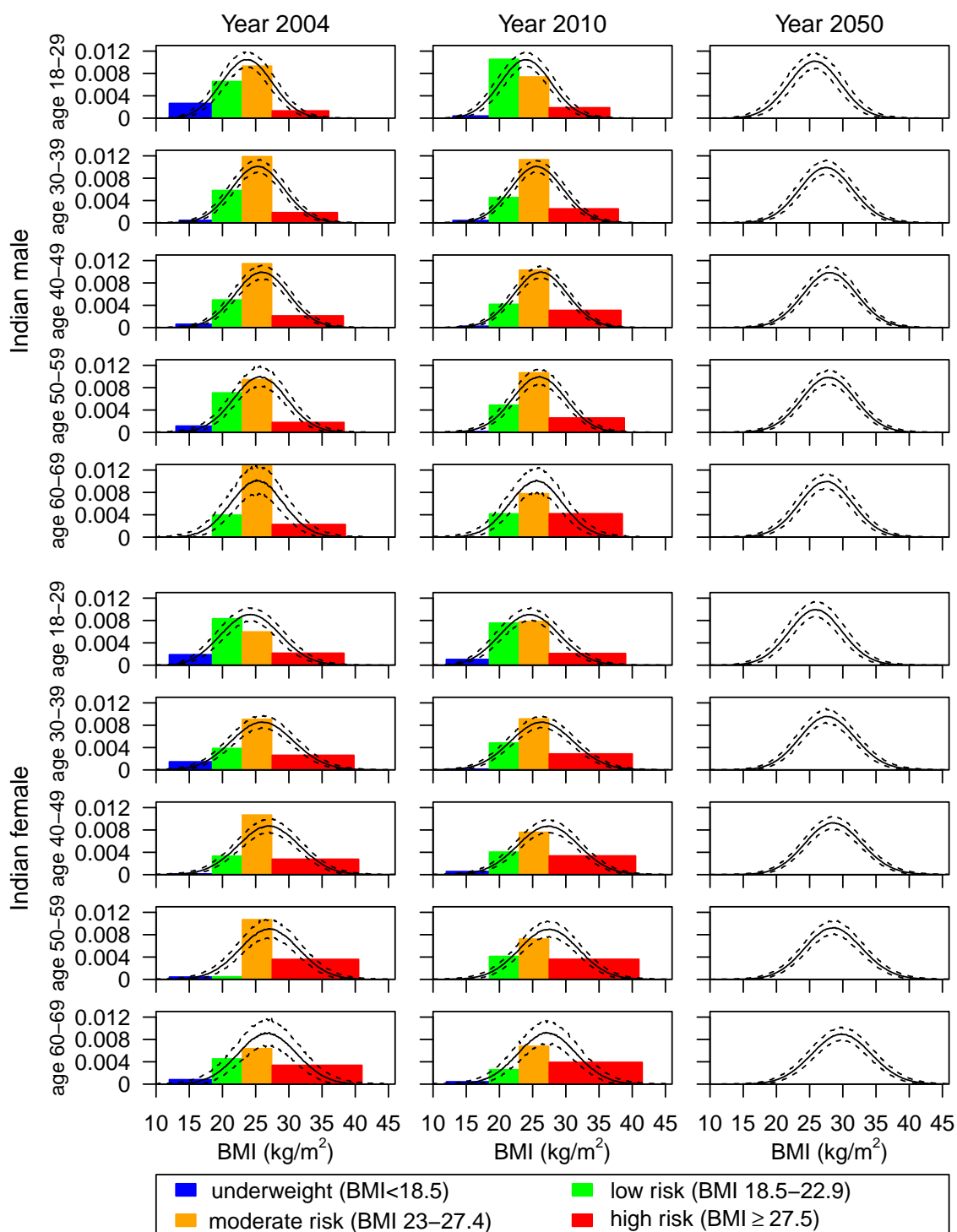
Supplementary Figure S6: Estimated and forecasted BMI distribution of Malay by age groups and gender and aggregated statistics of 4 BMI risk category status. Mean and 95% prediction intervals are plotted. Data are presented as bars.



Supplementary Figure S7: Estimated and forecasted BMI distribution of Indian by age groups and gender and aggregated statistics of 4 weight status.
Mean and 95% prediction intervals are plotted. Data are presented as bars.

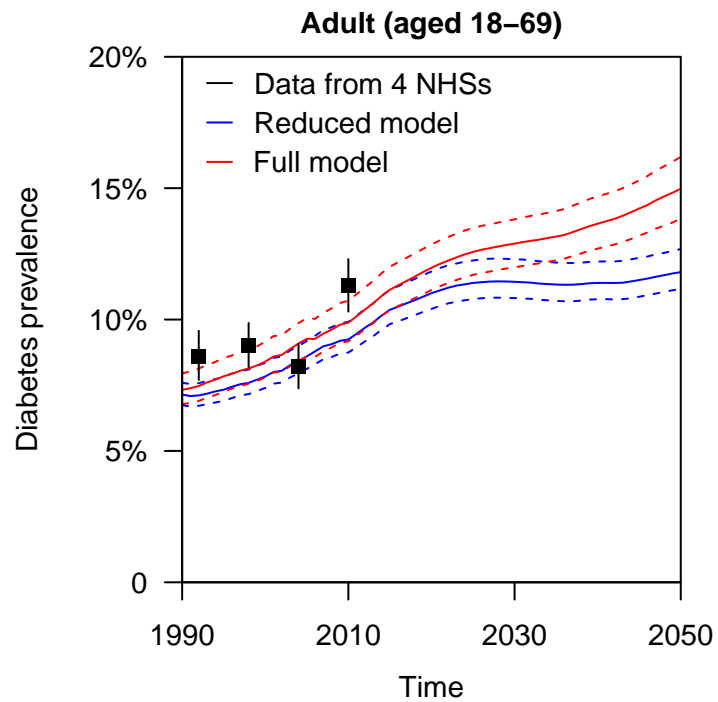


Supplementary Figure S8: Estimated and forecasted BMI distribution of Indian by age groups and gender and aggregated statistics of 4 BMI risk category status. Mean and 95% prediction intervals are plotted. Data are presented as bars.



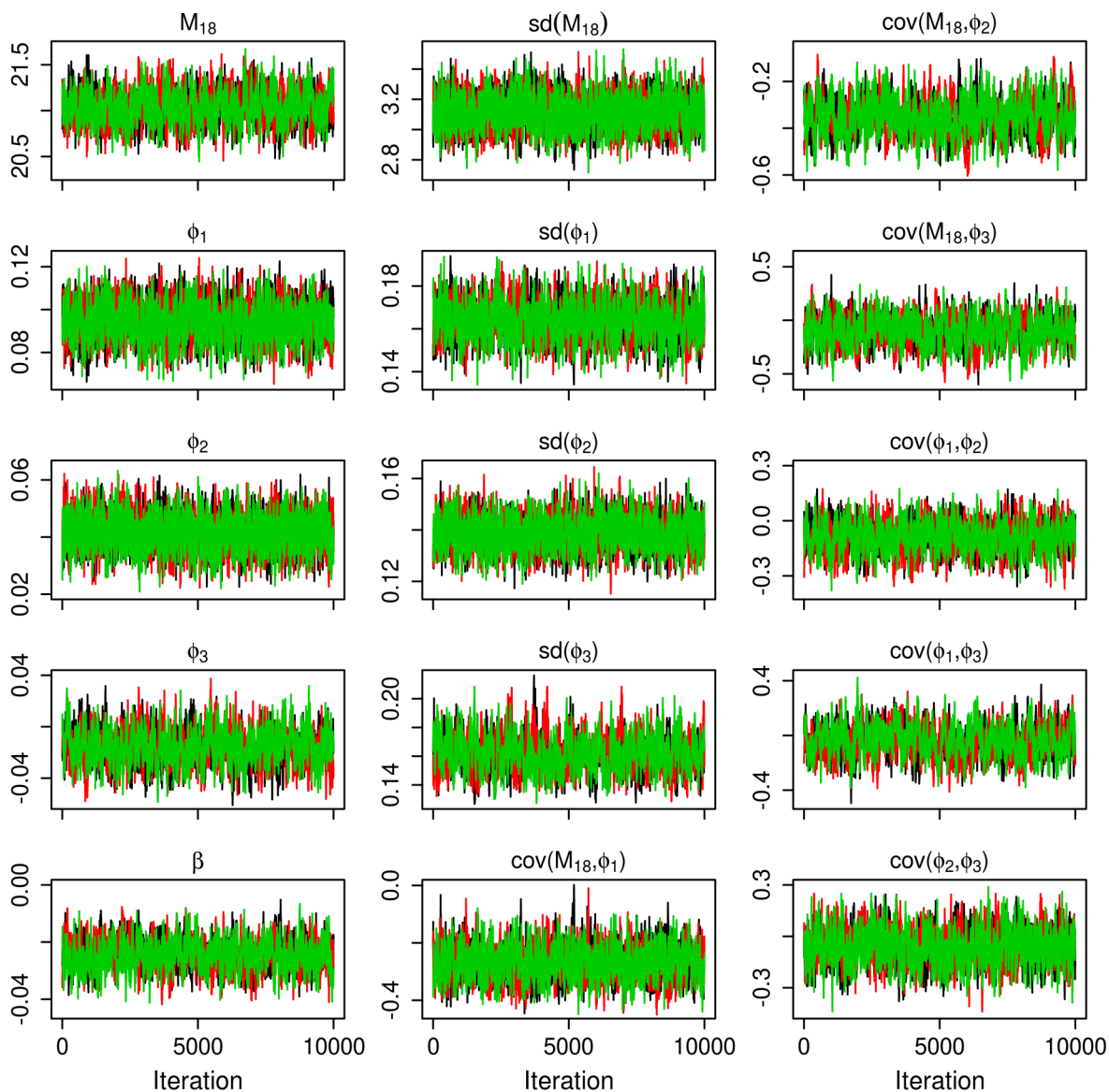
Sensitivity analysis

Supplementary Figure S9: Estimated and forecasted T2DM prevalence among working age adults from 2 models. Mean and 95% prediction intervals are plotted.

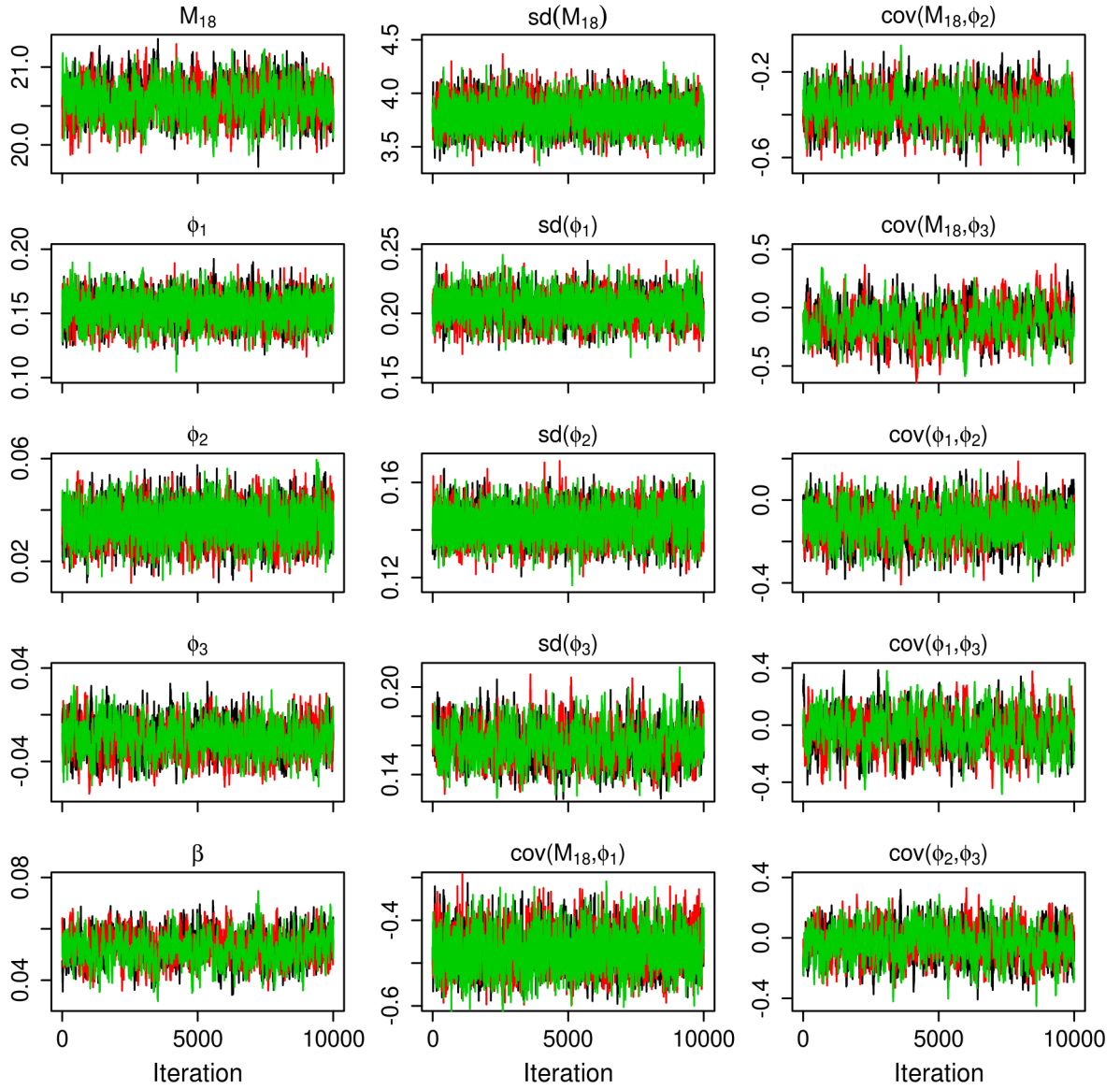


Trace plots for parameter estimation

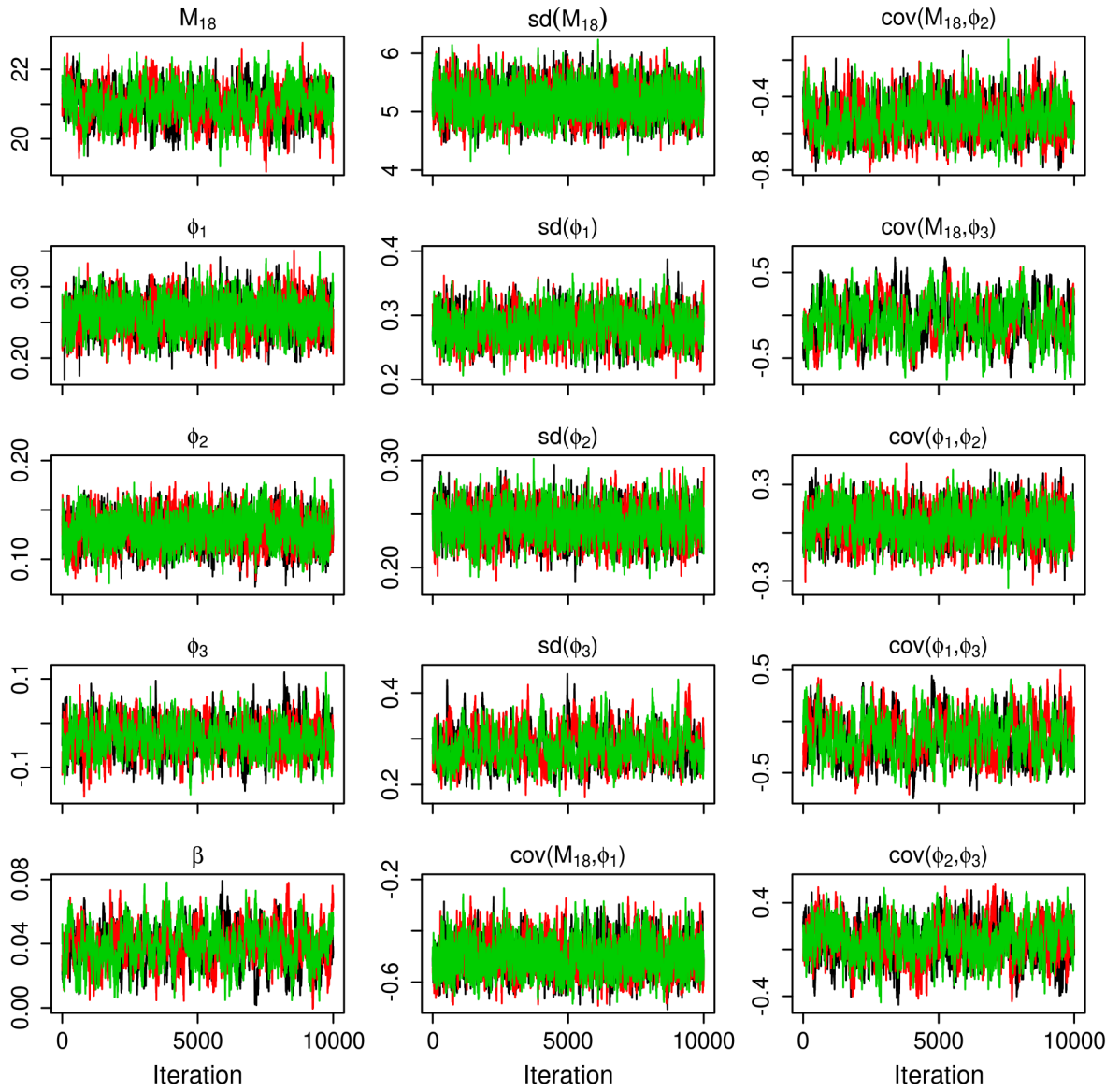
Supplementary Figure S10: Trace plots for parameters of BMI model in Chinese females.



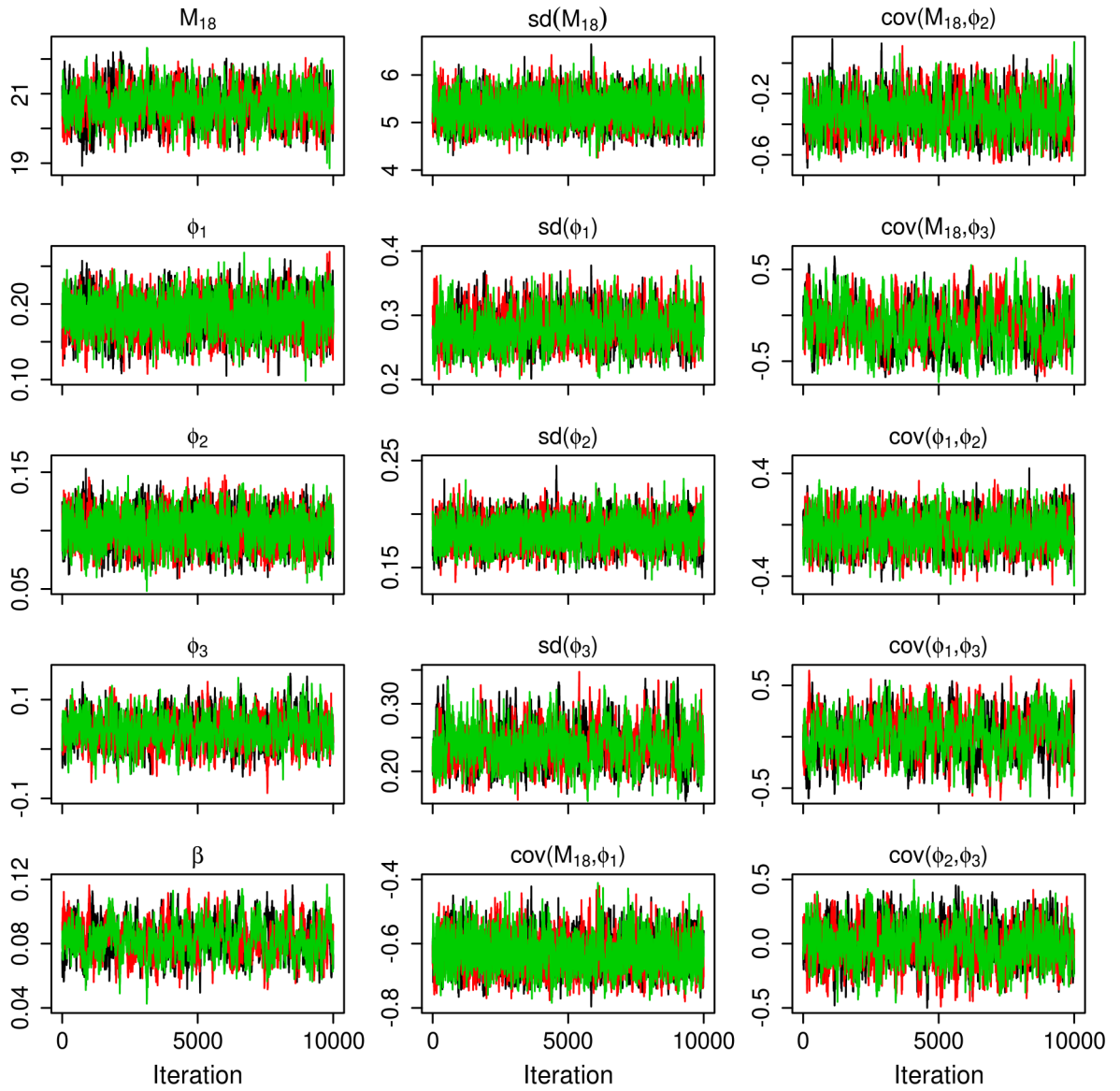
Supplementary Figure S11: Trace plots for parameters of BMI model in Chinese males.



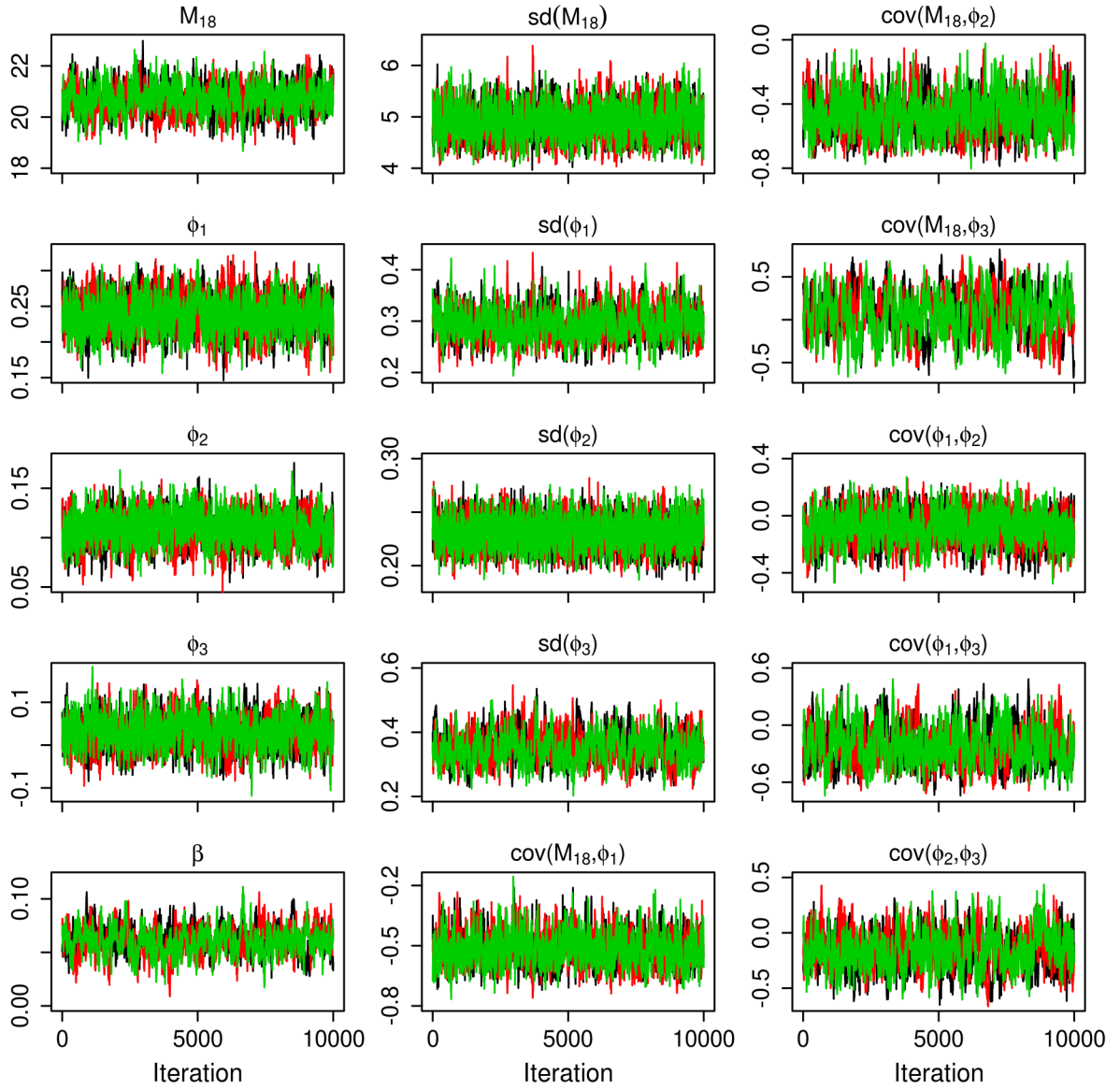
Supplementary Figure S12: Trace plots for parameters of BMI model in Malay females.



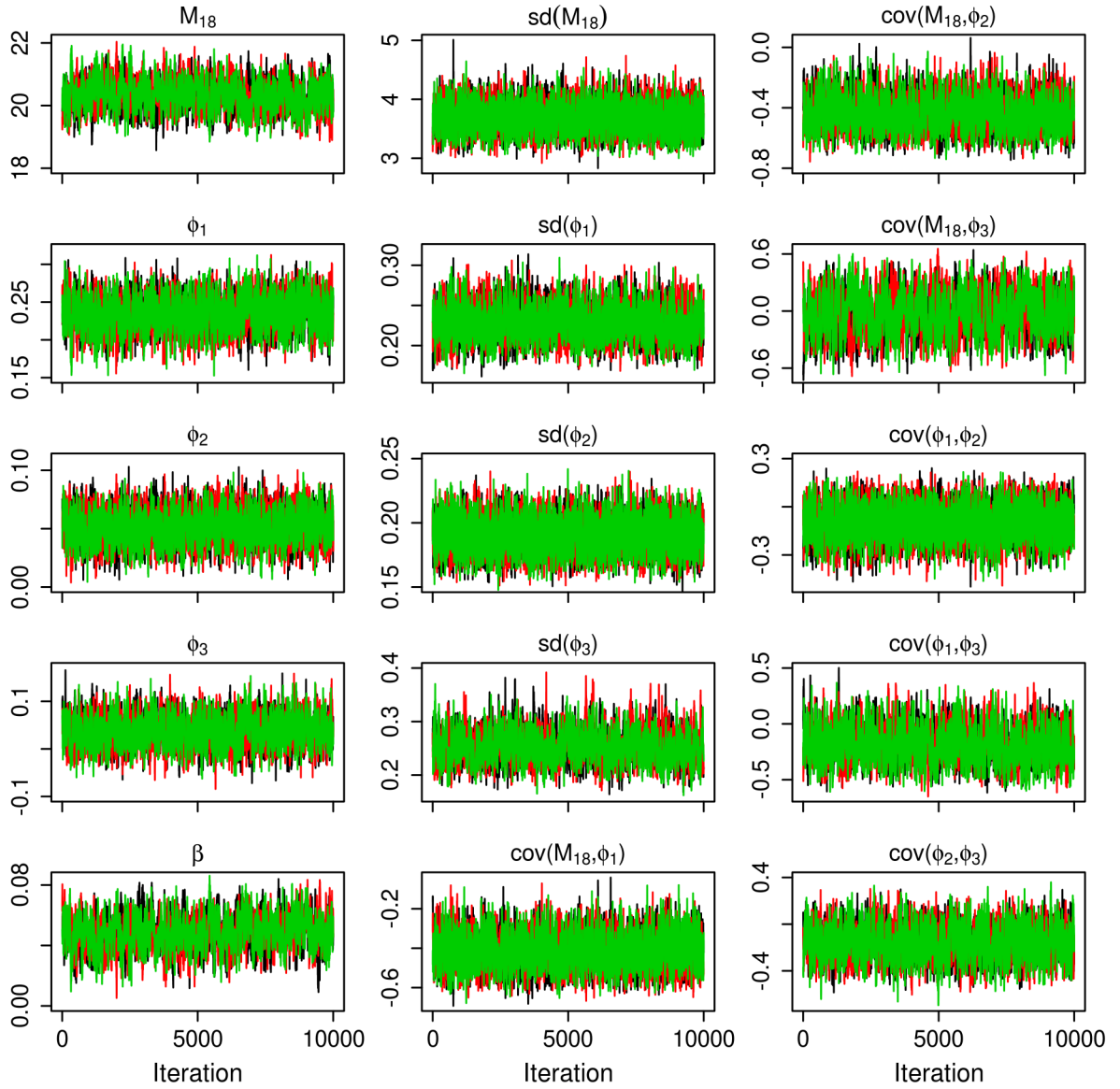
Supplementary Figure S13: Trace plots for parameters of BMI model in Malay males.



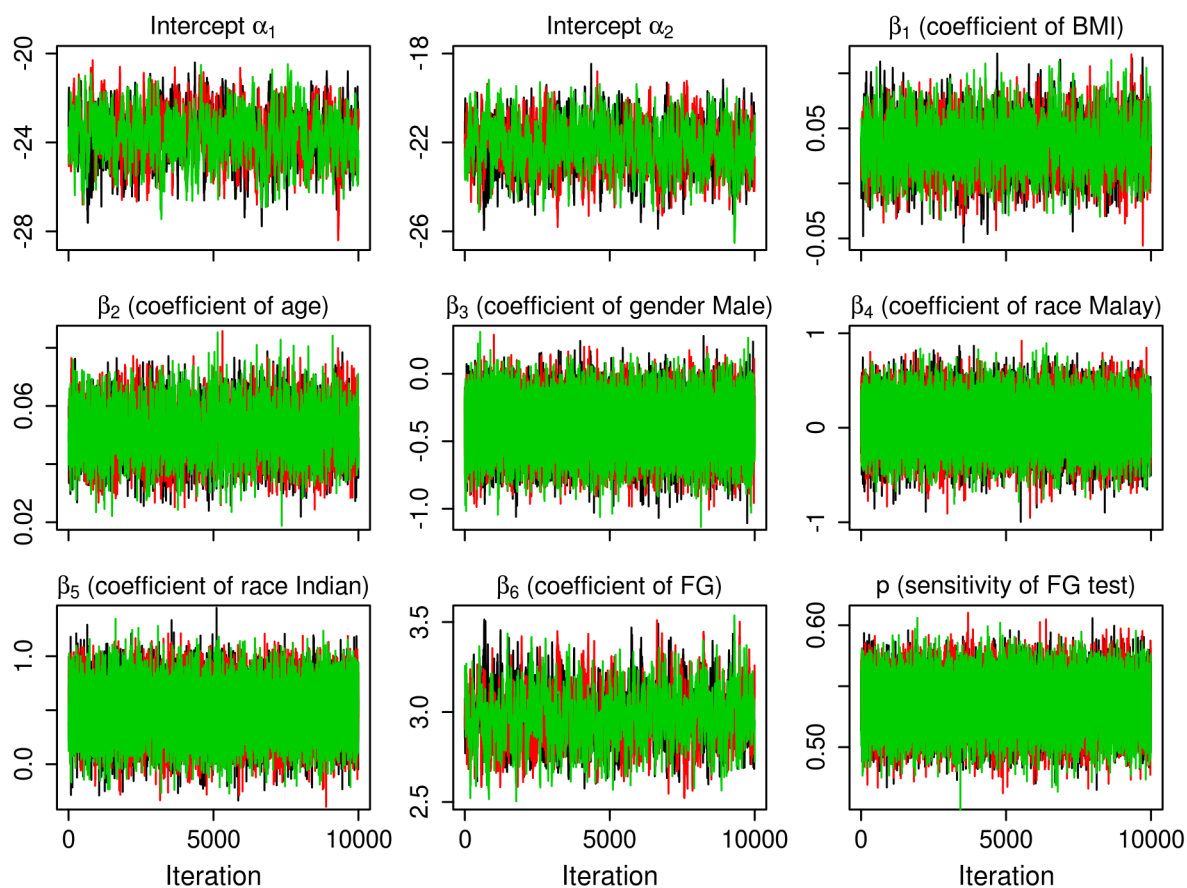
Supplementary Figure S14: Trace plots for parameters of BMI model in Indian females.



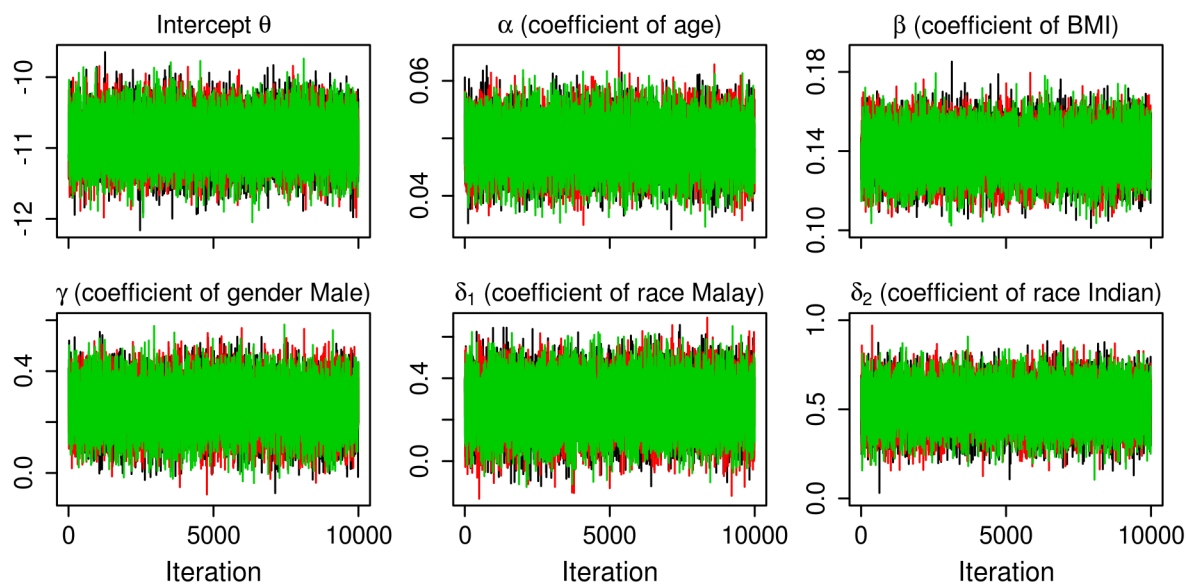
Supplementary Figure S15: Trace plots for parameters of BMI model in Indian males.



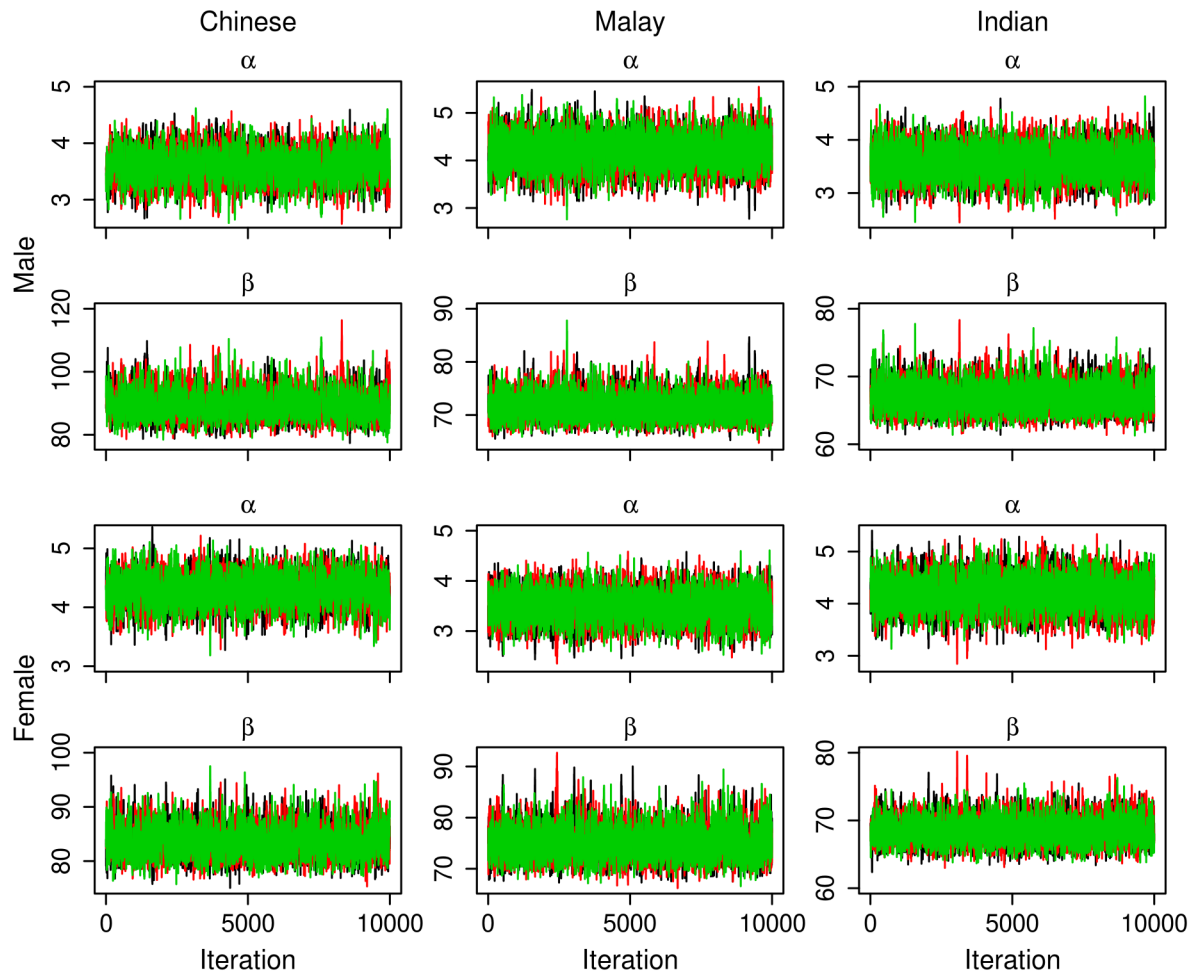
Supplementary Figure S16: Trace plots for parameters of model for correcting diabetic status.



Supplementary Figure S17: Trace plots for parameters of Diabetes Mellitus model.



Supplementary Figure S18: Trace plots for parameters of simple Diabetes Mellitus model without BMI and genetic risk.



REFERENCE

1. Lee RD, Carter LR. Modeling and forecasting US mortality. *J Am Stat Assoc.* 1992;**87**(419):659-81.
2. Pedroza C. A Bayesian forecasting model: predicting US male mortality. *Biostatistics.* 2006;**7**(4):530-50.
3. Castro LJ, Rogers A. Model migration schedules. *International Institute for Applied System Anaalysis, Laxenburg, Austria.*1981.
4. Nang EEK, Khoo CM, Tai ES, et al. Is there a clear threshold for fasting plasma glucose that differentiates between those with and without neuropathy and chronic kidney disease? The Singapore Prospective Study Program. *Am J Epidemiol.* 2009;**169**(12):1454-62.
5. Marjoram P, Molitor J, Plagnol V et al. Markov chain Monte Carlo without likelihoods. *Proc Nat Acad Sci USA.* 2003;**100**(26):15324-8.
6. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003).* March. 2003;20-22.
7. Plummer M. JAGS version 3.1.0 user manual. 2011.
8. R Development Core Team. R: A language and environment for statistical computing. 2013. Available from: <http://www.R-project.org/>.
9. Plummer M. rjags: Bayesian graphical models using MCMC. *R package version 2.0.* 2011.
10. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika.* 1970;**57**(1):97-109.
11. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys.* 1953;**21**:1087.
12. Ma S, Cutter J, Tan CE et al. Associations of diabetes mellitus and ethnicity with mortality in a multiethnic Asian population: data from the 1992 Singapore National Health Survey. *Am J Epidemiol.* 2003;**158**(6):543-52.