

Disentangling age, gender, and racial/ethnic disparities in multiple myeloma burden: a modeling study

Received: 17 December 2022

Accepted: 29 August 2023

Published online: 20 September 2023

 Check for updates

John H. Huber¹✉, Mengmeng Ji¹, Yi-Hsuan Shih², Mei Wang¹,
Graham Colditz¹ & Su-Hsin Chang¹

Multiple myeloma (MM) is a hematological malignancy that is consistently preceded by an asymptomatic condition, monoclonal gammopathy of undetermined significance (MGUS). Disparities by age, gender, and race/ethnicity in both MGUS and MM are well-established. However, it remains unclear whether these disparities can be explained by increased incidence of MGUS and/or accelerated progression from MGUS to MM. Here, we fit a mathematical model to nationally representative data from the United States and showed that the difference in MM incidence can be explained by an increased incidence of MGUS among male and non-Hispanic Black populations. We did not find evidence showing differences in the rate of progression from MGUS to MM by either gender or race/ethnicity. Our results suggest that screening for MGUS among high-risk groups (e.g., non-Hispanic Black men) may hold promise as a strategy to reduce the burden and MM health disparities.

Multiple myeloma (MM) is a malignant transformation of plasma cells that caused an estimated 34,920 diagnoses and 12,410 deaths in the United States in 2021 alone¹. MM is consistently preceded by a pre-malignant condition, monoclonal gammopathy of undetermined significance (MGUS)². A prior cohort study of a predominately white population in Olmsted County, Minnesota suggested that MGUS progresses to MM at a rate of approximately 1% per year with the 20-year cumulative risk totaling 18%³. However, due to the absence of population-based screening and treatment recommendations for MGUS^{4,5}, most MGUS cases are detected incidentally⁶, and, thus, our understanding of MGUS and MM are primarily informed by clinical studies. Further work is needed to understand the natural history of MM and how it varies across age, gender, and race/ethnicity.

One unresolved aspect of the natural history of MM concerns the observed disparities in MM incidence that exist by gender and race/ethnicity⁷. There is a greater burden of MM among men compared to women and non-Hispanic Black people compared to non-Hispanic white people. Additionally, both men and non-Hispanic Black people have a higher prevalence of MGUS^{8–10} and develop MM earlier than women and non-Hispanic white people^{11,12}. However, it remains unclear

whether the increased incidence of MM can be attributed to an increased incidence of MGUS, to an increased rate of progression from MGUS to MM, or a combination of both⁷. An improved understanding of the cause of these observed disparities can potentially guide future screening and treatment strategies aiming to reduce these disparities¹³.

Prior studies have aimed to understand racial and ethnic disparities in MM incidence^{10,11}. Landgren et al.⁹ leveraged the National Health and Nutritional Examination Survey and found that high-risk features of MGUS were more common among non-Hispanic Black people as compared to non-Hispanic white people, suggesting an increased rate of progression from MGUS to MM among non-Hispanic Black people. However, this study lacked long-term outcomes on MM progression and was unable to confirm this finding. Using the Veterans Health Administration database, Landgren et al.¹⁰ examined veterans diagnosed with MGUS during 1980–1996 and found no difference in rate of progression from MGUS to MM among non-Hispanic white people and non-Hispanic Black people. Therefore, they concluded that the higher MM incidence among non-Hispanic Black people could be explained by an increased incidence of MGUS. A more recent analysis

¹Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, USA. ²Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO, USA. ✉ e-mail: huber.j.h@wustl.edu

using the same database demonstrated progression to MM occurred at a younger age among non-Hispanic Black people as compared to non-Hispanic white people in patients diagnosed with MGUS, which may be attributed to the higher incidence of MGUS, the higher progression rate of MGUS to MM, or both in non-Hispanic Black people¹³. Nevertheless, reconciling the conclusions between these two studies is challenging, and, importantly, both studies relied upon study populations with clinically diagnosed MGUS. However, MGUS is primarily asymptomatic, so examining the progression of MM among patients with incidentally diagnosed MGUS excludes those with undiagnosed MGUS and therefore may provide an incomplete picture of the natural history of MM.

To address this, we leveraged two nationally representative databases from the United States on MGUS prevalence using data from the National Health and Nutritional Examination Survey (NHANES) and MM incidence using data from the Surveillance, Epidemiology, and End Results (SEER) program. Importantly, the NHANES tested all study participants for MGUS using serum protein electrophoresis irrespective of underlying comorbidities, providing more representative measures of MGUS prevalence than would be obtained using clinically diagnosed MGUS^{14–16}. We constructed a compartmental model of the natural history of MM and fit this model using the aforementioned MGUS prevalence and MM incidence. We then used the fitted model to isolate the contributions of age, gender, and race/ethnicity to the observed disparities in MM incidence. Finally, we predicted how the preclinical dwell time, defined as the time from MGUS onset to MM onset, is likely to vary by age, gender, and race/ethnicity.

Results

Model fit

The fitted mathematical model of the natural history of MM shows that five independent MCMC chains were well-mixed (Fig. S2), and the Gelman-Rubin statistics for each parameter were 1.0 (Table S2),

providing support that we converged upon on the posterior distribution.

The fitted model captured the trends in MGUS prevalence and MM incidence across age, gender, and race/ethnicity (Fig. 1). Consistent with the data, the fitted model predicted higher MGUS prevalence and MM incidence among non-Hispanic Black people compared to non-Hispanic white people and among men compared to women. Furthermore, the fitted model reproduced the data with appropriate uncertainty. The 95% posterior prediction interval contained all but two data points with most data points falling near the posterior median prediction. Taken together, these results show that our fitted model can explain the patterns in the data.

Contributions of age, gender, and race/ethnicity to the development of MGUS and progression to MM

Independent of gender and race/ethnicity, the rates of development of MGUS and progression to MM increased nonlinearly with age. The rate of MGUS development among healthy individuals monotonically increased with age (Fig. 2a, black line), nearly tripling from 0.0012 (95% Credible Interval (CI): 0.00099–0.0015) yr⁻¹ at age 60 to 0.0034 (0.0022–0.0049) yr⁻¹ by age 80. By comparison, the rate of progression to MM among MGUS-positive individuals monotonically increased up to age 71 (95% CI: 67–77) and then monotonically decreased. Modeling the rate of MM progression as a quadratic relationship with age allowed us to capture the declines in MM incidence seen in the 80–84 and 85+ age groups in the SEER data (Fig. 1, bottom row). Supplementary analyses found that including a quadratic term produced a better model fit to the data on the basis of DIC as compared to an alternative model in which the quadratic term was not included (see Supplementary Information for more details).

We estimated that gender and race/ethnicity modified the rates of development of MGUS and progression to MM with age. For healthy individuals, female gender reduced the rate of MGUS development by

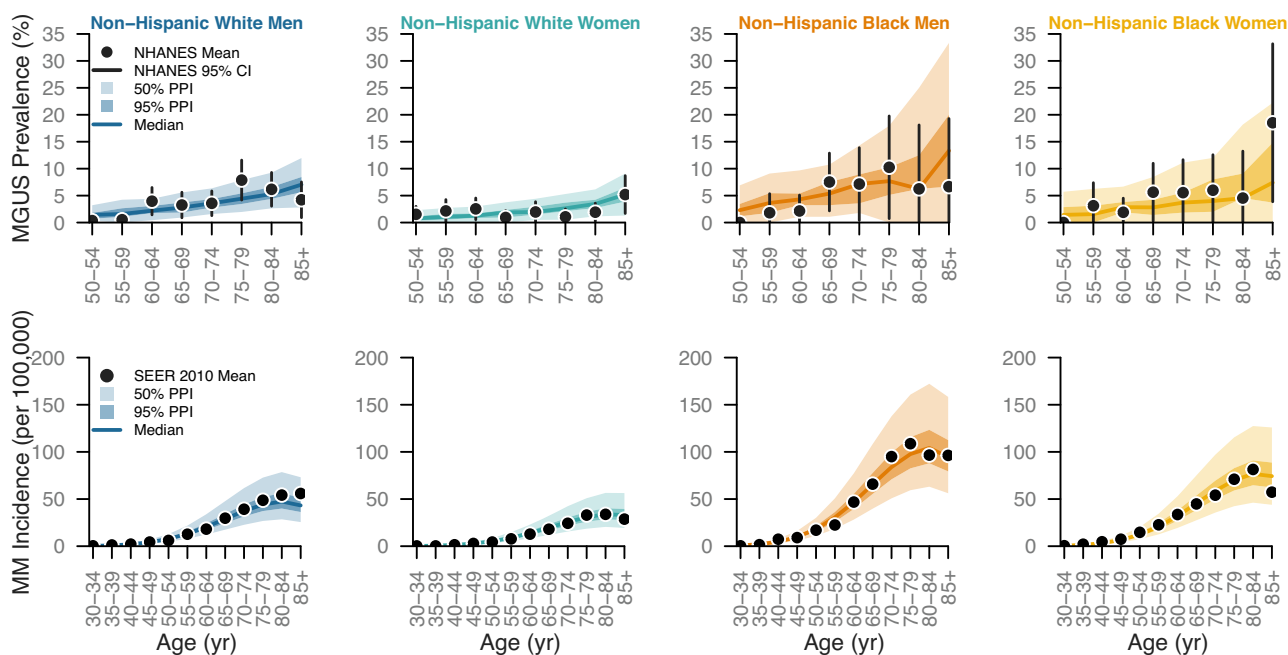


Fig. 1 | Comparison of fitted model to SEER and NHANES data. The model predictions for age-stratified MGUS prevalence and MM incidence are compared against the continuous NHANES 1999–2004 MGUS data (top row) and SEER 2010 MM incidence data (bottom row) for non-Hispanic white men (first column), non-Hispanic white women (second column), non-Hispanic Black men (third column), and non-Hispanic Black women (fourth column). Black points are the data used to fit the model, and the black vertical segments are the 95% confidence intervals on

the data. The 95% confidence intervals were calculated for each sub-population using the following samples sizes: $n = 1735$ non-Hispanic white men, $n = 1,703$ non-Hispanic white women, $n = 454$ non-Hispanic Black men, and $n = 463$ non-Hispanic Black women. The solid line is median posterior model prediction, the darker shaded area is the 50% posterior prediction interval (PPI), and the lighter shaded area is the 95% PPI.

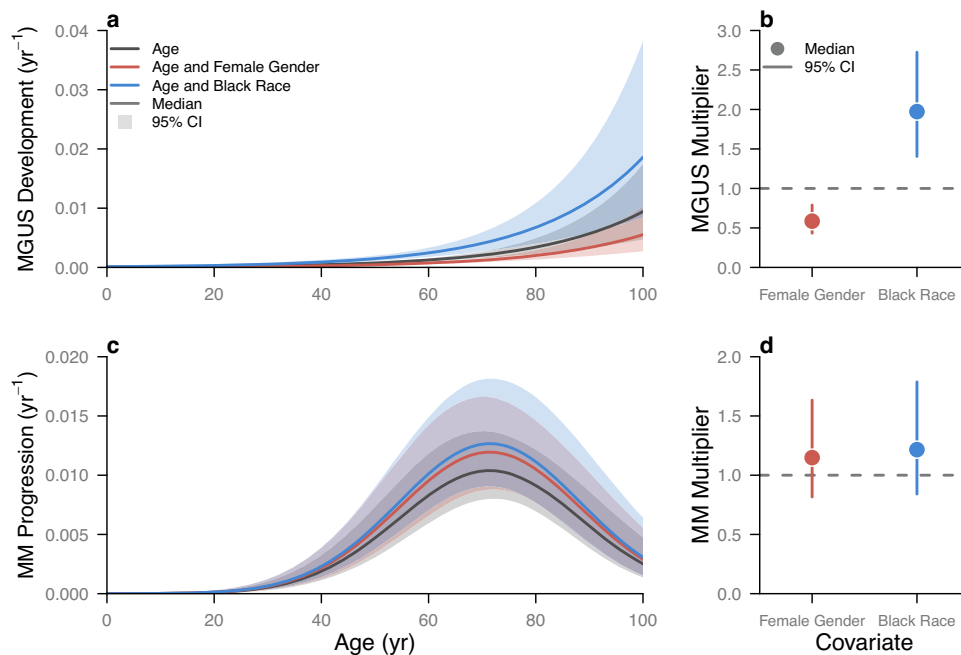


Fig. 2 | Contributions of age, gender, and race/ethnicity to the development of MGUS and progression to MM. **a** The isolated effects of age (black), age and female gender (red), and age and non-Hispanic Black race/ethnicity (blue) are shown for the rate of MGUS development (yr^{-1}). **b** The posterior estimate of the MGUS multiplier is shown for female gender (red) and non-Hispanic Black race/ethnicity (blue). **c** The isolated effect of age (black), age and female gender (red), and age and non-Hispanic Black race/ethnicity (blue) are shown for the rate of MM progression (yr^{-1}). **d** The posterior estimate of the MM multiplier is shown for

female gender (red) and non-Hispanic Black race/ethnicity (blue). In **(a)** and **(c)**, the line is the median estimate, and the shaded region is the 95% credible interval. In **(b)** and **(d)**, the point is the median estimate, and the vertical line segment is the 95% credible interval. Each 95% credible interval is calculated from the $n = 50,010$ posterior samples. The horizontal dotted line is the reference multiplier of one. Values above the reference multiplier suggest that the covariate increases the rate of development/progression, whereas values below the reference multiplier suggest that the covariate decreases the rate of development/progression.

a multiplier of 0.59 (95% CI: 0.43–0.79) (Fig. 2a, red line; Fig. 2b). Similarly, non-Hispanic Black race/ethnicity increased the rate of MGUS development among healthy individuals by a multiplier of 2.0 (95% CI: 1.4–2.7) (Fig. 2a, blue line; Fig. 2b). That the 95% CI for both of these multipliers did not include 1.0 implies that female gender and non-Hispanic Black race/ethnicity significantly affect the rate of MGUS development across all age groups. By comparison, female gender and non-Hispanic Black race/ethnicity did not significantly affect the rate of progression to MM among MGUS-positive individuals (Fig. 2d). The multipliers on progression to MM were 1.1 (95% CI: 0.82–1.6) for female gender and 1.2 (0.84–1.8) for non-Hispanic Black race/ethnicity. Supplementary analyses in which we modified the relationships of the rates of MGUS development and MM progression with age or when we used SEER data from 2004 consistently found a significant effect of female gender and non-Hispanic Black race/ethnicity on MGUS development (see the Supplementary Information for more details).

Duration of the preclinical dwell time

The preclinical dwell time, defined as the time from MGUS onset to MM onset, decreased with increasing age of MGUS onset (Fig. 3). For non-Hispanic white men, the expected preclinical dwell time was 16 (95% CI: 14–17) years at age 50, compared to 1.7 (95% CI: 1.6–1.8) years at age 90. As the age of onset increased, the variation in the expected preclinical dwell time similarly declined. This effect can be explained in part by the proportion of each cohort that survived to develop MM. For MGUS onset at age 50, 20% (95% CI: 16–26%) of non-Hispanic white men were expected to develop MM during their lifetime. By contrast, for MGUS onset at age 90, only 1.7% (95% CI: 1.1–2.6%) of non-Hispanic white men were expected to develop MM prior to death.

We estimated that the preclinical dwell time was affected by the gender and race/ethnicity of each cohort. Female gender was

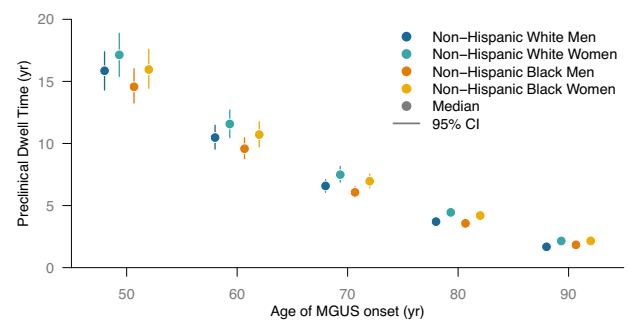


Fig. 3 | Duration of the preclinical dwell time. The posterior median (point) and the 95% credible interval (vertical line segment) is shown for the expected preclinical dwell time (i.e., the time from MGUS onset to MM onset) as a function of age for non-Hispanic white men (dark blue), non-Hispanic white women (light blue), non-Hispanic Black men (dark orange), and non-Hispanic Black women (light orange). Each 95% credible interval is calculated from $n = 50,010$ posterior samples.

associated with an increased dwell time across all ages of MGUS onset. For example, the expected dwell time at age 50 was 17 (95% CI: 15–19) years for non-Hispanic white women versus 16 (95% CI: 14–17) years for non-Hispanic white men and 16 (95% CI: 14–18) years for non-Hispanic Black women versus 15 (95% CI: 13–16) years for non-Hispanic Black men. Furthermore, independent of gender, non-Hispanic Black race/ethnicity was associated with a shorter preclinical dwell time. For instance, at age 70, the expected dwell times were 6.6 (95% CI: 6.1–7.1) years for non-Hispanic white men compared to 6.1 (95% CI: 5.6–6.5) years for non-Hispanic Black men and 7.5 (95% CI: 6.9–8.2) years for non-Hispanic white women compared to 7.0 (95% CI: 6.4–7.5) years for non-Hispanic Black women.

Discussion

Obtaining a detailed understanding of the drivers of MM disparities is challenging, because MGUS, the preclinical state, is asymptomatic and most often detected incidentally². Therefore, clinical studies, which examine progression to MM among patients with diagnosed MGUS exclude those with undiagnosed MGUS, likely biasing their analyses towards MGUS-positive patients with the greatest number of comorbidities^{10,13}. By leveraging nationally representative data on MGUS prevalence^{14–16}, our study calibrated a discrete-time, multistate compartmental model of the natural history of MM that was able to uncover whether the higher incidence of MGUS, the progression rate of MGUS to MM, or both contributed to MM health disparities across age, gender, and race/ethnicity.

Prior studies have revealed that male gender and non-Hispanic Black race/ethnicity are risk factors for MGUS and MM, independent of other factors, such as age and socioeconomic status^{8,9,17,18}. Our fitted model suggests that these disparities in MM incidence can be explained by an increased incidence of MGUS among healthy men and non-Hispanic Black people. Importantly, we found no statistically significant difference in the rate of progression from MGUS to MM, and these results were robust to multiple supplementary analyses that considered alternative models, including one in which there was no effect of MGUS on mortality, as well as alternative years of SEER MM incidence data. That the disparities in MM incidence can be explained by differential rate of development of MGUS suggests that strategies aiming to reduce disparities by gender and race/ethnicity should emphasize interventions that reduce the development of MGUS among high-risk groups. Although we identified an increased rate of development of MGUS amongst men and non-Hispanic Black people, our study cannot provide a mechanistic explanation for this phenomenon. It has been previously suggested that greater background plasma cell activity among non-Hispanic Black people may predispose them to developing MGUS and ultimately MM¹⁹, and mutational signatures may be detectable in the early decades of life²⁰. Alternatively, differences by race/ethnicity may be explained by socio-contextual factors⁷ and differences in the distribution of known risk factors, such as obesity¹³. Future investigation that accounts for these and other hypotheses may eliminate the practice of essentializing race/ethnicity in cancer risk prediction models²¹.

Independent of gender and race/ethnicity, we estimated the rates of developing MGUS and progressing from MGUS to MM increase nonlinearly with age. The rate of MGUS development monotonically increased with increasing age, suggesting that the observed increase in MGUS prevalence with age reflects a concomitant increase in MGUS incidence. This confirms a finding from a prior modeling study fitted to a predominately white cohort in Olmsted County, Minnesota and reveals that this finding is maintained across race/ethnicity as well as gender²². Furthermore, we estimated that the rate of progression from MGUS to MM peaked at approximately 71 years of age and subsequently declined, which mirrors the observed decline in MM incidence at higher age groups and may reflect a subset of older individuals with a more indolent presentation of MGUS and thus lower overall risk of progression to MM. Taken together, these results suggest that, if implemented, prevention strategies for MGUS may be cost-effective at all age groups, whereas prevention of MM among MGUS-positive individuals using pharmacological management, such as metformin²³ or aspirin²⁴, and non-pharmacological management, such as weight loss¹³, may not be cost-effective beyond 71 years of age.

We estimated that the preclinical dwell time, defined as the time from MGUS onset to MM onset, declined nonlinearly with increasing age of MGUS onset. Because the preclinical dwell time requires that individuals with MGUS survive long enough to develop MM, we attribute this phenomenon to two competing effects: (1) the rate of progression from MGUS to MM and (2) the baseline mortality rate. The rate of progression from MGUS to MM increases nonlinearly up to age

71, resulting in a concomitant decline in the preclinical dwell time. Additionally, as individuals age, they are subject to a greater competing risk of mortality. This implies that, at higher ages of MGUS onset, those individuals that survive to progress to MM do so much more quickly, thereby shortening the average preclinical dwell time. This explains apparent differences in the preclinical dwell time between non-Hispanic Black people and non-Hispanic white people. Although we identified no statistically significant difference in the rate of progression from MGUS to MM across race/ethnicity, non-Hispanic Black people are subject to a higher mortality, resulting in shorter preclinical dwell times as compared to non-Hispanic white people.

Our study is subject to a number of limitations. First, we lacked data on smoldering multiple myeloma (SMM), an intermediate state between MGUS and MM, so we are unable to assess how progression to SMM varies across age, gender, and race/ethnicity²⁵. Second, the data from NHANES and SEER are not collected from a single cohort. In order to utilize these distinct data sources within our modeling framework, we assumed that these data sources are nationally representative and thus reflect samples from the true MGUS prevalence and MM incidence within the U.S. population. Given the extensive geographic coverage and large samples sizes of these data sources^{14–16,26}, we believe the assumption is valid. However, the NHANES database is subject to non-response bias, which could affect our conclusions if the prevalence of MGUS was significantly different among those that responded as compared to those that did not respond. MGUS prevalence and MM incidence could also be underreported due to ascertainment bias. This would bias our estimates of the rates of MGUS development and progression from MGUS to MM downward and cause us to underestimate the preclinical dwell time. Additionally, extensions to our modeling framework could simulate multiple cohorts to account for period effects that likely shape MGUS prevalence and MM incidence. We considered only a single year of MGUS prevalence and MM incidence data. Our results were robust to the year of MM incidence data considered, but future work could assess whether increases in MM incidence over time can be attributed to an aging population or instead reflect an increase in MM risk over time. Finally, although we identified differences in MGUS development by gender and race/ethnicity, future work is needed to identify whether these differences can be explained by differences in the distribution of known risk factors, such as obesity¹⁸, across gender and race/ethnicity.

This study found that disparities in MM incidence can be explained by an increased incidence of MGUS, not an increased rate of progression to MM, among healthy men and non-Hispanic Black people. Future studies are needed to identify whether these differences can be explained by differences in the distribution of known risk factors.

Methods

Ethical approval was obtained from Washington University School of Medicine in St. Louis (IRB 202110041). All analyses were performed in accordance with relevant guidelines and regulations.

Model overview

Compartmental model. To model the natural history of multiple myeloma²², we constructed a discrete-time, multistate compartmental model consisting of four health states: healthy (H), monoclonal gammopathy of undetermined significance (MGUS), multiple myeloma (MM), and death (D) (Fig. 4).

For a given birth cohort of gender s and race/ethnicity r , the proportion P of the cohort in each of these states at a given age a is defined by the following set of differential equations:

$$\frac{dP_H}{da} = -\lambda_{\text{MGUS}}(a, s, r)P_H - \mu_H(a, s, r)P_H, \quad (1)$$

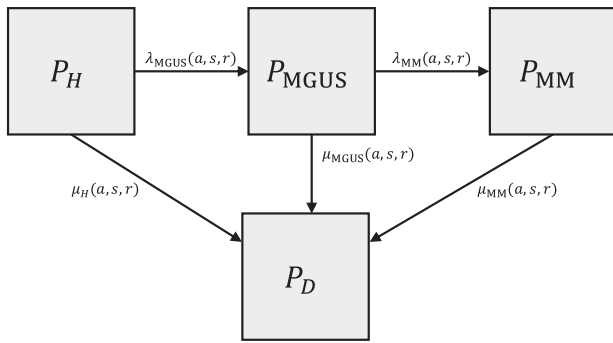


Fig. 4 | Schematic of compartment model. The schematic of the compartmental model of the natural history of multiple myeloma for a birth cohort is shown. Boxes represent the compartments, and arrows represent flows between compartments. P_H is the proportion of the birth cohort that is healthy, P_{MGUS} is the proportion of the birth cohort that has MGUS, P_{MM} is the proportion of the birth cohort that has MM, and P_D is the proportion of the both cohort that has died. $\lambda_{MGUS}(a, s, r)$ is the rate that a healthy individual of age a , gender s , and race/ethnicity r develops MGUS, and $\lambda_{MM}(a, s, r)$ is the rate that an individual of age a , gender s , and race/ethnicity r with MGUS develops MM. $\mu_H(a, s, r)$ is the mortality rate for a healthy individual of age a , gender s , and race/ethnicity r . $\mu_{MGUS}(a, s, r)$ is the mortality rate for an individual of age a , gender s , and race/ethnicity r with MGUS, and $\mu_{MM}(a, s, r)$ is the mortality rate for an individual of age a , gender s , and race/ethnicity r with MM.

$$\frac{dP_{MGUS}}{da} = \lambda_{MGUS}(a, s, r)P_H - \lambda_{MM}(a, s, r)P_{MGUS} - \mu_{MGUS}(a, s, r)P_{MGUS}, \tag{2}$$

$$\frac{dP_{MM}}{da} = \lambda_{MM}(a, s, r)P_{MGUS} - \mu_{MM}(a, s, r)P_{MM}, \tag{3}$$

$$\frac{dP_D}{da} = \mu_H(a, s, r)P_H + \mu_{MGUS}(a, s, r)P_{MGUS} + \mu_{MM}(a, s, r)P_{MM}. \tag{4}$$

In Eq. (1), $\lambda_{MGUS}(a, s, r)$ is the rate that a healthy individual of age a , gender s , and race/ethnicity r develops MGUS, which we computed as

$$\lambda_{MGUS}(a, s, r) = e^{\gamma_{MGUS} + \beta_{MGUS,a}a + \beta_{MGUS,s}s + \beta_{MGUS,r}r}. \tag{5}$$

In Eq. (5), γ_{MGUS} is an intercept term, such that $e^{\gamma_{MGUS}}$ denotes the baseline rate of developing MGUS independent of all other covariates. Additionally, $\beta_{MGUS,a}$, $\beta_{MGUS,s}$, and $\beta_{MGUS,r}$ are the coefficients that modulate the respective effects of age, gender, and race/ethnicity on the rate of MGUS development. In Eq. (2), $\lambda_{MM}(a, s, r)$ is the rate that an individual of age a , gender s , and race/ethnicity r with MGUS develops MM and is calculated as $\lambda_{MM}(a, s, r) = e^{\lambda_{MM} + \beta_{MM,a}a + \beta_{MM,a^2}a^2 + \beta_{MM,s}s + \beta_{MM,r}r}$. Finally, in Eqs. (1–4), $\mu_H(a, s, r)$ is the mortality rate for healthy individuals, $\mu_{MGUS}(a, s, r)$ is the mortality rate for individuals with MGUS, and $\mu_{MM}(a, s, r)$ is the mortality rate for individuals with MM. We followed Therneau et al.²² by defining distinct mortality rates for individuals with MGUS and MM.

Equations (1–4) cannot be solved analytically and must be simulated forward in time numerically. Doing so yields $P_H(a, s, r)$, $P_{MGUS}(a, s, r)$, $P_{MM}(a, s, r)$, and $P_D(a, s, r)$, representing the proportion of the birth cohort of individuals of gender s and race/ethnicity r that occupies each state at age a .

Prevalence and incidence. The quantities $P_{MGUS}(a, s, r)$ and $P_{MM}(a, s, r)$ obtained from Eqs. (1–4) do not represent the respective prevalence of MGUS and MM, because the denominator includes individuals in the birth cohort that previously died. To calculate age-

Table 1 | Characteristics of the data used for modeling fitting

	Dataset	
	Continuous NHANES	SEER
Measure	MGUS Prevalence	MM Incidence
Study Design	Prevalence Survey	Observation
Sample Size	4,355 individuals	17 catchment areas
Date	1999–2004	2010
Age Bins (years)	5	5
Stratification	Gender, Race/ Ethnicity	Gender, Race/ Ethnicity

stratified prevalence p of MGUS and MM among individuals of gender s and race/ethnicity r , we conditioned upon the proportion of the birth cohort that was alive at age a , such that

$$p_{MGUS}(a, s, r) = \frac{P_{MGUS}(a, s, r)}{1 - P_D(a, s, r)}, \tag{6}$$

$$p_{MM}(a, s, r) = \frac{P_{MM}(a, s, r)}{1 - P_D(a, s, r)}. \tag{7}$$

Similarly, calculating age-stratified incidence i of MGUS and MM among individuals of gender s and race/ethnicity r required that we condition upon the proportion of the birth cohort that was alive at age a . Therefore, we calculated $i_{MGUS}(a, s, r)$ and $i_{MM}(a, s, r)$ from Eqs. (1–7) as

$$i_{MGUS}(a, s, r) = \lambda_{MGUS}(a, s, r)p_H(a, s, r), \tag{8}$$

$$i_{MM}(a, s, r) = \lambda_{MM}(a, s, r)p_{MGUS}(a, s, r). \tag{9}$$

In Eq. (8), $p_H(a, s, r)$ is the prevalence of healthy individuals of age a , gender s , and race/ethnicity r , which we computed as $1 - p_{MGUS}(a, s, r) - p_{MM}(a, s, r)$ from Eqs. (6–7).

Model fitting

Data. Using the continuous National Health and Nutritional Examination Surveys (NHANES), 1999–2004^{14–16}, we obtained empirical estimates of MGUS prevalence for 4,355 individuals 50 years of age and older stratified by age, gender, and race/ethnicity (Table 1). We aggregated the continuous NHANES into 5-year age bins in order to increase the number of MGUS-positive samples within each group. This data is the most current nationally representative survey on MGUS prevalence within the United States. Additionally, we obtained age-, gender-, and race/ethnicity-stratified estimates of MM incidence in 2010 from the Surveillance, Epidemiology, and End Results (SEER) Program²⁶. MM incidence from 2010 was chosen specifically because the 6-year gap between the SEER and NHANES datasets approximates the gap between the 50–54 age group (i.e., the youngest age group in continuous NHANES for which MGUS data was available) and the 55–59 age group (i.e., the approximate age group for which MM incidence begins to substantially increase in SEER). To confirm that our conclusions were robust to the year of MM incidence data, we performed an alternative analysis in which used SEER data from 2004 (see sensitivity analysis in the Supplementary Information).

We obtained population estimates by age, gender, and race/ethnicity in 2010 from Centers for Disease Control (CDC) WONDER database²⁷. For individuals 85 years of age or greater, population estimates were aggregated. To disaggregate population estimates for 85+ years, we assumed that the population distribution of individuals 85–99 years of age in 2010 was equivalent to the distribution in 2014, the first year in which CDC WONDER did not aggregate this age group²⁸.

For healthy individuals, we made use of age-, gender-, and race/ethnicity-stratified mortality rates in 2010 from the CDC Life Tables²⁹. For MGUS-positive individuals, we followed Therneau et al.²² and assumed that mortality in MGUS-positive individuals was 1.25 times greater than the baseline age- and race/ethnicity-specific mortality rate for men and 1.11 times greater than the baseline age- and race/ethnicity-specific mortality rate for women (see sensitivity analysis in the Supplementary Information). Finally, for individuals with MM, we estimated gender- and race/ethnicity-stratified all-cause mortality rates by fitting an exponential survival distribution to MM all-cause survival data provided by SEER (Fig. S1). Because these survival curves are derived from the cohort of all MM individuals within SEER, they provide an estimate of the mean all-cause mortality rate among MM individuals, irrespective of treatment characteristics.

Likelihood function. A crucial component of model fitting is the likelihood, which calculates the probability of the model given the data. In this study, the likelihood is equal to the product of the likelihoods for the two data types: (1) MGUS prevalence and (2) MM incidence. Because we adopt a fully Bayesian approach, we do not specify a weight for each data type during model fitting.

MGUS prevalence. We modeled the probability that $y_{a,s,r}$ individuals of age a , gender s , and race/ethnicity r were MGUS-positive among $n_{a,s,r}$ surveyed individuals of age a , gender s , and race/ethnicity r from NHANES as a binomial process, such that

$$\Pr(y_{a,s,r} | n_{a,s,r}, p_{\text{MGUS}}(a, s, r)) = p_{\text{MGUS}}(a, s, r)^{y_{a,s,r}} (1 - p_{\text{MGUS}}(a, s, r))^{n_{a,s,r} - y_{a,s,r}} \tag{10}$$

In Eq. (10), $p_{\text{MGUS}}(a, s, r)$ is the model-predicted prevalence of MGUS among individuals of age a , gender s , and race/ethnicity r from Eqs. (1–4) and Eq. (6).

We aggregated the samples from the NHANES data into 5-year age bins in order to increase the number of MGUS-positive samples within each group. Accordingly, we calculated a weighted prevalence $\bar{p}_{\text{MGUS}}([a_1, a_2], s, r)$ between ages a_1 and a_2 as

$$\bar{p}_{\text{MGUS}}([a_1, a_2], s, r) = \frac{\sum_{a=a_1}^{a_2} N_{a,s,r} p_{\text{MGUS}}(a, s, r)}{\sum_{a=a_1}^{a_2} N_{a,s,r}} \tag{11}$$

where $N_{a,s,r}$ is the size of the subpopulation of age a , gender s , and race/ethnicity r . For age-binned MGUS prevalence, the probability of observing $y_{[a_1, a_2], s, r}$ individuals between the ages of a_1 and a_2 among $n_{[a_1, a_2], s, r}$ total individuals surveyed is then calculated equivalently to Eq. (10) using the weighted predicted prevalence from Eq. (11). Therefore, the likelihood of the model given the NHANES data can be expressed as

$$\begin{aligned} \mathcal{L}(\vec{y}, \vec{n}, \vec{a}_l, \vec{a}_u | \vec{\theta}) \\ = \prod_{i=1}^n \prod_{s \in \{M, F\}} \prod_{r \in \{\text{NHW, NHB}\}} \text{Binomial}(y_{[a_{l,i}, a_{u,i}], s, r} | n_{[a_{l,i}, a_{u,i}], s, r}, \bar{p}_{\text{MGUS}}([a_{l,i}, a_{u,i}], s, r)) \end{aligned} \tag{12}$$

In Eq. (12), $\vec{\theta}$ is the vector of parameters to be estimated, and \vec{y} and \vec{n} are the vectors of NHANES MGUS prevalence data where for each age bin the lower bound is defined by \vec{a}_l and the upper bound is defined by \vec{a}_u .

MM incidence. Because MM incidence was reported as a continuous rate, we modeled the logarithm of MM incidence from SEER in individuals of age a , gender s , and race/ethnicity r as a normal distribution with mean $\log(i_{\text{MM}}(a, s, r))$ and variance τ^2 where $i_{\text{MM}}(a, s, r)$ is the model-predicted MM incidence among individuals of age a , gender s ,

and race/ethnicity r calculated from Eq. (9). We estimated τ^2 as a parameter in our model and assumed that it did not depend upon age, gender, or race/ethnicity.

Similar to the MGUS prevalence data, MM incidence was binned by age. To accommodate this in our likelihood framework, we first calculated a weighted predicted incidence $\bar{i}([a_1, a_2], s, r)$ for individuals between ages a_1 and a_2 as

$$\bar{i}([a_1, a_2], s, r) = \frac{\sum_{a=a_1}^{a_2} N_{a,s,r} i_{\text{MM}}(a, s, r)}{\sum_{a=a_1}^{a_2} N_{a,s,r}} \tag{13}$$

where $N_{a,s,r}$ is the population of individuals of age a , gender s , and r . The likelihood of the SEER MM incidence data can be expressed as

$$\mathcal{L}(\vec{x}, \vec{a}_l, \vec{a}_u | \vec{\theta}) = \prod_{i=1}^n \prod_{s \in \{M, F\}} \prod_{r \in \{\text{NHW, NHB}\}} \text{Normal}(\log(x_{[a_{l,i}, a_{u,i}], s, r}) | i_{[a_{l,i}, a_{u,i}], s, r}, \tau^2) \tag{14}$$

In Eq. (14), $\vec{\theta}$ is the vector of estimated parameters, and \vec{x} is the vector of SEER MM incidence where for each age bin the lower bound is defined by \vec{a}_l and the upper bound is defined by \vec{a}_u .

Priors. We assumed uniform prior distributions for all model parameters (Table 2). The choice of the upper and lower bound for the prior distribution for each parameter was informed by the plausible ranges that yielded real model output. In general, we specified wide upper and lower bounds to allow the inference algorithm the flexibility to explore the parameter space. However, we restricted the prior distribution of $\beta_{\text{MGUS}, a}$ to [0,1]. The lower bound of this distribution was chosen because MGUS prevalence has been observed to increase with age³. Nevertheless, a sensitivity analysis was performed to evaluate how the choice of prior distribution affected the parameter inferences.

Markov chain Monte Carlo. We estimated the parameters of our model from the NHANES and SEER data using a Markov chain Monte Carlo (MCMC) algorithm³⁰. We ran the MCMC for 1,000,000 samples, applied a burn-in of 500,000 samples, and thinned every 50 samples to reduce autocorrelation, thereby obtaining a posterior distribution of 10,000 samples. We assessed convergence by running five chains in parallel and computing the Gelman-Rubin statistic for each parameter, where values less than 1.1 provide statistical support for convergence³¹. Converged chains were pooled to yield a final posterior distribution of 50,000 samples.

Analyses

After fitting the model and comparing the model predictions to the NHANES and SEER data, we used the fitted model to explore the epidemiology of MGUS and MM by age, gender, and race/ethnicity. First, we analyzed the estimated model parameters to isolate the contributions of age, gender, and race/ethnicity to the rates of progression

Table 2 | Parameter definitions and prior distributions

Parameter	Definition	Prior
γ_{MGUS}	Intercept for rate of MGUS development	[-20, 0]
$\beta_{\text{MGUS}, a}$	Age coefficient for rate of MGUS development	[0, 1]
$\beta_{\text{MGUS}, s}$	Gender coefficient for rate of MGUS development	[-15, 5]
$\beta_{\text{MGUS}, r}$	Race/ethnicity coefficient for rate of MGUS development	[-15, 5]
γ_{MM}	Intercept for rate of MM development	[-20, 0]
$\beta_{\text{MM}, a}$	Age coefficient for rate of MM development	[-15, 1]
β_{MM, a^2}	Quadratic age coefficient for rate of MM development	[-15, 1]
$\beta_{\text{MM}, s}$	Gender coefficient for rate of MM development	[-15, 5]
$\beta_{\text{MM}, r}$	Race/ethnicity coefficient for rate of MM development	[-15, 5]
τ^2	Variance of MM incidence	[0, 100]

from healthy to MGUS and from MGUS to MM. Next, we computed the expected duration of the preclinical dwell time (i.e., time from MGUS onset to MM onset) by age, gender, and race/ethnicity. The preclinical dwell time depends upon the rate of progression from MGUS to MM and the competing baseline mortality rate, both of which depend upon age, gender, and race/ethnicity. This relationship between the preclinical dwell time and the baseline mortality rate exists because individuals with MGUS must survive sufficiently long to progress to MM.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Databases used: National Health and Nutritional Examination Survey (NHANES) 1999–2004 (<https://www.cdc.gov/nchs/nhanes/index.htm>); CDC Life Tables (https://www.cdc.gov/nchs/products/life_tables.htm); CDC Wonder Population Database (<https://wonder.cdc.gov/Bridged-Race-v2020.HTML>); SEER*STAT Cancer Incidence (<https://seer.cancer.gov/seerstat/>); All data used in this analysis can be found at <https://zenodo.org/record/8244914>³².

Code availability

All code to reproduce the analyses can be found at <https://zenodo.org/record/8244914>³². Analyses were conducted in R using the BayesianTools and pomp R packages.

References

- Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer Statistics, 2021. *CA Cancer J. Clin.* **71**, 7–33 (2021).
- Cowan, A. J. et al. Diagnosis and Management of Multiple Myeloma: A Review. *JAMA* **327**, 464–477 (2022).
- Kyle, R. A. et al. Prevalence of Monoclonal Gammopathy of Undetermined Significance. *N. Engl. J. Med.* **354**, 1362–1369 (2006).
- Kyle, R. A. et al. Monoclonal gammopathy of undetermined significance (MGUS) and smoldering (asymptomatic) multiple myeloma: IMWG consensus perspectives risk factors for progression and guidelines for monitoring and management. *Leukemia* **24**, 1121–1127 (2010).
- Bories, C. & Jagannath, S. Asymptomatic monoclonal gammopathies. *Clin. Lymphoma Myeloma Leuk.* **14**(Suppl), S78–86 (2014).
- Madan, S. & Greipp, P. R. The incidental monoclonal protein: Current approach to management of monoclonal gammopathy of undetermined significance (MGUS). *Blood Rev.* **23**, 257–265 (2009).
- Marinac, C. R., Ghobrial, I. M., Birmann, B. M., Soiffer, J. & Rebbeck, T. R. Dissecting racial disparities in multiple myeloma. *Blood Cancer J.* **10**, 1–8 (2020).
- Vachon, C. M. et al. Prevalence of heavy chain MGUS by race and family history risk groups using a high-sensitivity screening method. *Blood Adv.* **6**, 3746–3750 (2022).
- Landgren, O. et al. Racial disparities in the prevalence of monoclonal gammopathies: a population-based study of 12 482 persons from the National Health and Nutritional Examination Survey. *Leukemia* **28**, 1537–1542 (2014).
- Landgren, O. et al. Risk of Monoclonal Gammopathy of Undetermined Significance (MGUS) and Subsequent Multiple Myeloma among African-American and White Veterans in the U.S. *Blood* **106**, 1541 (2005).
- Ailawadhi, S. et al. Outcome disparities in multiple myeloma: a SEER-based comparative analysis of ethnic subgroups. *Br. J. Haematol.* **158**, 91–98 (2012).
- Waxman, A. J. et al. Racial disparities in incidence and outcome in multiple myeloma: a population-based study. *Blood* **116**, 5501–5506 (2010).
- Chang, S. H. et al. Obesity and the Transformation of Monoclonal Gammopathy of Undetermined Significance to Multiple Myeloma: A Population-Based Cohort Study. *JNCI J. Natl Cancer Inst.* **109**, djw264 (2017).
- Centers for Disease Control and Prevention (CDC). National Health and Nutrition Examination Survey Data. Published online 1999. <https://www.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=1999>.
- Centers for Disease Control and Prevention (CDC). National Health and Nutrition Examination Survey Data. Published online 2001. <https://www.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2001>.
- Centers for Disease Control and Prevention (CDC). National Health and Nutrition Examination Survey Data. Published online 2003. <https://www.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2003>.
- Benjamin, M., Reddy, S. & Brawley, O. W. Myeloma and race: a review of the literature. *Cancer Metastasis Rev.* **22**, 87–93 (2003).
- Landgren, O. et al. Obesity is associated with an increased risk of monoclonal gammopathy of undetermined significance among black and white women. *Blood* **116**, 1056–1059 (2010).
- Bunce, C. M. & Drayson, M. T. Dissecting racial disparities in multiple myeloma—clues from differential immunoglobulin levels. *Blood Cancer J.* **10**, 1–2 (2020).
- Rustad, E. H. et al. Timing the initiation of multiple myeloma. *Nat. Commun.* **11**, 1917 (2020).
- Waters, E. A., Colditz, G. A. & Davis, K. L. Essentialism and Exclusion: Racism in Cancer Risk Prediction Models. *JNCI J. Natl Cancer Inst.* **113**, 1620–1624 (2021).
- Therneau, T. M. et al. Incidence of Monoclonal Gammopathy of Undetermined Significance and Estimation of Duration Before First Clinical Recognition. *Mayo Clin. Proc.* **87**, 1071–1079 (2012).
- Chang, S. H. et al. Association between metformin use and progression of monoclonal gammopathy of undetermined significance to multiple myeloma in US veterans with diabetes mellitus: a population-based retrospective cohort study. *Lancet Haematol.* **2**, e30–36 (2015).
- Birmann, B. M., Giovannucci, E. L., Rosner, B. A. & Colditz, G. A. Regular aspirin use and risk of multiple myeloma: a prospective analysis in the health professionals follow-up study and nurses' health study. *Cancer Prev. Res Philos. Pa.* **7**, 33–41 (2014).
- Kyle, R. A. et al. Clinical course and prognosis of smoldering (asymptomatic) multiple myeloma. *N. Engl. J. Med.* **356**, 2582–2590 (2007).
- National Cancer Institute. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER Research Data, Nov 2021 Sub (1975-2019) - Linked To County Attributes - Time Dependent (1990-2019) Income/Rurality, 1969-2020 Counties. Published online April 2022.
- National Center for Health Statistics (NCHS). Bridged-Race Population Estimates, United States July 1st resident population by state, county, age, sex, bridged-race, and Hispanic origin. Compiled from 1990-1999 bridged-race intercensal population estimates (released by NCHS on 7/26/2004); revised bridged-race 2000-2009 intercensal population estimates (released by NCHS on 10/26/2012); and bridged-race Vintage 2020 (2010-2020) postcensal population estimates (released by NCHS on 9/22/2021). <http://wonder.cdc.gov/bridged-race-v2020.html>.
- National Center for Health Statistics (NCHS). National Population Projections: United States by Age, Gender, Ethnicity and Race for years 2014-2060, released by the U.S. Census Bureau on December 10, 2014, on CDC WONDER Online Database. Published online 2015. <http://wonder.cdc.gov/population-projections-2014-2060.html>.

29. Arias, E. United States Life Tables, 2010. *Natl Vital Stat Rep* **63**, https://www.cdc.gov/nchs/data/nvsr/nvsr63/nvsr63_07.pdf (2014).
30. Hartig, F., Minunno, F., Paul, S., Cameron, D. & Ott, T. BayesianTools: General-purpose MCMC and SMC samplers and tools for Bayesian statistics. Published online 2019.
31. Gelman, A., & Rubin, D. B. Inference from Iterative Simulation Using Multiple Sequences. *Stat Sci.* **7** <https://doi.org/10.1214/ss/1177011136> 1992
32. Huber, J. et al. Disentangling age, gender, and racial/ethnic disparities in multiple myeloma burden: a modeling study. <https://doi.org/10.5281/zenodo.8244914>.

Acknowledgements

This work was funded by the National Institute of Health's National Cancer Institute Grant Number R01CA453475 (S-H.C.) and as part of the Cancer Intervention and Surveillance Modeling Network (CISNET), Grant Number U01CA265735 (S-H.C. and G.C.). The funder had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication. The article's contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute. The authors thank the Washington University Information Technology's Research Infrastructure Services for computing support.

Author contributions

J.H.H., G.C., and S-H.C. conceived of the study. S-H.C. and G.C. obtained the funding. M.W. and M.J. collected the data. J.H.H. and S-H.C. developed the methods and conducted the analyses. J.H.H., M.J., M.W., Y.S., and S-H.C. analyzed the results. J.H. and S-H.C. wrote the initial draft. All authors provided input into the manuscript and approved of its final form.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-41223-8>.

Correspondence and requests for materials should be addressed to John H. Huber.

Peer review information *Nature Communications* thanks Junjie Huang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023