



# Epidemiological and Economic Evaluation of Disease Burden in the United States: Data, Models, and Applications

## Citation

Li, Yunfei. 2020. Epidemiological and Economic Evaluation of Disease Burden in the United States: Data, Models, and Applications. Doctoral dissertation, Harvard T.H. Chan School of Public Health.

## Permanent link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37365692>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

**EPIDEMIOLOGICAL AND ECONOMIC EVALUATION OF DISEASE BURDEN  
IN THE UNITED STATES: DATA, MODELS, AND APPLICATIONS**

Yunfei LI

A Dissertation Submitted to the Faculty of  
The Harvard T.H. Chan School of Public Health  
in Partial Fulfilment of the Requirements  
For the Degree of *Doctor of Science*  
In the Department of *Global Health and Population*  
Harvard University  
Boston, Massachusetts  
November 2020

## **Epidemiological and Economic Evaluation of Disease Burden in the United States: Data, Models, and Applications**

### **Abstract**

This dissertation is comprised of three studies that evaluate disease burden in the United States, for specific chronic and infectious diseases. These studies use mathematical modelling to synthesize empirical evidence and estimate both epidemiological and economic outcomes. In the first paper, I investigate trends in the prevalence and incidence of diabetes and diabetes diagnosis among adults ages 20 years and older in the United States, over the period 2000-2016. Using an age-stratified Markov model of undiagnosed and diagnosed diabetes, I examine trends in true incidence of diabetes and the diagnosis rates. The model is estimated using repeated cross-sectional survey data on the prevalence of undiagnosed and diagnosed diabetes from the National Health and Examination Survey (NHANES) 1988-1994 and 1999-2016. In the second paper, I develop a novel model to assess the 10-year risk of fatal-plus-non-fatal cardiovascular disease (CVD) for patients with type 2 diabetes mellitus in the United States. This model is constructed as a sex-and-cohort stratified Cox proportional-hazards model using pooled data on fatal-plus-non-fatal CVD outcomes from 5 prospective cohorts. The resulting risk prediction equation provides more accurate predictions of total CVD risk compared to current risk scores and can be used in future comparative effectiveness and cost-effectiveness analyses to simulate outcomes of primary intervention policies targeted toward diabetic populations. In the third paper, I investigate disparities in health and economic outcomes associated with *N. gonorrhoeae* infection in the US in 2015. With probability tree models, I quantify the lifetime quality-adjusted life-years and costs associated with gonorrhea and its sequelae in the US and examine disparities in burden across race/ethnicity. These three studies report findings of substantive importance within each disease area. They also illustrate the utility of mathematical models for synthesizing data, estimating outcomes that would be difficult or impossible to measure empirically, and answering questions directly relevant to the goals of planning and prioritizing prevention policies.

## Contents

<b>Abstract</b> .....	<b>ii</b>
<b>List of Figures</b> .....	<b>iv</b>
<b>List of Tables</b> .....	<b>v</b>
<b>Acknowledgements</b> .....	<b>vii</b>
<b>Chapter 1. Introduction</b> .....	<b>1</b>
References .....	6
<b>Chapter 2. Prevalence and incidence trends for diabetes and diabetes diagnosis among adults ages 20 years and older, United States, 2000-2016: results of a mathematical model</b> .....	<b>9</b>
Abstract.....	10
Background .....	12
Methods.....	13
Results.....	18
Discussion.....	27
References .....	31
Supplementary Material .....	36
Supplementary Material References .....	45
<b>Chapter 3. Risk score to predict cardiovascular disease risk for patients with type 2 diabetes mellitus in the United States: a pooled analysis of prospective cohorts</b> .....	<b>46</b>
Abstract.....	47
Background .....	49
Methods.....	50
Results.....	56
Discussion.....	60
References .....	63
Supplementary Material .....	71
<b>Chapter 4. Disparities in health and economic outcomes associated with <i>N. gonorrhoeae</i> infection in the United States: costs and quality-adjusted life-years lost in 2015</b> .....	<b>77</b>
Abstract.....	78
Background .....	80
Methods.....	81
Results.....	90
Discussion.....	97
References .....	99
Supplementary Material .....	106
Supplementary Material References .....	119

<b>Chapter 5. Conclusion .....</b>	<b>122</b>
------------------------------------	------------

## List of Figures

<b>Chapter 1. Introduction.....</b>	<b>1</b>
<b>Chapter 2. Prevalence and incidence trends for diabetes and diabetes diagnosis among adults ages 20 years and older, United States, 2000-2016: results of a mathematical model .....</b>	<b>9</b>
Figure 2.1. Disease states in the Markov model of undiagnosed and diagnosed diabetes.* .....	14
Figure 2.2. Trends in age-standardized prevalence of undiagnosed, diagnosed, and all diabetes, incidence of diabetes and diagnosed diabetes, and rates of diabetes diagnosis in the overall population, men and women ages 20 years and older, 2000-2016. ....	24
Figure 2.3. Trends in prevalence of undiagnosed, diagnosed, and all diabetes, incidence of diabetes and diagnosed diabetes, and rates of diabetes diagnosis in men ages 20 years and older, by four age groups, 2000-2016. ....	25
Figure 2.4. Trends in prevalence of undiagnosed, diagnosed, and all diabetes, incidence of diabetes and diagnosed diabetes, and rates of diabetes diagnosis in women ages 20 years and older, by four age groups, 2000-2016. ....	26
Figure S2.1. A four-state Markov model of undiagnosed and diagnosed diabetes (Panel A) and states transition matrix for individuals at single year of age (i) and single calendar year (t).....	40
Figure S2.2. Example of B-spline basis with three different levels of penalty $\theta = 0, \theta = 1, \theta = 55$ . ..	41
Figure S2.5. Model fits evaluation: proportion of observed NHANES data covered by percentage of model posterior predictive intervals by outcomes and sex.....	44
<b>Chapter 3. Risk score to predict cardiovascular disease risk for patients with type 2 diabetes mellitus in the United States: a pooled analysis of prospective cohorts.....</b>	<b>46</b>
Figure 3.1. Flowchart of inclusion and exclusion of cohort participants .....	52
Figure 3.2. Observed and predicted 10-year risk of a cardiovascular disease event in risk score internal validation, by deciles of risk and sex. ....	58
<b>Chapter 4. Disparities in health and economic outcomes associated with <i>N. gonorrhoeae</i> infection in the United States: costs and quality-adjusted life-years lost in 2015.....</b>	<b>77</b>
Figure 4.1. Probability tree for sequelae following gonococcal infection among women. Abbreviations: PID = pelvic inflammatory disease, CPP = chronic pelvic pain, EP = ectopic pregnancy, TI = tubal infertility. ....	84
Figure 4.2. Probability tree for sequelae following gonococcal infection among MSW (Panel A) and MSM (Panel B), with complications of untreated urethral infections shown in Panel C. Abbreviations: DGI = disseminated gonococcal infection. ....	84
Figure 4.3. Population-level numbers of QALYs lost, QALYs lost per 1,000 population, total costs (US\$), total costs per 1,000 population, for women, by age and race/ethnicity, in 2015. ....	93
Figure 4.4. Population-level numbers of QALYs lost, QALYs lost per 1,000 population, total costs (US\$), total costs per 1,000 population, for heterosexual men, by age and race/ethnicity, in 2015. ....	94

Figure 4.5. Population-level numbers of QALYs lost, QALYs lost per 1,000 population, total costs (US\$), total costs per 1,000 population, for men who have sex with men (MSM), by age, in 2015. ....	94
Figure 4.6. Decomposition of total lifetime QALYs lost and costs, by main complications and by broad cost categories.....	96
<b>Chapter 5. Conclusion .....</b>	<b>122</b>

## List of Tables

<b>Chapter 1. Introduction.....</b>	<b>1</b>
<b>Chapter 2. Prevalence and incidence trends for diabetes and diabetes diagnosis among adults ages 20 years and older, United States, 2000-2016: results of a mathematical model .....</b>	<b>9</b>
Table 2.1. Prevalence per 100 persons of undiagnosed, diagnosed, and all diabetes in the U.S. general population ages 20 years and older in 2016, and annual percentage changes of the prevalence of undiagnosed, diagnosed, and all diabetes over 2000-2016, by age group and sex. .	19
Table 2.2. Incidence per 1,000 person-years of diabetes, rates of diabetes diagnosis, and incidence of diagnosed diabetes for ages 20 years and older in the U.S. general population in 2016 and annual percentage changes of incidence of diabetes, rates of diabetes diagnosis, and incidence of diagnosed diabetes over 2000-2016, by age group and sex.....	21
Table S2.1. Description of participants in the analytic sample of interview questionnaire and laboratory examination for NHANES III (1988-1994) and Continuous NHANES (1999-2016).....	36
Table S2.2. The risk ratios for all-cause mortality of undiagnosed and diagnosed diabetes compared to the general population without diabetes, by age groups and sex.....	41
<b>Chapter 3. Risk score to predict cardiovascular disease risk for patients with type 2 diabetes mellitus in the United States: a pooled analysis of prospective cohorts.....</b>	<b>46</b>
Table 3.1. Information about cohorts used in estimating the coefficients of the risk score.....	54
Table 3.2. Coefficients and hazard ratios from the Cox proportional hazard model used to predict the 10-year risk of fatal-plus-non-fatal CVD for patients with T2DM.....	57
Table 3.3. Validation results of the CVD risk model for patients with T2DM.....	58
Table S3.1. Test of proportional-hazards assumptions for risk factors in CVD risk model for patients with T2DM.....	71
Table S3.2. Coefficients and hazard ratios from the Cox proportional hazard model used to predict the 10-year risk of CVD for our study populations, with all sex and risk factors interactions .....	72
Table S3.3. Coefficients and hazard ratios from the Cox proportional hazard model used to predict the 10-year risk of CVD for our study populations, when hypertension and treatment with hypertension medications were included as a four categorical variable. ....	73
Table S3.4. Coefficients and hazard ratios from the Cox proportional hazard model used to predict the 10-year risk of CVD for our study populations, when elevated systolic blood pressure and treatment with hypertension medications were included as a four categorical variable.....	74
Table S3.5. Coefficients and hazard ratios from the Cox proportional hazard model used to predict the 10-year risk of CVD for our study populations, with TC/HDL-C ratio covariate. ....	75

Table S3.6. Comparison of 10-year CVD observed risk with 10-year ASCVD predicted risk by the 2013 ACC/AHA Pooled Cohort Risk Equation <sup>1</sup> in our study population from the five prospective cohorts, by race and sex. ....	76
Table S3.7. Validation results of the 10-year CVD risk model for participants with T2DM in the pooled four prospective cohorts (MRFIT excluded). ....	76
<b>Chapter 4. Disparities in health and economic outcomes associated with <i>N. gonorrhoeae</i> infection in the United States: costs and quality-adjusted life-years lost in 2015.....</b>	<b>77</b>
Table 4.1. Model input parameters describing gonorrhea and sequelae probabilities, proportions of testing and treatment for gonorrhea in women, MSW and MSM. ....	87
Table 4.1. Model input parameters describing gonorrhea and sequelae probabilities, proportions of testing and treatment for gonorrhea in women, MSW and MSM (Continued).....	88
Table 4.1. Model input parameters describing gonorrhea and sequelae probabilities, proportions of testing and treatment for gonorrhea in women, MSW and MSM (Continued).....	89
Table 4.2. Estimated number of discounted lifetime QALYs lost associated with gonorrhea, per incident gonococcal infection in 2015, by sex and age-group.....	91
Table 4.3. Estimated discounted lifetime costs associated with gonorrhea, per incident gonococcal infection in 2015 (in 2018 US dollars), by sex and age-group. ....	91
Table S4.1. Estimated probabilities of sequelae. ....	109
Table S4.2. Quality weight and duration of PID and its sequelae.....	110
Table S4.3. Weighted utilities and durations of inpatient treatment for PID.....	111
Table S4.4. Discounted life years lived and QALYs lost due to CPP with lifetime duration, for a woman who acquires gonorrhea age at 20 years in 2015 and experiences CPP beginning at age 25 years.....	113
Table S4.5-(a). Estimated probabilities of diagnosis and treatment for asymptomatic gonococcal infections for heterosexual women by age and race/ethnicity over 2000-2015. ....	114
Table S4.5-(b). Estimated probabilities of diagnosis and treatment for asymptomatic gonococcal infections for heterosexual men by age and race/ethnicity, and for MSM by age, over 2000-2015. ....	115
Table S4.6. Estimated number of discounted lifetime QALYs lost and costs associated with gonorrhea, per incident infection in 2015 by race/ethnicity groups.....	116
Table S4.7. Total number of discounted lifetime QALYs lost, and discounted lifetime QALYs lost per 1,000 person-years associated with gonococcal infections that occurred in 2015, by sex, age-groups and race/ethnicity.....	117
Table S4.8. Total discounted lifetime costs, and lifetime costs per 1,000 person-years associated with gonococcal infections that occurred in 2015 (in 2018 US dollars), by sex, age-groups and race/ethnicity.....	118
<b>Chapter 5. Conclusion .....</b>	<b>122</b>

## Acknowledgements

First and foremost, I would like to express my profound gratitude to my advisor, Professor Joshua A. Salomon, for guiding me through my master and doctoral graduate study at Harvard. I met almost all of the difficult but important challenges in my life at the early stage of my doctoral study, but one of the luckiest things happened to me is having Professor Salomon as my advisor. He introduced me to the world of global health and decision science and led my way to becoming an independent researcher. He guided me from combing through broad research questions to acquiring detailed research skillsets. His unreserved mentorship not only benefited my study and research work, but will keep illuminating my future career trajectories.

Secondly, I would like to express my truly appreciation to Professor Goodarz Danaei for being an invaluable source of advice and help on both research and career planning. His broad and deep knowledge of non-communicable disease research helped me develop research ideas and proposals that contributed to my thesis. In addition, I would like to sincerely thank Professor Nicolas A. Menzies for his exceptional support during my doctoral studies. His broad and deep understanding of disease modelling, with many insightful comments helped me improve and polish my thesis by a great extent. I could not have gotten this far without the advice and guidance from all my committee members.

I would like to express my special thanks to Barbara Heil and Allison Conary for keeping me on track and making sure I get things done. I would like to thank my outside reader, Professor Philip Clarke for his time and efforts in reviewing this thesis and providing constructive and detailed comments. I am truly grateful to Professor Winnie Chi-Man Yip, Professor Ankur Pandya, and Professor Mary C. Smith-Fawzi, for many thought-provoking conversations during my graduate study and assistant teaching experience at Harvard.

I would particularly like to thank collaborators Minttu M. Rönn, Yelena Malyuta, Christian Testa, Meghan E. Bellerose, Lin Zhu from Harvard Prevention Policy Modeling Lab and Harrell W. Chesson



and Tomas L. Gift from Division of STD Prevention, for their work and valuable comments on the last project of my thesis.

It was a great honor to have been awarded the China Medical Board and the Institute of International Education Next Generation Fellowship for my graduate study at Harvard. I sincerely appreciate Dr. Lincoln Chen and Zhe Zhang's support and advice on my research and career planning. I would like to give my sincere gratitude to my undergraduate study advisors: Professor Juan Chen, Professor Yangfeng Wu, Professor Yang Ke at Peking University. Though they did not directly contribute to this thesis, it could not have happened without their continued encouragement and believing in my pursuit of a doctoral degree with deeper training in public health at Harvard.

I am thankful to have these wonderful and supportive lab mates Rasouli Bahareh, Blakstad Mia, Rana Custodio Alejandro, Boyer Christopher Brian, and Monica Deng and doctoral cohort members Dani Poole, Katrina Ortblad and Annie Haakenstad. I also want to give many thanks to my dear friends in Boston: Xiaoxiao Jiang Kwete, Lingrui Liu, Simiao Chen, Boyu Ren, Zilin Li, Siyuan Ma, Jinyi Zhu, Xingchao Peng, Ping Gao, Da Zhang, Kai Wang, for all the joyful time and support you have brought to me. Many of you have made exciting new achievements in your own study area and will keep motivating me to explore in my research field. I would like to give my special thanks to Yi ZHENG, for coming into my life during the Boston Dragon Boat Festival and for always being there for me.

At last, I am deeply indebted to my parents Changlin LI, Jufang JI, my sister Yunge LI, my brother in law Zhengwei HOU, and my niece Congyi HOU for their unconditional love and continuous support all the time. There are not even enough words to describe how much I love and miss them. I strongly hope that we can have our reunion soon in the future when the Covid-19 epidemic gets under control.

Dedicated to the memory of my father, Changlin LI, who always encouraged me to pursue the things I love and always believed in my ability to be successful in my career. Your selfless and ever-growing love will continue support and brave me in the rest of my life.

## Chapter 1. Introduction

Estimates of the health and economic implications of disease are needed to inform policymaking, evaluate health system performance, and quantify health disparities. Disease burden estimates aim to capture the impact of diseases, injuries and risk factors in a given population.<sup>1</sup> Disease burden is measured using a variety of indicators that can describe the impact of disease from its early stages to the final clinical manifestation and health consequences: sickness, recovery, disability, or death. Epidemiological measures of disease burden include mortality, morbidity, and trends in these indicators over time. To facilitate comparison of the burden of different diseases and to consider both mortality and morbidity in a single measure, several summary health measures have been devised, including quality-adjusted life-years (QALYs)<sup>2</sup> and disability-adjusted life-years (DALYs).<sup>3-5</sup> In addition to measures of the population health impact of disease, assessing the economic consequences of health conditions provides complementary information for policy-making. Economic evaluation focuses on the direct and indirect costs of diseases for individuals, households, healthcare systems, and societies.<sup>6</sup>

Both epidemiological and economic evaluation of disease impact at population-level rely on estimating measures of disease frequency such as prevalence or incidence. In addition, a comprehensive assessment of disease burden requires estimation of prevalence and/or incidence of health consequences and complications secondary to the disease. For comparative analysis across diseases, mapping from occurrence of disease-specific events into generic measures can be accomplished using additional information such as duration of disability, magnitude of health losses associated with different disabilities, and relevant costs.

Information about the prevalence or incidence of disease typically provide the starting point for measuring disease burden. Prevalence and incidence of disease in the United States are commonly estimated from nationally-representative health surveys, surveillance and disease registration systems, administrative databases or prospective epidemiological studies. While direct analysis of

these data sources can provide policy makers and researchers an intuitive and timely description of epidemiological patterns of disease, direct estimation from these sources will often lead to biased measures. For example, there have been several studies that used national surveillance data to estimate the prevalence and incidence of diabetes from 1980 to 2017.<sup>7-11</sup> However, overall diabetes prevalence was underestimated by studies based on self-reported diagnosis in the National Health Interview Survey (NHIS) or Behavioral Risk Factor Surveillance (BRFSS), as almost a quarter of diabetes cases remain undiagnosed.<sup>12</sup> The incidence of diagnosed diabetes, estimated using the NHIS or the National Health and Nutrition Examination Surveys (NHANES), will be a biased estimate of the true incidence of diabetes without considering the effects of changes in diabetes diagnosis rates.<sup>13</sup> Several cohort studies have attempted to estimate diabetes incidence, but these studies also have important limitations.<sup>14,15</sup> Estimation of current incidence is challenging in cohort studies that rely on long duration of follow-up and therefore track incidence over an extended period. Incidence estimates from cohort studies using claims data may lack generalizability to the general population, because study participants enrolled in commercial insurance programs will not be representative of the national populations.

Prevalence and incidence of disease complications and sequelae are important components of a comprehensive disease burden evaluation, especially when the disease complications and sequelae are long-term chronic conditions that may have significant health effects and economic costs. For example, cardiovascular disease (CVD) causes 70% of deaths among patients with diabetes in the United States, and contributes substantially to the total health and economic burden of diabetes.<sup>16,17</sup> The prevalence and incidence of CVD among diabetic populations can be directly observed using disease registration data linked with hospital records or vital registration. However, national diabetes registries (NDRs) are only available in limited number of developed countries (Denmark, Sweden, Australia) worldwide. In the absence of such data in the US, one can indirectly estimate CVD risk among diabetics by applying a CVD risk prediction equation to data from the most recent nationally representative datasets such as NHANES. However, current CVD risk prediction equations for diabetic

populations have limitations. CVD risk prediction equations developed in the general population including diabetes as a predictor, such as the 2013 ACC/AHA Pooled Cohort Risk Equation<sup>18</sup>, tend to underestimate CVD risk for diabetic populations, due to differences in CVD incidence between the diabetic and general populations.<sup>19–21</sup> Diabetes-specific prediction models, such as the United Kingdom Prospective Diabetes Study Outcomes Model 2 (UKPDS OM2)<sup>22</sup> and the Risk Equations for Complications Of type 2 Diabetes (RECODE)<sup>23</sup>, were developed and validated in populations of people with diabetes outside of the United States or in clinical trial populations, and may lack generalizability to populations in the United States.

Measuring long-term health and economic burden associated with a defined disease requires an overarching framework to aggregate measures of disease occurrence, measures of downstream health consequences, and summary measures such as QALYs lost and total related costs. QALYs lost and costs of disease and sequelae require estimation of disutilities, durations, and costs associated various health outcomes that are relevant to the natural history of disease, and interventions and policies relating to the disease. In the context of gonorrhea and other sexually transmitted diseases (STDs) in racial and ethnic minorities, disparities have been measured using incidence and prevalence in several previous studies.<sup>24–27</sup> However, incidence and prevalence of gonorrhea only represent disease frequency without assessing downstream complications secondary to infection. Measuring disparities in gonorrhea-associated burden requires assessment of both short- and long-term health outcomes and costs relating to the sequelae of gonococcal infections.

In this dissertation, I report three studies that employ mathematical modeling to evaluate the epidemiological and economic disease burden of specific health conditions in the United States. In each study, I develop models that incorporate observed data and other relevant evidence specifically in the disease area of diabetes and diabetic cardiovascular complications, and in the disease area of gonorrhea. The overall goal is to address some of the challenges in estimation of disease burden described above, and to answer the following three questions:

- I. How has the epidemiological burden of diabetes changed over recent decades in the United States?
- II. What is the 10-year risk of cardiovascular disease (CVD) for patients with type 2 diabetes in the United States? What is the epidemiological burden of CVD among diabetic populations in recent years in the United States?
- III. What is the health and economic burden of gonorrhea, and related disparities, in recent years in the United States?

In the second chapter of this dissertation, *“Dynamic Modeling of Prevalence and Incidence Trends for Diabetes and Diabetes Diagnosis among Adults Aged 20 Years or Older, United States, 2000-2016”*, I examine how the epidemiological burden of diabetes in the United States has changed over recent decades. To do so, I develop a dynamic mathematical model that simulates progression of individuals from no diabetes, to undiagnosed diabetes, to diagnosed diabetes. This model allows the estimation of incidence and diagnosis rates from cross-sectional data. I fit this model using repeated cross-sectional survey data on the prevalence of undiagnosed and diagnosed diabetes from the National Health and Examination Survey (NHANES) 1988-1994 and 1999-2016. Using the fitted model, I estimate trends in the incidence of diabetes and diagnosis rates, both of which determine trends in incidence of diagnosed diabetes. Overall, I estimate a 17-year period of decreasing prevalence of undiagnosed diabetes and increasing prevalence of diagnosed and total diabetes, and decreasing incidence of diabetes, decreasing rates of diabetes diagnosis, and decreasing incidence of diagnosed diabetes over the same period.

In the third chapter, *“Risk score to predict cardiovascular disease (CVD) risk for patients with type 2 diabetes mellitus (T2DM) in the United States: a pooled analysis of prospective cohorts”*, I develop a novel CVD risk prediction model specifically for populations with T2DM in the US, and compare the performance of this model to published CVD risk scores that can be applied to diabetic patients. The model is constructed as a multivariable risk factor model estimated using pooled data on fatal-plus-

non-fatal CVD outcomes from 5 prospective cohorts. I find that this model predicts an accurate 10-year fatal-plus-non-fatal CVD risk for patients with T2DM in the US, with good discrimination and calibration performance. This performance is superior to the 2013 ACC/AHA Pooled Cohort Risk Equation, which underestimates 10-year atherosclerotic CVD risk in diabetic populations. The model developed in this study can be used to quantify total diabetes-specific CVD burden in the US over the next decade, and to simulate health outcomes for future studies that evaluate the effectiveness and cost-effectiveness of primary prevention options for diabetic patients.

In the fourth chapter, "*Disparities in health and economic outcomes associated with N. gonorrhoeae infection in the United States: costs and quality-adjusted life-years lost in 2015*", I evaluate the health and economic burden of gonorrhea and assess disparities in this burden across race/ethnicity. To implement this analysis, I develop probability tree models that capture clinical outcomes of gonorrhea and sequelae. These models allow quantification of both short- and long-term consequences and costs of gonococcal infection and sequelae, and synthesize information on gonorrhea and sequelae from different sources into a consistent framework that allows an integrated analysis of disparities in gonorrhea disease burden. I use these models to estimate attributable disease burden in terms of the discounted lifetime costs and quality-adjusted life-years (QALYs) lost due to incident infections acquired during 2015. I report population-level disease burden, disaggregated by sex, age, race/ethnicity, and for men who have sex with men (MSM). This analysis shows the highest absolute burden of both QALYs and costs to be in Non-Hispanic Black women, and the highest per-capita burden to be in MSM and American Indian/Alaska Native women. In the final chapter of this dissertation I provide a synthesis of key themes across the three studies.

In summary, this dissertation uses mathematical models to evaluate the epidemiological and economic burden of three different diseases in the United States. These results illustrate the use of mathematical modelling to synthesize observed data to infer epidemiological trends that are not always directly observable, to quantify the short-term and long-term health outcomes associated with

disease and complications, and to measure disparities in the health and economic burden of disease across population subgroups.

## References

1. Hessel F. Burden of DiseaseBurdenof disease(s). In: Kirch W, ed. *Encyclopedia of Public Health*. Springer Netherlands; 2008:94-96. doi:10.1007/978-1-4020-5614-7\_297
2. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med*. 1977;296(13):716-721. doi:10.1056/NEJM197703312961304
3. Murray CJL, Lopez AD, Harvard School of Public Health, World Health Organization, World Bank. *The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020*. Published by the Harvard School of Public Health on behalf of the World Health Organization and the World Bank ; Distributed by Harvard University Press; 1996.
4. Lopez AD, Murray CCJL. The global burden of disease, 1990–2020. *Nat Med*. 1998;4(11):1241-1243. doi:10.1038/3218
5. Murray CJL, Salomon JA, Mathers C. A critical examination of summary measures of population health. *World Health Organization Bulletin of the World Health Organization; Geneva*. 2000;78(8):981-994.
6. World Health Organization. *WHO Guide to Identifying The Economic Consequences of Disease and Injury*. Department of health Systems Financing; Health Systems and Services [https://www.who.int/choice/publications/d\\_economic\\_impact\\_guide.pdf?ua=1](https://www.who.int/choice/publications/d_economic_impact_guide.pdf?ua=1)
7. Geiss LS, Wang J, Cheng YJ, et al. Prevalence and Incidence Trends for Diagnosed Diabetes Among Adults Aged 20 to 79 Years, United States, 1980-2012. *JAMA*. 2014;312(12):1218-1226. doi:10.1001/jama.2014.11494
8. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and Trends in Diabetes Among Adults in the United States, 1988-2012. *JAMA*. 2015;314(10):1021-1029. doi:10.1001/jama.2015.10029
9. Benoit SR, Hora I, Albright AL, Gregg EW. New directions in incidence and prevalence of diagnosed diabetes in the USA. *BMJ Open Diab Res Care*. 2019;7(1):e000657. doi:10.1136/bmjdr-2019-000657
10. McBean AM, Li S, Gilbertson DT, Collins AJ. Differences in Diabetes Prevalence, Incidence, and Mortality Among the Elderly of Four Racial/Ethnic Groups: Whites, Blacks, Hispanics, and Asians. *Diabetes Care*. 2004;27(10):2317-2324. doi:10.2337/diacare.27.10.2317



11. Cowie CC, Rust KF, Byrd-Holt DD, et al. Prevalence of Diabetes and Impaired Fasting Glucose in Adults in the U.S. Population: National Health and Nutrition Examination Survey 1999–2002. *Diabetes Care*. 2006;29(6):1263-1268. doi:10.2337/dc06-0062
12. Centers for Disease Control and Prevention. *National Diabetes Statistics Report, 2017*. Atlanta, GA: U.S. Dept of Health and Human Services, 2017
13. Selvin E, Ali MK. Declines in the Incidence of Diabetes in the U.S.—Real Progress or Artifact? *Diabetes Care*. 2017;40(9):1139-1143. doi:10.2337/dc16-2442
14. Nichols GA, Schroeder EB, Karter AJ, et al. Trends in Diabetes Incidence Among 7 Million Insured Adults, 2006–2011The SUPREME-DM Project. *Am J Epidemiol*. 2015;181(1):32-39. doi:10.1093/aje/kwu255
15. Weng W, Liang Y, Kimball ES, et al. Decreasing incidence of type 2 diabetes mellitus in the United States, 2007–2012: Epidemiologic findings from a large US claims database. *Diabetes Research and Clinical Practice*. 2016;117:111-118. doi:10.1016/j.diabres.2016.04.043
16. Haffner SM, Lehto S, Rönnemaa T, Pyörälä K, Laakso M. Mortality from Coronary Heart Disease in Subjects with Type 2 Diabetes and in Nondiabetic Subjects with and without Prior Myocardial Infarction. *New England Journal of Medicine*. 1998;339(4):229-234. doi:10.1056/NEJM199807233390404
17. Alva ML, Gray A, Mihaylova B, Leal J, Holman RR. The impact of diabetes-related complications on healthcare costs: new results from the UKPDS (UKPDS 84). *Diabet Med*. 2015;32(4):459-466. doi:10.1111/dme.12647
18. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology*. 2014;63(25, Part B):2889-2934. doi:10.1016/j.jacc.2013.11.002
19. Chamnan P, Simmons RK, Sharp SJ, Griffin SJ, Wareham NJ. Cardiovascular risk assessment scores for people with diabetes: a systematic review. *Diabetologia*. 2009;52(10):2001-2014. doi:10.1007/s00125-009-1454-0
20. Stephens JW, Ambler G, Vallance P, Betteridge DJ, Humphries SE, Hurel SJ. Cardiovascular risk and diabetes. Are the methods of risk prediction satisfactory? *European Journal of Cardiovascular Prevention & Rehabilitation*. 2004;11(6):521-528. doi:10.1097/01.hjr.0000136418.47640.bc
21. Bulughapitiya U, Siyambalapitiya S, Sithole J, Idris I. Is diabetes a coronary risk equivalent? Systematic review and meta-analysis. *Diabetic Medicine*. 2009;26(2):142-148. doi:10.1111/j.1464-5491.2008.02640.x
22. Hayes AJ, Leal J, Gray AM, Holman RR, Clarke PM. UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from

the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. *Diabetologia*. 2013;56(9):1925-1933. doi:10.1007/s00125-013-2940-y

23. Basu S, Sussman JB, Berkowitz SA, Hayward RA, Yudkin JS. Development and validation of Risk Equations for Complications Of type 2 Diabetes (RECODE) using individual participant data from randomised trials. *Lancet Diabetes Endocrinol*. 2017;5(10):788-798. doi:10.1016/S2213-8587(17)30221-8

24. Chesson HW, Patel CG, Gift TL, Aral SO. Trends in Selected Measures of Racial and Ethnic Disparities in Gonorrhea and Syphilis in the United States, 1981–2013: *Sexually Transmitted Diseases*. 2016;43(11):661-667. doi:10.1097/OLQ.0000000000000518

25. STDs in Racial and Ethnic Minorities - 2017 Sexually Transmitted Diseases Surveillance. Published June 17, 2019. Accessed September 22, 2019. <https://www.cdc.gov/std/stats17/minorities.htm>

26. Chesson HW, Kent CK, Owusu-Edusei K, Leichliter JS, Aral SO. Disparities in Sexually Transmitted Disease Rates Across the “Eight Americas”: *Sexually Transmitted Diseases*. 2012;39(6):458-464. doi:10.1097/OLQ.0b013e318248e3eb

27. Hoover KM, Bohm M, Keppel K. Measuring Disparities in the Incidence of Sexually Transmitted Diseases. [Review]. *Sexually Transmitted Diseases*. 2008;35(12). doi:10.1097/OLQ.0b013e3181886750

**Chapter 2. Prevalence and incidence trends for diabetes and diabetes  
diagnosis among adults ages 20 years and older, United States, 2000-2016:  
results of a mathematical model**

Yunfei Li<sup>1</sup>, Nicolas A Menzies<sup>1</sup>, Goodarz Danaei<sup>1,2</sup>, Joshua A Salomon<sup>3</sup>

1. Department of Global Health and Population, Harvard T. H. Chan School of Public Health
2. Department of Epidemiology, Harvard T. H. Chan School of Public Health
3. Center for Health Policy / Center for Primary Care and Outcomes Research, Stanford University

## **Abstract**

### **Background**

Although the prevalence and incidence of diagnosed diabetes are thought to have decreased in the United States after 2008, there is a lack of population-based estimates describing national trends in the prevalence of undiagnosed and all diabetes, and trends in the incidence of diabetes and rates of diabetes diagnosis.

### **Methods**

We developed an age-stratified Markov model of undiagnosed and diagnosed diabetes in the US adult population. In this model, smooth functions were used to represent year- and age-specific incidence of diabetes and rates of diabetes diagnosis. Using Bayesian methods, we calibrated this model separately for men and women, using nationally representative data on the prevalence of undiagnosed and diagnosed diabetes from the National Health and Nutrition Examination Survey (NHANES III and Continuous NHANES 1999-2016). We estimated the prevalence of all, undiagnosed, and diagnosed diabetes, as well as the incidence of diabetes and diagnosed diabetes over the period 2000-2016. Prevalence and incidence trends were summarized as the annual percentage change (APC).

### **Results**

Overall, the age-standardized prevalence of all diabetes per 100 adults was estimated to increase from 10.4 (95% uncertainty interval, 8.9 to 12.0) in 2000 to 12.0 (10.3 to 14.1) in 2016. Over the same period, the age-standardized prevalence of diagnosed diabetes per 100 adults increased from 6.8 (5.7 to 8.0) to 8.9 (7.5 to 10.5), and the age-standardized prevalence of undiagnosed diabetes per 100 adults decreased from 3.6 (2.7 to 4.7) to 3.2 (2.0 to 4.7). For diabetes-free adults, the age-standardized incidence of diabetes per 1,000 person-years decreased from 7.5 (6.6 to 8.5) in 2000 to 6.9 (4.3 to 10.7) in 2016. For adults with undiagnosed diabetes, the age-standardized rate of diabetes diagnosis per 1,000 person-years decreased from 168 (142 to 201) to 144 (98 to 208) during 2000-2016. Among

adults without diagnosed diabetes, the age-standardized incidence of diagnosed diabetes per 1,000 person-years decreased from 6.2 (5.5 to 6.9) to 5.7 (4.4 to 7.4) during 2000-2016. Trends within individual age categories were generally similar to these age-standardized results. However, the prevalence of undiagnosed diabetes and incidence of diabetes among men and women ages 20-49 years increased over 2000-2016. The age-standardized incidence of diagnosed diabetes in men ages 65 years and older increased with an APC of 1.5% (-2.2% to 5.2%) during 2000-2016.

## **Conclusions**

During 2000-2016, incidence of diabetes, rates of diabetes diagnosis, and incidence of diagnosed diabetes decreased in the US adult population. Continuing diabetes prevention efforts, monitoring and screening are needed to further reduce diabetes incidence and diabetes complications.

## Background

Diabetes is a leading cause of disease burden in the United States.<sup>1,2</sup> Between 1990 and 2016, diabetes rose from the sixth leading cause of disability-adjusted life-years (DALYs) to the fourth, with an estimated 66 thousand deaths, 2.5 million years lived with disability (YLDs) and 3.8 million DALYs attributable to diabetes in the U.S. in 2016.<sup>3</sup> Estimated economic costs due to diabetes increased by 26% from 2012 to 2017,<sup>4</sup> reaching \$327 billion, including \$237 billion in direct medical costs and \$90 billion in reduced productivity.<sup>4,5</sup> Systematic analyses of patterns and trends in diabetes are needed by policymakers to assess population health needs and to plan and evaluate different intervention strategies.

There have been several studies that used national surveillance data to describe the prevalence and incidence of diabetes from 1980 to 2017.<sup>6–11</sup> Some studies have reported a 20-year increase in prevalence of diagnosed diabetes nationally between the 1990s and the mid-2000s, followed by a plateau through 2017.<sup>6,8</sup> However, studies based on self-reported diagnosis in the National Health Interview Survey (NHIS)<sup>6,8</sup> or Behavioral Risk Factor Surveillance (BRFSS)<sup>12</sup> will underestimate overall diabetes prevalence, as undiagnosed diabetes accounted for an estimated 24% of all diabetes in 2017.<sup>13</sup> Prevalence of laboratory-confirmed diabetes, including undiagnosed cases, has been measured in the National Health and Nutrition Examination Surveys (NHANES).<sup>7,11</sup> In addition to prevalence, studies have also examined trends in the incidence of diagnosed diabetes,<sup>6,8</sup> which is determined by both trends in true incidence and trends in diagnosis rates.<sup>10,14</sup> Disentangling these two drivers of trends in incident diagnosed diabetes is challenging without directly measured, laboratory-confirmed diagnosis data that allows separating diagnosed and undiagnosed diabetes.

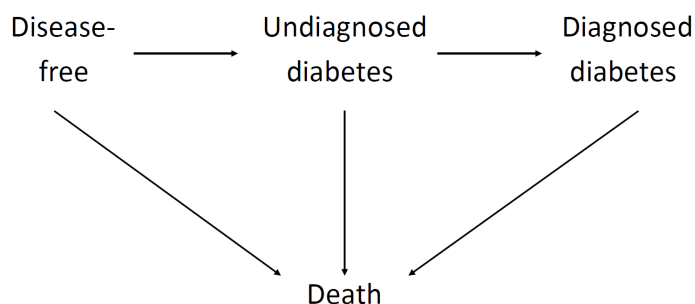
To overcome these limitations and provide contemporary, unbiased estimates of the epidemiological trends of diabetes incidence and diagnosis, we constructed a dynamic model that incorporated both undiagnosed and diagnosed diabetes. We estimated this model using nationally representative cross-sectional data, including information not only on self-reported diabetes diagnoses but also on

laboratory-measured glucose levels. Using the fitted model, we examined trends in rates of prevalence of all, undiagnosed and diagnosed diabetes, and in rates of incidence of diabetes, diagnosed diabetes, and diabetes diagnosis over the period 2000-2016.

## **Methods**

### **Analytic overview**

We developed a Markov model of undiagnosed and diagnosed diabetes in the U.S. population ages 20 years and older (Figure 2.1). Hazard rates for mortality among people without diabetes were estimated from the general all-cause mortality data reported in United States life tables, with adjustment for diabetes-related mortality using the Second National Health and Nutrition Examination Survey Mortality Study (SNHANES-MS).<sup>15,16</sup> Annual, age-specific transition rates from disease free to undiagnosed and from undiagnosed to diagnosed diabetes were specified using B-spline surfaces, which are flexible continuous functions in two dimensions (age and time), and separate functions were estimated for men and women. We calibrated this model to nationally representative NHANES III and Continuous NHANES data describing the prevalence of undiagnosed and diagnosed diabetes. A Bayesian approach was used to estimate model parameters, operationalized using incremental mixture importance sampling (IMIS).<sup>17</sup> We examined model fits using both visual inspection and posterior prediction interval checks.<sup>18</sup> Prevalence and incidence trends were summarized as the annual percentage change (APC) between 2000 and 2016. Analyses were undertaken in R (R-3.5.2).



**Figure 2.1. Disease states in the Markov model of undiagnosed and diagnosed diabetes.\***

\*Details on the states in the Markov model, with corresponding states transition matrix for individuals at single year of age (i) and single calendar year (t) are reported in Supplementary Material Figure S2.1.

## Data sources

We extracted nationally-representative cross-sectional data from the NHANES III (1988-1994) and Continuous NHANES (1999-2016), describing the prevalence of diagnosed and undiagnosed diabetes among the U.S. population ages 20 years and older. Conducted by the National Center for Health Statistics, NHANES is a national health and nutrition survey that uses a stratified, multistage cluster sample design to be representative of the U.S. civilian, noninstitutionalized population.<sup>19,20</sup> NHANES is unique in that it combines personal interviews with standardized physical examinations and laboratory tests.<sup>19,20</sup> For NHANES III, the overall responses rate was 86% in the interviewed sample and 78% in the examined sample. For Continuous NHANES 2015-2016 this decreased to 65% in the interviewed sample and 62% in the examined sample. Nonresponse bias analyses have suggested that the declining response rates did not produce substantial bias in the final survey estimates.<sup>21,22</sup>

We used data from the interview questionnaire and laboratory examination to provide a full accounting of diagnosed, undiagnosed, and all diabetes. The interview questionnaire was standardized to collect information on age, sex, pregnancy, and fasting status. Women who were pregnant were excluded from our analysis because pregnancy affects glucose measurements.<sup>7</sup> We excluded individuals who were randomly selected to participate in an afternoon or evening examination or had a total length of “food fast” hours less than 8 hours or longer than 24 hours.



Further details on the number of participants in the analytic sample can be found in Supplementary Material Table S2.1.

### **Definition of Diabetes**

Consistent with previous studies, 'diagnosed diabetes' was defined as a self-reported previous diagnosis of diabetes.<sup>6,7,11</sup> Participants were asked "other than during pregnancy for women, have you ever been told that you have diabetes or sugar diabetes" by a "doctor" (NHANES III) or "doctor or other health professionals" (Continuous NHANES). 'Undiagnosed diabetes' was defined as having a Fasting Plasma Glucose (FPG) level of 126mg/dL or higher without a self-reported previous diagnosis of diabetes. 'All diabetes' was defined as having either diagnosed or undiagnosed diabetes. Following the National Center for Health Statistics (NCHS) recommendations, FPG values in NHANES 2005-2016 were adjusted to be comparable to earlier values using a published CDC regression-based adjustment equation.<sup>20,23-28</sup> We did not use Hemoglobin A<sub>1c</sub> or 2-hour plasma glucose because they are not available in all NHANES rounds. In addition, previous analyses of NHANES data have indicated that, assuming universal screening of undiagnosed diabetes, the Hemoglobin A<sub>1c</sub> cut point of 6.5% identifies one-third fewer cases of undiagnosed diabetes than the FPG cut point of  $\geq 126$ mg/dL.<sup>29</sup> Two-hour plasma glucose is used less frequently in clinical care because it is relatively costly and burdensome for patients. Our definition of diabetes included both type 1 and type 2 diabetes, since NHANES cannot distinguish between the type of diabetes.

### **Statistical analysis**

We developed a Markov model to represent undiagnosed and diagnosed diabetes in the U.S. population ages 20 years and older between 1990-2016 (Figure 2.1). We stratified by sex, because observed patterns of diabetes incidence differ between men and women.<sup>6,30,31</sup> We assumed that people may only progress from disease-free to undiagnosed diabetes to diagnosed diabetes (with progression to death possible from any of the other three states), as diabetes is generally considered a progressive condition without remission. In addition, we assumed that there is no possibility of

progression from disease-free to diagnosed diabetes directly, as immediate diagnosis upon disease onset is unlikely.<sup>32</sup> The average duration between onset and clinical diagnosis for non-insulin-dependent diabetes in the U.S. has been estimated previously as 4–6 years.<sup>33,34</sup>

We modeled the transition probabilities between disease states by single year of age ( $i$ ) and single calendar year ( $t$ ) (Figure S2.1 in Supplementary Material). Yearly age- and sex-specific hazard rates from disease-free to death were estimated by subtracting diabetes-related mortality from all-cause mortality based on prevalence and risk ratios estimates (see Supplementary Material for details). Yearly age- and sex-specific all-cause mortality rates among the general populations were obtained from the annual United States Life Tables published by NCHS.<sup>15</sup> Prevalence of undiagnosed and diagnosed diabetes were estimated directly from NHANES III and Continuous NHANES. Risk ratios for all-cause mortality comparing those with undiagnosed or diagnosed diabetes to those without diabetes were estimated by fitting Cox proportional hazards models by age, sex and survey year to prospective mortality follow-up data from SNHANES-MS.<sup>16</sup> Details on distributions of risk ratios are reported in Supplementary Material Table S2.2.

The incidence of diabetes was represented as the transition rate from disease-free to undiagnosed diabetes. The diagnosis rate of diabetes was represented as the transition rate from undiagnosed to diagnosed diabetes. For each of these transition rates, we constructed individual B-spline surfaces in the dimensions of calendar year and age, allowing rates to change smoothly as a function of age and year. For the age dimension we allowed 5 control points, with three internal knots at the boundaries between age groups 20–34 years, 35–49 years, 50–64 years, and 65 years and older, and two additional boundary knots at ages 20 years and 99 years. For the year dimension we allowed 7 control points, with internal knots located in the middle year of NHANES III, and every two survey years of Continuous NHANES. This resulted in 48 parameters defining each spline surface. We specified weakly-informative priors for these parameters. We used a second-degree difference penalty on the time dimension to avoid overfitting. We conducted seven-fold cross-validation to determine the optimal value of the

penalty.<sup>35,36</sup> To implement the cross-validation, the NHANES data were divided into seven equal-sized folds for men and women separately. For each fold, we predicted the prevalence of undiagnosed and diagnosed diabetes from a model estimated using the data for the other six folds. This procedure was conducted for all seven folds. We selected the penalty value to maximize the log-likelihood of these out-of-sample predictions when compared to the original data.

We estimated the model separately for men and women, using incremental mixture importance sampling (IMIS). IMIS implements a Bayesian approach to estimate the joint posterior distribution of the model parameters, allowing for correlation and multi-modality of this distribution. All analyses were conducted in R.<sup>17</sup>

### **Outcome measures**

Primary model outcome measures included: the prevalence of undiagnosed, diagnosed, and all diabetes; rates of diabetes incidence and diagnosis (for disease-free and undiagnosed diabetes respectively); and incidence of diagnosed diabetes. These outcomes were estimated for each year during the period 2000-2016, stratified by sex and age group. Model outputs during the first ten-year period (1990-1999) were not reported due to lack of data from NHANES between 1995-1998, and were used to stabilize model conditions by the start of 2000. Results were summarized by age categories 20-34, 35-49, 50-64, and 65 years and older. To standardize results by age, we used the 2000 U.S. population. To quantify trends in each quantity of interest, we calculated APCs over the 17-year period by sex and age group. Uncertainty intervals were calculated as the 2.5th and 97.5th percentiles of the distribution of results for each outcome.

## Results

### Model checking

We used posterior predictive checks to evaluate model fit to the NHANES data.<sup>18</sup> Supplementary Material Figures S2.2 and S2.3 provide a comparison of modelled estimates to the empirical values. These comparisons show that the model predicted the prevalence of diabetes well: the 95% prediction intervals cover 96.8% and 98.4% of NHANES mean prevalence estimates for men and women separately. Full results of these posterior predictive checks (comparing modelled and empirical estimates for the prevalence of undiagnosed, diagnosed, and all diabetes) are reported in Supplementary Material Figure S2.5.

### Prevalence of undiagnosed, diagnosed, and all diabetes in 2016

In the overall population in 2016, the crude prevalence per 100 persons was 12.6 (95% Uncertainty Interval, 11.8 to 13.5) for all diabetes, 3.5 (2.9 to 4.1) for undiagnosed diabetes, and 9.1 (8.4 to 9.8) for diagnosed diabetes. The crude and age-standardized prevalence of all diabetes and diagnosed diabetes differed by sex and age groups (Table 2.1). The crude prevalence of all diabetes in 2016 was higher in men (14.0 per 100 persons [12.9 to 15.4]) than in women (11.2 per 100 persons [10.3 to 12.4]); age-standardized prevalence was also higher among men. The prevalence of diagnosed diabetes and all diabetes were highest among those ages 65 years and older, followed by ages 50-64 years, ages 35-49 years, and ages 20-34 years for both men and women.

**Table 2.1. Prevalence per 100 persons of undiagnosed, diagnosed, and all diabetes in the U.S. general population ages 20 years and older in 2016, and annual percentage changes of the prevalence of undiagnosed, diagnosed, and all diabetes over 2000-2016, by age group and sex.**

Population	Undiagnosed Diabetes		Diagnosed Diabetes		All Diabetes	
	Prevalence Per 100 persons (95% UI)	APC % (95% UI)	Prevalence Per 100 persons (95% UI)	APC % (95% UI)	Prevalence Per 100 persons (95% UI)	APC % (95% UI)
<b>Total</b>						
<b>Crude</b>	3.5 (2.9, 4.1)	-0.4 (-1.8, 1.0)	9.1 (8.4, 9.8)	2.4 (1.7, 3.1)	12.6 (11.8, 13.5)	1.5 (0.9, 2.1)
<b>Age- standardized*</b>	3.2 (2.0, 4.7)	-0.8 (-2.3, 0.6)	8.9 (7.5, 10.5)	1.7 (1.1, 2.4)	12.0 (10.3, 14.1)	0.9 (0.3, 1.5)
<b>Men</b>						
<b>Crude</b>	3.9 (3.1, 4.8)	-0.9 (-2.9, 1.4)	10.1 (8.2, 11.4)	3.2 (2.2, 4.3)	14.0 (12.9, 15.4)	1.8 (0.9, 2.7)
<b>Age- standardized*</b>	3.6 (2.2, 5.2)	-1.0 (-3.0, 1.4)	9.4 (8.0, 11.1)	2.3 (1.4, 3.3)	13.0 (11.1, 15.2)	1.2 (0.3, 2.1)
<b>Women</b>						
<b>Crude</b>	3.2 (2.5, 4.0)	0.1 (-1.8, 2.0)	8.1 (7.3, 9.0)	1.6 (0.7, 2.6)	11.2 (10.3, 12.4)	1.1 (0.3, 1.9)
<b>Age- standardized*</b>	2.8 (1.8, 4.1)	-0.7 (-2.7, 1.3)	8.3 (7.1, 9.8)	1.1 (0.3, 2.0)	11.1 (9.5, 13.1)	0.6 (-0.1, 1.4)
<b>Men, Age group</b>						
<b>20-34</b>	2.3 (1.4, 3.9)	2.8 (-1.4, 7.0)	1.2 (0.6, 1.9)	-0.8 (-5.8, 3.6)	3.5 (2.2, 5.2)	1.3 (-2.6, 4.6)
<b>35-49</b>	3.9 (2.5, 5.7)	1.2 (-2.6, 5.5)	5.6 (4.4, 7.0)	1.5 (-1.0, 3.9)	9.4 (7.6, 11.6)	1.4 (-0.9, 3.3)
<b>50-64</b>	4.2 (2.6, 5.7)	-3.0 (-6.3, -3.2)	15.4 (13.2, 18.0)	2.3 (0.9, 3.7)	19.6 (17.0, 22.5)	0.7 (-0.5, 2.0)
<b>65+</b>	4.7 (3.0, 6.4)	-4.0 (-6.0, -1.3)	26.7 (24.0, 29.6)	3.2 (2.1, 4.1)	31.5 (29.0, 34.0)	1.6 (0.7, 2.3)
<b>Women, Age group</b>						
<b>20-34</b>	1.7 (0.9, 2.8)	1.5 (-3.0, 5.9)	1.3 (0.8, 2.0)	-0.1 (-4.2, 4.3)	3.0 (2.0, 4.3)	0.8 (-2.3, 3.8)
<b>35-49</b>	2.5 (1.5, 3.8)	0.06 (-3.7, 3.9)	5.7 (4.6, 7.1)	1.7 (-2.8, 4.0)	8.2 (6.6, 9.8)	1.1 (-0.5, 3.0)
<b>50-64</b>	3.3 (2.1, 4.7)	-1.3 (-4.5, 1.7)	11.6 (10.1, 13.7)	0.9 (-0.7, 2.4)	14.9 (13.0, 17.4)	0.4 (-0.8, 1.7)
<b>65+</b>	4.4 (3.1, 6.1)	-1.8 (-4.3, 1.1)	19.0 (16.9, 21.1)	1.0 (-0.1, 2.2)	23.3 (20.9, 26.1)	0.4 (-0.5, 1.3)

Abbreviation: APC = Annual percentage change.

\*Age-standardized to the U.S. 2000 population based on age groups 20-34 years, 35-49 years, 50-64 years, and 65 years and older.

### **Incidence of diabetes, diagnosed diabetes, and rates of diabetes diagnosis in 2016**

Among adults without diagnosed diabetes, the crude incidence of diagnosed diabetes per 1,000 person-years was 6.2 (95% UI, 4.7 to 8.0) in 2016. For diabetes-free adults, the crude incidence of diabetes per 1,000 person-years was 7.1 (4.4 to 11.0) in 2016. For adults with undiagnosed diabetes, the crude rate of diabetes diagnosis per 1,000 person-years was 148 (101 to 213) in 2016. The crude incidence of diabetes and diagnosed diabetes were higher among men than women (Table 2.2). Among adults ages 50 years and older, the incidence of diabetes per 1,000 person-years was similar among men (7.7 [3.7 to 15.8] for ages 65 years and older) and women (7.7 [3.8 to 15.0] for ages 65 years and older). With a similar incidence of diabetes, the rate of diabetes diagnosis was higher for men than women ages 50 years and older, leading to a higher incidence of diagnosed diabetes for men than women.

**Table 2.2. Incidence per 1,000 person-years of diabetes, rates of diabetes diagnosis, and incidence of diagnosed diabetes for ages 20 years and older in the U.S. general population in 2016 and annual percentage changes of incidence of diabetes, rates of diabetes diagnosis, and incidence of diagnosed diabetes over 2000-2016, by age group and sex.**

Population	Diabetes		Diabetes Diagnosis <sup>†</sup>		Diagnosed Diabetes <sup>‡</sup>	
	Incidence Per 1,000 person-years (95% UI)	APC % (95% UI)	Rates Per 1,000 person-years (95% UI)	APC % (95% UI)	Incidence Per 1,000 person-years (95% UI)	APC % (95% UI)
<b>Total</b>						
<b>Crude</b>	7.1 (4.4, 11.0)	-0.7 (-2.8, 1.6)	148 (101, 213)	-0.5 (-2.7, 1.9)	6.2 (4.7, 8.0)	-0.1 (-2.0, 1.9)
<b>Age-standardized*</b>	6.9 (4.3, 10.7)	-0.8 (-2.9, 1.4)	144 (98, 208)	-0.6 (-2.9, 1.8)	5.7 (4.4, 7.4)	-0.6 (-2.5, 1.3)
<b>Men</b>						
<b>Crude</b>	7.2 (4.8, 11.3)	-1.2 (-3.6, 1.9)	160 (107, 233)	0.5 (-2.5, 3.8)	7.8 (5.2, 11.0)	0.9 (-2.0, 3.6)
<b>Age-standardized*</b>	7.1 (4.7, 11.0)	-1.3 (-3.8, 1.9)	152 (104, 223)	0.2 (-3.1, 3.5)	7.1 (4.8, 9.8)	0.3 (-2.5, 2.8)
<b>Women</b>						
<b>Crude</b>	6.3 (4.0, 9.7)	-0.2 (-3.5, 3.1)	146 (93, 226)	-1.4 (-4.7, 1.9)	4.6 (3.1, 6.9)	-1.6 (-4.2, 1.2)
<b>Age-standardized*</b>	6.2 (4.0, 9.3)	-0.4 (-3.6, 2.7)	148 (94, 231)	-1.3 (-4.6, 2.0)	4.5 (3.0, 6.6)	-1.9 (-4.5, 1.1)
<b>Men, Age group</b>						
<b>20-34</b>	6.0 (3.2, 10.0)	3.3 (-1.5, 8.2)	135 (66, 257)	-0.9 (-6.3, 5.2)	2.6 (1.2, 4.8)	0.5 (-5.0, 5.7)
<b>35-49</b>	7.6 (3.8, 13.4)	-0.3 (-4.3, 4.5)	126 (62, 226)	-1.6 (-6.1, 3.4)	5.5 (3.0, 9.8)	-0.3 (-4.6, 4.3)
<b>50-64</b>	7.9 (3.9, 15.9)	-4.9 (-9.2, -0.09)	208 (115, 347)	1.8 (-2.2, 6.4)	12.0 (7.2, 19.7)	-0.3 (-4.1, 3.5)
<b>65+</b>	7.7 (3.7, 15.8)	-2.3 (-6.6, 2.2)	172 (90, 275)	3.7 (-0.3, 8.4)	13.3 (8.1, 19.3)	1.5 (-2.2, 5.2)
<b>Women, Age group</b>						
<b>20-34</b>	4.8 (2.6, 7.9)	2.0 (-2.5, 6.7)	172 (88, 313)	-1.1 (-6.3, 4.1)	2.5 (1.2, 4.7)	-0.3 (-5.4, 5.7)
<b>35-49</b>	5.8 (2.8, 10.2)	-0.09 (-5.0, 4.3)	160 (75, 317)	-1.8 (-7.7, 3.5)	3.9 (2.0, 7.1)	-2.4 (-7.1, 1.9)
<b>50-64</b>	7.2 (3.4, 13.5)	-1.4 (-6.2, 3.4)	148 (71, 263)	-1.4 (-6.6, 3.2)	5.5 (2.8, 9.5)	-2.9 (-6.6, 1.4)
<b>65+</b>	7.7 (3.8, 15.0)	-1.6 (-6.1, 4.4)	98 (54, 158)	-0.9 (-4.9, 3.2)	7.1 (4.0, 11.4)	-1.5 (-5.7, 2.3)

<sup>†</sup>Diabetes Diagnosis was defined as detection and diagnosis of diabetes for adults with undiagnosed diabetes who having a Fasting Plasma Glucose (FPG) level of 126mg/dL or higher without a self-reported previous diagnosis of diabetes.

<sup>‡</sup>Diagnosed Diabetes was defined as a self-reported previous diagnosis of diabetes.

Abbreviation: APC = Annual percentage change.

\*Age-standardized to the U.S. 2000 population based on age groups 20-34 years, 35-49 years, 50-64 years, and 65 years and older.

### **Trends in the prevalence of undiagnosed, diagnosed, and all diabetes over 2000-2016**

During 2000-2016, the age-standardized prevalence of all diabetes and diagnosed diabetes increased. Estimates of APC were 0.9% (0.3% to 1.5%) for all diabetes and 1.7% (1.1% to 2.4%) for diagnosed diabetes in the overall population (Figure 2.2 and Table 2.1). Over the same period, the APC for the age-standardized prevalence of undiagnosed diabetes was -0.8% (-2.3% to 0.6%).

Similar results of an increasing trend in the prevalence of all and diagnosed diabetes and a decreasing trend in the prevalence of undiagnosed diabetes were estimated in many age groups in men and women (Figure 2.3 and Figure 2.4). However, the prevalence of undiagnosed diabetes among ages 20-34 years and ages 35-49 years men and women continued to increase during 2000-2016. For example, the prevalence of undiagnosed diabetes per 100 persons was 1.5 (0.9 to 2.1) in 2000 and 2.3 (1.4 to 3.9) in 2016 for men ages 20-34 years, with an estimated APC of 2.8% (-1.4% to 7.0%). The prevalence of undiagnosed diabetes per 100 persons for women ages 20-34 years increased from 1.3 (0.8 to 1.9) in 2000 to 1.7 (0.9 to 2.8) in 2016, with an APC of 1.5% (-3.0% to 5.9%).

The estimated increases in diagnosed and all diabetes prevalence and decrease in undiagnosed diabetes prevalence were driven primarily by the 50-64 years and 65 years and older populations, and among men in particular. For example, the prevalence of diagnosed diabetes per 100 men ages 65 years and older was estimated to increase substantially from 16.3 (14.5 to 18.1) in 2000 to 26.7 (24.0 to 29.6) in 2016, with an APC of 3.2% (2.1% to 4.1%). The APC of the prevalence of all diabetes for men ages 65 years and older was estimated to be 1.6% (0.7% to 2.3%) during 2000-2016.



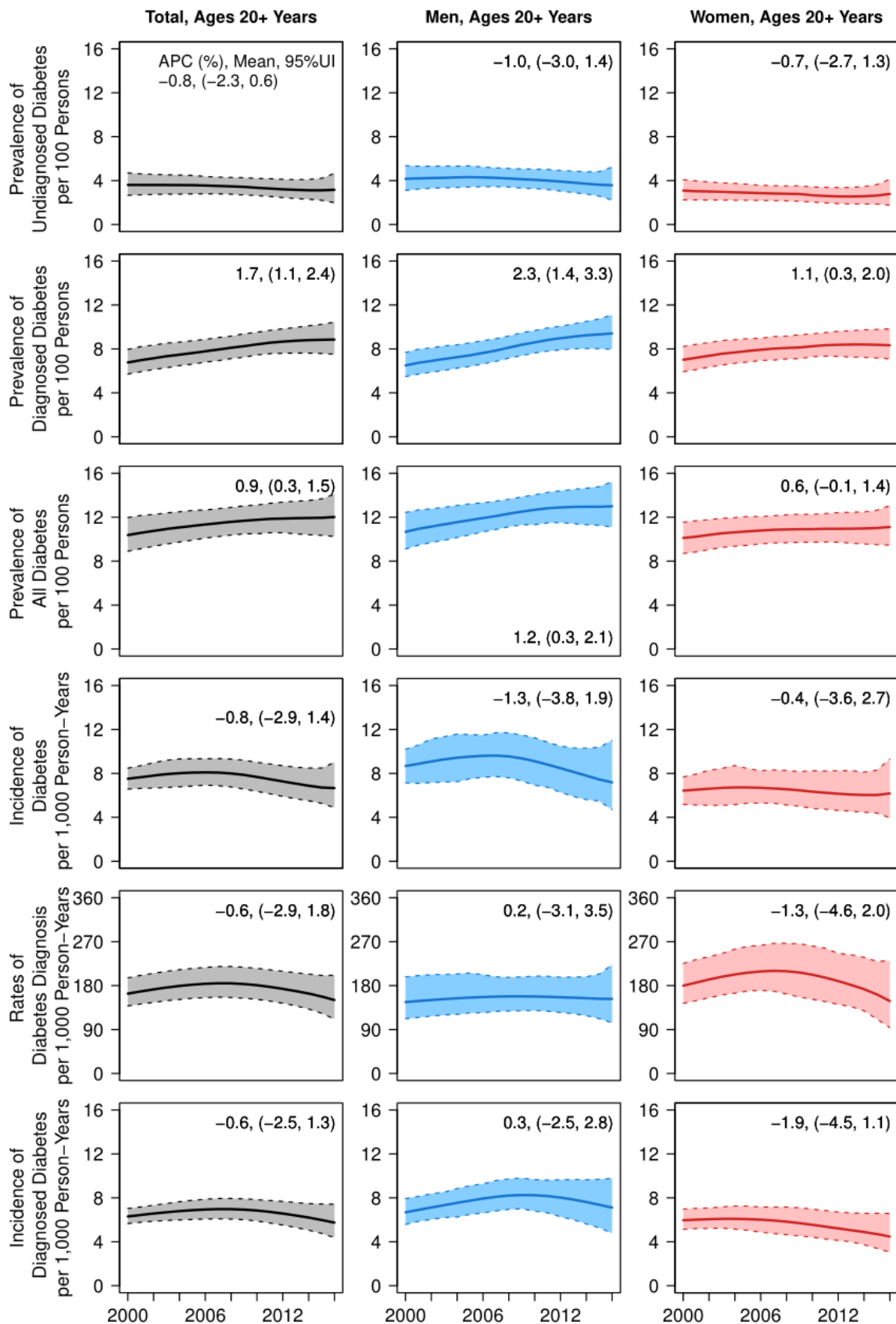
## **Trends in incidence of diabetes, diagnosed diabetes, and rates of diabetes diagnosis 2000-2016**

In the overall population, the age-standardized incidence of diagnosed diabetes per 1,000 person-years was similar in 2000 (6.2 [5.5 to 6.9]) and 2016 (5.7 [4.4 to 7.4]), corresponding to an estimated APC of -0.6% (-2.5% to 1.3%) (Figure 2.2 and Table 2.2). Decreasing trends were also estimated in the overall incidence of diabetes (APC -0.8% [-2.9% to 1.4%]) and rates of diabetes diagnosis (APC -0.6% [-2.9% to 1.8%]) during the study period.

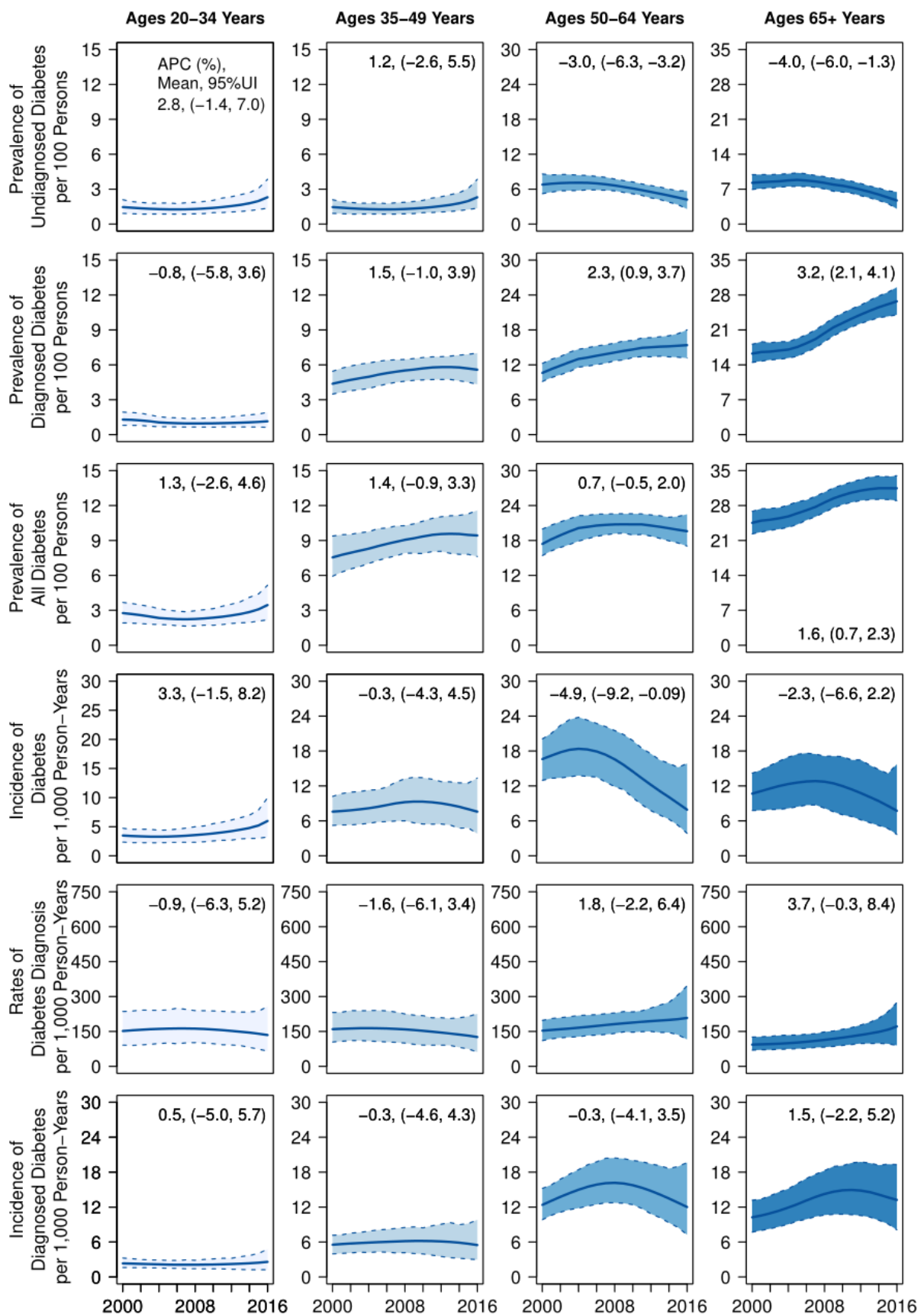
Whereas the trend in incidence of diabetes was relatively flat among women (APC of -0.36% [-3.6% to 2.7%]), incidence has decreased among men, with an estimated APC of -1.3% (-3.8% to 1.9%). Rates of diabetes diagnosis were estimated to have a minor increase in men and a substantial decrease in women with an APC of 0.2% (-3.1% to 3.5%) versus an APC of -1.3% (-4.6% to 2.0%) during the study period, but these changes were not statistically significant. Changes in the incidence of diagnosed diabetes were driven by both trends in incidence of diabetes and rates of diabetes diagnosis, leading to an increase in men but a decrease in women during 2000-2016.

Inspection of age-stratified trends in diabetes incidence suggested differential changes among younger vs. older adults. For example, the incidence of diabetes was estimated to increase in the age group 20-34 years for both men (APC 3.3% [-1.5% to 8.2%]) and women (APC 2.0% [-2.5% to 6.7%]), but decrease in the age group 65 and older years for men (APC -2.3% [-6.6% to 2.2%]) and women (APC -1.6% [-6.1% to 4.4%]). Rates of diabetes diagnosis were estimated to decline in all age groups for women and in ages below 50 years in men, but increased in men ages 50 years and older. Incidence of diagnosed diabetes were estimated to decrease in most age groups in men and women, except for men ages younger than 35 years or older than 65 years. The increase in diagnosed diabetes was mostly driven by the increasing trend in incidence of diabetes for men ages younger than 35 years, and by improving rates of diabetes diagnosis for men ages 65 years and older.

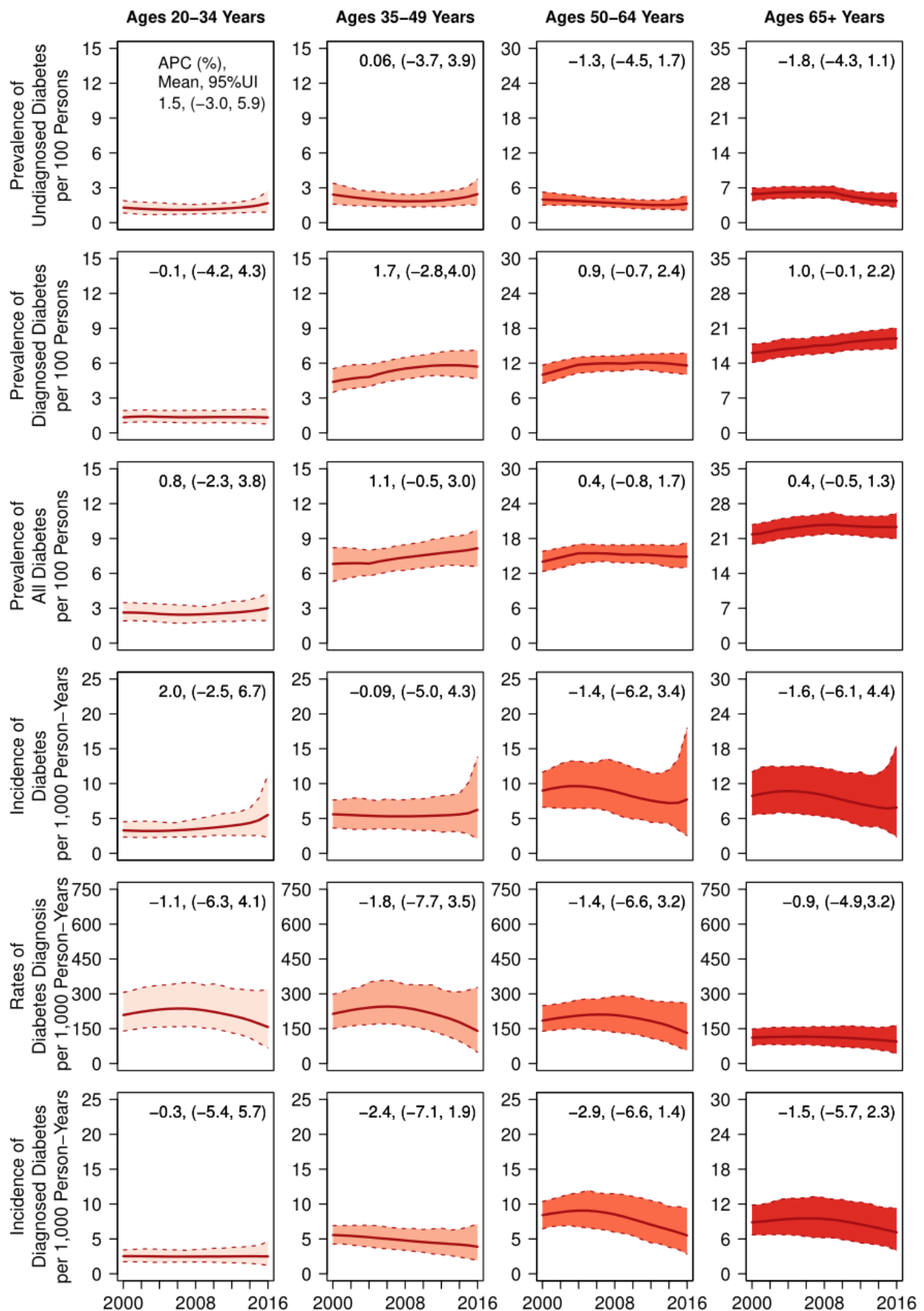
**Figure 2.2. Trends in age-standardized prevalence of undiagnosed, diagnosed, and all diabetes, incidence of diabetes and diagnosed diabetes, and rates of diabetes diagnosis in the overall population, men and women ages 20 years and older, 2000-2016.**



**Figure 2.3. Trends in prevalence of undiagnosed, diagnosed, and all diabetes, incidence of diabetes and diagnosed diabetes, and rates of diabetes diagnosis in men ages 20 years and older, by four age groups, 2000-2016.**



**Figure 2.4. Trends in prevalence of undiagnosed, diagnosed, and all diabetes, incidence of diabetes and diagnosed diabetes, and rates of diabetes diagnosis in women ages 20 years and older, by four age groups, 2000-2016.**



## Discussion

Overall, our analysis estimated a 17-year period of decreasing prevalence of undiagnosed diabetes and increasing prevalence of diagnosed and all diabetes during 2000-2016. The incidence of diabetes, rates of diabetes diagnosis, and incidence of diagnosed diabetes all decreased in the overall US population. In many age strata, trends in the prevalence and incidence of diabetes were similar to overall trends. However, the prevalence of undiagnosed diabetes and incidence of diabetes among young adults ages 20-34 years increased during 2000-2016. Incidence of diagnosed diabetes increased in men ages 65 years and older, driven by an improvement in rates of diabetes diagnosis.

Most previous studies using data from national surveys among U.S. adults (NHANES, NHIS, BRFSS) have focused on the prevalence and incidence of diagnosed diabetes.<sup>2,6-8,37,38</sup> Our estimated trends in the prevalence of undiagnosed, diagnosed and all diabetes, and trends in the incidence of diagnosed diabetes were comparable to these of previous studies.<sup>8,37</sup> However, little is known about the trends in the incidence of diabetes directly from nationally representative surveys. The Surveillance, Prevention, and Management of Diabetes Mellitus (SUPREME-DM) study, using electronic health records of inpatient care, laboratory and pharmacy data to identify diabetes in a diverse open cohort of 7 million insured adults in US, reported stabilization in incidence of overall diabetes over 2006-2010.<sup>39</sup> Another analysis based on claims data from all 50 US states suggested the annual rate of diagnosed diabetes declined substantially over 2007-2012, from 11.0 to 6.5 per 1,000 person-years.<sup>40</sup> Similar stable or decreasing trends in incidence of total or type 2 diabetes were reported by several studies for populations from Europe and Canada since 2006.<sup>14</sup>

Trends in the incidence of diabetes reported here are consistent with period changes in the four most important determinants of diabetes risk: age, overweight or obesity, smoking, and prediabetes. In our study, all outcome measures were age standardized, meaning that changes in age structure cannot explain the reported results. Regarding trends in prevalence of obesity, there were no significant changes during 2003-2010 among men and women adults, an increasing but not significant trend

during 2010-2016 among men adults, and a significant increasing trend during 2010-2016 among women adults in US.<sup>41,42</sup> Smoking prevalence was reported having significant declines after 2010 compared with 2003-2005, except for low socio-economic status (annual incomes of less than \$10,000).<sup>43</sup> Prediabetes, defined as FPG values of 100 to 126 mg/dL or Hemoglobin A<sub>1c</sub> values of 5.7% to 6.4%, was reported having stable prevalence among US adults during 2005-2016.<sup>37</sup> Trends in sugared beverage purchases, total food calories intake, and physical inactivity either plateaued or decreased since mid-2000s.<sup>44-47</sup> Taken together, current evidence on diabetes risk factors is consistent with the plateauing and slowing in diabetes incidence.

Trends in the incidence of diagnosed diabetes were not only related to trends in diabetes incidence but also to trends in the diagnosis rate, which has been affected by changes in diabetes diagnostic criteria and changes in testing and screening practices. In 1997 the American Diabetes Association (ADA) lowered the threshold for a diagnosis of diabetes from an FPG of 140 mg/dL to 126 mg/dL.<sup>48</sup> A subsequent increase in the diabetes diagnosis rates from 2000 to late-2000s was estimated in the overall and most population subgroups in our study. In 2010 the ADA added a Hemoglobin A<sub>1c</sub> of 6.5% or higher to the diagnostic criteria for diabetes.<sup>48</sup> Rates of diabetes diagnosis could increase if this change leads to increased awareness and detection of diabetes, as suggested by the SUPREME-DM study.<sup>39</sup> Alternatively, this change could lead to a lower rate of diabetes diagnosis because Hemoglobin A<sub>1c</sub> threshold is a less sensitive indicator than the FPG threshold.<sup>29,49</sup> In addition, it has been reported that Hemoglobin A<sub>1c</sub> is still used less frequently than glucose for screening during the first five years of this added recommendation in US.<sup>50</sup> As such, the balance of whether Hemoglobin A<sub>1c</sub> contributed to an increase or decrease in diabetes diagnosis rates remains uncertain. As suggested by our model results, estimated increases in the rates of diabetes diagnosis for men ages 65 years and older or decreases for women were not statistically significant. Practices around screening, testing, and diagnosis of diabetes were not monitored well by national surveillance systems, leaving limited information on diabetes screening changes associated with health reforms in the past few years.

The overall modest decrease in diabetes incidence and increase in diabetes prevalence in this study are consistent with substantial reductions in all-cause and cardiovascular disease specific mortality among individuals with diagnosed diabetes over the last three decades.<sup>51</sup> Extended life expectancies of individuals with diabetes after diagnosis would increase the number of people with diagnosed diabetes, leading to increased prevalence of diagnosed and all diabetes. Our results do not support the hypothesis of saturation of diagnosed diabetes, which postulates that recent declines in the incidence of diagnosed diabetes are due to depletion of the susceptible population through increased screening and diagnostic practices.<sup>52</sup> In our results, the prevalence of undiagnosed diabetes increased for adults ages 20-34 and 35-49 years, indicating that saturation has not been reached. The substantial decrease in undiagnosed diabetes prevalence for men aged 50 years and older was driven by both a decrease in diabetes incidence and an increase in diagnosis rates during 2000-2016.

To our knowledge, this study provides the first estimates of the incidence of diabetes, and rates of diabetes diagnosis among the US adult population from nationally representative surveys. Previous studies using national survey data only estimated the incidence of diagnosed diabetes from self-reports and were not verified by medical records, which underestimate disease incidence. However, there are several limitations. First, there were a substantial number of model parameters governing spline surfaces that described the annual age-specific incidence of diabetes and rates of diabetes diagnosis. Computational complexity would have increased substantially if we had further stratified by race/ethnicity, levels of income and education, or other social economic factors, so we did not extend the analysis to this level of detail. Second, the overall response rates for the Continuous NHANES has decreased over recent survey years. Although not introducing bias in the final survey estimates, the declining response rates do increase sampling error, especially in populations with relatively low prevalence such as young adults ages 20-34 years. Third, our analysis combined type 1 and type 2 diabetes because national surveys NHANES, NHIS, BRFSS do not distinguish the subtypes of diabetes. Prevalence of diagnosed type 1 diabetes was approximately estimated using data from NHANES by defining type 1 diabetes with self-report age of diabetes diagnosis and self-report

treatment with insulin.<sup>53</sup> Prevalence of undiagnosed type 1 diabetes cannot be estimated from NHANES due to lack of laboratory confirmation on cases of type 1 diabetes. Since type 2 diabetes is understood to account for around 95% of all diabetes for adults, the results from our estimates are most likely to be representative of trends in type 2 diabetes.

We estimated an overall decrease in diabetes incidence and an increase in diabetes prevalence for US adults ages 20 years and older during 2000-2016. Rates of diabetes diagnosis did not increase over the same interval. The findings in our study have important implications for monitoring the diabetes epidemic and guiding prevention efforts in the future. Although the incidence of diabetes decreased, declining mortality among diagnosed diabetic populations, and high and increasing prevalence of all diabetes have resulted in continued high total burden of diabetes. Diabetes prevention efforts, and improved screening and diagnosis, are needed to reduce the burden of both diabetes and diabetes complications.



## References

1. Mokdad AH, Ballestros K, Echko M, et al. The State of US Health, 1990-2016: Burden of Diseases, Injuries, and Risk Factors Among US States. *JAMA*. 2018;319(14):1444-1472. doi:10.1001/jama.2018.0158
2. Centers for Disease Control and Prevention. Long-term Trends in Diabetes. [https://www.cdc.gov/diabetes/statistics/slides/long\\_term\\_trends.pdf](https://www.cdc.gov/diabetes/statistics/slides/long_term_trends.pdf)
3. Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualization. Accessed February 18, 2020. <http://vizhub.healthdata.org/gbd-compare>.
4. Economic Costs of Diabetes in the U.S. in 2012. *Diabetes Care*. 2013;36(4):1033-1046. doi:10.2337/dc12-2625
5. Economic Costs of Diabetes in the U.S. in 2017. *Diabetes Care*. 2018;41(5):917-928. doi:10.2337/dci18-0007
6. Geiss LS, Wang J, Cheng YJ, et al. Prevalence and Incidence Trends for Diagnosed Diabetes Among Adults Aged 20 to 79 Years, United States, 1980-2012. *JAMA*. 2014;312(12):1218-1226. doi:10.1001/jama.2014.11494
7. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and Trends in Diabetes Among Adults in the United States, 1988-2012. *JAMA*. 2015;314(10):1021-1029. doi:10.1001/jama.2015.10029
8. Benoit SR, Hora I, Albright AL, Gregg EW. New directions in incidence and prevalence of diagnosed diabetes in the USA. *BMJ Open Diab Res Care*. 2019;7(1):e000657. doi:10.1136/bmjdr-2019-000657
9. State-Specific Incidence of Diabetes Among Adults --- Participating States, 1995--1997 and 2005--2007. Accessed May 12, 2020. <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5743a2.htm>
10. McBean AM, Li S, Gilbertson DT, Collins AJ. Differences in Diabetes Prevalence, Incidence, and Mortality Among the Elderly of Four Racial/Ethnic Groups: Whites, Blacks, Hispanics, and Asians. *Diabetes Care*. 2004;27(10):2317-2324. doi:10.2337/diacare.27.10.2317
11. Cowie CC, Rust KF, Byrd-Holt DD, et al. Prevalence of Diabetes and Impaired Fasting Glucose in Adults in the U.S. Population: National Health and Nutrition Examination Survey 1999--2002. *Diabetes Care*. 2006;29(6):1263-1268. doi:10.2337/dc06-0062
12. Dwyer-Lindgren L, Mackenbach JP, van Lenthe FJ, Flaxman AD, Mokdad AH. Diagnosed and Undiagnosed Diabetes Prevalence by County in the U.S., 1999--2012. *Dia Care*. 2016;39(9):1556-1562. doi:10.2337/dc16-0678

13. Centers for Disease Control and Prevention. *National Diabetes Statistics Report, 2017*. Atlanta, GA: U.S. Dept of Health and Human Services, 2017
14. Magliano DJ, Islam RM, Barr ELM, et al. Trends in incidence of total or type 2 diabetes: systematic review. *BMJ*. Published online September 11, 2019;I5003. doi:10.1136/bmj.I5003
15. Arias E. *United States Life Tables, 2000. National Vital Statistics Reports, Vol. 51, No. 3*. National Center for Health Statistics; 2002.
16. United States Department Of Health And Human Services. National Center For Health Statistics. National Health and Nutrition Examination Survey II: Mortality Study, 1992: Version 1. Published online March 30, 2006. doi:10.3886/ICPSR02631.V1
17. Raftery AE, Bao L. Estimating and Projecting Trends in HIV/AIDS Generalized Epidemics Using Incremental Mixture Importance Sampling. *Biometrics*. 2010;66(4):1162-1173.
18. Gelman A, Hill J. Chapter 8. Simulation for checking statistical procedures and model fits. In: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press; 2007.
19. National Center for Health Statistics. Plan and operation of the Third National Health and nutrition Examination Survey, 1988-94. *Vital Health Stat*. 1994;1(32).
20. Zipf G, Chiappa M, Porter KS. National Health and Nutrition Examination Survey: Plan and Operations, 1999-2010. *National Center for Health Statistics Vital Health Stat*. 2013;1(56).
21. Khare M, Mohadjer LK, Ezzati-Rice TM, Waksberg J. An Evaluation of Nonresponse Bias in NHANES III (1988-91). :6.
22. Groves RM, Peytcheva E. The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly*. 2008;72(2):167-189. doi:10.1093/poq/nfn011
23. NHANES 2005-2006: Plasma Fasting Glucose & Insulin Data Documentation, Codebook, and Frequencies. Accessed April 6, 2019. [https://wwwn.cdc.gov/Nchs/Nhanes/2005-2006/GLU\\_D.htm](https://wwwn.cdc.gov/Nchs/Nhanes/2005-2006/GLU_D.htm)
24. NHANES 2007-2008: Plasma Fasting Glucose & Insulin Data Documentation, Codebook, and Frequencies. Accessed April 6, 2019. [https://wwwn.cdc.gov/Nchs/Nhanes/2007-2008/GLU\\_E.htm](https://wwwn.cdc.gov/Nchs/Nhanes/2007-2008/GLU_E.htm)
25. NHANES 2009-2010: Plasma Fasting Glucose & Insulin Data Documentation, Codebook, and Frequencies. Accessed April 6, 2019. [https://wwwn.cdc.gov/Nchs/Nhanes/2009-2010/GLU\\_F.htm](https://wwwn.cdc.gov/Nchs/Nhanes/2009-2010/GLU_F.htm)
26. NHANES 2011-2012: Plasma Fasting Glucose & Insulin Data Documentation, Codebook, and Frequencies. Accessed April 6, 2019. [https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/GLU\\_G.htm](https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/GLU_G.htm)
27. NHANES 2013-2014: Plasma Fasting Glucose Data Documentation, Codebook, and Frequencies. Accessed April 6, 2019. [https://wwwn.cdc.gov/Nchs/Nhanes/2013-2014/GLU\\_H.htm](https://wwwn.cdc.gov/Nchs/Nhanes/2013-2014/GLU_H.htm)

28. NHANES 2015-2016: Plasma Fasting Glucose & Insulin Data Documentation, Codebook, and Frequencies. Accessed April 6, 2019. [https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/GLU\\_I.htm](https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/GLU_I.htm)
29. Association AD. Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*. 2014;37(Supplement 1):S81-S90. doi:10.2337/dc14-S081
30. Söderberg S, Zimmet P, Tuomilehto J, et al. High incidence of type 2 diabetes and increasing conversion rates from impaired fasting glucose and impaired glucose tolerance to diabetes in Mauritius. *Journal of Internal Medicine*. 2004;256(1):37-47. doi:10.1111/j.1365-2796.2004.01336.x
31. Abouzeid M, Wikström K, Peltonen M, et al. Secular trends and educational differences in the incidence of type 2 diabetes in Finland, 1972–2007. *Eur J Epidemiol*. 2015;30(8):649-659. doi:10.1007/s10654-015-0008-7
32. Sagesaka H, Sato Y, Someya Y, et al. Type 2 Diabetes: When Does It Start? *J Endocr Soc*. 2018;2(5):476-484. doi:10.1210/js.2018-00071
33. Harris MI, Klein R, Welborn TA, Knudman MW. Onset of NIDDM occurs at Least 4–7 yr Before Clinical Diagnosis. *Diabetes Care*. 1992;15(7):815-819. doi:10.2337/diacare.15.7.815
34. Porta M, Curletto G, Cipullo D, et al. Estimating the Delay Between Onset and Diagnosis of Type 2 Diabetes From the Time Course of Retinopathy Prevalence. *Diabetes Care*. 2014;37(6):1668-1674. doi:10.2337/dc13-2101
35. Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in R. *BMC Medical Research Methodology*. 2019;19(1):46. doi:10.1186/s12874-019-0666-3
36. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. Vol 103. Springer New York; 2013. doi:10.1007/978-1-4614-7138-7
37. National Diabetes Statistics Report 2020. Estimates of diabetes and its burden in the United States. Published online 2020:32.
38. Selvin E, Parrinello CM, Sacks DB, Coresh J. Trends in Prevalence and Control of Diabetes in the United States, 1988–1994 and 1999–2010. *Ann Intern Med*. 2014;160(8):517. doi:10.7326/M13-2411
39. Nichols GA, Schroeder EB, Karter AJ, et al. Trends in Diabetes Incidence Among 7 Million Insured Adults, 2006–2011The SUPREME-DM Project. *Am J Epidemiol*. 2015;181(1):32-39. doi:10.1093/aje/kwu255
40. Weng W, Liang Y, Kimball ES, et al. Decreasing incidence of type 2 diabetes mellitus in the United States, 2007–2012: Epidemiologic findings from a large US claims database. *Diabetes Research and Clinical Practice*. 2016;117:111-118. doi:10.1016/j.diabres.2016.04.043

41. Flegal KM, Carroll MD, Kit BK, Ogden CL. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. *JAMA*. 2012;307(5):491-497. doi:10.1001/jama.2012.39
42. Hales CM, Fryar CD, Carroll MD, Freedman DS, Ogden CL. Trends in Obesity and Severe Obesity Prevalence in US Youth and Adults by Sex and Age, 2007-2008 to 2015-2016. *JAMA*. 2018;319(16):1723-1725. doi:10.1001/jama.2018.3060
43. Pernenkil V, Wyatt T, Akinyemiju T. Trends in smoking and obesity among US adults before, during, and after the great recession and Affordable Care Act roll-out. *Preventive Medicine*. 2017;102:86-92. doi:10.1016/j.ypmed.2017.07.001
44. Ng SW, Slining MM, Popkin BM. Turning point for US diets? Recessionary effects or behavioral shifts in foods purchased and consumed. *Am J Clin Nutr*. 2014;99(3):609-616. doi:10.3945/ajcn.113.072892
45. Kit BK, Fakhouri TH, Park S, Nielsen SJ, Ogden CL. Trends in sugar-sweetened beverage consumption among youth and adults in the United States: 1999–2010. *Am J Clin Nutr*. 2013;98(1):180-188. doi:10.3945/ajcn.112.057943
46. Ussery EN, Carlson SA, Whitfield GP, Watson KB, Berrigan D, Fulton JE. Walking for Transportation or Leisure Among U.S. Women and Men - National Health Interview Survey, 2005-2015. *MMWR Morbidity and mortality weekly report*. 2017;66(25):657–662. doi:10.15585/mmwr.mm6625a1
47. Valizadeh P, Popkin BM, Ng SW. Distributional Changes in U.S. Sugar-Sweetened Beverage Purchases, 2002–2014. *American Journal of Preventive Medicine*. 2020;59(2):260-269. doi:10.1016/j.amepre.2020.02.002
48. The Expert Committee on the Diagnosis and Classification of Diabetes. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*. 1997;20(7):1183-1197. doi:10.2337/diacare.20.7.1183
49. The International Expert Committee. International Expert Committee Report on the Role of the A1C Assay in the Diagnosis of Diabetes. *Diabetes Care*. 2009;32(7):1327-1334. doi:10.2337/dc09-9033
50. Evron JM, Herman WH, McEwen LN. Changes in Screening Practices for Prediabetes and Diabetes Since the Recommendation for Hemoglobin A1c Testing. *Diabetes Care*. 2019;42(4):576-584. doi:10.2337/dc17-1726
51. Gregg EW, Cheng YJ, Srinivasan M, et al. Trends in cause-specific mortality among adults with and without diagnosed diabetes in the USA: an epidemiological analysis of linked national survey and vital statistics data. *Lancet*. 2018;391(10138):2430-2440. doi:10.1016/S0140-6736(18)30314-3

52. Selvin E, Ali MK. Declines in the Incidence of Diabetes in the U.S.—Real Progress or Artifact? *Diabetes Care*. 2017;40(9):1139-1143. doi:10.2337/dc16-2442
53. Menke A, Orchard TJ, Imperatore G, Bullard KM, Mayer-Davis E, Cowie CC. The Prevalence of Type 1 Diabetes in the United States: *Epidemiology*. 2013;24(5):773-774. doi:10.1097/EDE.0b013e31829ef01a

## Supplementary Material

### 1. Analytic sample

Table S2.1 shows numbers of participants and participant characteristics in the analytic samples for the interview and laboratory examination, by survey cycle years for NHANES III (1988-1994) and Continuous NHANES (1999-2016).

**Table S2.1. Description of participants in the analytic sample of interview questionnaire and laboratory examination for NHANES III (1988-1994) and Continuous NHANES (1999-2016).**

	NHANES III					Continuous NHANES					
Survey years	1988-1991	1992-1994	1999-2000	2001-2002	2003-2004	2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	2015-2016
	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
<b>Interview sample (n)</b>	9488	9337	2168	2479	2299	2191	2901	3118	2781	2897	2814
<b>Laboratory sample (n)</b>	3227	3399	2168	2479	2299	2191	2901	3118	2781	2897	2814
<b>Women (%)</b>	52.4 (50.1, 54.7)	53.8 (50.8, 55.3)	50.6 (48.1, 53.2)	51.4 (49.9, 53.0)	51.3 (49.6, 53.1)	50.8 (48.5, 53.1)	51.3 (48.9, 53.7)	51.2 (49.2, 53.1)	51.7 (49.0, 54.3)	51.3 (49.2, 53.4)	51.3 (49.7, 52.9)
<b>Age (mean, years)</b>	44.6 (43.8, 45.4)	44.9 (43.7, 46.1)	45.6 (44.3, 46.9)	45.7 (44.0, 47.4)	46.5 (45.2, 47.7)	47.0 (45.4, 48.7)	46.9 (45.8, 48.0)	47.2 (46.1, 48.3)	47.6 (46.3, 48.9)	47.6 (46.4, 48.9)	47.9 (46.7, 49.1)
<b>Ages 20-34 (%)</b>	35.6 (33.5, 37.8)	33.7 (31.0, 36.5)	29.7 (26.1, 33.7)	28.7 (24.2, 33.7)	27.9 (25.0, 31.1)	26.3 (23.1, 29.9)	27.0 (23.8, 30.5)	27.8 (24.9, 30.9)	26.1 (22.4, 30.3)	27.2 (24.5, 30.1)	27.2 (23.8, 30.8)
<b>Ages 35-49 (%)</b>	29.2 (27.2, 31.4)	31.3 (29.0, 33.6)	33.1 (29.1, 37.2)	33.2 (29.6, 37.0)	32.0 (28.6, 35.7)	32.5 (27.6, 37.8)	30.7 (26.7, 35.1)	28.7 (26.8, 30.1)	28.7 (25.4, 32.2)	27.2 (23.8, 30.9)	26.3 (23.3, 29.7)
<b>Ages 50-64 (%)</b>	18.3 (17.6, 19.1)	17.9 (16.2, 19.8)	21.3 (19.0, 23.8)	23.2 (19.4, 27.6)	23.6 (20.3, 27.2)	23.8 (20.8, 27.1)	25.4 (22.3, 28.8)	25.5 (22.6, 28.7)	27.8 (24.7, 31.1)	26.8 (24.2, 29.7)	25.9 (23.1, 28.8)
<b>Ages ≥ 65 (%)</b>	16.8 (15.1, 18.7)	17.1 (14.7, 19.7)	15.9 (14.0, 18.0)	14.8 (13.0, 16.8)	16.5 (14.2, 19.1)	17.4 (14.0, 21.3)	16.8 (14.9, 19.0)	18.3 (16.8, 19.8)	17.4 (15.8, 19.1)	18.7 (16.3, 21.4)	20.6 (18.1, 23.4)

## 2. Markov model

We developed a Markov model to represent undiagnosed and diagnosed diabetes in the U.S. population ages 20 year and older between 2000-2016. Four disease states were distinguished: disease-free (state 0), undiagnosed diabetes (state1), diagnosed diabetes (state 2), and death (state 3) (Figure S2.1 Panel A).

We modeled the transition probabilities between the four states by single year of age (i) and single calendar year (t) with the transition matrix presented in Figure S2.1 Panel B. The transition probabilities include mortality rates (noted as  $\mu$ ) and incidence rates (noted as  $\lambda$ ). We estimated mortality rates by from national life tables.<sup>1</sup> We estimated incidence rates by constructing individual B-spline surfaces with weakly-informative priors for parameters of the surfaces (see main manuscript).

Age- and sex-specific yearly mortality rates for each of the three diabetes states ( $\mu_{df}$ ,  $\mu_{ud}$ ,  $\mu_{dd}$ ) were estimated using 1) all-cause mortality rates (M) among the general population, 2) prevalence of undiagnosed and diagnosed diabetes ( $pr_{ud}$ ,  $pr_{dd}$ ), 3) and risk ratios ( $rr_{ud}$ ,  $rr_{dd}$ ) of all-cause mortality for undiagnosed and diagnosed diabetes compared to those without diabetes.

The mortality rates for disease-free ( $\mu_{df}$ ) were estimated based on subtracting diabetes-related mortality from all-cause mortality based on prevalence and risk ratios estimates, using equation (1).

$$\mu_{df} = M / (1 + (rr_{ud} - 1) * pr_{ud} + (rr_{dd} - 1) * pr_{dd}) \quad (1)$$

where

M is the all-cause mortality rates among the general population, which was obtained from the annual United States Life Tables by NCHS.<sup>1</sup>

$rr_{ud}$  is the risk ratio of all-cause mortality of undiagnosed diabetes to all-cause mortality of disease free, which was estimated from the SNHANES-MS<sup>2</sup> and reported in Table S2.2.

$pr_{ud}$  is the prevalence of undiagnosed diabetes, which was estimated from the NHANES III and Continuous NHANES.

$rr_{dd}$  is the risk ratio of all-cause mortality of diagnosed diabetes to all-cause mortality of disease free, which was estimated from the SNHANES-MS<sup>2</sup> and reported in Table S2.2.

$pr_{dd}$  is the prevalence of diagnosed diabetes, which was estimated from the NHANES III and Continuous NHANES.

Mortality rates for undiagnosed diabetes ( $\mu_{ud}$ ) and diagnosed diabetes ( $\mu_{dd}$ ) were estimated based on the mortality rates for disease-free ( $\mu_{df}$ ) and corresponding risk ratios ( $rr_{ud}$ ,  $rr_{dd}$ ), using equation (2) and equation (3).

$$\mu_{ud} = rr_{ud} * \mu_{df} \tag{2}$$

$$\mu_{dd} = rr_{dd} * \mu_{df} \tag{3}$$

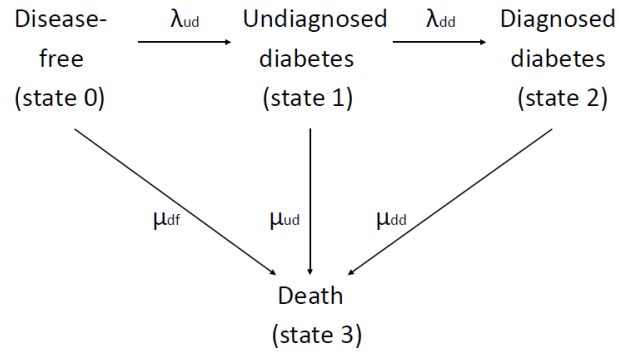
The total number of parameters in our model was 324. In each sex, there were 48 parameters defining the B-spline surface for incidence of diabetes ( $\lambda_{ud}$ ), 48 parameters defining the B-spline surface for diagnosis rate of diabetes ( $\lambda_{dd}$ ), 33 parameters for the risk ratios of undiagnosed diabetes to disease-free ( $rr_{ud}$ ), 33 parameters for the risk ratios of diagnosed diabetes to disease-free ( $rr_{dd}$ ).

To avoid overfitting and computational complexity, it is usually not necessary to place a knot at every interval (calendar year) in a spline function. A penalized spline that uses a reduced set of equidistant knots based on the B-spline basis is suggested as a general class of spline regression to approximate smoothing.<sup>3-5</sup> Penalty level, denoted as  $\theta$ , controls the trade-off between smoothness and model fit.  $\theta=0$  implies that a smoothing spline interpolates the data, while  $\theta = \infty$  implies a linear function.<sup>3</sup> Figure S2.2. shows the B-splines basis when applying different levels of penalty:  $\theta = 0$ ,  $\theta = 1$ ,  $\theta = 55$  to the B-spline surface of incidence of diabetes at year dimension.



Each B-spline surface had 5 knots with 2 degrees of freedom in the age dimension and 7 knots with 2 degrees of freedom in the calendar year dimension, resulting in a total of  $(5+2-1)*(7+2-1)=48$  number of parameters. For the risk ratios ( $rr_{ud}$ ), each had 33 parameters as independent means of the estimated risk ratios for 11 survey cycle years (2 for NHANES III and 9 for Continuous NHANES) by 3 age groups (20-49, 50-64, and  $\geq 65$  years). We specified weakly informative priors for 96 parameters of B-spline surfaces. We specified log-normal distributions for the risk ratios, with priors of means estimated by fitting Cox proportional hazards models on data from the Second National Health and Nutrition Examination Survey Mortality Study (SNHANES-MS)<sup>2</sup> and reported in Table S2.2. The most recent available data of SNHANES-MS was survey cycle year of 2009-2010. We set the means of risk ratios after 2010 equal to the estimates in 2009-2010 and allowed broad standard deviations.

**A**



**B**

	Disease free	Undiagnosed diabetes	Diagnosed diabetes	Death
	0	1	2	3
0	$1 - \lambda_{ud(i,t)} - \mu_{df(i,t)}$	$\lambda_{ud(i,t)}$	0	$\mu_{df(i,t)}$
1	0	$1 - \lambda_{dd(i,t)} - \mu_{ud(i,t)}$	$\lambda_{dd(i,t)}$	$\mu_{ud(i,t)}$
2	0	0	$1 - \mu_{dd(i,t)}$	$\mu_{dd(i,t)}$
3	0	0	0	1

**Figure S2.1. A four-state Markov model of undiagnosed and diagnosed diabetes (Panel A) and states transition matrix for individuals at single year of age (i) and single calendar year (t).**

df=disease free; ud=undiagnosed diabetes; dd=diabetes diagnosis.

$\lambda_{ud}$ : diabetes incidence rate, representing the transition rate from disease-free to undiagnosed diabetes.

$\lambda_{dd}$ : diagnosis rate of diabetes, representing the transition rate from undiagnosed to diagnosed diabetes.

$\mu_{df}$ : all-cause mortality rate for disease-free individuals.

$\mu_{ud}$ : all-cause mortality rate for undiagnosed diabetic individuals, representing the transition rate from undiagnosed diabetes to death,  $\mu_{ud} = rr_{ud} * \mu_{df}$ .

$\mu_{dd}$ : all-cause mortality rate for diagnosed diabetic individuals, representing the transition rate from diagnosed diabetes to death,  $\mu_{dd} = rr_{dd} * \mu_{df}$ .

$rr_{ud}$ : Risk ratio of all-cause mortality of undiagnosed diabetes to all-cause mortality of disease-free.

$rr_{dd}$ : Risk ratio of all-cause mortality of diagnosed diabetes to all-cause mortality of disease-free.

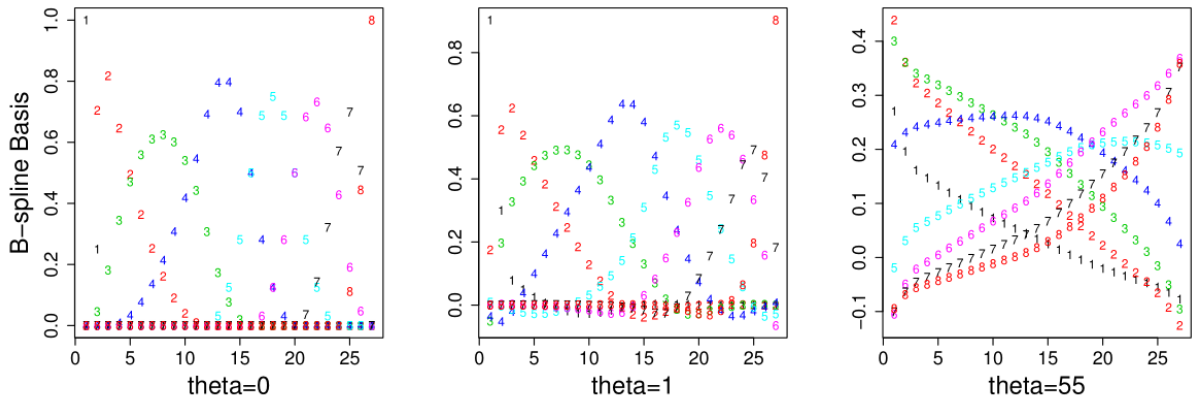


Figure S2.2. Example of B-spline basis with three different levels of penalty  $\theta = 0, \theta = 1, \theta = 55$ .

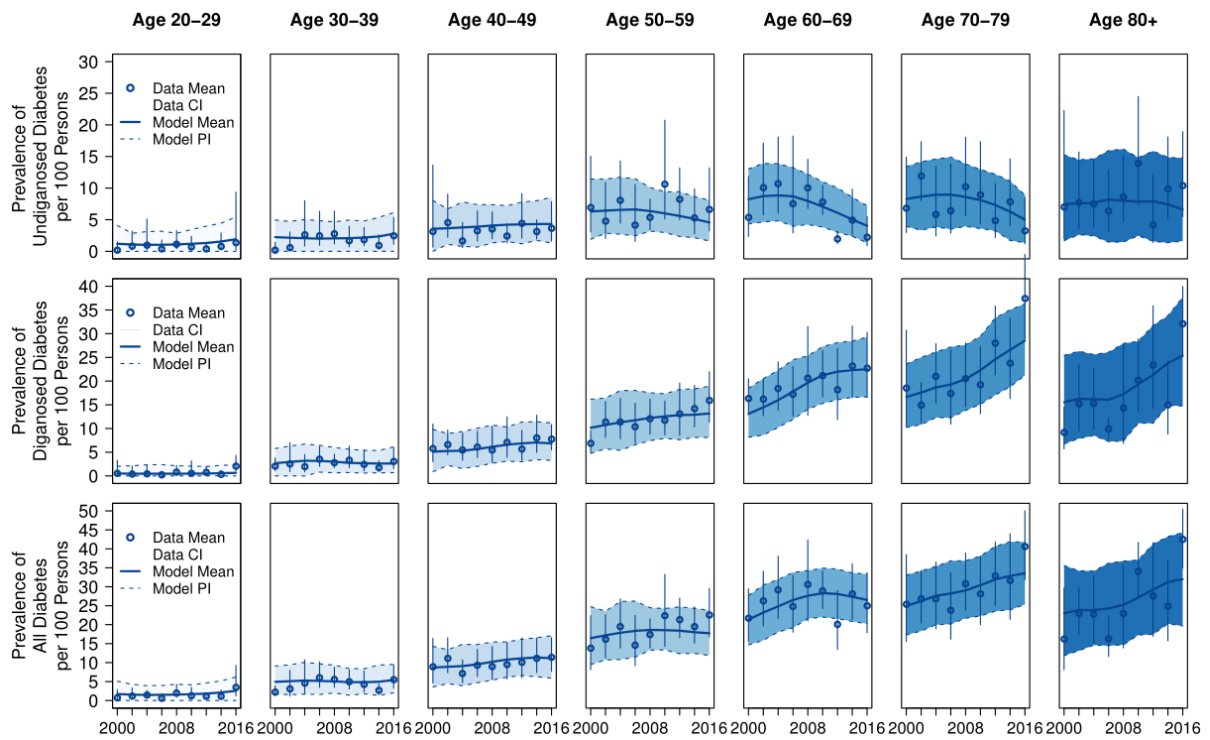
Table S2.2. The risk ratios for all-cause mortality of undiagnosed and diagnosed diabetes compared to the general population without diabetes, by age groups and sex.

	NHANES III				Continuous NHANES			
Sex and age group	1988-1990	1991-1994	1999-2000	2001-2002	2003-2004	2005-2006	2007-2008	2009-2010
<b>Risk ratios for all-cause mortality comparing undiagnosed diabetes to those without diabetes (<math>rr_{ud}</math>)</b>								
<b>Men</b>								
Ages 20-49	5.06	4.65	3.98	5.60	4.50	6.35	6.55	3.91
Ages 50-64	1.19	1.16	1.12	1.32	1.21	1.49	1.54	1.15
Ages $\geq 65$	1.09	1.05	1.06	1.20	1.08	1.37	1.41	1.07
<b>Women</b>								
Ages 20-49	3.73	3.19	3.13	2.16	2.53	2.57	2.60	3.32
Ages 50-64	2.19	1.87	1.84	1.27	1.48	1.51	1.53	2.29
Ages $\geq 65$	1.66	1.42	1.39	1.02	1.12	1.14	1.16	1.25
<b>Risk ratios for all-cause mortality comparing diagnosed diabetes to those without diabetes (<math>rr_{d}</math>)</b>								
<b>Men</b>								
Age 20-49	7.59	7.21	6.42	8.34	8.42	8.13	6.39	7.09
Age 50-64	2.70	2.57	2.29	2.97	3.00	2.90	2.28	2.53
Age $\geq 65$	1.37	1.30	1.16	1.50	1.52	1.47	1.15	1.28
<b>Women</b>								
Age 20-49	4.38	5.09	4.28	3.36	4.11	4.27	4.68	3.28
Age 50-64	3.15	3.66	3.08	2.41	2.95	3.07	3.36	2.35
Age $\geq 65$	1.99	2.31	1.94	1.52	1.86	1.93	2.12	1.49

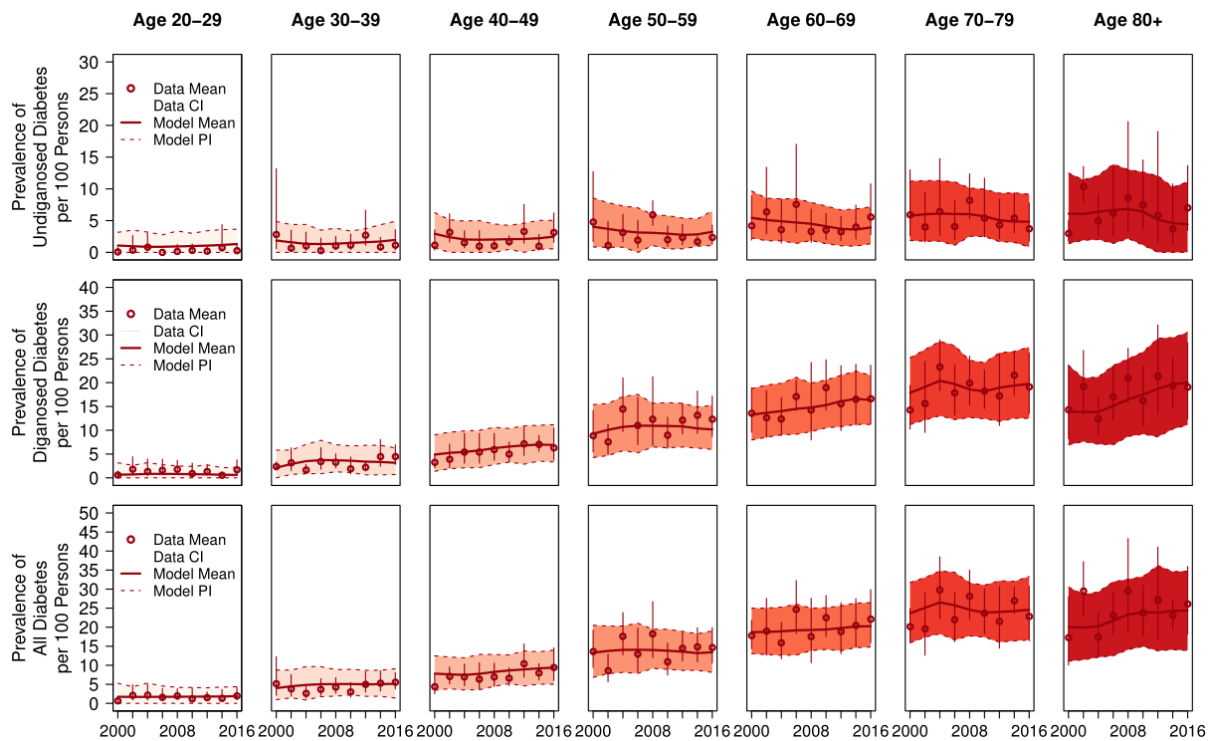
### **3. Evaluation of model fits with posterior predictive checks**

We used posterior predictive checks to evaluate the model fits to prevalence trends implied by NHANES data.<sup>6</sup> Visual comparison of modelled estimates (mean and 95% predictive intervals [95% PI]) to empirical values (mean and confidence intervals) for prevalence of undiagnosed, diagnosed, and all diabetes are presented in Figure S2.3 and Figure S2.4, by sex and age groups.

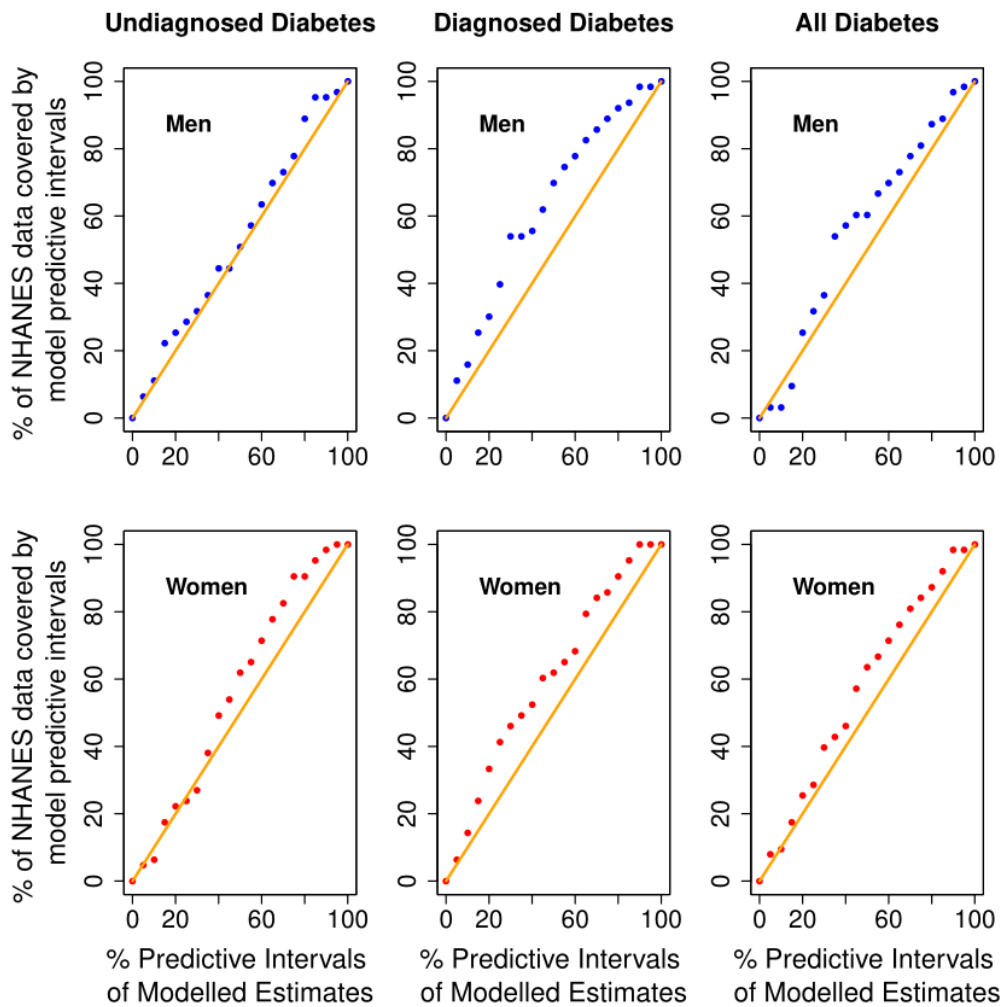
Model fits with posterior predictive checks were evaluated based on the proportion of empirical values observed from NHANES data being covered by the percentage of predictive intervals of modelled estimates. The proportion of observed NHANES data covered by the percentage of PI of modelled estimates for undiagnosed, diagnosed, and all diabetes, from 0% PI to 100% PI by 0.05 increment of the sequence, were plotted in Figure S2.5 for men and women separately. In details, in each panel of Figure S2.5, the x-axis represented the percentage from 0% to 100% PI of modelled estimates for each of the prevalence quantity by 5% increment of the sequence. The y-axis represented the proportion of observed data for each of the prevalence quantity from NHANES covered by the PI of modelled estimates. For example, the 95% PI covered 96.8% of observed prevalence of diagnosed diabetes for men and 98.4% of diagnosed diabetes for women, respectively. Most of the points were slightly higher or close to the 45-degree line, indicating that nearly all of the model estimated PIs covered slightly more or equal to the corresponding proportion of empirical values from the observed data. In summary, the posterior predictive checks showed that our model predicted the prevalence of diabetes well.



**Figure S2.3. Model fits of observed NHANES estimates (mean and confidence intervals) versus model predictions (mean and predictive intervals) of prevalence of undiagnosed, diagnosed and all diabetes, by age group in men during 2000-2016.**



**Figure S2.4. Model fits of observed NHANES estimates (mean and confidence intervals) versus model predictions (mean and predictive intervals) of prevalence of undiagnosed, diagnosed and all diabetes, by age group in women during 2000-2016.**



**Figure S2.5. Model fits evaluation: proportion of observed NHANES data covered by percentage of model posterior predictive intervals by outcomes and sex.**

#### **4. Wide uncertainty intervals (UIs) of incidence estimates**

In these analyses, the uncertainty intervals (UIs) were wider for estimates of incidence than estimates of prevalence (Figure 2.3 and Figure 2.4). This is due to our model being calibrated to prevalence of undiagnosed, diagnosed, and all diabetes using data from NHANES. NHANES doesn't have measures for incidence of diabetes, and incidence of diabetes diagnosis among undiagnosed adults. In addition, our analysis did not consider claims data or other cohort studies with measures of incidence of diabetes. Therefore, our model was not fit to incidence estimates but only prevalence estimates from empirical data, leading to wider UIs of model estimated incidence than model estimated prevalence, across all age groups.

Second, when only examining UIs of incidence, absolute width of UIs was not larger among old age groups than young age groups. For example, absolute width of UIs of incidence of diabetes diagnosis for women adults ages 50-64 years, and ages 65 years and older were smaller than that for women ages 20-49 years. Moreover, the relative width of UIs was also not larger among old age group than young age groups. For example, the UIs of incidence of diabetes were similar twice of the mean estimates for women ages 20-34 years and women ages 50-64 years.

## Supplementary Material References

1. Arias E. *United States Life Tables, 2000. National Vital Statistics Reports, Vol. 51, No. 3.* National Center for Health Statistics; 2002.
2. United States Department Of Health And Human Services. National Center For Health Statistics. National Health and Nutrition Examination Survey II: Mortality Study, 1992: Version 1. Published online March 30, 2006. doi:10.3886/ICPSR02631.V1
3. Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in R. *BMC Medical Research Methodology.* 2019;19(1):46. doi:10.1186/s12874-019-0666-3
4. Eilers PHC, Marx BD. Splines, knots, and penalties. *WIREs Comp Stat.* 2010;2(6):637-653. doi:10.1002/wics.125
5. Eilers PHC, Marx BD. Flexible Smoothing with B-splines and Penalties. *Statistical Science.* 1996(11):89-121.
6. Gelman A, Hill J. Chapter 8. Simulation for checking statistical procedures and model fits. In: *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press; 2007.

# **Chapter 3. Risk score to predict cardiovascular disease risk for patients with type 2 diabetes mellitus in the United States: a pooled analysis of prospective cohorts**

Yunfei Li<sup>1</sup>, Nicolas A Menzies<sup>1</sup>, Joshua A Salomon<sup>2</sup>, Goodarz Danaei<sup>1,3</sup>

1. Department of Global Health and Population, Harvard T.H. Chan School of Public Health
2. Center for Health Policy / Center for Primary Care and Outcomes Research, Stanford University
3. Department of Epidemiology, Harvard T.H. Chan School of Public Health



## **Abstract**

### **Background**

Previous research has shown that most cardiovascular disease (CVD) prediction models developed for the general population tend to underestimate CVD risk for diabetic patients. Current prediction equations specifically designed for patients with type 2 diabetes mellitus (T2DM) were developed using data from randomized clinical trials, and may lack generalizability to diabetic populations without intensive treatment and routine management. We developed a model to predict CVD risk (defined as fatal-plus-non-fatal coronary heart disease and stroke) among diabetic patients in the United States.

### **Methods**

We used data from diabetic patients with no history of CVD enrolled in five prospective cohort studies in the United States. With these data, we estimated the coefficients of a risk equation for fatal-plus-non-fatal CVD, using a sex-and-cohort stratified Cox proportional-hazards model. The risk prediction equation included age, body mass index (BMI), current smoking, systolic blood pressure (SBP), total cholesterol (TC), high-density lipoproteins cholesterol (HDL-C), fasting plasma glucose (FPG), treatment with hypertension medications, and treatment with diabetes medications. We conducted internal validation and out-of-sample cross-validation to evaluate model discrimination and calibration.

### **Findings**

We included data from 3,723 participants with diabetes. The mean age at baseline was 60 years (standard deviation 9.6), and 48.6% of eligible participants were women. During 15 years of follow-up, 659 (34.5%) men and 451 (24.9%) women developed a first CVD event. The risk score demonstrated good discrimination (Harrell's C statistic, 70% [95% confidence interval, 68% to 72%]) and calibration (regression slope, 0.99 [0.98 to 1.00]). In the cross-validation, the median C statistic was 70%, with a

range of 68% to 71%. The median slope of the calibration regression was 0.99, with a range of 0.96 to 1.05.

## **Conclusions**

We developed a novel CVD risk prediction model specifically for populations with T2DM in the United States. Our risk equation demonstrated improved performance relative to current models, which underestimate the total CVD risk among diabetic patients.

## Background

Patients with type 2 diabetes mellitus (T2DM) have a two-to-three-fold increased risk of cardiovascular disease (CVD) and death compared to people without diabetes.<sup>1-6</sup> T2DM patients also have substantially greater health resource use and economic costs from CVD than people without T2DM.<sup>7,8</sup> Multifactorial interventions through the management of high blood pressure, hypercholesterolaemia and hyperglycaemia have been effective in reducing the risk of fatal-plus-non-fatal CVD among diabetic populations.<sup>9-12</sup> Most guidelines for CVD prevention considered diabetes as a “coronary risk equivalent” that assigns patients with T2DM without prior myocardial infarction (MI) a high risk of coronary heart disease (CHD) similar to non-diabetic individuals who have had a MI.<sup>13-19</sup> The potential benefits of such treatment is often estimated using models obtained from the general population including diabetes as a predictor.<sup>1,2,4,20-23</sup> These risk prediction models usually do not perform well in diabetic populations, with moderate to poor discrimination or calibration.<sup>24,25</sup> Previous research showed that most CVD prediction models developed in the general population tend to underestimate CVD risk for diabetic populations, due to differences in CVD incidence between the diabetic and general populations.<sup>25-29</sup> Several studies in recent years have suggested that this problem could be overcome by developing diabetes-specific prediction models using data from exclusively diabetic populations.<sup>25,28</sup>

However, current diabetes-specific CVD prediction models have their own limitations. First, most data for model formation and validation have been derived from a limited set of populations, mostly from developed countries or Caucasian diabetic populations outside of the US.<sup>30-36</sup> There is growing evidence that the risk equations developed in one population cannot be applicable to other populations or between subgroups of population within a country.<sup>37,38,38,39</sup> This is because the mean risk factor levels and other determinants of CVD and death, such as genetics, societal factors, health care system for diagnostic, treatment, and follow-up among diabetic populations are different, resulting in differences in the risk of CVD and death between different diabetic populations and over

time. Second, current available diabetes-specific models developed among US populations predict only an individual CVD component such as CHD, rather than the combination of major CVD events, which is of interest to patients and care providers.<sup>40,41</sup> Recently, two diabetes-specific models have been developed based on multi-center data of clinical trials from multiple countries.<sup>39,42,43</sup> These models were developed and validated in clinical trial populations, and may lack generalizability to diabetic populations without intensive treatment regimens. Compared to other CVD risk equations,<sup>1,36</sup> these models used data with fewer years of follow-up<sup>39,42</sup> or had inferior internal validation.<sup>43</sup>

Therefore, we developed a novel CVD risk prediction model specifically for populations with T2DM in the US. Our model can be used to quantify the CVD disease burden over the next decade, simulate treatment guidelines and evaluate effectiveness and cost-effectiveness of interventions strategies, and guide primary care prevention policies for US diabetic populations. Although there are only a few countries with diabetes registries that have CVD and death rates among diabetic populations at this moment, our approach allows for recalibration of the model for use in different countries when the data become available.

## **Methods**

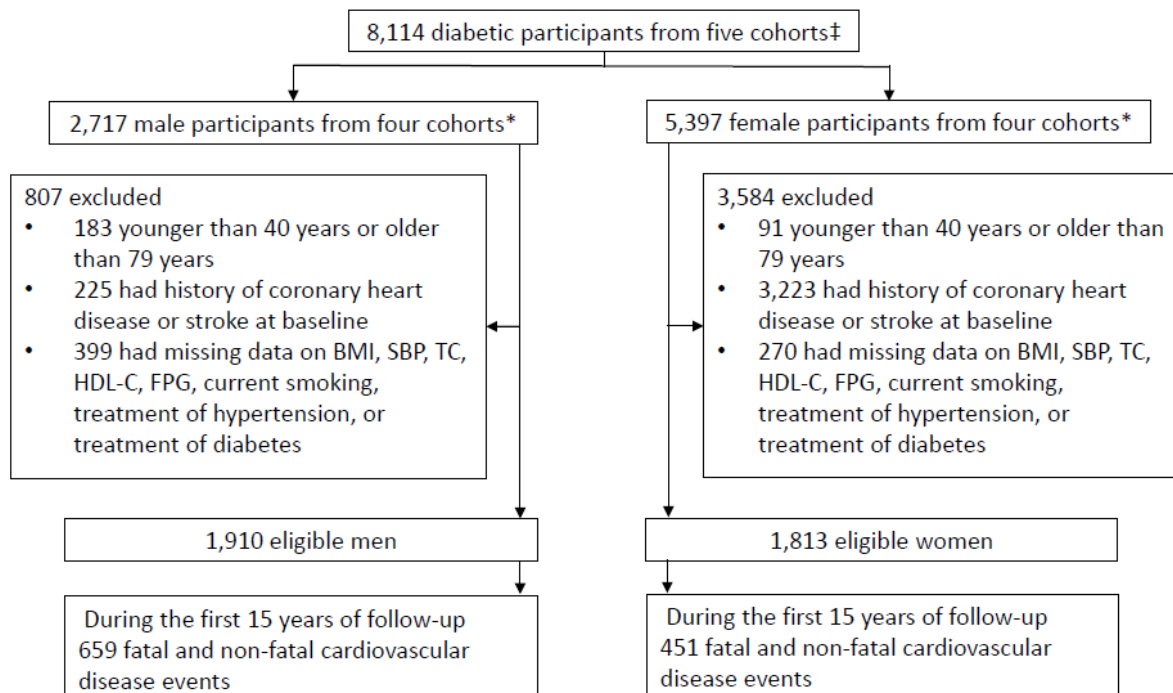
### **Data sources**

We collated individual-level data from 10 prospective cohorts: the Atherosclerosis Risk in Communities (ARIC), the Cardiovascular Health Study (CHS), the Framingham Heart Study original cohort (FHS-original), the Framingham Heart Study offspring cohort (FHS-offspring), the Honolulu Heart Program (HHP), the Puerto Rico Heart Health Program (PRHHP), the Jackson Heart Study (JHS), the Multi-Ethnic Study of Atherosclerosis (MESA), the Multiple Risk Factor Intervention Trial (MRFIT), and the Women's Health Initiative Clinical Trial and Observational Study (WHICTOS). We excluded the FHS-original and FHS-offspring cohorts because the total number of individuals with diabetes was low (less than 200 in each cohort). HHP, PRHHP and JHS were excluded due having to few CVD events (less

than 50 in each cohort) among eligible participants with diabetes who met our study inclusion criteria. We pooled individual-level data from the remaining 5 cohorts (ARIC, CHS, MESA, MRFIT, and WHICTOS). Pooling data from multiple cohorts enhanced statistical prediction power, which in turn allowed inclusion of interaction terms between age or sex and risk factors. Pooling data from multiple cohorts also reduced the effect of between-cohort variation on the coefficients.

Participants with diabetes were defined as individuals who had any one of (i) self-reported diabetes, (ii) medication with insulin or oral hypoglycaemic agents, (iii) fasting plasma glucose  $\geq 126$  mg/dL or 7mmol/L, (iv) random plasma glucose  $\geq 200$  mg/dL or 11.1 mmol/L, or (v) postprandial plasma glucose  $\geq 225$  mg/dL or 12.5 mmol/L at baseline, depending on the available data in each cohort. Our definition of diabetes did not distinguish between type 1 and type 2 diabetes, as most of the cohort studies did not distinguish the subtypes of diabetes. Since T2DM is generally considered as accounting for almost 95% of all diabetes for adults, risk prediction equation developed in this study will be more applicable to patients with T2DM. We included participants who were aged older than 40 years and younger than 80 years at baseline, due to the small number of events in younger participants and the heterogeneity and complexity of age-covariate interactions in the elderly population.<sup>1,44</sup> Participants were included if they did not have a history of CHD or stroke. The number of female participants with a history of CHD or stroke at baseline was 3,223, which was much larger than the number in male participants (225). This can be explained by that the substantial sex/gender differences in the prevalence and burden of CVDs among diabetic populations.<sup>45</sup> This was also consistent with findings from the previous Globorisk study that the association of CVD risk and diabetes was stronger in women than in men.<sup>2</sup> In addition, this can be explained by the inclusion of male participants from the MRFIT cohorts. Male participants from MRFIT were aged between 40-57 years at baseline, compared with female participants from the four perspectives were aged from 45-79 years. Female participants accounted for a larger proportion of the elderly population in which the prevalence of baseline CVD is greatest.<sup>46</sup> A total of 669 participants were excluded if they had missing data for the risk factors used in the risk

prediction equation. After exclusion, 3,723 participants with diabetes (mean age, 60 years; 1,813 women) were included in the analysis (Figure 3.1).



**Figure 3.1. Flowchart of inclusion and exclusion of cohort participants.**

‡Diabetic participants were defined as individuals who had any one of (i) self-reported diabetes, (ii) medication with insulin or oral hypoglycaemic agents, (iii) fasting plasma glucose  $\geq 126$  mg/dL or 7 mmol/L, (iv) random plasma glucose  $\geq 200$  mg/dL or 11.1 mmol/L, or (v) postprandial plasma glucose  $\geq 225$  mg/dL or 12.5 mmol/L at baseline, depending on the available data in each cohort.

\*The Multiple Risk Factor Intervention Trial include only men, and the Women’s Health Initiative Clinical Trial includes only women.

Abbreviations: BMI = body mass index, SBP = systolic blood pressure, TC = total cholesterol, HDL-C = high density lipoprotein-cholesterol, FPG = fasting plasma glucose

## Statistical analysis

Cox proportional hazards regression was used to estimate the coefficients of the risk equation to predict 10-year risk of a first CVD event for patients with T2DM in the United States.<sup>47</sup> Similar to previous cohort pooling studies, the baseline CVD hazard was allowed to vary by sex and cohort. Importantly, this formulation of the risk prediction equation also facilitates recalibration of the risk equation in other countries where CVD incidence and mortality data among diabetic populations becomes available, e.g. through linked data from diabetes registries.

The primary outcome for our model was fatal-plus-non-fatal CVD, which was defined any one of non-fatal myocardial infarction (MI, ICD10 codes I21-I22) or stroke (ICD10 codes I60-I69), death from ischaemic heart disease (IHD), sudden cardiac death (ICD10 codes I20-I25), or death from stroke. Outcomes were assessed over a 10-year period. Events data defined by ICD10 codes were not in use during the follow-up period for most these cohorts. We defined model outcome events using ICD10 codes and mapped them to clinical endpoints defined by each cohort's event ascertainment and adjudication committee. Heart failure was not included in the outcome events because the adjudication of heart failure varied substantially between study cohorts. We truncated participants with follow-up longer than 15 years, because baseline risk factors might have little value in the prediction of CVD events after a long period of time.

Covariates included in our model were age, current smoking, body mass index (BMI), systolic blood pressure (SBP), total cholesterol (TC), high-density lipoproteins cholesterol (HDL-C), fasting plasma glucose (FPG), treatment with hypertension medications, and treatment with diabetes medications. These covariates were chosen according to their causal effects on CVD and relevance to diabetes history or intensity.<sup>48-58</sup> We considered but did not select other variables such as low-density lipoprotein cholesterol, hemoglobin A1c, duration of diabetes or age at onset of diabetes because they were not measured in the baseline examination cycle of one or more of the cohorts. Baseline participant characteristics included in the risk prediction model are summarized in Table 3.1. We tested and confirmed that the assumption of proportionality of hazards of included covariates was met (Table S3.1 in the Supplementary Material). We included interaction terms between sex and SBP, and age and SBP, due to significant sex differences in CVD risk and gradient of CVD risk associated with increasing SBP.<sup>59,60</sup> We included an interaction term between sex and age because evidence suggests the existence of sex differences in the proportional changes in CVD risk with age.<sup>61-63</sup> We included an interaction term between age and BMI because prospective studies have shown that the CVD hazard ratio of BMI often decreases with age.<sup>52</sup>

**Table 3.1. Information about cohorts used in estimating the coefficients of the risk score.**

<b>Cohort</b>	<b>ARIC</b>	<b>CHS</b>	<b>MESA</b>	<b>MRFIT</b>	<b>WHICTOS</b>	<b>Pooled</b>
<b>Baseline date</b>	1986-1990	1989-1993	2001	1973-1976	1993	1986-2001
<b>Median follow-up years</b>	14.9	8.8	8.4	6.8	8.0	8.3
<b>Number of participants</b>	1242	662	830	647	342	3723
<b>Percent female</b>	58.6	51.9	47.8	0	100	48.6
<b>Age range at baseline</b>	45-66	65-79	45-79	40-57	50-79	40-79
<b>Mean (SD) age at baseline</b>	55.9 (5.7)	71.4 (3.9)	63.9 (8.8)	49.0 (4.7)	63.7 (6.7)	59.9 (9.6)
<b>Number of fatal-plus-non-fatal CVD</b>	423	315	112	206	54	1110
<b>Percent current smoker</b>	22.9	11.2	13.7	53.6	8.2	22.7
<b>Mean (SD) systolic blood pressure (mm Hg)</b>	129.8 (20.6)	140.3 (21.3)	132.7 (21.8)	141.7 (15.9)	134.4 (17.7)	134.8 (20.6)
<b>Mean (SD) body mass index</b>	31.0 (5.8)	28.7 (4.2)	30.7 (5.8)	28.7 (3.5)	32.1 (6.2)	30.2 (5.4)
<b>Mean (SD) total cholesterol (mmol/L)</b>	5.7 (1.2)	5.4 (1.1)	4.9 (1.0)	6.0 (1.1)	5.8 (1.1)	5.5 (1.2)
<b>Mean (SD) high-density lipoproteins cholesterol (mmol/L)</b>	1.2 (0.4)	1.2 (0.3)	1.2 (0.3)	1.0 (0.3)	1.3 (0.3)	1.2 (0.4)
<b>Mean (SD) fasting plasma glucose (mmol/L)</b>	9.7 (3.9)	9.4 (3.4)	8.3 (3.1)	7.2 (2.0)	9.2 (3.2)	8.8 (3.4)
<b>Percent treated with hypertension medications*</b>	51.0	62.4	62.7	58.4	48.2	56.7
<b>Percent treated with diabetes medications</b>	40.8	53.0	81.0	20.9	55.8	49.9

\*Treatment with hypertension medication was a binary covariate. The reference group included individuals who had blood pressure levels not being diagnosed as hypertension, and/or who had hypertension but were not taking hypotension lowering medications at baseline.

‡Abbreviation: ARIC = atherosclerosis risk in communities; CHS = cardiovascular health study; MESA = multi-ethnic study of atherosclerosis; MRFIT = multiple risk factor intervention trial; WHICTOS = women’s health initiative clinical trial and observational study. Detailed information on these cohorts are provided elsewhere. SD = standard deviation.

We fitted several alternative models and evaluated if they have improvement in model fit. We fitted a sex-specific multivariable risk factor algorithm by adding interaction terms between sex and risk factors due to the findings of prospective studies that said sex differences in CVD risk with risk



factors.<sup>64-66</sup> We included a covariate indicating firstly, whether a diabetic patient was under treatment with hypertension medications, and secondly, whether his/her SBP level match the target set by the recent Guideline for the Prevention, Detection, Evaluation and Management of High Blood Pressure in Adults.<sup>67,68</sup> This covariate was defined in two ways with different SBP cutoff points: elevated blood pressure with SBP  $\geq 120$  mm Hg; or hypertension with SBP  $\geq 140$  mm Hg. For example, the covariate with a cutoff point of 140 mm Hg of SBP was defined as a four-level categorical variable: “hypertension (SBP  $\geq 140$  mm Hg) without antihypertensive drugs”, “hypertension (SBP  $\geq 140$  mm Hg) with antihypertensive drugs”, “SBP controlled ( $< 140$  mm Hg) with antihypertensive drugs”, using “SBP controlled ( $< 140$  mm Hg) without antihypertensive drugs” as the reference group. According to the work from the Prospective Study Collaboration that TC/HDL-C ratio is more informative than HDL or TC in predicting cardiovascular risk, we fitted another alternative model by replacing HDL-C and TC as separate covariates to TC/HDL-C ratio as one covariate.<sup>64</sup> We used improvement in model fit, which was defined as a relative discrimination improvement on Harrell’s C statistic of 2% or more, as the criteria for model selection.

We conducted internal validation to examine the discrimination and calibration of the fitted model. Discrimination, which compares the predicted survival time between cases and non-cases and examines whether the cases have shorter survival time, was evaluated using Harrell’s C statistic.<sup>69-71</sup> Calibration, which measures the ability of the risk equation to correctly quantify the absolute 10-year CVD risk, was evaluated by comparing the predicted number of events during 10 years of follow-up with the observed number of events (corrected for censoring with the Kaplan-Meier estimator), stratified by sex and decile of risk.

We also examined out-of-sample predictive performance via cross-validation, as a more rigorous test of model fit. To do so, we first randomly split the data into five folds. We then iteratively withheld each two of the five folds as a testing set and used the other three folds to estimate the risk factor

coefficients; we then recalibrated the model to the testing set by replacing the mean risk factor levels and the baseline hazard and estimated discrimination and calibration statistics for each comparison.

We compared our fitted model to other recent risk models developed in the general population in the US, the atherosclerotic cardiovascular disease (ASCVD) risk equations published in the 2013 ACC/AHA Guidelines,<sup>1</sup> with a validation analysis in our study population. To examine whether CVD risk prediction models developed in the general population underestimate the real CVD risk in patients with T2DM, we compared the actual observed 10-year risk with the predicted 10-year risk of ASCVD by the 2013 ACC/AHA Pooled Cohort Risk Equation in our study population.

All analyses were performed with R (R-3.5.2). The study protocol was approved by the institutional review board at the Harvard T.H. Chan School of Public Health (Boston, MA, USA).

## Results

We included 3,723 participants with T2DM in the estimation of the proportional hazards model (Figure 3.1, Table 3.1). The mean age at baseline was 60 years (standard deviation 9.6) and 48.6% of eligible participants were women. During 15 years of follow-up, 659 (34.5%) men and 451 (24.9%) women had a first CVD event. Without controlling for differences in covariates, women had a lower risk of fatal-plus-non-fatal CVD (the 10-year risk was 21.3% in women versus 38.4% in men,  $p < 0.0001$ ).

The coefficients and hazard ratios (HRs) of the multivariable-adjusted regression for prediction of the 10-year risk of a first CVD event are presented in Table 2. Current smoking (HR: 1.36 [95% CI, 1.17 to 1.58]), increased TC (HR: 1.14 [1.08 to 1.21]), increased FPG with diabetes medications (HR: 1.04 [1.02 to 1.05]) had positive and statistically significant associations with CVD risk ( $p < 0.0001$ ). Increased HDL-C was associated with decreased fatal-plus-non-fatal CVD risk (HR: 0.60 [0.49 to 0.73]). HRs of risk factors that were not statistically significant at the 5% level ranged from 1.00 to 1.17. The magnitude of the association of SBP with an increased CVD risk was stronger in women than in men (HR: 1.07 [1.00 to 1.31]) (Table 3.2).

The Harrell's C statistic for the final model was 70% (68% to 72%) in the total study population, 66% (62% to 69%) in men and 75% (95% CI, 72% to 78%) in women (Table 3.3). The calibration regression slopes were 0.99 (95% CI, 0.98 to 1.00) in the total study population, 0.98 (95% CI, 0.95 to 1.01) in men, and 1.00 (95% CI, 0.99 to 1.02) in women (Figure 3.2). In the cross-validation, the median C statistic was 70%, with a range of 68% to 71% and the median slope of calibration regression was 0.99, with a range of 0.96 to 1.05 (Table 3.3).

**Table 3.2. Coefficients and hazard ratios from the Cox proportional hazard model used to predict the 10-year risk of fatal-plus-non-fatal CVD for patients with T2DM.**

<b>Risk Factors</b>	<b>Coefficient (95% CI)</b>	<b>Hazard Ratio (95% CI)</b>
<b>Age (year)</b>	0.0249 (-0.0788 to 0.1286)	1.03 (0.92 to 1.14)
<b>Age squared</b>	0.0003 (-0.0004 to 0.0011)	1.00 (0.99 to 1.01)
<b>Current smoking</b>	0.3065 (0.1557 to 0.4572)	1.36 (1.17 to 1.58)
<b>Body mass index (per 1 unit)</b>	0.0194 (-0.0637 to 0.1025)	1.02 (0.94 to 1.11)
<b>Systolic blood pressure (per 10 mm Hg)</b>	0.1602 (0.0356 to 0.3561)	1.17 (0.97 to 1.43)
<b>Total cholesterol (per 1mmol/L)</b>	0.1334 (0.0801 to 0.1868)	1.14 (1.08 to 1.21)
<b>High-density lipoproteins cholesterol (per 1mmol/L)</b>	-0.5180 (-0.7221 to -0.3139)	0.60 (0.49 to 0.73)
<b>Fasting plasma glucose (per 1mmol/L)</b>	0.0034 (-0.0183 to 0.0251)	1.00 (0.98 to 1.03)
<b>Treatment with hypertension medications*</b>	0.0686 (-0.0580 to 0.1951)	1.07 (0.94 to 1.22)
<b>Treatment with diabetes medications and fasting plasma glucose (per 1mmol/L)</b>	0.0348 (0.0213 to 0.0483)	1.04 (1.02 to 1.05)
<b>Age and female</b>	0.0028 (-0.0207 to 0.0264)	1.00 (0.98 to 1.03)
<b>Systolic blood pressure (per 10 mm Hg) and female</b>	0.0633 (0.0047 to 0.1220)	1.07 (1.00 to 1.13)
<b>Age and systolic blood pressure (per 10 mm Hg)</b>	-0.0017 (-0.0048 to 0.0013)	1.00 (0.99 to 1.00)
<b>Age and body mass index (per 1 unit)</b>	-0.0004 (-0.0018 to 0.0010)	1.00 (0.99 to 1.00)

\*Treatment with hypertension medication was a binary covariate. The reference group included individuals who had blood pressure being diagnosed as not having hypertension, and/or who had hypertension but were not taking hypotension lowering medication at baseline.

Abbreviation: CVD = cardiovascular disease; FPG = fasting plasma glucose; T2DM = type 2 diabetes mellitus.

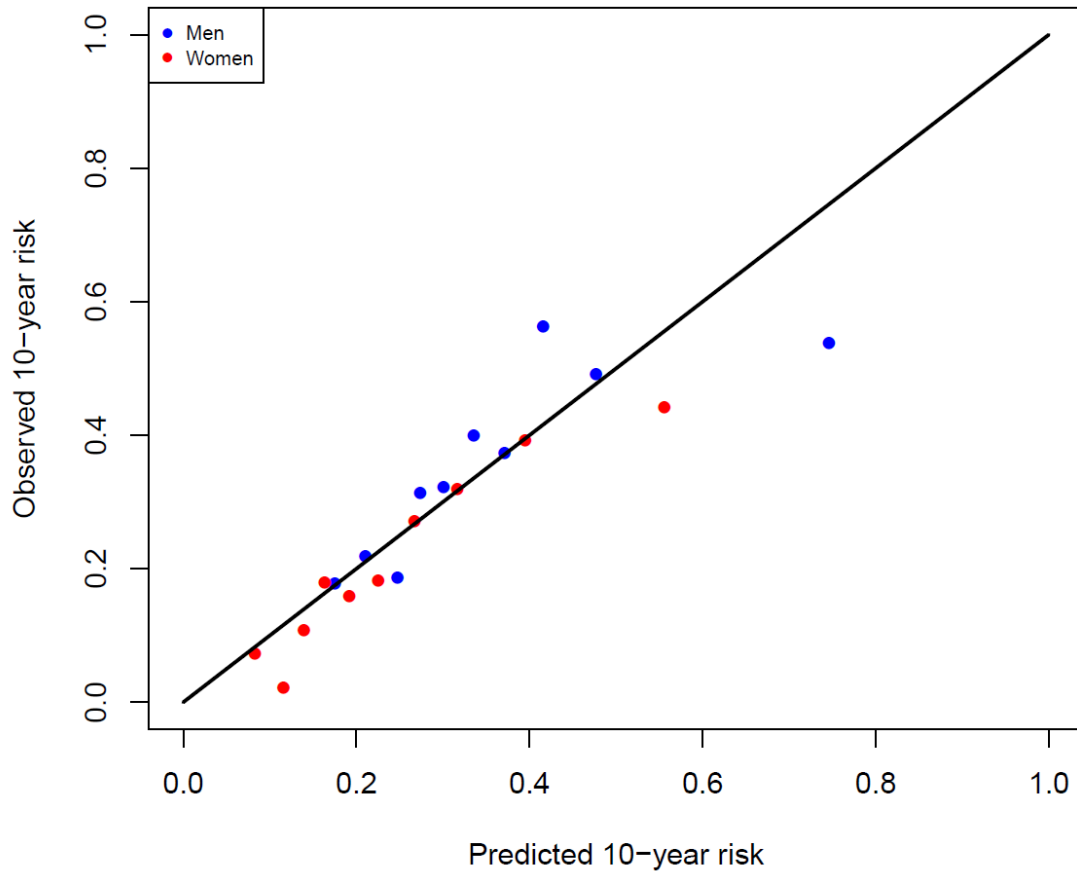


Figure 3.2. Observed and predicted 10-year risk of a cardiovascular disease event in risk score internal validation, by deciles of risk and sex.

Table 3.3. Validation results of the CVD risk model for patients with T2DM.

CVD Risk Model	Discrimination		Calibration regression slope
	Cohort name	Harrell's C statistic (95% CI)	Mean (95% CI)
<b>Internal validation</b>	Total	70 (68 to 72)	0.99 (0.98 to 1.00)
	Men	66 (62 to 69)	0.98 (0.95 to 1.01)
	Women	75 (72 to 78)	1.00 (0.99 to 1.02)
<b>Cross-validation in each of the ten sets of testing data*</b>	1 <sup>st</sup> testing set	69 (66 to 73)	0.99 (0.95 to 1.03)
	2 <sup>nd</sup> testing set	68 (65 to 72)	1.03 (0.99 to 1.07)
	3 <sup>rd</sup> testing set	70 (66 to 73)	1.01 (0.99 to 1.03)
	4 <sup>th</sup> testing set	70 (66 to 73)	0.96 (0.92 to 0.99)
	5 <sup>th</sup> testing set	68 (64 to 72)	1.05 (1.01 to 1.08)
	6 <sup>th</sup> testing set	71 (68 to 74)	0.97 (0.93 to 1.01)
	7 <sup>th</sup> testing set	70 (67 to 73)	0.96 (0.92 to 1.00)
	8 <sup>th</sup> testing set	71 (68 to 74)	1.01 (0.99 to 1.04)
	9 <sup>th</sup> testing set	70 (67 to 73)	0.96 (0.93 to 0.98)
	10 <sup>th</sup> testing set	70 (67 to 74)	0.98 (0.95 to 1.01)

\*The pooled cohorts were randomly split into five folds. We iteratively withheld every two of the five folds as the testing set from the Cox model and used the other three folds to estimate the coefficients; we then validated the obtained model against the withheld testing set data.

Abbreviation: CVD = cardiovascular disease; T2DM = type 2 diabetes mellitus.

### **Alternative model specifications**

Fitting separate models for each sex produced a similar C statistics (69% [67% to 71%]) to the main model but lower calibration regression slopes (0.90 [0.89 to 0.91]). In addition, all the sex and risk factors interaction variables (except for SBP and female) were not statistically significant ( $p > 0.05$ ; Supplementary Material Table S3.2). Model with the covariate replacing the binary hypertension treatment variable with a four-level categorical variable defined as SBP levels ( $\geq 140$  mm Hg or not) and hypertension treatment (with antihypertensive drugs or not) performed worse than the selected final model, with a C statistic of 68% (66% to 70%) and calibration regression slope of 0.96 (0.95 to 0.97) (Supplementary Material Table S3.3). Model with the covariate replacing the binary hypertension treatment variable with a four-level categorical variable defined as SBP levels ( $\geq 120$  mm Hg or not) and hypertension treatment (with antihypertensive drugs or not) had even worse validation, with a C statistic of 66% (64% to 68%) and calibration regression slope of 1.05 (1.04 to 1.06) (Supplementary Material Table S3.4). Model with the covariate replacing TC and HDL-C with the TC/HDL-C ratio as one covariate had good discrimination with a C statistic of 70% (68% to 72%) but worse calibration with a regression slope of 0.94 (0.93 to 0.95) (Supplementary Material Table S3.5).

### **Comparison to other risk equations**

We evaluated the performance of the 2013 ACC/AHA Pooled Cohort Risk Equation<sup>1</sup> in our study population. The predicted 10-year risk of ASCVD by the 2013 ACC/AHA Pooled Cohort Risk Equation underestimated the observed risk for all race-sex groups of participants with T2DM, with the percentage point differences ranging from 4.5% in White men to 11.7% in White women (Supplementary Material Table S3.6). Since ASCVD events accounted for 94.7% of CVD events in our study population, these results demonstrate consistent underestimation when applying the 2013 ACC/AHA Pooled Cohort Risk Equation in populations with T2DM. *Both* the 2013 ACC/AHA Pooled

Cohort Risk Equation and our risk prediction models were developed based on (partial) available prospective cohorts with pre-date baseline recruitment years from the National Heart, Lung, and Blood Institute (NHLBI). The difference between these two models is that the 2013 ACC/AHA Pooled Cohort Risk Equation used both diabetics and non-diabetics versus in our model we restricted the population to diabetes patients. Therefore, the comparison with model performance of their risk equations with ours by applying their risk equation in our study population is reasonable. When applying the two sets of risk models to current US diabetic population, it is necessary to recalibrate the baseline CVD incidence and mortality to current rates.

## **Discussion**

We developed a CVD risk prediction model for patients with T2DM in the United States, using pooled individual-level data from five prospective cohort studies. The risk prediction equation demonstrated good discrimination and calibration in internal validation, with Harrell's C statistics generally 70% or more and calibration regression slopes close to 1.

The discrimination for the risk prediction equation in men was not as good as it was in women. A potential cause of this was the use of data from the MRFIT cohort, which enrolled participants over an earlier period than the other cohorts with male participants. The risk factor characteristics were substantially different among men from the MRFIT and men from other prospective cohorts in our sample (Table 3.1). For example, the participants from MRFIT had a much lower level of FPG with a smaller proportion treated with diabetes medications than men from other cohorts (mean FPG: 7.2 mmol/L versus 9.1 mmol/L,  $p < 0.0001$ ; percent receiving diabetes medications: 20.9% versus 62.6%,  $p < 0.0001$ ). When excluding participants from MRFIT, the discrimination of risk prediction in men was substantially improved to 70% (95% CI, 67% to 73%) (Supplementary Material Table S3.7). We did not select the model excluding MRFIT as the main model, because the number of male participants would be smaller than current, resulting in poor model performance in cross-validation.

We compared our risk prediction equation with the 2013 ACC/AHA Pooled Cohort Risk Equation,<sup>1</sup> the United Kingdom Prospective Diabetes Study Outcomes Model 2 (UKPDS OM2),<sup>36</sup> and the Risk Equations for Complications Of type 2 Diabetes (RECODE).<sup>43</sup> The ACC/AHA risk model underestimated 10-year ASCVD risk for all race-sex groups in participants with T2DM. With calibration regression slopes all close to 1, our model predicted an accurate 10-year fatal-plus-non-fatal CVD risk that was highly comparable to the observed 10-year total CVD risk for participants with T2DM. We were unable to quantitatively compare our risk prediction equation with the UKPDS OM2 and the RECODE in our study population, because many parameters (e.g. age at diagnosis of diabetes, HbA1c, estimated glomerular filtration rate, albuminuria) of UKPDS OM2 equations and RECODE equations were not measured in the baseline examination of most of our study cohorts. UKPDS OM2 is one of the most commonly used risk equations presently for evaluating microvascular and cardiovascular risk among individuals with T2DM in many policy simulation studies.<sup>72-76</sup> UKPDS OM2 was developed in patients with T2DM from both clinical trials and observational data. UKPDS OM2 covariates included not only time invariant factors (e.g. sex, ethnicity, age at diagnosis of diabetes) but also time varying clinical risk factors (e.g. SBP, HDL-C, LDL-C) and time varying comorbidities (e.g. history of stroke, history of amputation). UKPDS OM2 can be used to assess the risk of individual components of cardiovascular events (e.g. IHD, MI, stroke) and death. By considering a wide range of covariates and health outcomes, the UKPDS OM2 is able to provide evidence for personalized treatment strategies during clinical practice and risk distributions by patient subgroups. UKPDS OM2 can also be appropriately applied to secondary prevention for diabetic patients with clinical manifest CVD. Our model does not include covariates of time varying risk factors or comorbidities, due to absence of relevant data or lack of inclusion in the appropriate examination cycle of one or more of the study cohorts. Our risk prediction equation predicts a general risk of a first fatal-or-non-fatal CVD. The application of our risk prediction equation is targeted to the adult diabetic population without clinical signs or symptoms of CVD, who merit evaluation for the primary prevention of CVD. UKPDS OM2 has fixed baseline CVD hazards that cannot be recalibrated when applied to diabetic populations in other countries. When applied to the

US diabetic populations, UKPDS OM2 may not have good discrimination with limited calibration.<sup>43</sup> For example, one previous study evaluated that the mortality predicted by the UKPDS OM was comparable to the observed mortality in US adults with T2DM using the National Health and Nutrition Examination Surveys (NHANES) 1988-1994.<sup>77</sup> However, their results were limited to a small sample size of 156 adults with diabetes who had characteristics similar to the UKPDS cohorts of newly-diagnosed diabetes. The RECODE equations were developed and validated in patients with T2DM in clinical trials, limiting their generalizability to diabetic populations without intensive treatment regimens. The RECODE equations also had inferior interval validation for CVD compared with the ACC/AHA risk equations and our model.<sup>1,43</sup>

Our study pooled multiple high-quality prospective cohorts, demonstrated good performance in both internal validation and cross-validation. The inclusion of multiethnic community-based cohorts (ARIC and MESA) offers a racially and geographically diverse population with T2DM in our model.<sup>78,79</sup> We conducted harmonization of data regarding to the different measures of glucose and risk factors across cohorts. The definition and adjudication of outcome events also varied across cohorts. We used the ICD10 codes to define the model outcome, which were mapped to clinical endpoints ascertained and adjudicated by each cohort. We developed the model stratified by sex and cohort and calibrated the sex- and cohort-specific baseline hazards during validation, allowing the sex-specific CVD incidence and mortality to vary across cohorts. Our study also has some limitations. First, although pooling data from multiple cohorts provides enhanced statistical power compared with using a single cohort, our cohorts were all from United States and territories. Validity of the risk prediction equation needs to be assessed for other diabetic populations, which may have different cardiovascular complications profiles.<sup>80,81</sup> Second, we developed a single general CVD risk score for any initial fatal-or-non-fatal CVD events as opposed to modeling CHD and stroke separately. Stratified by cohort and sex, the number of outcomes events defined as CHD and stroke separately was smaller than the number of outcomes events defined as total CVD, resulting in a poorer model performance. Lastly, the number of participants, duration of follow-up years, and candidate covariates were much smaller in this study



than used for other studies such as UKPDS OM2, which limited our ability to develop an enhanced model including additional risk factors.

In summary, we developed a novel CVD risk prediction model specifically for populations with T2DM in the US. Our risk equation can be utilized to simulate outcomes of intervention policies targeted on patients with T2DM in comparative effectiveness and cost-effectiveness research. As a more accurate risk assessment tool than the ACC/AHA risk equations, our model allows recalibration of baseline CVD incidence and mortality to target diabetic populations. With our risk model, the estimated effectiveness for high-risk diabetic patients would be even larger from intervention strategies such as “benefit-based tailored treatment”, medication adherence and lifestyle interventions, or adding pharmacists to health care management, compared with benefits quantified by other risk equations.<sup>73,74,76</sup> This novel risk equation may be useful for future studies that evaluate the effectiveness and cost-effectiveness of primary prevention options for diabetic patients.

## References

1. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology*. 2014;63(25, Part B):2889-2934. doi:10.1016/j.jacc.2013.11.002
2. Hajifathalian K, Ueda P, Lu Y, et al. A novel risk score to predict cardiovascular disease risk in national populations (GloboRisk): a pooled analysis of prospective cohorts and health examination surveys. *The Lancet Diabetes & Endocrinology*. 2015;3(5):339-355. doi:10.1016/S2213-8587(15)00081-9
3. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ*. 2008;336(7659):1475-1482. doi:10.1136/bmj.39609.449676.25
4. D’Agostino RB, Vasan RS, Pencina MJ, et al. General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study. *Circulation*. 2008;117(6):743-753. doi:10.1161/CIRCULATIONAHA.107.699579

5. Arima H, Yonemoto K, Doi Y, et al. Development and validation of a cardiovascular risk prediction model for Japanese: the Hisayama study. *Hypertension Research*. 2009;32(12):1119-1122. doi:10.1038/hr.2009.161
6. Collaboration TERF. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *The Lancet*. 2010;375(9733):2215-2222. doi:10.1016/S0140-6736(10)60484-9
7. Association AD. Economic Costs of Diabetes in the U.S. in 2017. *Diabetes Care*. 2018;41(5):917-928. doi:10.2337/dci18-0007
8. Nichols GA, Brown JB. The Impact of Cardiovascular Disease on Medical Care Costs in Subjects With and Without Type 2 Diabetes. *Diabetes Care*. 2002;25(3):482-486. doi:10.2337/diacare.25.3.482
9. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. UK Prospective Diabetes Study Group. *BMJ*. 1998;317(7160):703-713.
10. Gaede P, Lund-Andersen H, Parving H-H, Pedersen O. Effect of a Multifactorial Intervention on Mortality in Type 2 Diabetes. *New England Journal of Medicine*. 2008;(358):580:591.
11. Gæde P, Vedel P, Larsen N, Jensen GVH, Parving H-H, Pedersen O. Multifactorial Intervention and Cardiovascular Disease in Patients with Type 2 Diabetes. *New England Journal of Medicine*. 2003;348(5):383-393. doi:10.1056/NEJMoa021778
12. Zoungas S, Chalmers J, Neal B, et al. Follow-up of Blood-Pressure Lowering and Glucose Control in Type 2 Diabetes. *N Engl J Med*. 2014;371(15):1392-1406. doi:10.1056/NEJMoa1407963
13. Arnett Donna K., Blumenthal Roger S., Albert Michelle A., et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. 2019;140(11):e596-e646. doi:10.1161/CIR.0000000000000678
14. Piepoli MF, Hoes AW, Agewall S, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *Eur Heart J*. 2016;37(29):2315-2381. doi:10.1093/eurheartj/ehw106
15. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97(18):1837-1847. doi:10.1161/01.cir.97.18.1837
16. Haffner SM, Lehto S, Rönnemaa T, Pyörälä K, Laakso M. Mortality from Coronary Heart Disease in Subjects with Type 2 Diabetes and in Nondiabetic Subjects with and without Prior Myocardial

Infarction. *New England Journal of Medicine*. 1998;339(4):229-234. doi:10.1056/NEJM199807233390404

17. Evans JMM, Wang J, Morris AD. Comparison of cardiovascular risk between patients with type 2 diabetes and those who had had a myocardial infarction: cross sectional and cohort studies. *BMJ*. 2002;324(7343):939.

18. Eberly LE, Cohen JD, Prineas R, Yang L, Intervention Trial Research group. Impact of incident diabetes and incident nonfatal cardiovascular disease on 18-year mortality: the multiple risk factor intervention trial experience. *Diabetes Care*. 2003;26(3):848-854. doi:10.2337/diacare.26.3.848

19. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Final Report. *Circulation*. 2002;106(25):3143-3143. doi:10.1161/circ.106.25.3143

20. Conroy RM, Pyörälä K, Fitzgerald AP, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J*. 2003;24(11):987-1003. doi:10.1016/S0195-668X(03)00114-3

21. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol*. 1979;110(3):281-290. doi:10.1093/oxfordjournals.aje.a112813

22. Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Münster (PROCAM) study. *Circulation*. 2002;105(3):310-315. doi:10.1161/hc0302.102575

23. Balkau B, Hu G, Qiao Q, et al. Prediction of the risk of cardiovascular mortality using a score that includes glucose as a risk factor. The DECODE Study. *Diabetologia*. 2004;47(12):2118-2128. doi:10.1007/s00125-004-1574-5

24. Brindle P, Beswick A, Fahey T, Ebrahim S. Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review. *Heart*. 2006;92(12):1752-1759. doi:10.1136/hrt.2006.087932

25. Chamnan P, Simmons RK, Sharp SJ, Griffin SJ, Wareham NJ. Cardiovascular risk assessment scores for people with diabetes: a systematic review. *Diabetologia*. 2009;52(10):2001-2014. doi:10.1007/s00125-009-1454-0

26. Stephens JW, Ambler G, Vallance P, Betteridge DJ, Humphries SE, Hurel SJ. Cardiovascular risk and diabetes. Are the methods of risk prediction satisfactory? *European Journal of Cardiovascular Prevention & Rehabilitation*. 2004;11(6):521-528. doi:10.1097/01.hjr.0000136418.47640.bc

27. Coleman RL, Stevens RJ, Retnakaran R, Holman RR. Framingham, SCORE, and DECODE Risk Equations Do Not Provide Reliable Cardiovascular Risk Estimates in Type 2 Diabetes. *Diabetes care*. 2007;30(5):1292-1293. doi:http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.2337/dc06-1358

28. Bulughapitiya U, Siyambalapitiya S, Sithole J, Idris I. Is diabetes a coronary risk equivalent? Systematic review and meta-analysis. *Diabet Med.* 2009;26(2):142-148. doi:10.1111/j.1464-5491.2008.02640.x
29. Echouffo-Tcheugui JB, Kengne AP. On the importance of global cardiovascular risk assessment in people with type 2 diabetes. *Primary Care Diabetes.* 2013;7(2):95-102. doi:10.1016/j.pcd.2013.03.002
30. Kothari V, Stevens RJ, Adler AI, et al. UKPDS 60: risk of stroke in type 2 diabetes estimated by the UK Prospective Diabetes Study risk engine. *Stroke.* 2002;33(7):1776-1781. doi:10.1161/01.str.0000020091.07144.c7
31. Stevens R, Kothari V, Adler AI, Stratton I, Holman R. The UKPDS risk engine: a model for the risk of coronary heart disease in Type II diabetes (UKPDS 56) (vol 101, pg 671, 2001). *Clinical Science.* 2002;102(6):679–679. doi:10.1042/cs1020679
32. Davis WA, Knuiaman MW, Davis TME. An Australian cardiovascular risk equation for type 2 diabetes: the Fremantle Diabetes Study. *Intern Med J.* 2010;40(4):286-292. doi:10.1111/j.1445-5994.2009.01958.x
33. Elley CR, Robinson E, Kenealy T, Bramley D, Drury PL. Derivation and validation of a new cardiovascular risk score for people with type 2 diabetes: the new zealand diabetes cohort study. *Diabetes Care.* 2010;33(6):1347-1352. doi:10.2337/dc09-1444
34. Cederholm J, Eeg-Olofsson K, Eliasson B, Zethelius B, Nilsson PM, Gudbjörnsdottir S. Risk Prediction of Cardiovascular Disease in Type 2 Diabetes: A risk equation from the Swedish National Diabetes Register. *Diabetes Care.* 2008;31(10):2038-2043. doi:10.2337/dc08-0662
35. Donnan PT, Donnelly L, New JP, Morris AD. Derivation and validation of a prediction score for major coronary heart disease events in a U.K. type 2 diabetic population. *Diabetes Care.* 2006;29(6):1231-1236. doi:10.2337/dc05-1911
36. Hayes AJ, Leal J, Gray AM, Holman RR, Clarke PM. UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. *Diabetologia.* 2013;56(9):1925-1933. doi:10.1007/s00125-013-2940-y
37. Chowdhury MZI, Yeasmin F, Rabi DM, Ronksley PE, Turin TC. Predicting the risk of stroke among patients with type 2 diabetes: a systematic review and meta-analysis of C-statistics. *BMJ Open.* 2019;9(8):e025579. doi:10.1136/bmjopen-2018-025579
38. Laxy M, Schöning VM, Kurz C, et al. Performance of the UKPDS Outcomes Model 2 for Predicting Death and Cardiovascular Events in Patients with Type 2 Diabetes Mellitus from a German Population-Based Cohort. *Pharmacoeconomics.* 2019;37(12):1485-1494. doi:10.1007/s40273-019-00822-4

39. Kengne AP, Patel A, Colagiuri S, et al. The Framingham and UK Prospective Diabetes Study (UKPDS) risk equations do not reliably estimate the probability of cardiovascular events in a large ethnically diverse sample of patients with diabetes: the Action in Diabetes and Vascular Disease: Preterax and Diamicron-MR Controlled Evaluation (ADVANCE) Study. *Diabetologia*. 2010;53(5):821-831. doi:10.1007/s00125-010-1681-4
40. Folsom AR, Chambless LE, Duncan BB, Gilbert AC, Pankow JS, Atherosclerosis Risk in Communities Study Investigators. Prediction of coronary heart disease in middle-aged adults with diabetes. *Diabetes Care*. 2003;26(10):2777-2784. doi:10.2337/diacare.26.10.2777
41. Yudkin JS, Chaturvedi N. Developing risk stratification charts for diabetic and nondiabetic subjects. *Diabet Med*. 1999;16(3):219-227. doi:10.1046/j.1464-5491.1999.00026.x
42. Kengne AP, Patel A, Marre M, et al. Contemporary model for cardiovascular risk prediction in people with type 2 diabetes. *Eur J Cardiovasc Prev Rehabil*. 2011;18(3):393-398. doi:10.1177/1741826710394270
43. Basu S, Sussman JB, Berkowitz SA, Hayward RA, Yudkin JS. Development and validation of Risk Equations for Complications Of type 2 Diabetes (RECODE) using individual participant data from randomised trials. *Lancet Diabetes Endocrinol*. 2017;5(10):788-798. doi:10.1016/S2213-8587(17)30221-8
44. Bell SP, Saraf A. Risk stratification in very old adults: How to best gauge risk as the basis of management choices for patients aged over 80. *Prog Cardiovasc Dis*. 2014;57(2):197-203. doi:10.1016/j.pcad.2014.08.001
45. Wang Y, O'Neil A, Jiao Y, et al. Sex differences in the association between diabetes and risk of cardiovascular disease, cancer, and all-cause and cause-specific mortality: a systematic review and meta-analysis of 5,162,654 participants. *BMC Medicine*. 2019;17(1):136. doi:10.1186/s12916-019-1355-0
46. Mosca L, Barrett-Connor E, Wenger NK. Sex/Gender Differences in Cardiovascular Disease Prevention What a Difference a Decade Makes. *Circulation*. 2011;124(19):2145-2154. doi:10.1161/CIRCULATIONAHA.110.968792
47. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B (Methodological)*. 1972;34(2):187-220.
48. Singh GM, Danaei G, Farzadfar F, et al. The Age-Specific Quantitative Effects of Metabolic Risk Factors on Cardiovascular Diseases and Diabetes: A Pooled Analysis. *PLOS ONE*. 2013;8(7):e65174. doi:10.1371/journal.pone.0065174
49. Tzoulaki Ioanna, Elliott Paul, Kontis Vasilis, Ezzati Majid. Worldwide Exposures to Cardiovascular Risk Factors and Associated Health Effects. *Circulation*. 2016;133(23):2314-2333. doi:10.1161/CIRCULATIONAHA.115.008718

50. Doyle JT, Dawber TR, Kannel WB, Heslin AS, Kahn HA. Cigarette Smoking and Coronary Heart Disease. *New England Journal of Medicine*. 1962;266(16):796-801. doi:10.1056/NEJM196204192661602
51. Thun MJ, Apicella LF, Henley SJ. Smoking vs Other Risk Factors as the Cause of Smoking-Attributable Deaths: Confounding in the Courtroom. *JAMA*. 2000;284(6):706-712. doi:10.1001/jama.284.6.706
52. Fliotsos Michael, Zhao Di, Rao Vishal N., et al. Body Mass Index From Early-, Mid-, and Older-Adulthood and Risk of Heart Failure and Atherosclerotic Cardiovascular Disease: MESA. *Journal of the American Heart Association*. 2018;7(22):e009599. doi:10.1161/JAHA.118.009599
53. Nordestgaard BG, Palmer TM, Benn M, et al. The effect of elevated body mass index on ischemic heart disease risk: causal estimates from a Mendelian randomisation approach. *PLoS Med*. 2012;9(5):e1001212. doi:10.1371/journal.pmed.1001212
54. Adams TD, Gress RE, Smith SC, et al. Long-Term Mortality after Gastric Bypass Surgery. *New England Journal of Medicine*. 2007;357(8):753-761. doi:10.1056/NEJMoa066603
55. Holmes MV, Asselbergs FW, Palmer TM, et al. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J*. 2015;36(9):539-550. doi:10.1093/eurheartj/ehv571
56. Amarenco P, Steg PG. The paradox of cholesterol and stroke. *Lancet*. 2007;370(9602):1803-1804. doi:10.1016/S0140-6736(07)61751-6
57. Curb JD, Pressel SL, Cutler JA, et al. Effect of diuretic-based antihypertensive treatment on cardiovascular disease risk in older diabetic patients with isolated systolic hypertension. Systolic Hypertension in the Elderly Program Cooperative Research Group. *JAMA*. 1996;276(23):1886-1892.
58. Alderman Michael H., Cohen Hillel, Madhavan Shantha. Diabetes and Cardiovascular Events in Hypertensive Patients. *Hypertension*. 1999;33(5):1130-1134. doi:10.1161/01.HYP.33.5.1130
59. Franklin SS, Wong ND. Hypertension and Cardiovascular Disease: Contributions of the Framingham Heart Study. *Global Heart*. 2013;8(1):49-57. doi:10.1016/j.gheart.2012.12.004
60. Wei Y-C, George NI, Chang C-W, Hicks KA. Assessing Sex Differences in the Risk of Cardiovascular Disease and Mortality per Increment in Systolic Blood Pressure: A Systematic Review and Meta-Analysis of Follow-Up Studies in the United States. *PLoS One*. 2017;12(1). doi:10.1371/journal.pone.0170218
61. Jousilahti Pekka, Vartiainen Erkki, Tuomilehto Jaakko, Puska Pekka. Sex, Age, Cardiovascular Risk Factors, and Coronary Heart Disease. *Circulation*. 1999;99(9):1165-1172. doi:10.1161/01.CIR.99.9.1165

62. Daly Caroline, Clemens Felicity, Lopez Sendon Jose L., et al. Gender Differences in the Management and Clinical Outcome of Stable Angina. *Circulation*. 2006;113(4):490-498. doi:10.1161/CIRCULATIONAHA.105.561647
63. Bots SH, Peters SAE, Woodward M. Sex differences in coronary heart disease and stroke mortality: a global assessment of the effect of ageing between 1980 and 2010. *BMJ Global Health*. 2017;2(2):e000298. doi:10.1136/bmjgh-2017-000298
64. Prospective Studies Collaboration, Lewington S, Whitlock G, et al. Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. *Lancet*. 2007;370(9602):1829-1839. doi:10.1016/S0140-6736(07)61778-4
65. Peters SAE, Huxley RR, Woodward M. Diabetes as a risk factor for stroke in women compared with men: a systematic review and meta-analysis of 64 cohorts, including 775 385 individuals and 12 539 strokes. *The Lancet*. 2014;383(9933):1973-1980. doi:10.1016/S0140-6736(14)60040-4
66. Huxley RR, Woodward M. Cigarette smoking as a risk factor for coronary heart disease in women compared with men: a systematic review and meta-analysis of prospective cohort studies. *The Lancet*. 2011;378(9799):1297-1305. doi:10.1016/S0140-6736(11)60781-2
67. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension*. 2018;71(6). doi:10.1161/HYP.0000000000000065
68. James PA, Oparil S, Carter BL, et al. 2014 Evidence-Based Guideline for the Management of High Blood Pressure in Adults: Report From the Panel Members Appointed to the Eighth Joint National Committee (JNC 8). *JAMA*. 2014;311(5):507-520. doi:10.1001/jama.2013.284427
69. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4
70. Newson RB. Comparing the Predictive Powers of Survival Models Using Harrell's C or Somers' D. *The Stata Journal*. 2010;10(3):339-358. doi:10.1177/1536867X1001000303
71. Schmid M, Potapov S. A comparison of estimators to evaluate the discriminatory power of time-to-event models. *Statistics in Medicine*. 2012;31(23):2588-2609. doi:10.1002/sim.5464
72. Genuth S. The UKPDS and its global impact. *Diabetic Medicine*. 2008;25(s2):57-62. doi:10.1111/j.1464-5491.2008.02504.x

73. Basu S, Shankar V, Yudkin JS. Comparative effectiveness and cost-effectiveness of target-versus benefit-based treatment of type 2 diabetes in low- and middle-income countries. *Lancet Diabetes Endocrinol.* 2016;4(11):922-932. doi:10.1016/S2213-8587(16)30270-4
74. Yu J, Shah BM, Ip EJ, Chan J. A Markov Model of the Cost-Effectiveness of Pharmacist Care for Diabetes in Prevention of Cardiovascular Diseases: Evidence from Kaiser Permanente Northern California. *JMCP.* 2013;19(2):102-114. doi:10.18553/jmcp.2013.19.2.102
75. Critselis E, Vlahou A, Stel VS, Morton RL. Cost-effectiveness of screening type 2 diabetes patients for chronic kidney disease progression with the CKD273 urinary peptide classifier as compared to urinary albumin excretion. *Nephrol Dial Transplant.* 2018;33(3):441-449. doi:10.1093/ndt/gfx068
76. Nerat T, Locatelli I, Kos M. Type 2 diabetes: cost-effectiveness of medication adherence and lifestyle interventions. *Patient Prefer Adherence.* 2016;10:2039-2049. doi:10.2147/PPA.S114602
77. Song M, Alexander CM, Mavros P, et al. Use of the UKPDS Outcomes Model to predict all-cause mortality in U.S. adults with type 2 diabetes mellitus: comparison of predicted versus observed mortality. *Diabetes Res Clin Pract.* 2011;91(1):121-126. doi:10.1016/j.diabres.2010.10.011
78. unav. THE ATHEROSCLEROSIS RISK IN COMMUNITIES (ARIC) STUDY: DESIGN AND OBJECTIVES. *American journal of epidemiology.* 1989;129(4):687-702. doi:10.1093/oxfordjournals.aje.a115184
79. Burke G, Lima J, Wong ND, Narula J. The Multiethnic Study of Atherosclerosis. *Global Heart.* 2016;11(3):267-268. doi:10.1016/j.ghheart.2016.09.001
80. Chan JCN, Malik V, Jia W, et al. Diabetes in Asia: epidemiology, risk factors, and pathophysiology. *JAMA.* 2009;301(20):2129-2140. doi:10.1001/jama.2009.726
81. Ma RC, Chan JC. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. *Ann N Y Acad Sci.* 2013;1281(1):64-91. doi:10.1111/nyas.12098



## Supplementary Material

**Table S3.1. Test of proportional-hazards assumptions for risk factors in CVD risk model for patients with T2DM.**

Covariates	Slope	Chi2	df	Prob>chi2
Age (year)	-0.02	0.30	1	0.58
Age squared	-0.01	0.09	1	0.76
Current smoking	0.03	1.27	1	0.26
Body mass index (per 1 unit)	-0.04	1.68	1	0.20
Systolic blood pressure (per 10 mm Hg)	-0.01	0.07	1	0.78
Total cholesterol (per 1 mmol/L)	-0.04	1.98	1	0.16
High-density lipoproteins cholesterol (per 1mmol/L)	0.01	0.12	1	0.73
Fasting plasma glucose (per 1 mmol/L)	0.01	0.05	1	0.82
Treatment with hypertension medications*	0.04	1.93	1	0.17
Treatment with diabetes medications and fasting plasma glucose (per 1 mmol/L)	-0.03	1.08	1	0.30
Age and female	-0.04	1.41	1	0.58
Systolic blood pressure (per 10 mm Hg) and female	-0.02	0.30	1	0.24
Age and systolic blood pressure (per 10 mm Hg)	0.02	0.00	1	0.96
Age and body mass index (per 1 unit)	0.04	1.98	1	0.16
<b>Global test</b>		10.87	15	0.91

\* Treatment with hypertension medication was a binary covariate. The reference group included individuals who had blood pressure being diagnosed as not having hypertension, and/or who had hypertension but were not taking hypotension lowering medications at baseline.

Abbreviation: T2DM = type 2 diabetes mellitus.

**Table S3.2. Coefficients and hazard ratios from the Cox proportional hazard model used to predict the 10-year risk of CVD for our study populations, with all sex and risk factors interactions.**

	<b>Harrell's C statistic (95% CI) 69% (67% to 71%)</b>	<b>Calibration slope (95% CI) 0.90 (0.89 to 0.91)</b>
<b>Risk Factors</b>	<b>Coefficient (95% CI)</b>	<b>Hazard Ratio (95% CI)</b>
<b>Age (year)</b>	0.0103 (-0.0939 to 0.1146)	1.01 (0.91 to 1.12)
<b>Age squared</b>	0.0004 (-0.0004 to 0.0011)	1.00 (0.99 to 1.00)
<b>Current smoking</b>	0.2474 (0.0609 to 0.4339)	1.28 (1.06 to 1.54)
<b>Body mass index (per 1 unit)</b>	0.0156 (-0.0672 to 0.0985)	1.02 (0.94 to 1.10)
<b>Systolic blood pressure (per 10 mm Hg)</b>	0.1594 (-0.0372 to 0.3559)	1.17 (0.96 to 1.43)
<b>Total cholesterol (per 1 mmol/L)</b>	0.1103 (0.03751 to 0.1831)	1.12 (1.04 to 1.20)
<b>High-density lipoproteins cholesterol (per 1 mmol/L)</b>	-0.6269 (-0.9167 to -0.3372)	0.53 (0.40 to 0.71)
<b>Fasting plasma glucose (per 1 mmol/L)</b>	0.0127 (-0.0177 to 0.0430)	1.01 (0.98 to 1.04)
<b>Treatment with hypertension medications*</b>	0.0155 (-0.1482 to 0.1793)	1.02 (0.86 to 1.20)
<b>Treatment with diabetes medications and fasting plasma glucose (per 1mmol/L)</b>	0.0356 (0.0173 to 0.0540)	1.04 (1.02 to 1.06)
<b>Age and female</b>	-0.0014 (-0.0254 to 0.0226)	1.00 (0.97 to 1.02)
<b>Current smoking and female</b>	0.1694 (-0.1459 to 0.4846)	1.18 (0.86 to 1.62)
<b>Body mass index (per 1 unit) and female</b>	-0.0192 (-0.0456 to 0.0073)	0.98 (0.96 to 1.01)
<b>Systolic blood pressure (per 10 mm Hg) and female</b>	0.0628 (0.0026 to 0.1230)	1.06 (1.00 to 1.13)
<b>Total cholesterol (per 1 mmol/L) and female</b>	0.0522 (-0.0548 to 0.1592)	1.05 (0.95 to 1.17)
<b>High-density lipoprotein cholesterol (per 1 mmol/L) and female</b>	0.2566 (-0.1528 to 0.6660)	1.29 (0.86 to 1.95)
<b>Fasting plasma glucose (per 1 mmol/L) and female</b>	-0.0176 (-0.0613 to 0.0261)	0.98 (0.94 to 1.03)
<b>Treatment with hypertension medications* and female</b>	0.1327 (-0.1254 to 0.3909)	1.14 (0.88 to 1.48)
<b>Fasting plasma glucose (per 1 mmol/L) and treatment with diabetes medications and female</b>	-0.0002 (-0.0273 to 0.0270)	1.00 (0.97 to 1.03)
<b>Age and systolic blood pressure (per 10 mm Hg)</b>	-0.0017 (-0.0048 to 0.0013)	1.00 (0.99 to 1.00)
<b>Age and body mass index (per 1 unit)</b>	-0.0001 (-0.0015 to 0.0012)	1.00 (0.99 to 1.00)

\*Treatment with hypertension medication was a binary covariate. The reference group included individuals who had blood pressure being diagnosed as not having hypertension, and/or who had hypertension but were not taking hypotension lowering medication at baseline.

Abbreviation: CVD = cardiovascular disease.

**Table S3.3. Coefficients and hazard ratios from the Cox proportional hazard model used to predict the 10-year risk of CVD for our study populations, when hypertension and treatment with hypertension medications were included as a four categorical variable.**

	<b>Harrell's C statistic (95% CI)</b> <b>68% (66% to 70%)</b>	<b>Calibration slope (95% CI)</b> <b>0.96 (0.95 to 0.97)</b>
<b>Risk Factors</b>	<b>Coefficient (95% CI)</b>	<b>Hazard Ratio (95% CI)</b>
<b>Age (year)</b>	0.0250 (-0.0791 to 0.1290)	1.03 (0.92 to 1.14)
<b>Age squared</b>	0.0003 (-0.0004 to 0.0011)	1.00 (0.99 to 1.00)
<b>Current smoking</b>	0.3066 (0.1559 to 0.4574)	1.36 (1.17 to 1.58)
<b>Body mass index (per 1 unit)</b>	0.0168 (-0.0665 to 0.1002)	1.02 (0.94 to 1.11)
<b>Systolic blood pressure (per 10 mm Hg)</b>	0.1822 (-0.0207 to 0.3851)	1.20 (0.98 to 1.47)
<b>Total cholesterol (per 1mmol/L)</b>	0.1335 (0.0799 to 0.1867)	1.14 (1.08 to 1.21)
<b>High-density lipoproteins cholesterol (per 1mmol/L)</b>	-0.5185 (-0.7226 to -0.3143)	0.60 (0.49 to 0.73)
<b>Fasting plasma glucose (per 1mmol/L)</b>	0.0031 (-0.0186 to 0.0248)	1.00 (0.98 to 1.03)
<b>Systolic blood pressure controlled with treatment ‡</b>	-0.0225 (-0.2582 to 0.2132)	0.98 (0.77 to 1.24)
<b>Hypertension without treatment*</b>	0.0844 (-0.0750 to 0.2437)	1.09 (0.93 to 1.28)
<b>Hypertension with treatment †</b>	-0.0653 (-0.3176 to 0.1869)	0.94 (0.73 to 1.21)
<b>Treatment with diabetes medications and fasting plasma glucose (per 1mmol/L)</b>	0.0350 (0.0214 to 0.0485)	1.04 (1.02 to 1.05)
<b>Age and female</b>	0.0028 (-0.0208 to 0.0263)	1.00 (0.98 to 1.03)
<b>Systolic blood pressure (per 10 mm Hg) and female</b>	0.0623 (0.0038 to 0.1207)	1.06 (1.00 to 1.13)
<b>Age and systolic blood pressure (per 10 mm Hg)</b>	-0.0018 (-0.0049 to 0.0013)	1.00 (0.99 to 1.00)
<b>Age and body mass index (per 1 unit)</b>	-0.0003 (-0.0017 to 0.0010)	1.00 (0.99 to 1.00)

\* Hypertension without treatment were diabetic individuals who had SBP  $\geq$ 140 mm Hg and were not taking any hypertension medications at baseline. The reference group was individuals who had SBP <140 mm Hg and were not taking any hypertension medications at baseline.

† Hypertension with treatment were diabetic individuals who had SBP  $\geq$ 140 mm Hg and were taking any hypertension medications at baseline. The reference group was individuals who had SBP <140 mm Hg and were not taking any hypertension medications at baseline.

‡ Systolic blood pressure controlled with treatment were diabetic individuals who had SBP <140 mm Hg and were taking any hypertension medications at baseline. The reference group was individuals who had SBP <140 mm Hg and were not taking any hypertension medications at baseline.

Abbreviation: CVD = cardiovascular disease.

**Table S3.4. Coefficients and hazard ratios from the Cox proportional hazard model used to predict the 10-year risk of CVD for our study populations, when elevated systolic blood pressure and treatment with hypertension medications were included as a four categorical variable.**

	<b>Harrell's C statistic (95% CI) 66% (64% to 68%)</b>	<b>Calibration slope (95% CI) 1.05 (1.04 to 1.06)</b>
<b>Risk Factors</b>	<b>Coefficient (95% CI)</b>	<b>Hazard Ratio (95% CI)</b>
<b>Age (year)</b>	0.0190 (-0.0852 to 0.1232)	1.02 (0.92 to 1.13)
<b>Age squared</b>	0.0003 (-0.0004 to 0.0011)	1.00 (0.99 to 1.01)
<b>Current smoking</b>	0.3074 (0.1566 to 0.4583)	1.36 (1.17 to 1.58)
<b>Body mass index (per 1 unit)</b>	0.0202 (-0.0630 to 0.1035)	1.02 (0.94 to 1.11)
<b>Systolic blood pressure (per 10 mm Hg)</b>	0.1077 (-0.0985 to 0.3139)	1.11 (0.91 to 1.37)
<b>Total cholesterol (per 1mmol/L)</b>	0.1335 (0.0802 to 0.1868)	1.14 (1.08 to 1.21)
<b>High-density lipoproteins cholesterol (per 1mmol/L)</b>	-0.5182 (-0.7220 to -0.3143)	0.60 (0.49 to 0.73)
<b>Fasting plasma glucose (per 1mmol/L)</b>	0.0030 (-0.0187 to 0.0248)	1.00 (0.98 to 1.03)
<b>Systolic blood pressure controlled with treatment ‡</b>	0.3054 (0.0530 to 0.5579)	1.36 (1.05 to 1.75)
<b>Elevated systolic blood pressure without treatment*</b>	0.1131 (-0.1725 to 0.3986)	1.12 (0.84 to 1.49)
<b>Elevated systolic blood pressure with treatment †</b>	0.2472 (0.0002 to 0.4941)	1.28 (1.00 to 1.64)
<b>Treatment with diabetes medications and fasting plasma glucose (per 1mmol/L)</b>	0.0350 (0.0215 to 0.0486)	1.04 (1.02 to 1.05)
<b>Age and female</b>	0.0017 (-0.0219 to 0.0252)	1.00 (0.98 to 1.03)
<b>Systolic blood pressure (per 10 mm Hg) and female</b>	0.0674 (0.0069 to 0.1279)	1.07 (1.01 to 1.14)
<b>Age and systolic blood pressure (per 10 mm Hg)</b>	-0.0013 (-0.0045 to 0.0018)	1.00 (0.99 to 1.00)
<b>Age and body mass index (per 1 unit)</b>	-0.0004 (-0.0018 to 0.0010)	1.00 (0.99 to 1.00)

\*Elevated systolic blood pressure without treatment were diabetic individuals who had SBP  $\geq$ 120 mm Hg and were not taking any hypertension medications at baseline. The reference group was individuals who had SBP <120 mm Hg and were not taking any hypertension medications at baseline.

† Elevated systolic blood pressure with treatment were diabetic individuals who had SBP  $\geq$ 120 mm Hg and were taking any hypertension medications at baseline. The reference group was individuals who had SBP <120 mm Hg and were not taking any hypertension medications at baseline.

‡ Systolic blood pressure controlled with treatment were diabetic individuals who had SBP <120 mm Hg and were taking any hypertension medications at baseline. The reference group was individuals who had SBP <120 mm Hg and were not taking any hypertension medications at baseline.

Abbreviation: CVD = cardiovascular disease.

**Table S3.5. Coefficients and hazard ratios from the Cox proportional hazard model used to predict the 10-year risk of CVD for our study populations, with TC/HDL-C ratio covariate.**

	<b>Harrell's C statistic (95% CI) 70% (68% to 72%)</b>	<b>Calibration slope (95% CI) 0.94 (0.93 to 0.95)</b>
<b>Risk Factors</b>	<b>Coefficient (95% CI)</b>	<b>Hazard Ratio (95% CI)</b>
<b>Age (year)</b>	0.0347 (-0.0689 to 0.1383)	1.04 (0.93 to 1.15)
<b>Age squared</b>	0.0003 (-0.0005 to 0.0010)	1.00 (0.99 to 1.00)
<b>Current smoking</b>	0.2912 (0.1409 to 0.4416)	1.34 (1.16 to 1.56)
<b>Body mass index (per 1 unit)</b>	0.0179 (-0.0648 to 0.1007)	1.02 (0.94 to 1.11)
<b>Systolic blood pressure (per 10 mm Hg)</b>	0.1768 (-0.0196 to 0.3731)	1.19 (0.98 to 1.45)
<b>Total cholesterol/high-density lipoproteins cholesterol</b>	0.1157 (0.0939 to 0.1475)	1.12 (1.09 to 1.16)
<b>Fasting plasma glucose (per 1 mmol/L)</b>	0.0015 (-0.0202 to 0.0232)	1.00 (0.98 to 1.02)
<b>Treatment with hypertension medications*</b>	0.0630 (-0.0636 to 0.1895)	1.07 (0.84 to 1.21)
<b>Treatment with diabetes medications and fasting plasma glucose (per 1mmol/L)</b>	0.0352 (0.0217 to 0.0486)	1.04 (1.02 to 1.05)
<b>Age and female</b>	0.0032 (-0.0204 to 0.0268)	1.00 (0.98 to 1.03)
<b>Systolic blood pressure (per 10 mm Hg) and female</b>	0.0634 (0.0049 to 0.1220)	1.07 (1.00 to 1.13)
<b>Age and systolic blood pressure (per 10 mm Hg)</b>	-0.0020 (-0.0050 to 0.0011)	1.00 (0.99 to 1.00)
<b>Age and body mass index (per 1 unit)</b>	-0.0003 (-0.0017 to 0.0010)	1.00 (0.99 to 1.00)

\*Treatment with hypertension medication was a binary covariate. The reference group included individuals who had blood pressure being diagnosed as not having hypertension, and/or who had hypertension but were not taking hypotension lowering medication at baseline.

Abbreviation: CVD = cardiovascular disease.

**Table S3.6. Comparison of 10-year CVD observed risk with 10-year ASCVD predicted risk by the 2013 ACC/AHA Pooled Cohort Risk Equation<sup>1</sup> in our study population from the five prospective cohorts, by race and sex.**

<b>Diabetic Populations</b>	<b>10-Year risk of ASCVD predicted by ACC/AHA model (%)</b>	<b>10-Year risk of CVD<sup>†</sup> observed</b>	<b>Underestimates (percentage point difference)</b>
<b>Women African American</b>	13.92	18.97	5.05
<b>Women White</b>	11.58	23.23	11.65
<b>Men African American</b>	23.10	30.93	7.83
<b>Men White</b>	33.47	37.99	4.52
<b>Total</b>	25.11	29.13	4.02

Abbreviations: CVD = cardiovascular disease; ASCVD = atherosclerotic cardiovascular disease; ACC/AHA model = American College of Cardiology/American Heart Association.

<sup>†</sup>94.7% of CVD events were ASCVD event in our study population.

**Table S3.7. Validation results of the 10-year CVD risk model for participants with T2DM in the pooled four prospective cohorts (MRFIT excluded).**

<b>CVD Risk Model (MRFIT excluded)</b>	<b>Study Population</b>	<b>Discrimination</b>	<b>Calibration regression slope</b>
		Harrell's C statistic (%) (95% CI)	Mean (95% CI)
<b>Internal validation</b>	Total	73 (71-75)	1.01 (1.00-1.03)
	Men	70 (67-73)	1.01 (0.98-1.04)
	Women	75 (72-77)	1.00 (0.98-1.02)

Abbreviation: CVD = cardiovascular disease; MRFIT: multiple risk factor intervention trial; T2DM = type 2 diabetes mellitus.

**Chapter 4. Disparities in health and economic outcomes associated with *N. gonorrhoeae* infection in the United States: costs and quality-adjusted life-years lost in 2015**

Yunfei Li<sup>1</sup>, Minttu M Rönn<sup>1</sup>, Ashleigh R Tuite<sup>1</sup>, Harrell W Chesson<sup>2</sup>, Thomas L Gift<sup>2</sup>, Thomas A Trikalinos<sup>3</sup>, Christian Testa<sup>1</sup>, Meghan Bellerose<sup>1</sup>, Katherine Hsu<sup>4</sup>, Andrés A Berruti<sup>2</sup>, Yelena Maljuta<sup>1</sup>, Nicolas A Menzies<sup>1</sup>, Joshua A Salomon<sup>5</sup>

1. Department of Global Health and Population, Harvard T. H. Chan School of Public Health
2. Division of STD Prevention, Centers for Disease Control and Prevention
3. Department of Health Services Policy & Practice, Brown University
4. Division of STD Prevention & HIV/AIDS Surveillance, Massachusetts Department of Public Health
5. Center for Health Policy / Center for Primary Care and Outcomes Research, Stanford University

## **Abstract**

### **Background**

Disparities in the health and economic burden of gonorrhoea have not been systematically quantified. We estimated population-level health losses and costs associated with gonococcal infection and sequelae in the United States.

### **Methods**

We used probability tree models to capture gonorrhoea sequelae and to estimate attributable disease burden in terms of the discounted lifetime costs and quality-adjusted life-years (QALYs) lost due to incident infections acquired during 2015. Numbers of infections in 2015 were estimated using a gonorrhoea transmission model. We evaluated population-level disease burden, disaggregated by sex, age, race/ethnicity, and for men who have sex with men (MSM). We conducted a multivariate sensitivity analysis for key parameters.

### **Findings**

Discounted lifetime QALYs lost per incident gonococcal infection were estimated as 0.093 (95% uncertainty interval [UI] 0.022-0.22) for women, 0.0020 (95% UI 0.0015-0.0024) for heterosexual men, and 0.0017 (95% UI 0.0013-0.0022) for MSM. Discounted lifetime costs per incident infection were USD 254 (95% UI 109-460), 306 (95% UI 157-479), and 281 (95% UI 145-443), respectively. Total discounted lifetime QALYs lost due to infections acquired during 2015 were 53,293 (95% UI 12,326-125,366) for women, 623 (95% UI 431-876) for heterosexual men, and 1,286 (632-2,074) for MSM. Total discounted lifetime costs were USD 145 million (63 to 266 million), 97 million (46-168 million), and 206 million (82-410 million), respectively. The highest absolute burden of both QALYs and costs was observed in Non-Hispanic Black women, and highest relative burden was identified in MSM and American Indian/Alaska Native women.



**Interpretation**

Gonorrhea causes significant health losses and costs in the United States. The results can inform planning and prioritizing prevention policies.

**Funding**

Centers for Disease Control and Prevention, Charles A. King Trust

## Background

In the United States, there were 583,405 gonorrhea diagnoses reported in 2018, making *Neisseria gonorrhoeae* the second most common notifiable infection.<sup>1</sup> Along with chlamydia, gonorrhea is an important cause of pelvic inflammatory disease (PID), chronic pelvic pain (CPP), ectopic pregnancy (EP), and tubal infertility (TI) in women and epididymitis in men.<sup>2-4</sup> Gonococcal infection may also increase the risk of human immunodeficiency virus (HIV) acquisition and transmission.<sup>5,6</sup>

In the United States, reported gonorrhea rates have increased since 2009,<sup>1,7-9</sup> with a pronounced increase among men who have sex with men (MSM).<sup>1,10-12</sup> Although disparities in gonorrhea and other STIs are often measured using diagnosis rates,<sup>2,7,8,10</sup> diagnosed infections represent only a subset of all infections. Moreover, a comprehensive assessment of gonorrhea-associated burden requires evaluation of downstream health consequences secondary to infection.

To date, the population-level disparities in costs and sequelae associated with gonococcal infection have not been systematically quantified, reflecting a knowledge gap in our understanding of gonorrhea burden in the United States. This study estimated lifetime quality-adjusted life-years (QALYs) lost and costs associated with gonococcal infections acquired in 2015 by age, sex, race/ethnicity, and for MSM in the United States. We developed probability tree models to quantify long-term health and economic consequences associated with gonorrhea. We combined these estimates with estimates of gonococcal infection incidence derived from a gonorrhea transmission model.<sup>13</sup> The aim of this study was to present a comprehensive analysis of gonorrhea burden and disparities in the context of current prevention efforts.

## Methods

### Analytic overview

We used probability tree models to represent clinical outcomes following gonococcal infection among women, men who have sex with women (MSW), and MSM. Next, we estimated costs and health losses associated with the clinical endpoints of each distinct path along the probability tree and aggregated these into the total numbers of expected discounted lifetime QALYs lost and expected total discounted lifetime costs per each incident gonococcal infection. We combined these estimates with the total numbers of estimated incident infections acquired in 2015 by age, sex and race/ethnicity to estimate overall population-level health and economic disease burden and disparities. Model inputs and parameter values were derived from a variety of sources, including: (1) estimates of gonorrhoea incidence, and probabilities of symptomatic infection, testing and treatment from a gonorrhoea transmission modeling study<sup>13</sup>; and (2) probabilities of gonorrhoea sequelae, durations, disutilities and costs from synthesis of published literature. Analyses were undertaken in R (R-3.5.2).

### Model structure and probabilities of key clinical outcomes

We developed probability trees to model clinical outcomes following gonococcal infection, adapted from prior decision analysis studies.<sup>14-16</sup> Separate trees were specified for women, MSW and MSM (Figure 4.1, Figure 4.2).

For women, gonococcal infections were categorized as symptomatic or asymptomatic with different durations of infections, leading to distinct probabilities of developing PID and subsequent complications (Figure 4.1).<sup>14</sup> Although treatment for gonorrhoea is not explicitly distinguished as a separate branch in Figure 4.1, the estimated durations of infections accounted for the fraction of cases treated. The possible complications following PID were CPP, EP, and TI (Figure 4.1B). Although we included disseminated gonococcal infection (DGI) as a sequela for men (see below), we excluded this sequela for women in the interest of parsimony, given that it occurs with lower probability than the

other included sequelae. The probability of symptomatic infection and durations of either symptomatic or asymptomatic infection were derived from a prior transmission modelling study. Duration of asymptomatic infection differed across age groups due to differences in screening rates.<sup>13</sup> For the probability of developing PID, we synthesized data from previously published studies<sup>17–22</sup> on rates of developing PID secondary to chlamydial infection,<sup>23</sup> and translated the rates into associated probabilities of PID given durations of symptomatic or asymptomatic gonococcal infection. The conditional probabilities of CPP, EP, and TI given PID were derived by pooling estimates from previous longitudinal studies on sequelae among women infected with chlamydia.<sup>24–30</sup>

For MSW, infections were categorized as symptomatic or asymptomatic urethral gonococcal infections (Figure 4.2A), followed by possible sequelae of epididymitis and DGI (Figure 4.2C). We assumed that all symptomatic infections are diagnosed and successfully treated, and that treatment prevents any further complications in men; this simplifying assumption implies that epididymitis and DGI occur only among untreated asymptomatic cases. There is sparse evidence on the relationship between duration of infection and sequelae development for men, and epididymitis and DGI were modeled as probabilities that were independent of duration. We assumed that urethritis was present in all symptomatic infections for the duration of the infection, and that urethritis was not present in asymptomatic infections.<sup>31</sup> The probabilities of symptomatic and asymptomatic infections were obtained from the gonorrhea transmission model.<sup>13</sup> We estimated the probability of epididymitis among untreated asymptomatic infections based on pooled estimates from longitudinal studies on sequelae among men infected with chlamydia.<sup>32,33</sup> The probability of DGI was derived from a previous Institute of Medicine (IOM) study.<sup>16</sup>

For MSM, the model treated urethral infections as described for MSW (Figure 4.2B and Figure 4.2C). In addition, rectal and pharyngeal infections, which are typically asymptomatic, were assumed to incur no direct health utility losses, consistent with assumptions made in a study on rectal chlamydia.<sup>34</sup> This was a simplifying assumption that also implied that only urethral infections lead to epididymitis and

DGI. The probability of symptomatic urethral infection for both MSW and MSM was obtained from the gonorrhea transmission model.<sup>13</sup> We estimated the probability of asymptomatic urethral infection among MSM by assuming that the odds of symptoms given urethral infection were the same for MSM as for MSW. Further details can be found in Supplementary Material.

Diverging from the approach taken in prior related studies using probability trees, we have explicitly modeled both independent sequelae and combinations of sequelae following PID for women and following urethral infections for MSW and MSM, whereas prior studies have assumed that all complications were independent and additive.<sup>13,15,16</sup> Further details on the values, ranges, distributions, and sources of probabilities for key clinical outcomes can be found in Table 4.1 and in Supplementary Material. In addition to the pathways of clinical outcomes shown in Figure 4.1 and Figure 4.2, we further stratified some health states by treatment status for the purpose of computing costs and utility losses. For gonococcal infections, we distinguished treated and untreated infections; for sequelae following infection, we assumed that all cases are treated and distinguished inpatient from outpatient treatment for PID, EP, epididymitis and DGI. Yearly probabilities of testing and treatment for gonococcal infections for women, MSW and MSM were obtained from the gonorrhea transmission model,<sup>13</sup> and were stratified by age and race/ethnicity (Table S4.5-(a) and Table S4.5-(b) in Supplementary Material). We estimated the probabilities of inpatient (vs. outpatient) treatment for PID, EP, epididymitis and DGI as 0.10, 0.15, 0.0054 and 0.29, respectively.<sup>16,35-37</sup>

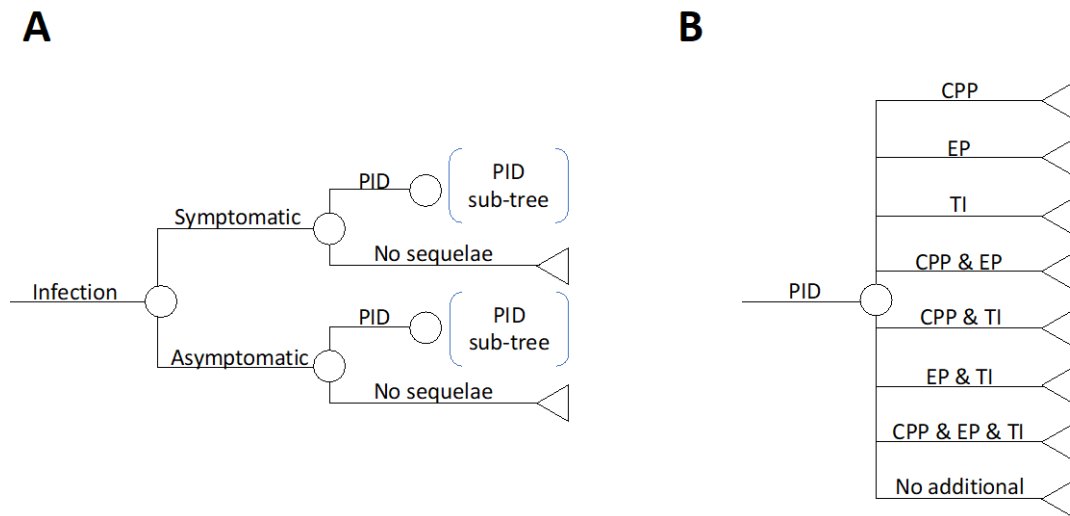


Figure 4.1. Probability tree for sequelae following gonococcal infection among women. Abbreviations: PID = pelvic inflammatory disease, CPP = chronic pelvic pain, EP = ectopic pregnancy, TI = tubal infertility.

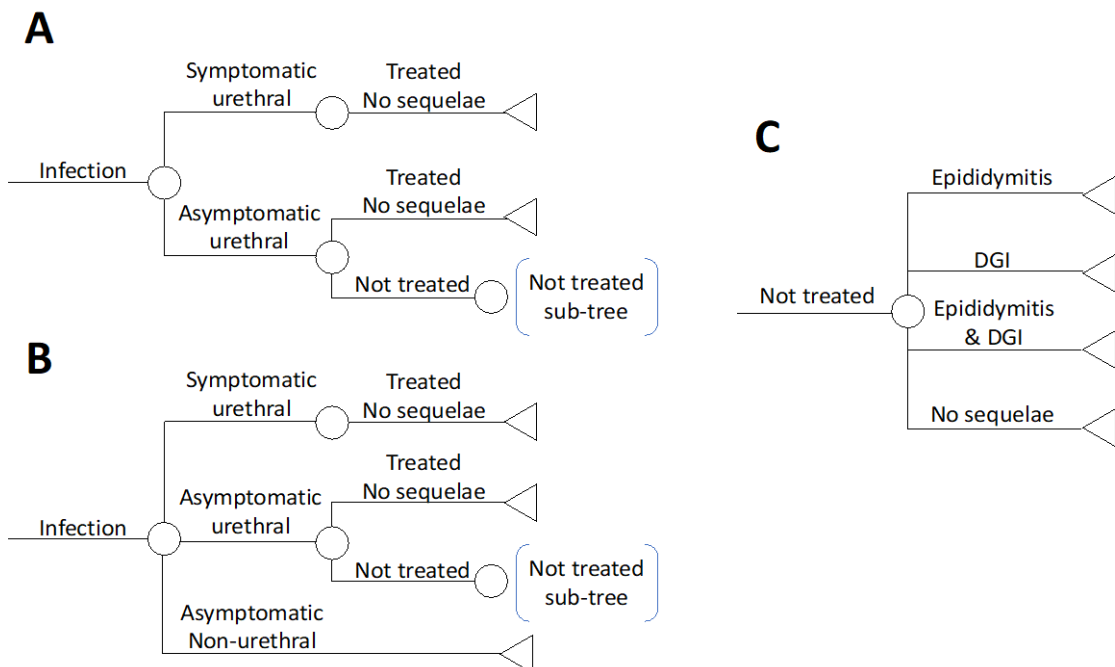


Figure 4.2. Probability tree for sequelae following gonococcal infection among MSW (Panel A) and MSM (Panel B), with complications of untreated urethral infections shown in Panel C. Abbreviations: DGI = disseminated gonococcal infection.

## **Lifetime QALYs lost and costs per incident infection**

Utilities and durations for clinical outcomes related to gonorrhoea and sequelae were derived from the prior IOM study, which used an expert panel to estimate durations of sequelae and health-state weights measured via the Health Utilities Index (HUI).<sup>16</sup> The utility weights for gonorrhoea and sequelae were multiplied by background utilities reflecting chronic comorbidities by age, based on nationally representative EQ-5D index scores.<sup>38</sup> CPP and TI were treated as chronic sequelae, and durations were estimated using a life table approach. Age-specific mortality rates for women were derived from National Center for Health Statistics (NCHS) data.<sup>39</sup> Costs were estimated from the healthcare perspective and included all direct medical costs regardless of payer. All costs were inflated to 2018 U.S. dollars using the medical care component of the consumer price index.<sup>40</sup> Costs and QALYs incurred in years after the incidence of infection were discounted to the year of infection, using a discount rate of 3% per year.<sup>41</sup> Further details can be found in Table 4.1 and Supplementary Material.

Based on the probability tree models in Figure 4.1 and Figure 4.2, discounted lifetime QALYs lost per incident infection were estimated by summing the expected losses for each unique sequela or sequelae combination (the product of the probability, duration, and disutility). Lifetime costs per incident infection were estimated analogously for diagnosis and treatment of symptomatic cases and asymptomatic cases. For symptomatic infection, all cases were assumed to be diagnosed and treated. For asymptomatic infection, the proportion of diagnosis and treatment can be found in Table 4.1 and Supplementary Material Table S4.5. Details on the discounted costs of diagnosis and treatment for gonorrhoea and sequelae can be found in Table 4.1.

## **Disease burden and disparities at population level**

Estimates for the incidence of gonococcal infection were obtained from a transmission modelling study.<sup>13</sup> The study used a model that was calibrated using a Bayesian approach to synthesize information from several large national datasets over the period 2000-2015, including reported gonorrhoea diagnoses, prevalence estimates from the National Health and Nutrition Examination

Survey, and other sources.<sup>13</sup> This national-level gonorrhea meta-population model described a population aged 15-24 and 25-39 years, with the heterosexual population of men and women stratified into the following race/ethnicity groups: Non-Hispanic Black, Hispanic, and “White and Other.” The population of MSM was stratified by age, but not by race/ethnicity due to limited data available at the national level.

In our model, the population was stratified by age and sex, and men were further stratified into MSW and MSM. Among the heterosexual population (women and MSW), the population in the present study was additionally divided into five race/ethnicity categories following the classification of the NCHS bridged-race categories: Non-Hispanic Black (NHB), Hispanic, American Indian or Alaska Native (AI/AN), Asian, Native Hawaiian or Other Pacific Islander (A/NH/OPI), and Non-Hispanic White (NHW).<sup>42,43</sup> To compute incidence of gonococcal infection among heterosexual AI/AN, A/NH/OPI and NHW as sub-populations of incidence estimated among “White and Other,” we assumed that incidence rates for the three subgroups followed the same relative rates as those observed in reported case rates in these subgroups, which were obtained from the NCHHSTP AtlasPlus data.<sup>44</sup>

To estimate the population size in the United States by sex, age, and race/ethnicity, we used Bridged-Race Population Estimates from CDC WONDER.<sup>45</sup>

### **Outcome measures**

The health and economic burden of disease associated with gonorrhea was measured as the numbers of discounted lifetime QALYs lost and costs due to gonococcal infections that occurred in 2015, by age and race/ethnicity for women, MSW, and MSM. These aggregate costs and QALYs were computed as the product of population size, incidence rates per population, and QALYs lost and costs per incident infection. Results are summarized in terms of aggregate population-level counts, as well as measures per 1,000 person-years.



## Multivariate sensitivity analysis

We performed multivariate sensitivity analyses including uncertainty in (i) incidence of gonorrhoea in each sub-population, (ii) proportion of each sub-population tested for gonorrhoea yearly, (iii) proportion of infected persons within each sub-population receiving treatment; (iv) sequelae probabilities, and (v) utilities, durations and costs. 1,000 samples were used in the sensitivity analyses. The five groups of input parameters were drawn from the posterior distributions estimated by the previously published calibrated transmission model<sup>13</sup> (for input categories i-iii above) or from specified distributions describing uncertainty in sequelae probabilities (Table 4.1). We report all outcomes using the mean and 95% uncertainty intervals (95% UI).

The reporting in this study follows Consolidated Health Economic Evaluation Reporting Standards (CHEERS)<sup>46</sup> and Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER)<sup>47</sup> as applicable.

**Table 4.1. Model input parameters describing gonorrhoea and sequelae probabilities, proportions of testing and treatment for gonorrhoea in women, MSW and MSM.**

Parameter	Mean estimate	Uncertainty interval*	References
<i>Testing and treatment probabilities</i>			
Probability of diagnosis and treatment	Table S4.5 (a) and (b)		13
Probability of adverse reactions to ceftriaxone and azithromycin			
Mild to moderate	0.15	0.076 to 0.25	48–51
Severe	0.0016	0.00017 to 0.0047	48–51
Probability of inpatient treatment for PID	0.10	0.098 to 0.11	35
Probability of inpatient treatment for EP	0.15	0.13 to 0.18	35
Probability of inpatient treatment for epididymitis	0.0054	0.0028 to 0.0092	36
Probability of inpatient treatment for DGI	0.29	0.17 to 0.43	37
<i>Outcome probabilities (women)</i>			
Proportion of infections that are symptomatic	0.37	0.27 to 0.47	13
Probability of PID given symptomatic infection	0.0025	0.00092 to 0.0055	13,17–23

**Table 4.1. Model input parameters describing gonorrhea and sequelae probabilities, proportions of testing and treatment for gonorrhea in women, MSW and MSM (Continued)**

Probability of PID given asymptomatic infection (age 15-24y)	0.075	0.031 to 0.13	13,17–23
Probability of PID given asymptomatic infection (age 25-39y)	0.091	0.037 to 0.16	13,17–23
Probability of CPP given PID	0.26	0.23 to 0.29	24,25
Probability of EP given PID	0.071	0.049 to 0.098	25–27,29
Probability of TI given PID	0.17	0.12 to 0.23	25,26,28–30
<i>Outcome probabilities (men)</i>			
Probability of symptomatic urethral infection (MSW)	0.72	0.60 to 0.84	13
Probability of symptomatic urethral infection (MSM)	0.62	0.43 to 0.76	13
Probability of asymptomatic urethral infection (MSM)	0.22	0.12 to 0.34	Supplementary Material
Probability of urethral infection (MSW)	1	Fixed	
Probability of symptoms given urethral infection (MSW)	0.72	0.60 to 0.84	13
Probability of urethral infection (MSM)	0.83	0.62 to 0.96	Supplementary Material
Probability of symptoms given urethral infection (MSM)	0.74	0.61 to 0.85	13, Supplementary Material
Probability of epididymitis given untreated urethral infection	0.042	0.0012 to 0.14	32,33
Probability of DGI given untreated urethral infection	0.010	0.0075 to 0.013	16,52
<i>Durations (years)</i>			
Gonococcal infection (women, symptomatic)†	0.017	0.0099 to 0.031	10
Gonococcal infection (women, asymptomatic, ages 15-24y)†	0.53	0.39 to 0.66	10
Gonococcal infection (women, asymptomatic, ages 25-39y)†	0.65	0.47 to 0.83	10
PID (inpatient treatment)	0.036	0.019 to 0.054	16
PID (outpatient treatment)	0.027	0.014 to 0.041	16
CPP	†	5 to lifetime	16,39,53
EP (inpatient treatment)	0.085	0.044 to 0.13	16
EP (outpatient treatment)	0.076	0.040 to 0.11	16
TI	†	5 to lifetime	16,39,53
Urethritis	0.019	0.0010 to 0.028	16
Epididymitis (inpatient treatment)	0.0082	0.0043 to 0.012	16
Epididymitis (outpatient treatment)	0.019	0.0010 to 0.028	16
DGI (inpatient treatment)	0.030	0.016 to 0.044	16
DGI (outpatient treatment)	0.022	0.011 to 0.032	16

**Table 4.1. Model input parameters describing gonorrhea and sequelae probabilities, proportions of testing and treatment for gonorrhea in women, MSW and MSM (Continued)**

<i>State-specific utilities</i>			
Symptomatic gonococcal infection (women)	0.85	0.78 to 0.92	16
PID (inpatient treatment)‡	0.76	0.64 to 0.87	16
PID (outpatient treatment)‡	0.63	0.45 to 0.81	16
CPP	0.60	0.41 to 0.79	16
EP (inpatient treatment)‡	0.62	0.44 to 0.80	16
EP (outpatient treatment)‡	0.58	0.38 to 0.78	16
TI	0.82	0.73 to 0.91	16
Urethritis	0.84	0.76 to 0.92	16
Epididymitis (inpatient treatment)	0.30	0.012 to 0.59	16
Epididymitis (outpatient treatment)	0.46	0.20 to 0.72	16
DGI (inpatient treatment)‡	0.68	0.53 to 0.84	16
DGI (outpatient treatment)‡	0.60	0.41 to 0.79	16
<i>Costs (in 2018 US\$)</i>			
<u>Testing costs</u>			
Urine nucleic acid amplification and diagnosis procedure	60	31 to 89	15,16,54–56
<u>Treatment of gonorrhea</u>			
Azithromycin 1 g	28	15 to 41	15,48,55,57
Ceftriaxone 250 mg	22	11 to 32	48,55,56
Short clinic visit	37	19 to 54	15,55,56
<u>Side effect</u>			
Adverse reactions to ceftriaxone and azithromycin			
Mild to moderate	77	40 to 115	16,48
Severe	8,007	4,174 to 11,898	48,58–61
<u>Treatment of sequelae</u>			
PID (inpatient treatment)	10,637	5,544 to 15,805	15,16,35,48,57,62–65
PID (outpatient treatment)	582	303 to 864	15,16,35,48,57,62–65
CPP	1,191	622 to 1,769	16,35,48
EP (inpatient treatment)	11,978	6,243 to 17,798	16,35,48
EP (outpatient treatment)	3,229	1,683 to 4,798	16,35,48
TI	6,245	3,255 to 9,280	16,35,48,57,63–65
Urethritis	235	123 to 349	16,48,62–65
Epididymitis (inpatient treatment)	7,364	3,838 to 10,942	16,36,48,57,62–64
Epididymitis (outpatient treatment)	422	220 to 627	16,36,48,57,62–64
DGI (inpatient treatment)	6,849	3,570 to 10,176	16
DGI (outpatient treatment)	563	293 to 836	16

\*Uncertainty intervals given as 95% uncertainty intervals for probabilities and durations of gonococcal infections based on results from the previous transmission model,<sup>13</sup> for probabilities of sequelae based on beta

distributions, and for all other parameters based on uniform distributions (see Supplementary Material for details).

† The previous transmission model was used to estimate duration of asymptomatic infection, which varies in relation to the fraction of infections that are diagnosed. Differences over years and between different race/ethnicity groups were relatively small, while differences between age groups were larger. For parsimony, we therefore allowed durations to vary by age but not by year and race/ethnicity group. Symptomatic infection was assumed to be 100% diagnosed and treated, with a much shorter duration of infection.

‡ Mean duration taken from uniform distribution between 5 years and lifetime duration. While expert opinion suggests lifetime duration of these conditions,<sup>16</sup> previous analyses have assumed much shorter durations based on unspecified evidence.<sup>53</sup> Details on life table calculations used to compute discounted QALY losses for lifetime duration are provided in Supplementary Material.

‡ The utility values were higher for inpatient treatment than outpatient treatment for sequelae of PID and EP for women and DGI for men. The utility values shown for inpatient and outpatient treatment reflect the average values over the duration of the condition, which is longer for inpatient treatment than outpatient treatment for most outcomes.<sup>16</sup> Inpatient treatment is associated with a greater total utility loss than outpatient treatment for most outcomes. For example, for PID, the 0.63 utility value for outpatient treatment is applied for 10 days. For inpatient treatment, the 0.76 value is applied for an average of about 13.3 days and reflects a utility value of about 0.55 applied for 3.3 days of inpatient care plus a utility value of about 0.82 applied for 10 days of follow-up outpatient care after the inpatient visit. Details of calculations are described in Supplementary Material table S4.3.

PID = pelvic inflammatory disease; CPP = chronic pelvic pain; EP = ectopic pregnancy; TI = tubal infertility; DGI = disseminated gonococcal infection; MSM = men who have sex with men; MSW = men who have sex with women.

## Results

### Lifetime QALYs lost and costs per gonococcal Infection

The discounted lifetime QALYs lost per incident gonococcal infection in 2015 were estimated as 0.093 (95% UI 0.022-0.22) for women, 0.0020 (0.0015-0.0024) for MSW, and 0.0017 (0.0013-0.0022) for MSM in (Table 4.2). The QALYs lost per incident infection were highest for women aged 25-39 years and were higher for women than for MSW and MSM. QALYs lost from gonococcal infection in women were predominantly from chronic sequelae such as CPP and TI, both of which are long-term complications with duration from years to lifetime. QALYs lost in men were from short-term complications with durations of several days to weeks. QALYs lost per incident gonococcal infection were higher in MSW than in MSM, since the estimates only measured the QALYs lost per incident urethral gonococcal infection, and we assumed all infections in MSW were urethral.

**Table 4.2. Estimated number of discounted lifetime QALYs lost associated with gonorrhea, per incident gonococcal infection in 2015, by sex and age-group.**

Sex and age-group	QALYs lost per infection	95% uncertainty interval (UI)
<b>Women</b>		
Ages 15-24 years	0.089	(0.021 to 0.21)
Ages 25-39 years	0.11	(0.026 to 0.25)
Ages 15-39 years	0.093	(0.022 to 0.22)
<b>MSW</b>		
Ages 15-24 years	0.0020	(0.0015 to 0.0024)
Ages 25-39 years	0.0019	(0.0015 to 0.0024)
Ages 15-39 years	0.0020	(0.0015 to 0.0024)
<b>MSM</b>		
Ages 15-24 years	0.0018	(0.0013 to 0.0023)
Ages 25-39 years	0.0017	(0.0012 to 0.0022)
Ages 15-39 years	0.0017	(0.0013 to 0.0022)

Estimated discounted lifetime costs per incident infection in 2015 were \$254 (95% UI 109-460) for women, \$306 (157-479) for MSW, and \$281 (145-443) for MSM (Table 4.3). The costs per incident infection in the 15-24 year population were lower than for the 25-39 year population, for women, which can be attribute to a lower probability of PID and sequelae given asymptomatic infection among young than old age groups. Costs of treatment per incident infection by race/ethnicity groups are reported in Supplementary Material Table S4.6.

**Table 4.3. Estimated discounted lifetime costs associated with gonorrhea, per incident gonococcal infection in 2015 (in 2018 US dollars), by sex and age-group.**

Sex and age-group	Costs (2018 US dollars)	95% uncertainty interval (UI)
<b>Women</b>		
Ages 15-24 years	248	(108 to 446)
Ages 25-39 years	278	(115 to 511)
Ages 15-39 years	254	(109 to 460)
<b>MSW</b>		
Ages 15-24 years	307	(157 to 480)
Ages 25-39 years	306	(157 to 479)
Ages 15-39 years	306	(157 to 479)
<b>MSM</b>		
Ages 15-24 years	282	(145 to 444)
Ages 25-39 years	281	(144 to 444)
Ages 15-39 years	281	(145 to 443)

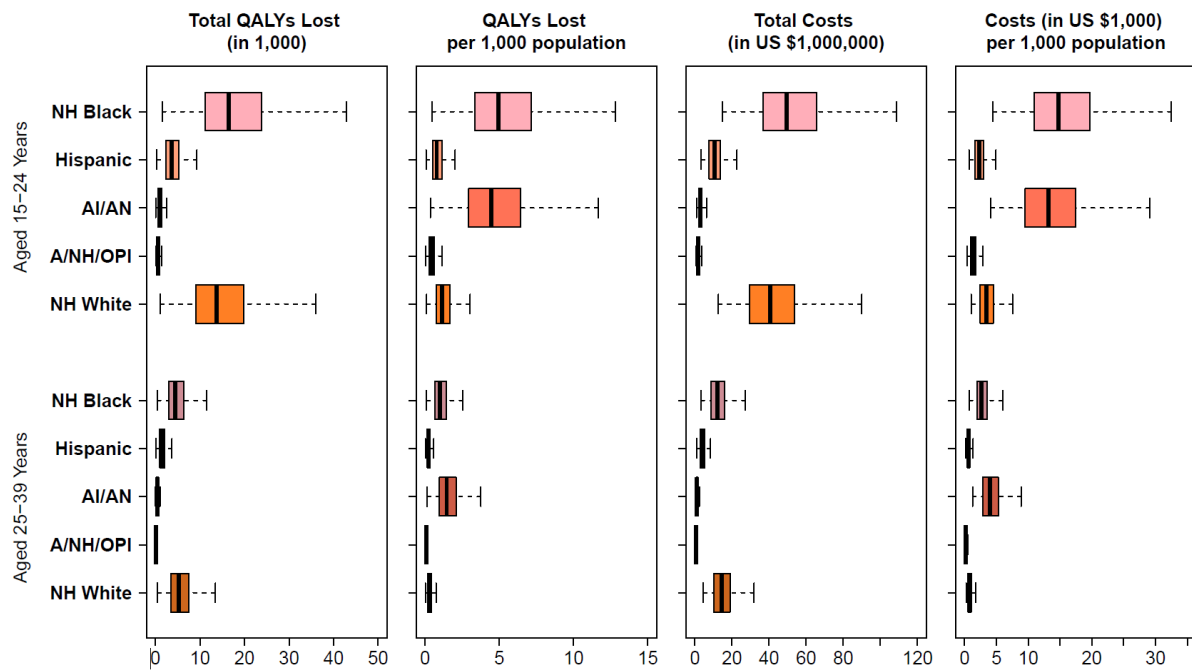
## Population-level QALYs lost and costs due to gonococcal infections

The population-level discounted lifetime QALYs lost associated with gonococcal infections in 2015 were 53,293 (95% UI 12,326-125,366) for women, 623 (431-876) for MSW, and 1,286 (632-2,074) for MSM (Table S4.7). The total population-level discounted lifetime costs associated with gonococcal infections in 2015 were \$145 million (63-266 million) for women, \$97 million (46-168 million) for MSW, and \$206 million (82-410 million) for MSM (Table S4.8). The total QALYs lost were highest for women aged 15-24 years and were higher for women than for MSW and MSM, which can be attributed to a higher number of QALYs lost per gonococcal infection for women than for men (Table 4.2, Figure 4.3, Figure 4.4, Figure 4.5). The total QALYs lost were higher for MSM than MSW, as a result of higher gonorrhea incidence among MSM than MSW (0.34, 95%UI 0.21-0.50 versus 0.0060, 0.0045-0.0080) in 2015.<sup>13</sup> With similar costs per gonococcal infection (Table 4.3), the total costs were highest for MSM than for women and MSW (Figure 4.3, Figure 4.4, Figure 4.5), primarily due to a much higher incidence for MSM (0.34, 95% UI 0.21-0.50) than women (0.011, 0.0087-0.013) and MSW (0.0060, 0.0045-0.0080) in 2015.<sup>13</sup>

Disparities in the total lifetime QALYs lost and costs reflect variation in the incidence of gonococcal infection across different populations, as well as different population sizes across groups. The total QALYs lost in 2015 were highest among NHB and NHW, followed by Hispanic, AI/AN, and A/NH/OPI, for both women and heterosexual men. For example, the total QALYs lost in 2015 were 18,694 (95% UI 4,437-44,119) for NHB, 15,583 (3,447-37,711) for NHW, 4,074 (962-9,504) for Hispanic, 1,090 (241-2,638) for AI/AN, and 632 (140-1,529) for A/NH/OPI among women aged 15-24 (Figure 4.3, Table S4.7). The total costs in 2015 were highest among NHB and NHW for women. For men, the highest total costs were among MSM, NHW, and NHB (Figure 4.4, Figure 4.5 and Table S4.8).

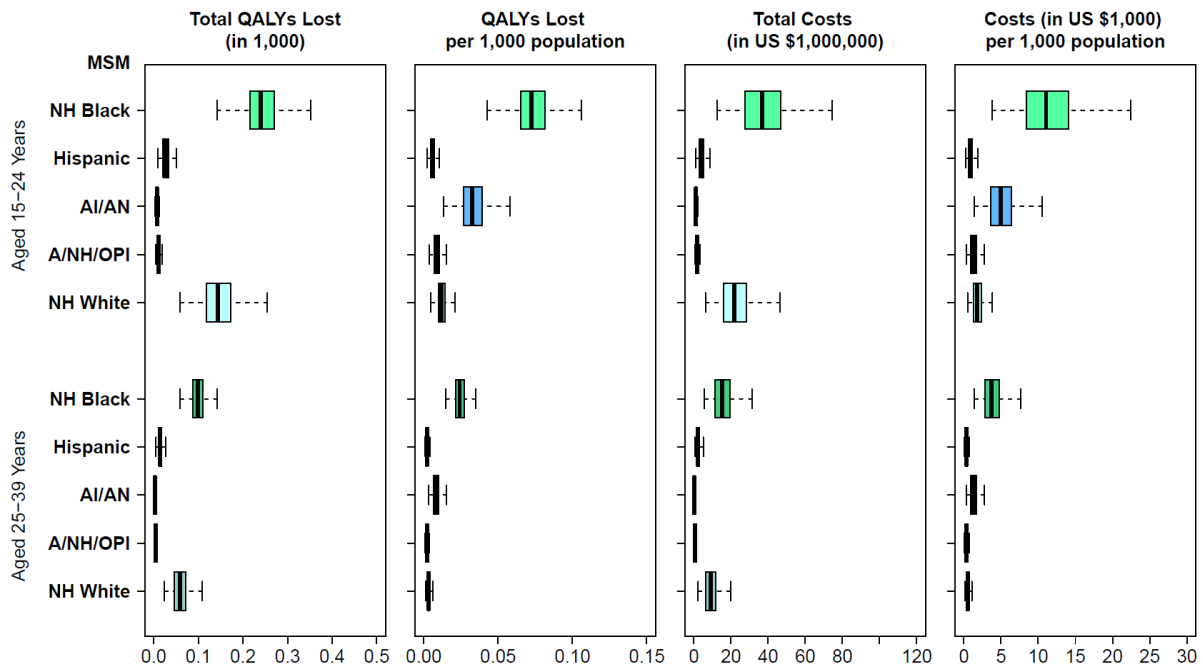
Disparities in lifetime QALYs lost and costs per 1,000 person-years reflect variation in incidence across different populations. QALYs lost per 1,000 person-years were highest among AI/AN and NHB, followed by NHW, Hispanic and A/NH/OPI among women. For men, the highest QALYs lost per 1,000

person-years were among MSM and NHB (Figure 4.4, Figure 4.5 and Table S4.7). Costs per 1,000 person-years in 2015 were higher among MSM, at \$96,352 (95% UI 38,412 -191,179) than MSW, at \$1,836 (879-3,182) (Table S4.8).



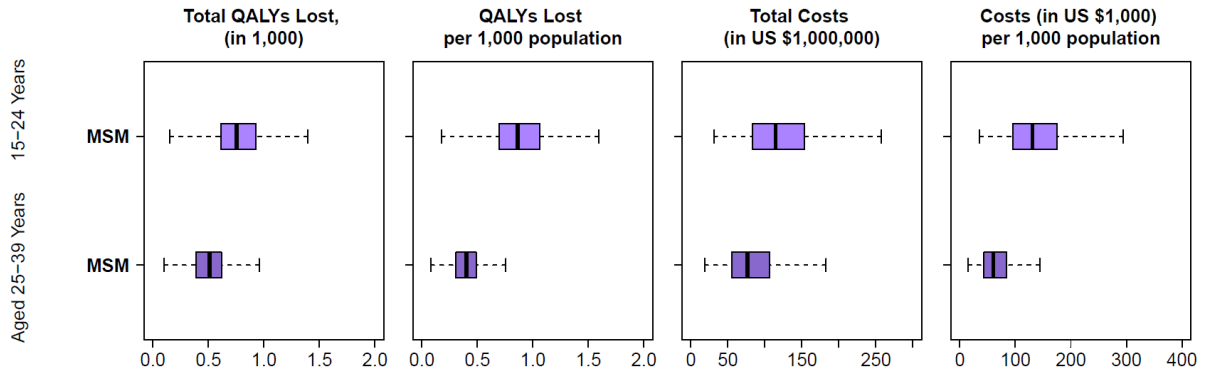
**Figure 4.3. Population-level numbers of QALYs lost, QALYs lost per 1,000 population, total costs (US\$), total costs per 1,000 population, for women, by age and race/ethnicity, in 2015.**

\*NH White (Non-Hispanic White), A/NH/OPI (Asian or Native Hawaiian or Other Pacific Islander), AI/AN (American Indian or Alaska Native), NH Black (Non-Hispanic Black). Note that x-axis differs between the plots



**Figure 4.4. Population-level numbers of QALYs lost, QALYs lost per 1,000 population, total costs (US\$), total costs per 1,000 population, for heterosexual men, by age and race/ethnicity, in 2015.**

\* NH White (Non-Hispanic White), A/NH/OPI (Asian or Native Hawaiian or Other Pacific Islander), AI/AN (American Indian or Alaska Native), NH Black (Non-Hispanic Black). Note that x-axis differs between the plots.



**Figure 4.5. Population-level numbers of QALYs lost, QALYs lost per 1,000 population, total costs (US\$), total costs per 1,000 population, for men who have sex with men (MSM), by age, in 2015.**

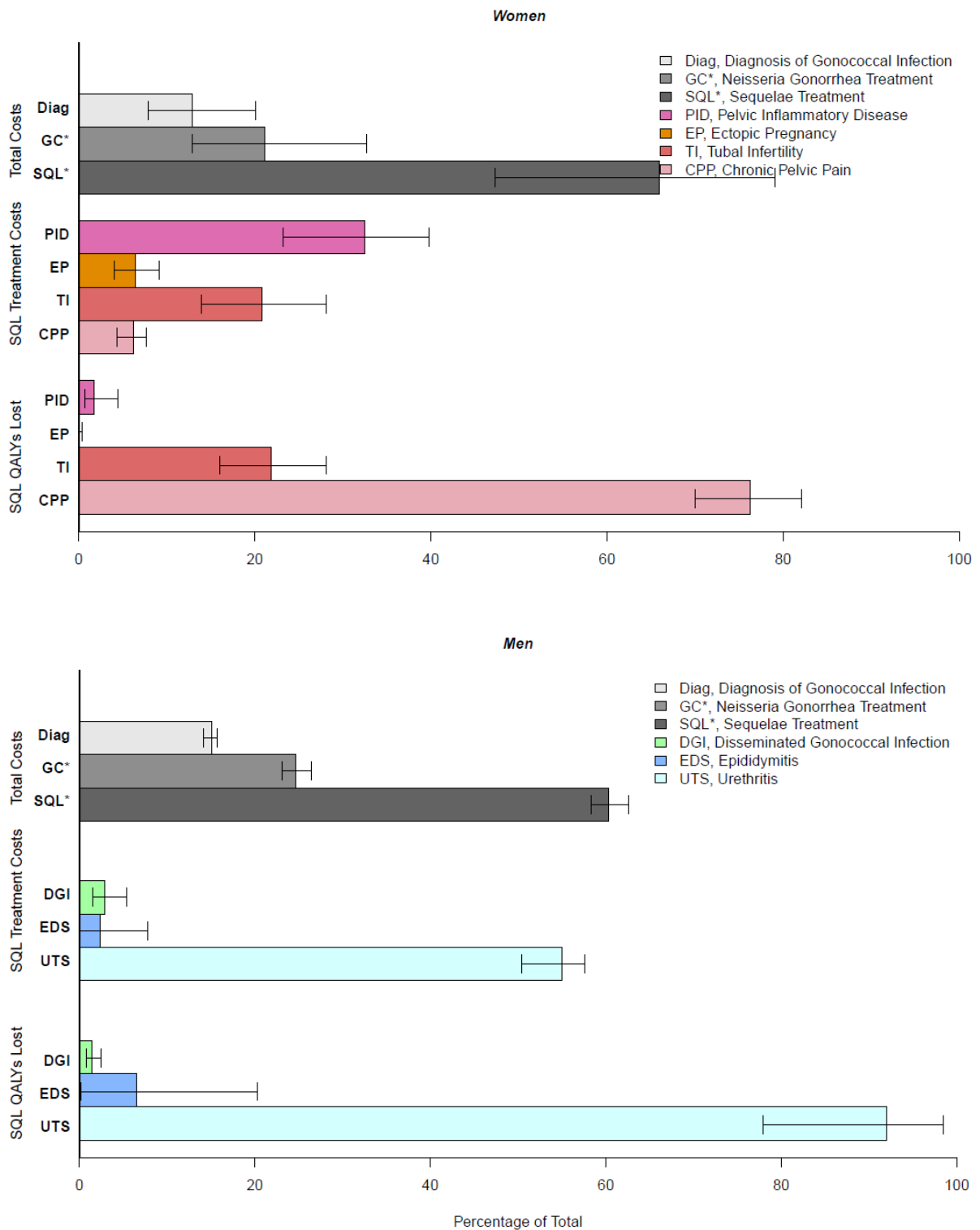
\* Note that x-axis differs between the plots.

### Composition of lifetime QALYs lost and costs

QALYs lost were higher in women than in MSW for all race/ethnicity groups. The duration of chronic sequelae in women was estimated to be longer than the duration of sequelae in men. Among women, the composition of total QALYs lost was estimated to be 76% (95% UI 70%-82%) due to CPP and 22%



(16%-28%) from TI (Figure 4.6), with negligible contributions from other outcomes. Although chronic complications of CPP and TI were the main source of total QALYs lost in women with gonococcal infection due to their long and lifetime durations, the direct costs from these two complications were smaller. TI was estimated to contribute 20% (13%-28%) of the total direct medical costs among women, with a further contribution of 6% (4%-8%) from CPP. The major contributor to the total costs were treatment of PID (32%, 95% UI 22%-39%) and treatment of gonorrhoea (23%, 14%-35%). Among men, the major contributor to the total QALYs lost was urethritis (92%, 78%-98%). Urethritis also was the largest component of costs for treating complications (53%, 48%-57%). The cost of test and diagnosis for gonorrhoea was the smallest component, contributing 13% (8%-19%) of the total costs for women, and 15% (13%-15%) for men. The main costs were from treatment for gonorrhoea sequelae, which contributed 64% (46%-78%) of the total costs for women, and 58% (55%-62%) for men.



**Figure 4.6. Decomposition of total lifetime QALYs lost and costs, by main complications and by broad cost categories.**

\*Note GC represented the costs of treatment for gonococcal infections, not including the costs of treatment for gonorrhea sequelae. Sql represented the costs of treatment gonorrhea sequelae.

## Discussion

In this study we quantified the burden of gonococcal infection in the United States using discounted lifetime QALYs lost and costs, and examined disparities in the population. Women were estimated to have higher QALYs lost and costs per incident gonococcal infection than men, and NHB women had the largest burden of disease of all subpopulations examined. QALYs lost and costs per 1,000 person-years were the highest among MSM and NHB women, reflecting the significant burden of gonococcal infection in these populations. QALYs lost per 1,000 person-years for NHB women were over 4 times those of NHW women, and for AI/AN over 3 times as high as for NHW women. QALYs lost per 1,000 person-years for MSM were over 80 times those of NHW MSW.

Gonococcal infection has relatively short duration, but its longer-term consequences are captured in our estimates of QALYs lost. These estimates point to unmet sexual and reproductive health needs and wider inequities within the population, which place people at differential risk of infection acquisition and sequelae. NHB and MSM have been noted as key populations for gonorrhea prevention also by analyses of surveillance data.<sup>2,8</sup> The high relative burden of gonorrhea among AI/AN has received less attention. Reports of a multi-state syphilis outbreak,<sup>66</sup> and high burden of chlamydia among young AI/AN,<sup>67</sup> signal that there are systemic disparities that need to be addressed. Discrimination, socio-economic status, segregation, institutional racism, and access to and utilization of health care are among the factors contributing to racial/ethnic disparities.<sup>68-70</sup>

We estimated the discounted lifetime costs per incident infection to be \$254 (95% UI 109-460) for women infected in 2015. This is lower than a previous estimate of 354 (95% UI 177-531) per incident infection in women.<sup>56</sup> The difference compared to estimate used in previous study<sup>71</sup> is primarily attributed to a lower probability of PID used in our model, which was estimated based on the duration of gonococcal infection from the previous transmission modeling study<sup>13</sup> and an estimated yearly probability of PID for chlamydial infections synthesized from previous studies.<sup>17-23</sup> It is also attributed to a lower rate of inpatient care of PID and sequelae used in our model, which was estimated from

claims data<sup>35</sup> on the costs of treatment for PID, compared estimates reported in previous studies.<sup>16,48,57,62,63,65</sup> Evidence of a lower rate of inpatient treatment for complications of urethral infection among men from recent studies<sup>36,37</sup> was also applied in our model.

Our study has a number of limitations. First, as with previous studies,<sup>56,62,72</sup> we have assumed that chlamydia and gonorrhea are associated with similar probabilities of developing sequelae. However, gonococcal infection may result in more severe PID than chlamydia.<sup>4</sup> In that case, the assumption that sequelae probabilities are transferrable from studies of outcomes secondary to chlamydial infections may underestimate the QALYs lost due to gonorrhea. Second, costs in our study reflect average costs. There are undoubtedly variations between states, and some costs may have changed over time. Third, as an economic analysis of national disease burden, it would be more consistent with standard methodological recommendations to use community-based rather than expert-based utility measures; however, we were not able to identify any existing studies that report community-based utilities for the range of outcomes associated with gonococcal infection among men. Finally, we did not include the potential increase for HIV acquisition or transmission in people with gonococcal infection. If bacterial STIs increase HIV acquisition and transmission, 10.2% of HIV infections could be attributable to gonorrhea and chlamydia among MSM.<sup>73</sup> This would result in larger total burden associated with gonococcal infection than our current model estimates.

The findings in our study have important implications for resource prioritization and planning, and for informing control policies. They underscore the longer-term burden of a short-term infection, and the continued disparities by race/ethnicity and for MSM within the United States. The QALYs lost and costs associated with gonococcal infection are likely to increase in the future with rising antibiotic resistance. Decreasing disparities by reducing the risk of gonorrhea within disproportionately affected populations would improve the overall health of the population. Measuring both short- and long-term consequences and costs of gonococcal infection provides a comprehensive framework for measuring and evaluating gonorrhea associated health outcomes.

## Source of funding

This work was supported by the U.S. Centers for Disease Control and Prevention, National Center for HIV, Viral Hepatitis, STD, and TB Prevention Epidemiologic and Economic Modeling Agreement (5NU38PS004644). MMR was also supported by Charles. A King Trust postdoctoral fellowship. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention or the authors' other affiliated institutions.

## Previous presentation

Poster at the 2018 National STD Prevention Conference in Washington DC, August 27-30, 2018.

## Conflict of interest

The authors declare no conflicts of interest.

## References

1. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2018*. Atlanta: U.S. Department of Health and Human Services; 2019. doi:10.15620/cdc.79370
2. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2016*. Atlanta: U.S. Department of Health and Human Services; 2016:164. [https://www.cdc.gov/std/stats16/CDC\\_2016\\_STDS\\_Report-for508WebSep21\\_2017\\_1644.pdf](https://www.cdc.gov/std/stats16/CDC_2016_STDS_Report-for508WebSep21_2017_1644.pdf)
3. Tracking the Hidden Epidemics: Trends in STDs in the United States 2000: (620032007-001). Published online 2000. doi:10.1037/e620032007-001
4. Reekie J, Donovan B, Guy R, et al. Risk of Pelvic Inflammatory Disease in Relation to Chlamydia and Gonorrhoea Testing, Repeat Testing, and Positivity: A Population-Based Cohort Study. *Clin Infect Dis*. 2018;66(3):437-443. doi:10.1093/cid/cix769
5. Fleming DT, Wasserheit JN. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. *Sex Transm Infect*. 1999;75(1):3-17. doi:10.1136/sti.75.1.3

6. Rottingen J-A, William Cameron DM, Garnett GP. A Systematic Review of the Epidemiologic Interactions Between Classic Sexually Transmitted Diseases and HIV: How Much Really Is Known? *Sex Transm Dis.* 2001;28(10):579-597.
7. Hoover KM, Bohm M, Keppel K. Measuring Disparities in the Incidence of Sexually Transmitted Diseases. [Review]. *Sex Transm Dis.* 2008;35(12). doi:10.1097/OLQ.0b013e3181886750
8. Chesson HW, Patel CG, Gift TL, Aral SO. Trends in Selected Measures of Racial and Ethnic Disparities in Gonorrhea and Syphilis in the United States, 1981–2013: *Sex Transm Dis.* 2016;43(11):661-667. doi:10.1097/OLQ.0000000000000518
9. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2013.* Atlanta: U.S. Department of Health and Human Services; 2014. Accessed June 23, 2020. <https://wonder.cdc.gov/wonder/help/STD/STDSurv2013.pdf>
10. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2017.* Atlanta: U.S. Department of Health and Human Services; 2018. Accessed September 21, 2019. <https://www.cdc.gov/std/stats17/default.htm>
11. Stenger MR, Pathela P, Anschuetz G, et al. Increases in the Rate of Neisseria gonorrhoeae Among Gay, Bisexual and Other Men Who Have Sex With Men—Findings From the Sexually Transmitted Disease Surveillance Network 2010–2015: *Sex Transm Dis.* 2017;44(7):393-397. doi:10.1097/OLQ.0000000000000623
12. Williams AM, Weston EJ, Gift TL, Torrone E. Increases in the Estimated Number of Reported Gonorrhea Cases Among Men Who Have Sex With Men: The Role of Testing. *Sex Transm Dis.* 2019;46(11):713-715. doi:10.1097/OLQ.0000000000001019
13. Tuite AR, Rönn MM, Wolf EE, et al. Estimated Impact of Screening on Gonorrhea Epidemiology in the United States: Insights From a Mathematical Model. *Sex Transm Dis.* 2018;45(11):713. doi:10.1097/OLQ.0000000000000876
14. Tuite AR, Jayaraman GC, Allen VG, Fisman DN. Estimation of the Burden of Disease and Costs of Genital Chlamydia trachomatis Infection in Canada: *Sex Transm Dis.* 2012;39(4):260-267. doi:10.1097/OLQ.0b013e31824717ae
15. Hu D, Hook EW, Goldie SJ. Screening for *Chlamydia trachomatis* in Women 15 to 29 Years of Age: A Cost-Effectiveness Analysis. *Ann Intern Med.* 2004;141(7):501. doi:10.7326/0003-4819-141-7-200410050-00006
16. Institute of Medicine (US) Committee to Study Priorities for Vaccine Development. *Vaccines for the 21st Century: A Tool for Decisionmaking.* (Stratton KR, Durch JS, Lawrence RS, eds.). National Academies Press (US); 2000. Accessed September 21, 2019. <http://www.ncbi.nlm.nih.gov/books/NBK233313/>

17. Morre SA, van Den Brule AJC, Rozendaal L, et al. The natural course of asymptomatic *Chlamydia trachomatis* infections: 45% clearance and no development of clinical PID after one-year follow-up. *Int J Std Aids*. 2002;13:12–18.
18. Oakeshott P, Kerry S, Aghaizu A, et al. Randomised controlled trial of screening for *Chlamydia trachomatis* to prevent pelvic inflammatory disease: the POPI (prevention of pelvic infection) trial. *BMJ*. 2010;340(7752):903-903.
19. Østergaard L, Andersen B, Møller JK, Olesen F. Home Sampling versus Conventional Swab Sampling for Screening of *Chlamydia trachomatis* in Women: A Cluster-Randomized 1-Year Follow-up Study. *Clin Infect Dis*. 2000;31(4):951-957. doi:10.1086/318139
20. Rees E. The treatment of pelvic inflammatory disease. *Am J Obstet Gynecol*. 1980;138(7):1042-1047. doi:10.1016/0002-9378(80)91105-9
21. Scholes D, Stergachis A, Heidrich FE, Andrilla H, Holmes KK, Stamm WE. Prevention of Pelvic Inflammatory Disease by Screening for Cervical Chlamydial Infection. *N Engl J Med*. 1996;334(21):1362-1366. doi:10.1056/NEJM199605233342103
22. Price MJ, Ades AE, De Angelis D, et al. Risk of Pelvic Inflammatory Disease Following *Chlamydia trachomatis* Infection: Analysis of Prospective Studies With a Multistate Model. *Am J Epidemiol*. 2013;178(3):484-492. doi:10.1093/aje/kws583
23. Trikalinos TA. *PROBABILITY OF PELVIC INFLAMMATORY DISEASE IN CHLAMYDIA TRACHOMATIS INFECTION.*; 2018. [https://www.brown.edu/public-health/cesh/sites/public-health-cesh/files/STI\\_probabilities.pdf](https://www.brown.edu/public-health/cesh/sites/public-health-cesh/files/STI_probabilities.pdf)
24. Ness RB, Soper DE, Holley RL, et al. Effectiveness of inpatient and outpatient treatment strategies for women with pelvic inflammatory disease: Results from the pelvic inflammatory disease evaluation and clinical health (peach) randomized trial. *Am J Obstet Gynecol*. 2002;186(5):929-937. doi:10.1067/mob.2002.121625
25. Weström L. Effect of acute pelvic inflammatory disease on fertility. *Am J Obstet Gynecol*. 1975;121(5):707-713. doi:10.1016/0002-9378(75)90477-9
26. Weström L. Incidence, prevalence, and trends of acute pelvic inflammatory disease and its consequences in industrialized countries. *Am J Obstet Gynecol*. 1980;138(7):880-892. doi:10.1016/0002-9378(80)91077-7
27. Weström L, Bengtsson LP, Mårdh PA. Incidence, trends, and risks of ectopic pregnancy in a population of women. *Br Med J Clin Res Ed*. 1981;282(6257):15-18. doi:10.1136/bmj.282.6257.15
28. Weström L. Gynecological Chlamydial infections. *Infection*. 1982;10(1):S40-S45. doi:10.1007/BF01640713

29. Westrom LM, Joesoef R, Reynolds G, Hagdu AM, Thompson SEM. Pelvic Inflammatory Disease and Fertility: A Cohort Study of 1,844 Women with Laparoscopically Verified Disease and 657 Control Women with Normal Laparoscopic Results. *Sex Transm Dis*. 1992;19(4):185-192.
30. Brunham RC, Maclean IW, Binns B, Peeling RW. Chlamydia trachomatis: Its Role in Tubal Infertility. *J Infect Dis*. 1985;152(6):1275-1282.
31. John J, Donald WH. Asymptomatic urethral gonorrhoea in men. *Br J Vener Dis*. 1978;54(5):322-323. doi:10.1136/sti.54.5.322
32. Paavonen J, Kousa M, Saikku P, Vartiainen E, Kanerva L, Lassus A. Treatment of nongonococcal urethritis with trimethoprim-sulphadiazine and with placebo. A double-blind partner-controlled study. *Br J Vener Dis*. 1980;56(2):101-104. doi:10.1136/sti.56.2.101
33. Geisler WM, Wang C, Morrison SG, Black CM, Bandea CI, Hook EW. The natural history of untreated Chlamydia trachomatis infection in the interval between screening and returning for treatment. *Sex Transm Dis*. 2008;35(2):119-123. doi:10.1097/OLQ.0b013e318151497d
34. Thanh NX, Akpınar I, Gratrix J, et al. Benefit of adjunct universal rectal screening for Chlamydia genital infections in women attending Canadian sexually transmitted infection clinics. *Int J STD AIDS*. 2017;28(13):1311-1324. doi:10.1177/0956462417704344
35. Rein D. Direct medical cost of pelvic inflammatory disease and its sequelae: decreasing, but still substantial. *Obstet Gynecol*. 2000;95(3):397-402. doi:10.1016/S0029-7844(99)00551-7
36. Gift TL, Owens CJ. The Direct Medical Cost of Epididymitis and Orchitis: Evidence From a Study of Insurance Claims. *Sex Transm Dis*. 2006;33(10):S84. doi:10.1097/01.olq.0000235149.41948.fa
37. Belkacem A, Caumes E, Ouanich J, et al. Changing patterns of disseminated gonococcal infection in France: cross-sectional data 2009–2011: Table 1. *Sex Transm Infect*. 2013;89(8):613-615. doi:10.1136/sextrans-2013-051119
38. Sullivan PW, Ghushchyan V. Preference-Based EQ-5D Index Scores for Chronic Conditions in the United States. *Med Decis Mak Int J Soc Med Decis Mak*. 2006;26(4):410-420. doi:10.1177/0272989X06290495
39. Centers for Disease Control and Prevention NC for HS. Multiple Cause of Death, 1999-2017 on CDC WONDER Online Database. Accessed September 22, 2019. <https://wonder.cdc.gov/mcd-icd10.html>
40. Bureau UC. Statistical Abstract of the United States: 2010. Accessed September 22, 2019. <https://www.census.gov/library/publications/2009/compendia/statab/129ed.html>
41. Lopez AD, Murray CCJL. The global burden of disease, 1990–2020. *Nat Med*. 1998;4(11):1241-1243. doi:10.1038/3218



42. STDs in Racial and Ethnic Minorities - 2017 Sexually Transmitted Diseases Surveillance. Published June 17, 2019. Accessed September 22, 2019. <https://www.cdc.gov/std/stats17/minorities.htm>
43. Appendix A - 2017 Sexually Transmitted Diseases Surveillance. Published January 11, 2019. Accessed September 22, 2019. <https://www.cdc.gov/std/stats17/appendix-a.htm>
44. AtlasPlus | NCHHSTP | CDC. Published February 14, 2019. Accessed September 22, 2019. <https://www.cdc.gov/nchhstp/atlas/index.htm>
45. U.S. Census Populations With Bridged Race Categories. Published June 20, 2019. Accessed September 22, 2019. [https://www.cdc.gov/nchs/nvss/bridged\\_race.htm](https://www.cdc.gov/nchs/nvss/bridged_race.htm)
46. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BJOG Int J Obstet Gynaecol*. 2013;120(6):765-770. doi:10.1111/1471-0528.12241
47. Stevens GA, Alkema L, Black RE, et al. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *The Lancet*. 2016;388(10062):e19-e23. doi:10.1016/S0140-6736(16)30388-9
48. Magid D. Doxycycline Compared with Azithromycin for Treating Women with Genital Chlamydia trachomatis Infections: An Incremental Cost-Effectiveness Analysis. *Ann Intern Med*. 1996;124(4):389. doi:10.7326/0003-4819-124-4-199602150-00002
49. Bai Z-G, Bao X-J, Cheng W-D, Yang K-H, Li Y-P. Efficacy and safety of ceftriaxone for uncomplicated gonorrhoea: a meta-analysis of randomized controlled trials. *Int J STD AIDS*. 2012;23(2):126-132. doi:10.1258/ijsa.2009.009198
50. Lau C-Y, Qureshi AK. Azithromycin Versus Doxycycline for Genital Chlamydial Infections: A Meta-Analysis of Randomized Clinical Trials. *Sex Transm Dis*. 2002;29(9):497-502.
51. Ross JDC, Brittain C, Cole M, et al. Gentamicin compared with ceftriaxone for the treatment of gonorrhoea (G-ToG): a randomised non-inferiority trial. *The Lancet*. 2019;393(10190):2511-2520. doi:10.1016/S0140-6736(18)32817-4
52. Birrell JM, Gunathilake M, Singleton S, Williams S, Krause V. Characteristics and Impact of Disseminated Gonococcal Infection in the "Top End" of Australia. *Am J Trop Med Hyg*. 2019;101(4):753-760. doi:10.4269/ajtmh.19-0288
53. Aledort JE, Hook EWI, Weinstein MC, Goldie SJM. The Cost Effectiveness of Gonorrhoea Screening in Urban Emergency Departments. *Sex Transm Dis*. 2005;32(7):425-436.
54. Gift TL, Kissinger P, Mohammed H, Leichliter JS, Hogben M, Golden MR. The Cost and Cost-Effectiveness of Expedited Partner Therapy Compared With Standard Partner Referral for the

Treatment of Chlamydia or Gonorrhoea. *Sex Transm Dis.* 2011;38(11):1067–1073. doi:10.1097/OLQ.0b013e31822e9192

55. Gift TL, Walsh CD, Haddix A, Irwin KLM. A Cost-Effectiveness Evaluation of Testing and Treatment of Chlamydia trachomatis Infection Among Asymptomatic Women Infected With Neisseria gonorrhoeae. *Sex Transm Dis.* 2002;29(9):542-551.

56. Owusu-Edusei K, Chesson HW, Gift TL, et al. The Estimated Direct Medical Cost of Selected Sexually Transmitted Infections in the United States, 2008: *Sex Transm Dis.* 2013;40(3):197-201. doi:10.1097/OLQ.0b013e318285c6d2

57. Petitta A, Hart SM, Bailey EM. Economic evaluation of three methods of treating urogenital chlamydial infections in the emergency department. *Pharmacotherapy.* 1999;19(5):648-654. doi:10.1592/phco.19.8.648.31534

58. Doxycycline compared with azithromycin for treating women with genital chlamydia trachomatis infections: an incremental cost effectiveness analysis: Magid D, Douglas JM, Schwartz JS. *Ann Intern Med.* 1996; 124:389–398. *J Emerg Med.* 1996;14(6):788. doi:10.1016/S0736-4679(97)85165-7

59. Classen DC, Pestotnik SL, Evans RS, Lloyd JF, Burke JP. Adverse Drug Events in Hospitalized Patients: Excess Length of Stay, Extra Costs, and Attributable Mortality. *JAMA.* 1997;277(4):301-306. doi:10.1001/jama.1997.03540280039031

60. Bates DW, Spell N, Cullen DJ, et al. The Costs of Adverse Drug Events in Hospitalized Patients. *JAMA.* 1997;277(4):307-311. doi:10.1001/jama.1997.03540280045032

61. White TJ, Arakelian A, Rho JP. Counting the Costs of Drug-Related Adverse Events. *Pharmacoeconomics.* 1999;15(5):445-458. doi:10.2165/00019053-199915050-00003

62. Mehta SD, Bishai D, Howell MRS, Rothman RE, Quinn TCM, Zenilman JM. Cost-Effectiveness of Five Strategies for Gonorrhoea and Chlamydia Control Among Female and Male Emergency Department Patients. *Sex Transm Dis.* 2002;29(2):83-91.

63. Randolph AG, Washington AE. Screening for Chlamydia Trachomatis in Adolescent Males: A Cost-Based Decision Analysis. 1990;80(5):6.

64. Washington AE, Johnson RE, Sanders LL. Chlamydia trachomatis Infections in the United States: What Are They Costing Us? *JAMA.* 1987;257(15):2070-2072. doi:10.1001/jama.1987.03390150086041

65. Ginocchio RHS, Veenstra DL, Connell FA, Marrazzo JM. The Clinical and Economic Consequences of Screening Young Men for Genital Chlamydial Infection: *Sex Transm Dis.* 2003;30(2):99-106. doi:10.1097/00007435-200302000-00001

66. Bowen VB, Peterman TA, Calles DL, Thompson ARM, Kirkcaldy RD, Taylor MM. Multistate Syphilis Outbreak Among American Indians, 2013 to 2015. [Miscellaneous Article]. *Sex Transm Dis*. 2018;45(10):690-695. doi:10.1097/OLQ.0000000000000809
67. Learner ER, Torrone EA, Fine JPS, Pence BW, Powers KA, Miller WC. Chlamydia Prevalence Trends Among Women and Men Entering the National Job Training Program From 1990 Through 2012. [Miscellaneous Article]. *Sex Transm Dis*. 2018;45(8):554-559. doi:10.1097/OLQ.0000000000000798
68. Hogben M, Leichliter JS. Social Determinants and Sexually Transmitted Disease Disparities. [Review]. *Sex Transm Dis*. 2008;35(12). doi:10.1097/OLQ.0b013e31818d3cad
69. Harling G, Subramanian S, Bärnighausen T, Kawachi I. Socioeconomic Disparities in Sexually Transmitted Infections Among Young Adults in the United States: Examining the Interaction Between Income and Race/Ethnicity. *Sex Transm Dis*. 2013;40(7):575-581. doi:10.1097/OLQ.0b013e31829529cf
70. Sources of racial disparities in HIV prevalence in men who have sex with men in Atlanta, GA, USA: a modelling study- ClinicalKey. Accessed September 21, 2019. <https://www-clinicalkey-com.ezp-prod1.hul.harvard.edu/#!/content/playContent/1-s2.0-S235230181730067X?returnurl=null&referrer=null>
71. Yeh JMM, Hook EWI, Goldie SJM. A Refined Estimate of the Average Lifetime Cost of Pelvic Inflammatory Disease. *Sex Transm Dis*. 2003;30(5):369-378.
72. Potential impact of vaccination against *Neisseria meningitidis* on *Neisseria gonorrhoeae* in the United States: Results from a decision-analysis model: Human Vaccines & Immunotherapeutics: Vol 10, No 12. Accessed September 22, 2019. <https://www.tandfonline.com/doi/full/10.4161/hv.36221>
73. Jones J, Weiss K, Mermin J, et al. Proportion of Incident Human Immunodeficiency Virus Cases Among Men Who Have Sex With Men Attributable to Gonorrhea and Chlamydia: A Modeling Analysis. *Sex Transm Dis*. 2019;46(6):357-363. doi:10.1097/OLQ.0000000000000980

## Supplementary Material

### Supplementary Methods

#### 1. Sequelae probabilities

The probability of symptomatic urethral infection (given a gonococcal infection) for both heterosexual men (MSW) and men who have sex with men (MSM) was obtained from a gonorrhea transmission modeling study.<sup>1</sup> The model did not stratify gonococcal infection by site of infection, and it estimated the overall probability of symptomatic infection in MSM and in MSW. The probability of asymptomatic urethral infection among MSM was estimated by assuming that the odds of symptoms given urethral infection were the same for MSM as for MSW (equation 1):

$$\text{Prob}^{\text{MSW}}(\text{SympUreth}) / \text{Prob}^{\text{MSW}}(\text{AsympUreth}) = \text{Prob}^{\text{MSM}}(\text{SympUreth}) / \text{Prob}^{\text{MSM}}(\text{AsympUreth}) \quad (1)$$

For MSW, we used equation 2, assuming they only have urethral infections (see also Figure 4.2A in main manuscript):

$$\text{Prob}^{\text{MSW}}(\text{SympUreth}) + \text{Prob}^{\text{MSW}}(\text{AsympUreth}) = 1 \quad (2)$$

We used equation 3 for MSM to allow for some infections not being urethral (see also Figure 4.2B in main manuscript):

$$\text{Prob}^{\text{MSM}}(\text{SympUreth}) + \text{Prob}^{\text{MSM}}(\text{AsympUreth}) + \text{Prob}^{\text{MSM}}(\text{Othersites}) = 1 \quad (3)$$

Combining equations 1-3, we have equation 4:

$$\text{Prob}^{\text{MSM}}(\text{SympUreth}) / \text{Prob}^{\text{MSW}}(\text{SympUreth}) + \text{Prob}^{\text{MSM}}(\text{Othersites}) = 1 \quad (4)$$

where:

$\text{Prob}^{\text{MSW}}(\text{SympUreth})$  is the probability of symptomatic urethral infection given gonococcal infection among MSW;

$\text{Prob}^{\text{MSW}}(\text{AsympUreth})$  is the probability of asymptomatic urethral infection given gonococcal infection among MSW;

$\text{Prob}^{\text{MSM}}(\text{SympUreth})$  is the probability of symptomatic urethral infection given gonococcal infection among MSM;

$\text{Prob}^{\text{MSM}}(\text{AsympUreth})$  is the probability of asymptomatic urethral infection given gonococcal infection among MSM;

$\text{Prob}^{\text{MSM}}(\text{Othersites})$  is the probability of other site infections given gonococcal infection among MSM.

Equation 4 implies a restriction that the ratio of the probability of symptomatic urethral infection among MSM to the probability of symptomatic urethral infection among MSW should be greater than 0 but equal to or less than 1. A beta distribution (mean 0.87, 95% UI 0.50-0.99, alpha=4.77, and beta=0.75) was fitted to this ratio estimated in the previous modeling study.<sup>1</sup> Then we calibrated the probability of symptomatic urethral infection among MSM with the fitted beta distribution.

Finally, rearranging equation 1, the probability of asymptomatic urethral infection among MSM was estimated as:

$$\text{Prob}^{\text{MSM}}(\text{AsympUreth}) = \text{Prob}^{\text{MSM}}(\text{SympUreth}) * \text{Prob}^{\text{MSW}}(\text{AsympUreth}) / \text{Prob}^{\text{MSW}}(\text{SympUreth})$$

The probabilities of developing pelvic inflammatory disease (PID), chronic pelvic pain (CPP), ectopic pregnancy (EP), tubal factor infertility (TI), and epididymitis were estimated by synthesizing evidence from primary studies identified through scoping literature reviews. Most relevant studies on these outcomes identified sequelae due to chlamydia, with a lack of direct evidence on sequelae secondary to gonococcal infection. Therefore, as in previous modeling analyses,<sup>2-8</sup> we assumed that probabilities of developing sequelae after gonococcal infection were the same as probabilities after chlamydial infection.

For PID, we used an evidence synthesis model, described elsewhere,<sup>8</sup> to estimate the annual probability that a *Chlamydia trachomatis* infection leads to PID, using information from prospective or retrospective population or clinic-based cohorts, or from RCTs of chlamydia screening. Four trials and one case cohort were included in the analysis.<sup>2-7</sup> Eligible studies ascertained the exposure (infection status) with a range of laboratory investigations (cultures, NAAT assays) and the outcome (PID) with clinical criteria.

We synthesized information using a continuous-time Markov model with four health states: "chlamydia positive symptomatic", "chlamydia positive asymptomatic", "chlamydia negative", and "PID". In each cohort, we modeled the number of patients with PID at the end of follow-up based on the vector of probabilities of being in each health state at the end of the follow-up (i.e., the Markov trace) and the initial conditions (the trace at time 0, known from the design of each cohort). In turn, the time-constant hazard rates for the transitions between states were implicit functions of the trace of the Markov process. The hazard rates for all transitions were allowed to vary by cohort, with the exception of the causal hazard rate for PID attributed to a chlamydial infection, which was assumed to be the same across all cohorts. The causal rate of PID secondary to chlamydial infection was not identifiable using only the available data, namely, from the number of patients in each state at baseline and the number of patients with PID at the end of follow-up. We specified informative priors for the clearance rate of chlamydia with and without treatment. We used external data to inform the clearance rates of chlamydia under treatment. We estimated the clearance rate of chlamydia without treatment assuming that the prevalence of chlamydial infections at baseline for a large population-based RCT corresponded to the stationary distribution (long-term equilibrium) of the Markov model, and solving for it. Estimation was done using Markov Chain Monte Carlo (MCMC), implemented in the JAGS modeling language. We used the multi-state modeling module in JAGS to obtain solutions to the implicit functions between the Markov trace and the hazard rates. Further details on the methodology are available in a technical report.<sup>8</sup>

There were relatively few studies on the probabilities of CPP, EP, TI, epididymitis, and DGI, which led to a simpler modeling framework for each of these outcomes. We modeled probabilities of CPP, EP and TI conditional on having PID, based on pooled studies on each sequela,<sup>11–17</sup> and using log-link functions. The probabilities of epididymitis were estimated among those with nongonococcal urethritis<sup>18</sup> or *Chlamydia trachomatis* infection.<sup>19</sup> We adopted a previous estimate of the probability of DGI among men with symptomatic *Neisseria gonorrhoea*, developed in an Institute of Medicine study (IOM).<sup>9</sup> That study did not report uncertainty around the estimate, so we defined a beta distribution with mean equal to the estimated probability of DGI among men from the IOM study, with the upper limit of the 95% UI equal to the estimate reported in the most recent study.<sup>10</sup>

Table S4.1 summarizes the estimated probabilities resulting from our review and synthesis of the evidence. In our probability trees, we allowed for all possible combinations of sequelae based on a simplifying assumption that the probabilities are independent, which leads to a multiplicative formulation for the joint probabilities.

**Table S4.1. Estimated probabilities of sequelae.**

Sequelae	Mean	Standard deviation	Uncertainty Interval	Distribution for uncertainty analysis
Chronic pelvic pain	0.261	0.0156	(0.231, 0.292)	Beta (alpha=203.608, beta=577.096)
Ectopic pregnancy	0.0708	0.0124	(0.0488, 0.0979)	Beta (alpha=29.048, beta=381.277)
Tubal infertility	0.169	0.0286	(0.117, 0.230)	Beta (alpha=28.186, beta=138.95)
Epididymitis	0.0422	0.0381	(0.00116, 0.137)	Beta (alpha=1.087, beta=24.614)
Disseminated gonococcal infection	0.0100	0.0014	(0.0075, 0.013)	Beta (alpha=48.08, beta=4,759)

## 2. Utilities

We used estimates of health state utilities (weights) derived from prior studies among experts, since we did not identify any estimates from the societal or patient perspective for men. Our method follows the same approach used in earlier studies, for example Tuite *et al.* (2012)<sup>20</sup> and Hu *et al.* (2004).<sup>21</sup>

Table S4.2 presents the utility estimates for women aged 15-24 years from expert and patient perspectives identified in the literature review. Estimates from the patient perspective result in higher utility values than those based on expert perspective.

**Table S4.2. Quality weight and duration of PID and its sequelae**

Health state	Utility (expert perspective) <sup>9</sup>	Utility (patient perspective) <sup>22</sup>	Duration (years) <sup>9,23-25</sup>
PID (inpatient treatment)	0.76	0.78	13 days (0.036)
PID (outpatient treatment)	0.63	0.82	10 days (0.027)
Chronic pelvic pain	0.60	0.76	5 years, to remaining lifetime (23.55)
Ectopic pregnancy (inpatient)	0.62	0.82	31 days (0.085)
Ectopic pregnancy (outpatient)	0.58	0.82	28 days (0.076)
Tubal infertility	0.82	0.84	5 years, to remaining lifetime (23.55)

\*Duration of remaining lifetime was estimated based on lifetable for females aged 15-24 years in 2015.<sup>25</sup>

## 3. Weighted utilities and durations of sequelae inpatient and outpatient care

There are multiple scenarios for inpatient and outpatient treatment for gonorrhoea sequelae of PID, EP and DGI in women and men, according to the IOM framework. For example, there were PID (salpingitis, perihepatitis) outpatient only, PID (salpingitis, perihepatitis) inpatient no surgery, PID (salpingitis, perihepatitis) inpatient with surgery, PID (salpingitis, perihepatitis) outpatient after inpatient, PID (bilateral salpingo-oophorectomy) inpatient, and PID (bilateral salpingo-oophorectomy) outpatient after inpatient. For simplification and matching evidence from other studies, we included inpatient treatment and outpatient treatment in our probability tree as weighted combination of the inpatient



and outpatient care scenarios. We illustrate the calculations for the example of inpatient treatment of PID in women, with information in Table S4.3. Table S4.3 column 1, 2, 3, and 4 are derived from the Table A17-2 disease scenarios for gonorrhea infection in women and men in the IOM framework.<sup>9</sup>

**Table S4.3. Weighted utilities and durations of inpatient treatment for PID.**

Scenarios	% cases among all scenarios <sup>9</sup>	Utilities <sup>9</sup>	Durations (years) <sup>9</sup>	% cases among PID inpatient only (weights)	Weighted durations	Weighted QALYs lost
PID (salpingitis, perihepatitis) inpatient no surgery	0.075	0.57	0.011	0.69	0.0076	0.0033
PID (salpingitis, perihepatitis) inpatient with surgery	0.025	0.46	0.0054	0.23	0.0012	0.00067
PID (salpingitis, perihepatitis) outpatient after inpatient	0.100	0.83	0.027	0.92	0.025	0.0042
PID (bilateral salpingo-oophorectomy) inpatient	0.008	0.40	0.0027	0.074	0.00020	0.00012
PID (bilateral salpingo-oophorectomy) outpatient after inpatient	0.008	0.76	0.027	0.074	0.0020	0.00048
Total	0.1008			1.00	0.036	0.0088

We first calculated the proportion of cases under each scenario among PID inpatient treatment only and reported in the fifth column in Table S4.3, as the second column represented the proportion of cases among all scenarios of gonorrhea and sequelae in the original IOM study. For example, proportion of PID (salpingitis, perihepatitis) inpatient no surgery =  $0.075 / (0.075 + 0.025 + 0.008) = 0.69$ . The total duration of PID inpatient care was calculated as a weighted duration of all these scenarios (column 6). That is,  $0.011 * 0.69 + 0.0054 * 0.23 + 0.027 * 0.92 + 0.0027 * 0.074 + 0.027 * 0.074 = 0.036$  (See also Table 4.1). Then weighted QALYs lost from each scenario were calculated in column 7, leading to a total QALYs lost from inpatient PID as 0.0088. The weighted utility of total inpatient PID treatment is  $1 - 0.0088 / 0.036 = 0.76$  (also reported in Table 4.1).

#### 4. Durations and discounting for long-term sequelae

Following conventions in an earlier study commissioned by the Institute of Medicine,<sup>9</sup> and in other similar studies, such as Regnier *et al.* (2014),<sup>26</sup> we applied time discounting to the durations of the two long-term sequelae (CPP and TI). Estimated durations for these two sequelae varied widely in the published literature. We therefore assumed a wide uncertainty range, between 5 years and remaining lifetime.<sup>9,23</sup> We also assumed that there was a five-year lag between infection and the occurrence of these sequelae.<sup>9</sup> When duration of CPP and TI were assumed to be remaining lifetime, we used a life table approach with age-specific mortalities to estimate discounted life expectancy for females in 2015, using a 3% annual discount rate. In a preliminary analysis, we compared results across race/ethnicity groups and over time, but given that differences in discounted life expectancy were relatively small in these comparisons, for parsimony we pooled across all race/ethnicity groups and used data from 2015 to apply to all years.<sup>25</sup>

In Table S4.4 we illustrate the calculations for the example of an infection among a woman aged 20 years in 2015. For each five-year age interval (beginning after the five-year lag from incidence), we computed the present value of the loss of quality-adjusted life-years during that interval. We first obtained the age-group-specific conditional probability of all-cause mortality (and its complement, survival) for women in 2015 from the National Center for Health Statistics.<sup>25</sup> We assumed that deaths occurred at the midpoint of each age interval up to 95-99 years and older. For the final, open interval (ages 100 years and more), we used estimated life expectancy at 100 years to define the average years lived in that open interval (2.284 years for females in 2015). The undiscounted total years lived by those who were alive at the start of an interval was calculated using the standard life table approach, accounting for years lived by those surviving the entire interval and years lived by those who died during the interval. For example, the undiscounted total years lived by those who survived and those who died between age 20-24 years was computed as  $(0.00232*2.5+(1-0.00232)*5)*1=4.994$  years. The undiscounted total years lived by those who survived or died between age 25-29 years was

computed as  $(0.00304*2.5+(1-0.00304)*5)*0.998=4.981$  years, and so on. We accounted for both condition-specific QALY adjustments as well as age-specific QALY adjustments (reflecting comorbidities) using a multiplicative model.

**Table S4.4. Discounted life years lived and QALYs lost due to CPP with lifetime duration, for a woman who acquires gonorrhea age at 20 years in 2015 and experiences CPP beginning at age 25 years.**

Age (x) at start of interval (years)	Conditional probability of dying between ages x and x+n	Probability of surviving from age 20 to exact age x	Years lived within each interval by each person alive at age 20	Discounted years lived within each interval by each person alive at age 20	Back-ground health-state weight <sup>27</sup>	Health state weight for CPP	Discounted QALYs lost due to CPP between ages x and x+n
20	0.00232	1.000	4.994	4.704	0.922	NA	NA
25	0.00304	0.998	4.981	4.030	0.922	0.600	1.486
30	0.00417	0.995	4.963	3.451	0.901	0.600	1.244
35	0.00549	0.990	4.939	2.954	0.901	0.600	1.065
40	0.00791	0.985	4.906	2.525	0.871	0.600	0.880
45	0.0120	0.977	4.857	2.155	0.871	0.600	0.751
50	0.0189	0.965	4.782	1.832	0.842	0.600	0.617
55	0.0275	0.947	4.671	1.549	0.842	0.600	0.522
60	0.0385	0.921	4.517	1.303	0.823	0.600	0.429
65	0.0566	0.886	4.304	1.085	0.823	0.600	0.357
70	0.0904	0.836	3.989	0.886	0.790	0.600	0.280
75	0.145	0.760	3.525	0.698	0.790	0.600	0.221
80	0.240	0.650	2.859	0.514	0.736	0.600	0.151
85	0.392	0.494	1.985	0.331	0.736	0.600	0.0976
90	0.590	0.300	1.059	0.169	0.736	0.600	0.0497
95	0.777	0.123	0.377	0.0587	0.736	0.600	0.0173
100	1.000	0.027	0.0628	0.0097	0.736	0.600	0.00286
<b>Total</b>	NA	NA	56.776	23.551	NA	NA	8.169

## 5. Probabilities of diagnosis and treatment

We derived year-, race/ethnicity-, and age-specific probabilities of diagnosis and treatment for asymptomatic gonococcal infections from a calibrated gonorrhea metapopulation-model.<sup>1</sup> In the study, heterosexual women and men were categorized as Non-Hispanic black, Hispanic, or “white and other,” whereas MSM were not stratified by race/ethnicity. Each population was stratified as two age groups (15-24 years and 25-39 years). As in the previous modeling study,<sup>1</sup> all symptomatic gonococcal

infections were assumed to be reported and have received treatment. That is, the probabilities of treatment for symptomatic gonococcal infections were assumed to be 100%.

Table S4.5 report model-estimated probabilities of diagnosis and treatment for asymptomatic gonococcal infections along the same strata.

**Table S4.5-(a). Estimated probabilities of diagnosis and treatment for asymptomatic gonococcal infections for heterosexual women by age and race/ethnicity over 2000-2015.**

Year	Non-Hispanic Black		Hispanic		White and Other	
	Mean	UI	Mean	UI	Mean	UI
Age 15-24 years						
2000	0.178	0.108, 0.260	0.166	0.084, 0.296	0.187	0.115, 0.278
2001	0.178	0.108, 0.260	0.166	0.084, 0.296	0.187	0.115, 0.278
2002	0.191	0.122, 0.276	0.180	0.098, 0.302	0.202	0.138, 0.290
2003	0.202	0.132, 0.286	0.190	0.109, 0.311	0.214	0.152, 0.302
2004	0.212	0.141, 0.297	0.198	0.116, 0.316	0.222	0.162, 0.308
2005	0.221	0.150, 0.305	0.204	0.120, 0.320	0.229	0.166, 0.313
2006	0.229	0.157, 0.314	0.208	0.121, 0.325	0.233	0.169, 0.318
2007	0.236	0.163, 0.319	0.211	0.123, 0.329	0.236	0.170, 0.323
2008	0.242	0.169, 0.328	0.212	0.124, 0.332	0.238	0.171, 0.325
2009	0.247	0.173, 0.337	0.213	0.125, 0.333	0.239	0.171, 0.325
2010	0.251	0.177, 0.343	0.213	0.125, 0.334	0.239	0.171, 0.325
2011	0.254	0.179, 0.345	0.213	0.125, 0.333	0.238	0.171, 0.325
2012	0.256	0.180, 0.349	0.212	0.124, 0.331	0.237	0.170, 0.324
2013	0.257	0.181, 0.351	0.211	0.124, 0.330	0.237	0.170, 0.321
2014	0.258	0.182, 0.351	0.210	0.123, 0.329	0.236	0.169, 0.320
2015	0.257	0.181, 0.349	0.210	0.123, 0.328	0.235	0.169, 0.319
Age 25-39 years						
2000	0.161	0.083, 0.278	0.123	0.056, 0.229	0.142	0.080, 0.225
2001	0.161	0.083, 0.278	0.123	0.056, 0.229	0.142	0.080, 0.225
2002	0.166	0.094, 0.276	0.146	0.069, 0.258	0.168	0.103, 0.247
2003	0.171	0.098, 0.278	0.163	0.080, 0.282	0.187	0.113, 0.269
2004	0.174	0.104, 0.279	0.175	0.087, 0.307	0.202	0.121, 0.291
2005	0.177	0.106, 0.279	0.185	0.091, 0.326	0.212	0.127, 0.307
2006	0.179	0.104, 0.281	0.191	0.094, 0.340	0.219	0.129, 0.321
2007	0.181	0.102, 0.286	0.196	0.096, 0.348	0.224	0.130, 0.327
2008	0.182	0.102, 0.289	0.198	0.097, 0.353	0.227	0.131, 0.330
2009	0.183	0.100, 0.293	0.199	0.097, 0.353	0.228	0.131, 0.333
2010	0.183	0.099, 0.294	0.199	0.097, 0.353	0.228	0.131, 0.334
2011	0.183	0.096, 0.293	0.199	0.097, 0.351	0.227	0.130, 0.333
2012	0.183	0.091, 0.292	0.197	0.097, 0.349	0.226	0.129, 0.331
2013	0.183	0.088, 0.292	0.196	0.097, 0.347	0.225	0.129, 0.330
2014	0.182	0.087, 0.291	0.195	0.096, 0.345	0.224	0.128, 0.329
2015	0.182	0.087, 0.290	0.195	0.096, 0.345	0.223	0.128, 0.329

**Table S4.5-(b). Estimated probabilities of diagnosis and treatment for asymptomatic gonococcal infections for heterosexual men by age and race/ethnicity, and for MSM by age, over 2000-2015.**

Year	Non-Hispanic Black		Hispanic		White and Other		MSM	
	Mean	UI	Mean	UI	Mean	UI	Mean	UI
<i>Age 15-24 years</i>								
2000	0.015	0.004, 0.032	0.039	0.016, 0.078	0.038	0.015, 0.076	0.138	0.070, 0.229
2001	0.015	0.004, 0.032	0.039	0.016, 0.078	0.038	0.015, 0.076	0.138	0.070, 0.229
2002	0.015	0.001, 0.034	0.045	0.021, 0.086	0.043	0.022, 0.079	0.136	0.067, 0.227
2003	0.017	0.001, 0.037	0.049	0.022, 0.100	0.047	0.023, 0.085	0.139	0.069, 0.230
2004	0.020	0.004, 0.042	0.052	0.021, 0.114	0.050	0.023, 0.091	0.145	0.072, 0.233
2005	0.024	0.008, 0.046	0.055	0.020, 0.122	0.052	0.022, 0.097	0.152	0.082, 0.239
2006	0.029	0.010, 0.052	0.057	0.019, 0.129	0.054	0.022, 0.103	0.162	0.093, 0.247
2007	0.034	0.013, 0.060	0.058	0.019, 0.133	0.055	0.022, 0.105	0.172	0.102, 0.262
2008	0.040	0.015, 0.070	0.058	0.019, 0.135	0.056	0.022, 0.106	0.182	0.110, 0.275
2009	0.046	0.017, 0.080	0.059	0.018, 0.136	0.056	0.022, 0.107	0.193	0.119, 0.289
2010	0.051	0.019, 0.090	0.058	0.018, 0.137	0.056	0.021, 0.106	0.202	0.123, 0.303
2011	0.056	0.021, 0.100	0.058	0.018, 0.136	0.056	0.021, 0.105	0.211	0.127, 0.314
2012	0.060	0.022, 0.108	0.058	0.018, 0.134	0.055	0.021, 0.104	0.218	0.131, 0.324
2013	0.062	0.022, 0.113	0.058	0.018, 0.133	0.055	0.021, 0.103	0.224	0.133, 0.333
2014	0.064	0.023, 0.116	0.057	0.018, 0.131	0.055	0.021, 0.101	0.228	0.134, 0.340
2015	0.064	0.023, 0.115	0.057	0.018, 0.131	0.055	0.021, 0.101	0.229	0.134, 0.342
<i>Age 25-39 years</i>								
2000	0.048	0.018, 0.121	0.055	0.016, 0.119	0.052	0.019, 0.101	0.222	0.121, 0.326
2001	0.048	0.018, 0.121	0.055	0.016, 0.119	0.052	0.019, 0.101	0.222	0.121, 0.326
2002	0.048	0.019, 0.115	0.054	0.017, 0.117	0.051	0.020, 0.095	0.220	0.122, 0.321
2003	0.048	0.020, 0.111	0.053	0.017, 0.113	0.051	0.020, 0.093	0.219	0.122, 0.326
2004	0.048	0.021, 0.104	0.053	0.018, 0.108	0.050	0.020, 0.090	0.218	0.117, 0.337
2005	0.047	0.021, 0.100	0.052	0.018, 0.101	0.049	0.020, 0.088	0.217	0.114, 0.342
2006	0.047	0.021, 0.099	0.051	0.018, 0.098	0.049	0.020, 0.083	0.216	0.110, 0.345
2007	0.047	0.020, 0.095	0.050	0.018, 0.094	0.048	0.020, 0.083	0.215	0.106, 0.351
2008	0.047	0.020, 0.095	0.049	0.018, 0.092	0.047	0.020, 0.086	0.214	0.105, 0.352
2009	0.046	0.018, 0.094	0.048	0.017, 0.091	0.047	0.020, 0.089	0.213	0.104, 0.352
2010	0.046	0.017, 0.095	0.048	0.017, 0.093	0.046	0.020, 0.095	0.213	0.104, 0.352
2011	0.046	0.017, 0.095	0.047	0.016, 0.095	0.045	0.020, 0.101	0.213	0.102, 0.351
2012	0.046	0.016, 0.094	0.047	0.016, 0.098	0.045	0.019, 0.105	0.212	0.102, 0.350
2013	0.045	0.016, 0.094	0.046	0.015, 0.099	0.045	0.018, 0.107	0.212	0.103, 0.348
2014	0.045	0.016, 0.094	0.046	0.015, 0.099	0.044	0.018, 0.108	0.212	0.104, 0.347
2015	0.045	0.015, 0.094	0.046	0.015, 0.099	0.044	0.018, 0.109	0.212	0.105, 0.347

## 6. Defining probabilities and costs of side effects from gonorrhea treatment

There are mild to moderate events and severe events as adverse reactions to gonorrhea treatment using Azithromycin (1g) and Ceftriaxone (250mg). Mild to moderate reactions usually include one or more of the following events: vomiting, reduction in hearing, dizziness or unsteadiness, skin rash,

injection pain, diarrhea and abdominal pain, fatigue, etc. Severe reactions usually include anaphylaxis, cholestatic jaundice, and interstitial nephritis. In our model, the probabilities of side effects were generated from beta distributions fitted in probabilities observed or estimated from a cost-effectiveness analysis and several meta-analyses of randomized controlled trials.<sup>28–31</sup> The costs of side effects were also driven from these studies,<sup>28–31</sup> with mean as average of costs estimated by each study, and 95% UI as symmetrically to the mean.

## Supplementary Results

### 1. Lifetime QALYs, costs per incident gonococcal infection by race/ethnicity

**Table S4.6. Estimated number of discounted lifetime QALYs lost and costs associated with gonorrhea, per incident infection in 2015 by race/ethnicity groups.**

	NHB	Hispanic	AI/AN	A/NH/OPI	NHW	MSM
<i>A. QALYs per incident infection in 2015, women</i>						
Mean	0.0921	0.0933	0.0935	0.0928	0.0931	NA
95% UI	(0.0221 to 0.217)	(0.0225 to 0.221)	(0.0226 to 0.222)	(0.0224 to 0.220)	(0.0224 to 0.221)	NA
<i>B. QALYs per incident infection in 2015, MSW and MSM</i>						
Mean	0.00200	0.00200	0.00200	0.00200	0.00200	0.00175
95% UI	(0.00153 to 0.00239)	(0.00153 to 0.00239)	(0.00153 to 0.00240)	(0.00153 to 0.00240)	(0.00153 to 0.00240)	(0.00125 to 0.00224)
<i>C. Costs per incident infection in 2015 (in 2018 USD), women</i>						
Mean	254	252	255	254	255	NA
95% UI	(109 to 459)	(109 to 460)	(109 to 462)	(109 to 459)	(109 to 460)	NA
<i>D. Costs per incident infection in 2015 (in 2018 USD), MSW and MSM</i>						
Mean	307	306	306	306	306	281
95% UI	(157 to 480)	(157 to 480)	(157 to 479)	(157 to 479)	(157 to 479)	(145 to 443)

\*NH White (Non-Hispanic White), A/NH/OPI (Asian or Native Hawaiian or Other Pacific Islander), AI/AN (American Indian or Alaska Native), NH Black (Non-Hispanic Black)

## 2. Population-level QALYs lost due to gonococcal infections

**Table S4.7. Total number of discounted lifetime QALYs lost, and discounted lifetime QALYs lost per 1,000 person-years associated with gonococcal infections that occurred in 2015, by sex, age-groups and race/ethnicity.**

Sex and age-group	Total number of QALYs lost	95% uncertainty interval	QALYs lost per 1000 person-years	95% uncertainty interval
<b>Women</b>				
Ages 15-24 y, all	40,073	(9,425 to 94,124)	1.87	(0.44 to 4.40)
Ages 25-39 y, all	13,221	(3,118 to 31,971)	0.41	(0.097 to 1.00)
<i>Ages 15-39 y, all</i>	53,293	(12,326 to 125,366)	1.00	(0.23 to 2.35)
Ages 15-24 y, NHB	18,694	(4,437 to 44,119)	5.59	(1.33 to 13.20)
Ages 15-24 y, Hispanic	4,074	(962 to 9,504)	0.88	(0.21 to 2.06)
Ages 15-24 y, AI/AN	1,090	(241 to 2,638)	5.05	(1.12 to 12.22)
Ages 15-24 y, A/NH/OPI	632	(140 to 1,529)	0.50	(0.11 to 1.20)
Ages 15-24 y, NHW	15,583	(3,447 to 37,711)	1.31	(0.29 to 3.16)
Ages 25-39 y, NHB	5,073	(1,185 to 12,009)	1.12	(0.26 to 2.64)
Ages 25-39 y, Hispanic	1,625	(387 to 3,890)	0.26	(0.061 to 0.61)
Ages 25-39 y, AI/AN	456	(103 to 1,120)	1.63	(0.37 to 4.01)
Ages 25-39 y, A/NH/OPI	215	(49 to 527)	0.086	(0.019 to 0.21)
Ages 25-39 y, NHW	5,852	(1,323 to 14,367)	0.32	(0.072 to 0.78)
<b>MSW</b>				
Ages 15-24 y, all	440	(302 to 623)	0.020	(0.014 to 0.029)
Ages 25-39 y, all	183	(124 to 261)	0.0059	(0.0040 to 0.0084)
<i>Ages 15-39 y, all</i>	623	(431 to 876)	0.012	(0.0082 to 0.017)
Ages 15-24 y, NHB	245	(173 to 343)	0.071	(0.050 to 0.099)
Ages 15-24 y, Hispanic	27.48	(14.86 to 45.24)	0.0056	(0.0030 to 0.0092)
Ages 15-24 y, AI/AN	7.26	(4.11 to 11.98)	0.032	(0.018 to 0.054)
Ages 15-24 y, A/NH/OPI	11.11	(6.29 to 18.34)	0.0085	(0.0048 to 0.014)
Ages 15-24 y, NHW	149	(84.41 to 246)	0.012	(0.0067 to 0.020)
Ages 25-39 y, NHB	100	(71.52 to 139)	0.024	(0.017 to 0.033)
Ages 25-39 y, Hispanic	15.29	(8.61 to 25.13)	0.0022	(0.0012 to 0.0036)
Ages 25-39 y, AI/AN	2.27	(1.20 to 3.76)	0.0082	(0.0043 to 0.014)
Ages 25-39 y, A/NH/OPI	4.65	(2.45 to 7.70)	0.0020	(0.0011 to 0.0034)
Ages 25-39 y, NHW	60.36	(31.84 to 99.99)	0.0032	(0.0017 to 0.0053)
<b>MSM</b>				
Ages 15-24 y, all	772	(363 to 1,286)	0.88	(0.41 to 1.47)
Ages 25-39 y, all	514	(224 to 860)	0.41	(0.18 to 0.68)
<i>Ages 15-39 y, all</i>	1,286	(632 to 2,074)	0.60	(0.29 to 0.97)

\*NH White (Non-Hispanic White), A/NH/OPI (Asian or Native Hawaiian or Other Pacific Islander), AI/AN (American Indian or Alaska Native), NH Black (Non-Hispanic Black), all is the total population of these five race/ethnicity groups.

### 3. Population-level costs due to gonococcal infections

**Table S4.8. Total discounted lifetime costs, and lifetime costs per 1,000 person-years associated with gonococcal infections that occurred in 2015 (in 2018 US dollars), by sex, age-groups and race/ethnicity.**

Sex and age-group	Total costs (2018 US dollars, millions)	95% uncertainty interval	Costs per 1000 person- years (2018 US dollars, thousands)	95% uncertainty interval
<b>Women</b>				
Ages 15-24 y	111.31	(48.44 to 202.84)	5.21	(2.27 to 9.49)
Ages 25-39 y	34.04	(13.90 to 64.65)	1.06	(0.43 to 2.02)
<i>Ages 15-39 y</i>	145.34	(62.79 to 265.92)	2.72	(1.18 to 4.98)
Ages 15-24 y, NHB	52.31	(22.18 to 97.46)	15.65	(6.64 to 29.16)
Ages 15-24 y, Hispanic	11.16	(4.69 to 20.86)	2.42	(1.02 to 4.52)
Ages 15-24 y, AI/AN	3.01	(1.28 to 5.63)	13.96	(5.94 to 26.07)
Ages 15-24 y, A/NH/OPI	1.75	(0.74 to 3.26)	1.37	(0.58 to 2.56)
Ages 15-24 y, NHW	43.08	(18.32 to 80.47)	3.61	(1.54 to 6.74)
Ages 25-39 y, NHB	12.94	(5.34 to 25.42)	2.85	(1.18 to 5.60)
Ages 25-39 y, Hispanic	4.17	(1.67 to 8.21)	0.66	(0.26 to 1.30)
Ages 25-39 y, AI/AN	1.18	(0.48 to 2.31)	4.24	(1.71 to 8.26)
Ages 25-39 y, A/NH/OPI	0.56	(0.22 to 1.09)	0.22	(0.090 to 0.43)
Ages 25-39 y, NHW	15.19	(6.12 to 29.60)	0.83	(0.33 to 1.61)
<b>MSW</b>				
Ages 15-24 y	68.06	(32.76 to 118.43)	3.15	(1.52 to 5.49)
Ages 25-39 y	28.89	(13.58 to 50.61)	0.93	(0.44 to 1.62)
<i>Ages 15-39 y</i>	96.94	(46.39 to 167.97)	1.84	(0.88 to 3.18)
Ages 15-24 y, NHB	38.02	(18.06 to 66.41)	11.04	(5.24 to 19.27)
Ages 15-24 y, Hispanic	4.24	(1.73 to 8.51)	0.86	(0.35 to 1.73)
Ages 15-24 y, AI/AN	1.12	(0.47 to 2.11)	5.00	(2.10 to 9.43)
Ages 15-24 y, A/NH/OPI	1.71	(0.72 to 3.23)	1.31	(0.55 to 2.47)
Ages 15-24 y, NHW	22.96	(9.65 to 43.29)	1.83	(0.77 to 3.44)
Ages 25-39 y, NHB	15.87	(7.61 to 26.98)	3.74	(1.79 to 6.36)
Ages 25-39 y, Hispanic	2.41	(0.97 to 4.81)	0.35	(0.14 to 0.70)
Ages 25-39 y, AI/AN	0.36	(0.14 to 0.71)	1.29	(0.50 to 2.56)
Ages 25-39 y, A/NH/OPI	0.73	(0.28 to 1.45)	0.32	(0.12 to 0.64)
Ages 25-39 y, NHW	9.51	(3.64 to 18.86)	0.51	(0.19 to 1.00)
<b>MSM</b>				
Ages 15-24 y	122.98	(48.85 to 244.58)	140.37	(55.75 to 279.14)
Ages 25-39 y	83.46	(31.68 to 167.38)	65.90	(25.02 to 132.17)
<i>Ages 15-39 y</i>	206.44	(82.30 to 409.62)	95.35	(38.41 to 191.18)

\*NH White (Non-Hispanic White), A/NH/OPI (Asian or Native Hawaiian or Other Pacific Islander), AI/AN (American Indian or Alaska Native), NH Black (Non-Hispanic Black), all is the total population of these five race/ethnicity groups.



## Supplementary Material References

1. Tuite AR, Rönn MM, Wolf EE, et al. Estimated Impact of Screening on Gonorrhoea Epidemiology in the United States: Insights From a Mathematical Model. *Sex Transm Dis*. 2018;45(11):713. doi:10.1097/OLQ.0000000000000876
2. Morre SA, van Den Brule AJC, Rozendaal L, et al. The natural course of asymptomatic Chlamydia trachomatis infections: 45% clearance and no development of clinical PID after one-year follow-up. *Int J Std Aids*. 2002;13:12–18.
3. Oakeshott P, Kerry S, Aghaizu A, et al. Randomised controlled trial of screening for Chlamydia trachomatis to prevent pelvic inflammatory disease: the POPI (prevention of pelvic infection) trial. *BMJ*. 2010;340(7752):903-903.
4. Østergaard L, Andersen B, Møller JK, Olesen F. Home Sampling versus Conventional Swab Sampling for Screening of *Chlamydia trachomatis* in Women: A Cluster-Randomized 1-Year Follow-up Study. *Clin Infect Dis*. 2000;31(4):951-957. doi:10.1086/318139
5. Rees E. The treatment of pelvic inflammatory disease. *Am J Obstet Gynecol*. 1980;138(7):1042-1047. doi:10.1016/0002-9378(80)91105-9
6. Scholes D, Stergachis A, Heidrich FE, Andrilla H, Holmes KK, Stamm WE. Prevention of Pelvic Inflammatory Disease by Screening for Cervical Chlamydial Infection. *N Engl J Med*. 1996;334(21):1362-1366. doi:10.1056/NEJM199605233342103
7. Price MJ, Ades AE, De Angelis D, et al. Risk of Pelvic Inflammatory Disease Following Chlamydia trachomatis Infection: Analysis of Prospective Studies With a Multistate Model. *Am J Epidemiol*. 2013;178(3):484-492. doi:10.1093/aje/kws583
8. Trikalinos TA. *PROBABILITY OF PELVIC INFLAMMATORY DISEASE IN CHLAMYDIA TRACHOMATIS INFECTION.*; 2018. [https://www.brown.edu/public-health/cesh/sites/public-health-cesh/files/STI\\_probabilities.pdf](https://www.brown.edu/public-health/cesh/sites/public-health-cesh/files/STI_probabilities.pdf).
9. Institute of Medicine (US) Committee to Study Priorities for Vaccine Development. *Vaccines for the 21st Century: A Tool for Decisionmaking*. (Stratton KR, Durch JS, Lawrence RS, eds.). Washington (DC): National Academies Press (US); 2000. <http://www.ncbi.nlm.nih.gov/books/NBK233313/>. Accessed September 21, 2019.
10. Birrell JM, Gunathilake M, Singleton S, Williams S, Krause V. Characteristics and Impact of Disseminated Gonococcal Infection in the “Top End” of Australia. *Am J Trop Med Hyg*. 2019;101(4):753-760. doi:10.4269/ajtmh.19-0288
11. Weström L. Effect of acute pelvic inflammatory disease on fertility. *Am J Obstet Gynecol*. 1975;121(5):707-713. doi:10.1016/0002-9378(75)90477-9

12. Ness RB, Soper DE, Holley RL, et al. Effectiveness of inpatient and outpatient treatment strategies for women with pelvic inflammatory disease: Results from the pelvic inflammatory disease evaluation and clinical health (peach) randomized trial. *Am J Obstet Gynecol.* 2002;186(5):929-937. doi:10.1067/mob.2002.121625
13. Weström L. Incidence, prevalence, and trends of acute pelvic inflammatory disease and its consequences in industrialized countries. *Am J Obstet Gynecol.* 1980;138(7):880-892. doi:10.1016/0002-9378(80)91077-7
14. Weström L, Bengtsson LP, Mårdh PA. Incidence, trends, and risks of ectopic pregnancy in a population of women. *Br Med J Clin Res Ed.* 1981;282(6257):15-18. doi:10.1136/bmj.282.6257.15
15. Weström E, Joesoef E, Reynolds E, Hagdu E, Thompson E. Pelvic Inflammatory Disease and Fertility: A Cohort Study of 1,844 Women with Laparoscopically Verified Disease and 657 Control Women with Normal Laparoscopic Results. *Sex Transm Dis.* 1992;19(4):185-192. doi:10.1097/00007435-199207000-00001
16. Weström L. Gynecological Chlamydial infections. *Infection.* 1982;10(1):S40-S45. doi:10.1007/BF01640713
17. Brunham RC, Maclean IW, Binns B, Peeling RW. Chlamydia trachomatis: Its Role in Tubal Infertility. *J Infect Dis.* 1985;152(6):1275-1282.
18. Paavonen J, Kousa M, Saikku P, Vartiainen E, Kanerva L, Lassus A. Treatment of nongonococcal urethritis with trimethoprim-sulphadiazine and with placebo. A double-blind partner-controlled study. *Br J Vener Dis.* 1980;56(2):101-104. doi:10.1136/sti.56.2.101
19. Geisler WM, Wang C, Morrison SG, Black CM, Bandea CI, Hook EW. The natural history of untreated Chlamydia trachomatis infection in the interval between screening and returning for treatment. *Sex Transm Dis.* 2008;35(2):119-123. doi:10.1097/OLQ.0b013e318151497d
20. Tuite AR, Jayaraman GC, Allen VG, Fisman DN. Estimation of the Burden of Disease and Costs of Genital Chlamydia trachomatis Infection in Canada: *Sex Transm Dis.* 2012;39(4):260-267. doi:10.1097/OLQ.0b013e31824717ae
21. Hu D, Hook EW, Goldie SJ. Screening for *Chlamydia trachomatis* in Women 15 to 29 Years of Age: A Cost-Effectiveness Analysis. *Ann Intern Med.* 2004;141(7):501. doi:10.7326/0003-4819-141-7-200410050-00006
22. Trent M, Lehmann HP, Qian Q, Thompson CB, Ellen JM, Frick KD. Adolescent and parental utilities for the health states associated with pelvic inflammatory disease. *Sex Transm Infect.* 2011;87(7):583-587. doi:10.1136/sextrans-2011-050187
23. Aledort JE, Hook EW, Weinstein MC, Goldie SJ. The Cost Effectiveness of Gonorrhoea Screening in Urban Emergency Departments. *Sex Transm Dis.* 2005;32(7):425-436.

24. Centers for Disease Control and Prevention NC for HS. Multiple Cause of Death, 1999-2017 on CDC WONDER Online Database. <https://wonder.cdc.gov/mcd-icd10.html>. Accessed September 22, 2019.
25. Arias E, Xu J. United States Life Tables, 2015. *Natl Vital Stat Rep.* 67(7). [https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67\\_07-508.pdf](https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67_07-508.pdf).
26. Régnier SA, Huels J. Potential impact of vaccination against *Neisseria meningitidis* on *Neisseria gonorrhoeae* in the United States: Results from a decision-analysis model. *Hum Vaccines Immunother.* 2014;10(12):3737-3745. doi:10.4161/hv.36221
27. Sullivan PW, Ghushchyan V. Preference-Based EQ-5D Index Scores for Chronic Conditions in the United States. *Med Decis Mak Int J Soc Med Decis Mak.* 2006;26(4):410-420. doi:10.1177/0272989X06290495
28. Magid D. Doxycycline Compared with Azithromycin for Treating Women with Genital *Chlamydia trachomatis* Infections: An Incremental Cost-Effectiveness Analysis. *Ann Intern Med.* 1996;124(4):389. doi:10.7326/0003-4819-124-4-199602150-00002
29. Bai Z-G, Bao X-J, Cheng W-D, Yang K-H, Li Y-P. Efficacy and safety of ceftriaxone for uncomplicated gonorrhoea: a meta-analysis of randomized controlled trials. *Int J STD AIDS.* 2012;23(2):126-132. doi:10.1258/ijsa.2009.009198
30. Lau C-Y, Qureshi AK. Azithromycin Versus Doxycycline for Genital Chlamydial Infections: A Meta-Analysis of Randomized Clinical Trials. *Sex Transm Dis.* 2002;29(9):497–502.
31. Ross JDC, Brittain C, Cole M, et al. Gentamicin compared with ceftriaxone for the treatment of gonorrhoea (G-ToG): a randomised non-inferiority trial. *The Lancet.* 2019;393(10190):2511-2520. doi:10.1016/S0140-6736(18)32817-4

## Chapter 5. Conclusion

### Summary of dissertation papers

This dissertation uses mathematical modelling to synthesize empirical evidence and evaluate the epidemiological and economic burden of specific chronic and infectious diseases in the United States. In the second chapter of this dissertation, "*Dynamic Modeling of Prevalence and Incidence Trends for Diabetes and Diabetes Diagnosis among Adults Aged 20 Years or Older, United States, 2000-2016*", I develop an age-stratified Markov model of undiagnosed and diagnosed diabetes and examine how the epidemiological burden of diabetes in the United States has changed over recent decades. The model estimates a 17-year period of decreasing prevalence of undiagnosed diabetes, increasing prevalence of diagnosed and all diabetes, and decreasing incidence of diabetes, rates of diabetes diagnosis, and incidence of diagnosed diabetes in the US adult population. This work provides the first estimates of the incidence of diabetes, and rates of diabetes diagnosis among the US adult population from nationally representative surveys.

In the third chapter, "*Risk score to predict cardiovascular disease (CVD) risk for patients with type 2 diabetes mellitus (T2DM) in the United States: a pooled analysis of prospective cohorts*", I develop a novel CVD risk prediction model specifically for populations with T2DM in the United States. The model is constructed as a multivariable risk factor model estimated using pooled data on fatal-plus-non-fatal CVD outcomes from 5 prospective cohorts. The model predicts a 10-year fatal-plus-non-fatal CVD risk for patients with T2DM in the US, with good discrimination and calibration performance. This performance is superior to the 2013 ACC/AHA Pooled Cohort Risk Equation, which underestimates 10-year atherosclerotic CVD risk in diabetic populations.

In the fourth chapter, "*Disparities in health and economic outcomes associated with *N. gonorrhoeae* infection in the United States: costs and quality-adjusted life-years lost in 2015*", I develop probability tree models that capture clinical outcomes of gonorrhea and sequelae, and use these models to

evaluate the health and economic burden of gonorrhea and disparities across race/ethnicity groups. I report population-level disease burden in terms of the discounted lifetime costs and quality-adjusted life-years (QALYs) lost due to incident infections acquired during 2015, disaggregated by sex, age, race/ethnicity, and for men who have sex with men (MSM). The findings suggest that the highest absolute burden of both QALYs and costs occurs in Non-Hispanic Black women, and the highest per-capita burden occurs in MSM and American Indian/Alaska Native women.

### **Policy implications and future research**

The results of this dissertation illustrate the use of mathematical modelling to synthesize observed data to infer epidemiological trends that are not always directly observable, to quantify the short-term and long-term health outcomes associated with disease and complications, and to measure disparities of health and economic burden of disease across population subgroups. The findings from our estimation of incidence trends for diabetes and diabetes diagnosis in the US adult population highlight the importance of continuing surveillance of trends in diabetes screening and testing and interventions to reduce diabetes incidence. Although the incidence of diabetes decreased, declining mortality among diagnosed diabetic populations, and high and increasing prevalence of all diabetes have resulted in continued high total number of diabetic individuals in the United States. In addition, according to our diabetes-specific CVD risk model, the estimated cardiovascular disease incidence and death rates among patients with type 2 diabetes in the United States may be higher than the estimates predicted and reported in the 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk. With decreasing all-cause and CVD-specific mortality but increasing long-term morbidity, the overall burden of diabetes and cardiovascular complications in the United States will remain high at least over the next decade. For these reasons, continued efforts to implement effective primary and secondary prevention for diabetes remain important public health priorities.

The third paper in this dissertation, while focusing on a different clinical condition, shares several key methodological themes with the first two papers. Our analysis of the health consequences and

economic costs related to gonococcal infection synthesizes evidence from various sources to answer broad questions on resource prioritization and planning, relevant to public health policies. Across all three studies, the findings have identified or can be used to identify subpopulations with high or increasing disease and complications that may have been under-appreciated in previous research. more specifically, findings from the first paper indicate an increasing trend in prevalence of undiagnosed diabetes in adults ages 20-49 years, with continuing increases likely. Patients with high CVD incidence and mortality over the next decade can be quantified when applying the CVD risk prediction model in the second paper to risk factor profiles of diabetic populations in the United States. Results from the third paper not only quantify the large total disease burden associated with gonorrhoea among non-Hispanic Black populations and among men who have sex with men, but also highlight the substantial relative burden among American Indian/Alaska Native populations. Identifying subpopulations with disproportionately large disease burden can provide evidence to enable decision making to target groups with the highest need.

In addition to informing decision-making and health policy, the mathematical models developed in this dissertation highlight gaps in information availability, such as limited surveillance data on changes in diabetes detection and screening, CVD incidence and mortality for high-risk diabetic populations, and disparities in gonorrhoea and sequelae diagnosis and treatment. By illuminating these limitations, modeling results can indicate directions for collecting new data to enable disease burden assessment. By presenting a comprehensive analysis framework of disease burden associated with gonococcal infection in the third study, we also provide guidance on how to collect and synthesize information that is most relevant to population disease burden and disparities evaluation.

In summary, this dissertation describes three evaluations of disease burden in the United States, in the areas of chronic and infectious disease, and considering both economic and epidemiological outcomes. In each study, modelling is used to synthesize a range of empirical data, describe the

relationship between these data and subsequent outcomes, and report results directly relevant to policy or clinical decision-making.