Three Essays on Analysis of U.S. Infant Mortality Using Systems and Data Science Approaches

Alireza Ebrahimvandi

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy
in
Industrial and Systems Engineering

Niyousha Hosseinichimeh, Chair
Navid Ghaffarzadegan
Zhenyu Kong
Konstantinos P Triantis

December 3, 2019

Blacksburg, VA

Keywords: infant mortality, preterm birth, risk factors, health disparities, predictive modeling, health interventions, data science, statistical analysis, machine learning, system dynamics, optimization, simulation, endogenous dynamics

Three Essays on Analysis of U.S. Infant Mortality Using Systems and Data Science Approaches

Alireza Ebrahimvandi

ABSTRACT

High infant mortality (IM) rates in the U.S. have been a major public health concern for decades. Many studies have focused on understanding causes, risk factors, and interventions that can reduce IM. However, death of an infant is the result of the interplay between many risk factors, which in some cases can be traced to the infancy of their parents. Consequently, these complex interactions challenge the effectiveness of many interventions. The long-term goal of this study is to advance the common understanding of effective interventions for improving health outcomes and, in particular, infant mortality. To achieve this goal, I implemented systems and data science methods in three essays to contribute to the understanding of IM causes and risk factors.

In the first study, the goal was to identify patterns in the leading causes of infant mortality across states that successfully reduced their IM rates. I explore the trends at the state-level between 2000 and 2015 to identify patterns in the leading causes of IM. This study shows that the main drivers of IM rate reduction is the preterm-related mortality rate. The second study builds on these findings and investigates the risk factors of preterm birth (PTB) in the largest obstetric population that has ever been studied in this field. By applying the latest statistical and machine learning techniques, I study the PTB risk factors that are both generalizable and identifiable during the early stages of pregnancy. A major finding of this study is that socioeconomic factors such as parent education are more important than generally known factors such as race in the prediction of PTB. This finding is significant evidence for theories like Lifecourse, which postulate that the main determinants of a health trajectory are the social scaffolding that addresses the upstream roots of health. These results point to the need for more comprehensive approaches that change the focus from medical interventions during pregnancy to the time where mothers become vulnerable to the risk factors of PTB. Therefore, in the third study, I take an aggregate approach to study the dynamics of population health that results in undesirable outcomes in major indicators like infant mortality. Based on these new explanations, I offer a systematic approach that can help

in addressing adverse birth outcomes—including high infant mortality and preterm birth rates—which is the central contribution of this dissertation.

In conclusion, this dissertation contributes to a better understanding of the complexities in infant mortality and health-related policies. This work contributes to the body of literature both in terms of the application of statistical and machine learning techniques, as well as in advancing health-related theories.

Three Essays on Analysis of U.S. Infant Mortality Using Systems and Data Science Approaches

Alireza Ebrahimvandi

GENERAL AUDIENCE ABSTRACT

The U.S. infant mortality rate (IMR) is 71% higher than the average rate for comparable countries in the Organization for Economic Co-operation and Development (OECD). High infant mortality and preterm birth rates (PBR) are major public health concerns in the U.S. A wide range of studies have focused on understanding the causes and risk factors of infant mortality and interventions that can reduce it. However, infant mortality is a complex phenomenon that challenges the effectiveness of the interventions, and the IMR and PBR in the U.S. are still higher than any other advanced OECD nation. I believe that systems and data science methods can help in enhancing our understanding of infant mortality causes, risk factors, and effective interventions.

There are more than 130 diagnoses—causes—for infant mortality. Therefore, for 50 states tracking the causes of infant mortality trends over a long time period is very challenging. In the first essay, I focus on the medical aspects of infant mortality to find the causes that helped the reduction of the infant mortality rates in certain states from 2000 to 2015. In addition, I investigate the relationship between different risk factors with infant mortality in a regression model to investigate and find significant correlations. This study provides critical recommendations to policymakers in states with high infant mortality rates and guides them on leveraging appropriate interventions.

Preterm birth (PTB) is the most significant contributor to the IMR. The first study showed that a reduction in infant mortality happened in states that reduced their preterm birth. There exists a considerable body of literature on identifying the PTB risk factors in order to find possible explanations for consistently high rates of PTB and IMR in the U.S. However, they have fallen short in two key areas: generalizability and being able to detect PTB in early pregnancy. In the second essay, I investigate a wide range of risk factors in the largest obstetric population that has ever been studied in PTB research. The predictors in this study consist of a wide range of

variables from environmental (e.g., air pollution) to medical (e.g., history of hypertension) factors. Our objective is to increase the understanding of factors that are both generalizable and identifiable during the early stage of pregnancy. I implemented state-of-the-art statistical and machine learning techniques and improved the performance measures compared to the previous studies. The results of this study reveal the importance of socioeconomic factors such as, parent education, which can be as important as biomedical indicators like the mother's body mass index in predicting preterm delivery.

The second study showed an important relationship between socioeconomic factors such as, education and major health outcomes such as preterm birth. Short-term interventions that focus on improving the socioeconomic status of a mother during pregnancy have limited to no effect on birth outcomes. Therefore, we need to implement more comprehensive approaches and change the focus from medical interventions during pregnancy to the time where mothers become vulnerable to the risk factors of PTB. Hence, we use a systematic approach in the third study to explore the dynamics of health over time. This is a novel study, which enhances our understanding of the complex interactions between health and socioeconomic factors over time. I explore why some communities experience the downward spiral of health deterioration, how resources are generated and allocated, how the generation and allocation mechanisms are interconnected, and why we can see significantly different health outcomes across otherwise similar states. I use Ohio as the case study, because it suffers from poor health outcomes despite having one of the best healthcare systems in the nation. The results identify the trap of health expenditure and how an external financial shock can exacerbate health and socioeconomic factors in such a community. I demonstrate how overspending or underspending in healthcare can affect health outcomes in a society in the long-term.

Overall, this dissertation contributes to a better understanding of the complexities associated with major health issues of the U.S. I provide health professionals with theoretical and empirical foundations of risk assessment for reducing infant mortality and preterm birth. In addition, this study provides a systematic perspective on the issue of health deterioration that many communities in the US are experiencing, and hope that this perspective improves policymakers' decision-making.

Acknowledgements

I have been blessed by the support of many great people over the years. I would like to start by thanking my smart, beautiful, fun and caring wife, Negar Darabi. We have been together throughout this journey. Negar has also been a busy PhD student in my grad life, however, she has always been supportive and helpful. I am eternally thankful to her. I am also thankful to my parents who have always supported and encouraged me to pursue higher education. I cannot thank them enough for their endless sacrifice and eternal support from thousands of miles away. I am blessed to have a wonderful family and I dedicate this work to them.

I would like to sincerely thank my advisor, Niyousha Hosseinichimeh. She has been a wonderful mentor and a friend. I cannot imagine obtaining my PhD without her support. Her calm, thoughtful, and supportive character has guided me through the most difficult times of my study. Our weekly meetings were like a breath of fresh air that would boost my energy and encouraged me to continue my research. She was so helpful, dedicated to conducting research, and inspiring that I will always remember her as a great role model. Her interesting and impactful research was helpful both in my personal and academic life. I had the great opportunity of starting my PhD as an assistant to Niyousha on an infant mortality project and after conducting such research, I have a different approach to life.

I would also like to sincerely thank my dissertation committee members for their insightful suggestions and guidance: Navid Ghaffarzadegan, Kostas Triantis, and James Kong. I was lucky to have Navid's insightful comments during my PhD and specifically for my last essay. He always made me think about important questions and problems in society. He also offered me innovative ways to tackle them. Navid is a nice, supportive, and helpful person that I will always remember as an exemplary figure. Kostas Triantis provided me great feedback and taught me a lot in approaching problems the right way, and explaining them in the best possible way, both in writing and presentation. These are great lessons and skills that I will carry forever. James Kong also provided me with great suggestions and insights about my second paper and my research benefited a great deal. Working with my committee members has been a pleasure, and I look forward to continuing to do so in the future.

I had the great opportunity of conducting research and interacting with many impactful scholars within and outside of the Virginia Tech community. I especially thank Mohammad Jalali and Navid Ghaffarzadegan for offering me a great opportunity of collaborative research that resulted in a turning point in my academic life. I would also like to thank Kenton Colvin from whom I have learned many teaching skills. I am also thankful to Majid Jaridi for teaching me both life and research lessons.

I am also thankful to Leah Johnson, Ebru Bish, Douglas Bish, Maury Nussbaum, Wafik Iskander, and Kimberly Ellis for what I have learned in their classes. My special thanks to Hannah Parks, who has always been supportive and helpful to me. She is a wonderful and kind person who went above and beyond to make sure that our research remains our only focus. I am also thankful to the supportive staff of the ISE department: Rhonda Hawley, Akiko Nakata, and Mike Montgomery for their support and always being available to help. I would also like to thank Alexa Garvoille and Lauren Garretson for their advice on professional writing. Their instruction made a big difference in my technical writing.

I have been lucky to have wonderful, smart, and fun friends and lab mates in my graduate studies. I would like to thank Arash Baghaei Lakeh, Ran Xu, Sarah Mostafavi, Andrew Thomas, Jose Guevara Maldonado, Maryam Andalib, Farshad Farkhondehmaal, Nastaran Khalili, Maede Maftouni, Hesam Mahmoudi, Ben Beach, Ngoc Nguyen, Hrayer Aprahamian, Saloumeh Sadeghzade, Hussein El Hajj, Saman Madinei, Ning-Yuan Liu, Taylan Topcu, and Tim Lutz. I should also thank Elina and Omeed as the two children of Niyousha! Elina had to attend many of our randomly scheduled meetings. She was so patient and quiet that I sometimes would forget that she was in the meetings!

Last but not least, I would like to thank my many wonderful friends here in the U.S. and in Iran. They have made my life lovely and fun.

Table of contents

List of Figures

List of tables

Chapter 1. Introduction

1.1 Problem context

The U.S. infant mortality rate (IMR) is 71% higher than the average rate for comparable countries in the Organization for Economic Co-operation and Development (1). High infant mortality and preterm birth rates (PBR) are major public health concerns in the U.S. A wide range of studies have focused on understanding the causes and risk factors of infant mortality and interventions that can reduce it (2-16). However, the infant mortality is a complex phenomenon that challenges the effectiveness of the interventions (17, 18). I believe that systems and data science methods can help in advancing our understanding of infant mortality causes, risk factors, and effective interventions.

In essay one, I exploit variations in infant mortality across all 50 states between 2000 and 2015 to identify causes that contributed to IMR reduction. I also examine factors associated with preterm related deaths at the state level. In this paper, the argument is that states with high infant mortality need to focus on reducing preterm-related deaths, as it is the largest contributor to the success of states with a high infant-death rate reduction. In essay two, I use statistical and machine learning techniques to enhance our understanding of the influential risk factors of preterm birth (PTB). In this study, I use the high-dimensional dataset of U.S. birth records in 2016 and combine it with two other major area-level datasets to increase the number of potential predictors of preterm birth that are present in early pregnancy—before the second trimester. Analyzing a national-level data with machine learning techniques increases the generalizability of the risk factors in the obstetric population, which has received little attention in the literature. A major finding of this study is that socioeconomic factors such as, parent education are more important than generally known factors such as race in the prediction of PTB. In the third essay, I take a more aggregate perspective to study the dynamics of population health that results in undesirable outcomes in major areas like infant mortality. I am particularly interested in understanding why some states are performing poorly despite having the best healthcare systems in the nation and allocating most of their resources to health. To do this, I use the system dynamics approach to capture interconnections between health and socioeconomic (SE) factors in a community in the long-term. I also consider the effects of resource allocation to health on SE status in the community. The novelty of this study lies in exploring the dynamic interactions of socioeconomic status, resource generation-depletion, and aggregate health outcomes in a community over time. More information about research contributions in each study are presented below.

1.2 Research contributions

1.2.1 Essay #1

In the first part of this study, I focused on the medical causes of infant mortality. There are more than 130 diagnoses—causes—for infant mortality. Therefore, tracking the trends of these causes for 50 states over a long period of time is very challenging. The goal was to find the causes that helped in the reduction of the infant mortality rates in certain states from 2000 to 2015. To achieve this goal, I defined three questions: (1) Which of the IMR subgroup(s) are most responsible for the reduction of IMR at the national level from 2000 to 2015?; (2) Which subgroup(s) are most responsible for the IMR reduction in states that have successfully reduced their IMRs from 2000 to 2015?; and (3) Are variations in teen pregnancy, multiple births, and prenatal care associated with state variations in the preterm-related mortality rate? To find the answers, I used statistical tools including significant test and fixed-effect regression model.

This study contributed to the research in infant mortality in several ways. First, it provided new insights into the impact of the different diagnoses that contributed to the reduction of infant mortality in successful states. Understanding drivers of IMR reduction in states that have achieved substantial rate reductions may help to improve high IMRs in other states. Many studies have focused on one state or a limited number of risk factors across states to investigate determinants of preterm births (19-25). However, there were no studies that compare preterm-related mortality rate (PMR) trends by states over time. In addition, it contributed to the literature by investigating whether the variations in teen pregnancy, prenatal care, and multiple births are associated with state variations in the PMR. This study also provided recommendations to policymakers in states with high infant mortality rates to leverage interventions targeting preterm-related mortalities such as reducing the percentage of pregnant women with inadequate care. This essay, co-authored with Niyousha Hosseinichimeh and Jay Iams, was published in the American Journal of Perinatology in 2018.

1.2.2 Essay #2

The first study showed that the successful states associated with IMR reduction have decreased their PMR significantly more than unsuccessful ones. In addition, a large number of studies in the literature focused on identifying preterm birth (PTB) risk factors in order to find possible explanations for consistently high rates of PTB in the U.S. Therefore, I defined the second study to investigate the risk factors of PTB in the largest obstetric population that has ever been studied in this field. The objective in the second study is to increase the understanding of the PTB risk factors that are both generalizable and identifiable during the early stages of pregnancy.

Past studies lack generalizability for several reasons (26, 27). First, some studies used a majority White population from only one geographical location to assess the PTB risk factors (26). The second shortcoming of the current frameworks for assessing the PTB risk is that about two-thirds of preterm deliveries happen to women with no risk factors (28). For example, a history of PTB is a significant predictor of a subsequent PTB (17, 29). However, this risk factor is not applicable to women without a prior birth (nulliparous). The nulliparous constitutes about one-third of the general obstetric population. Third, many of the proposed studies only considered the main effect of the individual risk factors of PTB while controlling for a limited number of confounding variables and interactions that are chosen manually (28, 30, 31).

This study builds on previous research in several innovative ways that creates new contributions to the overall literature of preterm birth. First, I applied state-of-the-art data science techniques on U.S. linked birth datasets that are representative of the national obstetric population in order to better understand the risk factors of PTB. I also included both nulliparous (mothers with no prior birth) and multiparous singleton pregnancies in the data to increase the generalizability of the results. Second, I only included the factors that are present early in pregnancy, because many interventions are only effective if administered before the 24th week of gestation (32). Third, this study provides the “level of importance” for each variable, which is the first time this comparison has been reported in a PTB study. This metric compares the relative predictive power of variables for the entire population. In addition, I report the effect sizes for each variable level, which are comparable to the traditional logistic regression results. Fourth, I use proper machine learning (ML) methods that have the capability of checking high-order interactions with minimal supervision. It is important to consider the interactions, because it enhances the capability of the model in capturing complex relationships. Combining ML with big data can increase our understanding of preterm birth risk factors, as it enables us to check high-order interactions between risk factors in the national obstetrical population.

In this study, I obtained the 2016 U.S. linked birth dataset and combined it with two other area-level datasets, Area Health Resources File and County Health Ranking, to increase the potential set of predictors for PTB. I then applied state-of-the-art ML techniques to a cohort of 3.4 million singleton deliveries for the first time. However, using the U.S. birth dataset had its own challenges, like the existence of anomalous observations and random errors. To mitigate this problem, I applied one of the advanced ML techniques, auto-encoders with deep neural nets, to perform data cleaning and preparation. The other strength of this study was the implementation of Bayesian Optimization in order to reach optimal hyper-parameters more efficiently. Building a predictive model using a large sample size provided insights into all possible individual risk factors and their interactions. The results of this study showed an improvement in the performance metrics compared to previous predictive models.

The results of our best models match the findings of previous studies. The variables like hypertension ("*hyper*"), interval since last live birth ("*interval*"), and a history of PTB ("*Previous_preterm*") are among the most important predictors of preterm birth, which is consistent with the findings of previous studies (17, 33). Also, the model shows that the 'effect size' of a *Previous_preterm* is the largest in the model. However, the variable importance plot (VIP) reveals a novel and insightful finding compared to the previous studies. While the VIP shows that a variable like *Previous_preterm* is still an important predictor for PTB, it suggests larger relative importance to factors like *hyper* or *interval* in the prediction of PTB in the general obstetric population. Another important finding of this study is that a known risk factor of PTB, race, loses its level of importance after I add socioeconomic factors like education level of parents and their interaction with race in the model.

1.2.3 Essay #3

The second study showed the important relationship of socioeconomic factors such as education level and major health outcomes like preterm birth. Short-term interventions that focus on improving the socioeconomic status of a mother during pregnancy have limited to no effect on birth outcomes. Therefore, we need to implement more comprehensive approaches and change the focus from medical interventions during pregnancy to the time where mothers become vulnerable to the risk factors of PTB. Hence, I use a systematic approach in the third study to explore the dynamics of health over time. This is a novel study, which enhances our understanding of the complex interactions between health and socioeconomic factors over time.

I explore why some communities experience the downward spiral of health deterioration, how resources are generated and allocated, how the generation and allocation mechanisms are interconnected, and why we can see significantly different health outcomes across otherwise similar states.

Previous studies offered explanations on why the U.S. health is in a poor and costly condition (34, 35). Many point to social and income inequality as the source of poor outcomes (36, 37). While some fault is associated with over-medicalization or private insurers for creating administrative burdens for providers (38, 39), others blame the unequal distribution of primary care providers that limits their ability in offering the highest-quality care (40). Some economists believe health drives socioeconomic status (41). Some social scientists look at the adverse health outcomes like overall mortality purely as a function of socioeconomic and behavioral factors that trigger physiological responses, resulting in poor functioning of the immune system in the long-run (42-44). Systems modelers tried to expand the boundaries of their model and include nonmedical variables that might affect health outcomes. Past system dynamics studies have investigated the impact of interventions on the prevalence and costs of chronic or acute diseases at the community and national levels (45-47). However, their model boundary is limited to the health system and does not include interactions of health with socioeconomic variables (48). In addition, resources are not generated endogenously.

These studies usually miss the link between the effect and interaction of resources allocated to healthcare and the socioeconomic status of the population in the long-term, especially for the disadvantaged population. In particular, these studies usually ignore the effects of resource constraints, time delays, and the endogenous dynamics of resource generation. System scientists have shown how complexities and time delays hinder our ability to reach correct conclusions about the impacts of our decisions on the whole systems (49-51). Reaching an optimal answer becomes particularly harder when there are various interacting agencies and actors involved who might have multiple and sometimes conflicting interests.

I use a systematic approach to bridge the gap in the literature and provide policymakers a comprehensive framework that can help them in making better and more robust decisions for advancing the health and well-being of their communities. In this novel approach, I consider the effect of health-related decisions on both the health outcomes and the social status of the population in the long-run. On the other hand, I include the effect of social status in shaping health outcomes. This framework is different from past studies because it adds the explicit connection and interaction of social status to the health outcomes and resource generation-depletion in a community. Overall, this systematic model combines these interconnected decisions to provide a new explanation associated with the attainment of health.

I use Ohio as our case study to ground our theoretical framework on empirical information. This improves our understanding of why such a state with one of the best healthcare systems in the U.S. struggles to improve the health outcomes despite high expenditures in medical care. My results identify the trap of health expenditure and how an external financial shock can exacerbate health and socioeconomic factors in such a society. I present how overspending or underspending in healthcare can affect health outcomes in a society in the long-term. The proposed framework supports critical thinking and more robust conclusions about how resource allocations are likely to affect the performance of the health system in the short and long term. I also discuss that the solutions for improvement in population health status does not necessarily lie in the healthcare system.

Overall, this dissertation contributes to a better understanding of the complexities in major health issues of the U.S. I provide health professionals with theoretical and empirical foundations of risk assessment for reducing infant mortality and preterm birth. In addition, this study provides a systematic perspective on the issue of health deterioration that many communities in the US are experiencing, and I hope that this perspective improves policymakers' decision-making.

References

1. How infant mortality rates in the United States compare to rates in other countries [Internet]. Peterson-Kaiser Health System Tracker. 2017 [cited Nov 12, 2017]. Available from: https://www.healthsystemtracker.org/chart-collection/infant-mortality-u-s-compare-countries/#item-infant-mortality-declined-u-s-comparable-countries.

2. Alio AP, Richman AR, Clayton HB, Jeffers DF, Wathington DJ, Salihu HM. An Ecological Approach to Understanding Black–White Disparities in Perinatal Mortality. Maternal and Child Health Journal. 2010;14(4):557-66.

3. McLeroy KR, Bibeau D, Steckler A, Glanz K. An Ecological Perspective on Health Promotion Programs. Health Education Quarterly. 1988;15(4):351-77.

4. Elder TE, Goddeeris JH, Haider SJ. Racial and ethnic infant mortality gaps and the role of socio-economic status. Labour economics. 2016;43:42-54.

5. Iams JD, Applegate MS, Marcotte MP, Rome M, Krew MA, Bailit JL, et al. A Statewide Progestogen Promotion Program in Ohio. Obstetrics & Gynecology. 2017;129(2):337-46.

6. Latendresse G. The interaction between chronic stress and pregnancy: preterm birth from a biobehavioral perspective. Journal of midwifery & women's health. 2009;54(1):8-17.

7. Lefmann T, Combs-Orme T. Prenatal stress, poverty, and child outcomes. Child and Adolescent Social Work Journal. 2014;31(6):577-90.

8. Class QA, Khashan AS, Lichtenstein P, Langstrom N, D'Onofrio BM. Maternal stress and infant mortality: the importance of the preconception period. Psychological science. 2013;24(7):1309-16.

9. Burguet A, Kaminski M, Abraham-Lerat L, Schaal JP, Cambonie G, Fresson J, et al. The complex relationship between smoking in pregnancy and very preterm delivery. BJOG: an International Journal of Obstetrics & Gynaecology. 2004;111(3):258-65.

10. Gage TB, Fang F, O'Neill E, Dirienzo G. Maternal education, birth weight, and infant mortality in the United States. Demography. 2013;50(2):615-35.

11. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. The Lancet. 2008;371(9608):261-9.

12. Katz J, Lee ACC, Kozuki N, Lawn JE, Cousens S, Blencowe H, et al. Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis. The Lancet. 2013;382(9890):417-25.

13. Mathews TJ, MacDorman MF, ME T. Infant mortality statistics from the 2013 period linked birth/infant death data set. National vital statistics reports. Hyattsville, MD: National Center for Health Statistics2015.

14. Iams JD. Identification of candidates for progesterone: why, who, how, and when? Obstetrics and gynecology. 2014;123(6):1317-26.

15. MacDorman MF, Mathews T. Recent trends in infant mortality in the United States. 2008.

16. Sommers BD, Baicker K, Epstein AM. Mortality and access to care among adults after state Medicaid expansions. The New England journal of medicine. 2012;367(11):1025-34.

17. Iams JD. Prevention of preterm parturition. New England Journal of Medicine. 2014;370(3):254-61.

18. Chen A, Oster E, Williams H. Why Is Infant Mortality Higher in the United States Than in Europe? American economic journal Economic policy. 2016;8(2):89-124.

19. Finer LB, Kost K. Unintended pregnancy rates at the state level. Perspectives on sexual and reproductive health. 2011;43(2):78-87.

20. Paul DA, Mackley A, Locke RG, Stefano JL, Kroelinger C. State infant mortality: an ecologic study to determine modifiable risks and adjusted infant mortality rates. Maternal and child health journal. 2009;13(3):343-8.

21. Reeve CL, Basalik D. A state level investigation of the associations among intellectual capital, religiosity and reproductive health. Intelligence. 2011;39(1):64-73.

22. Ananth CV, Joseph KS, Oyelese Y, Demissie K, Vintzileos AM. Trends in preterm birth and perinatal mortality among singletons: United States, 1989 through 2000. Obstetrics & Gynecology. 2005;105(5, Part 1):1084-91.

23. Hibbs S, Rankin KM, David RJ, Collins Jr JW. The Relation of Neighborhood Income to the Age-Related Patterns of Preterm Birth Among White and African-American Women: The Effect of Cigarette Smoking. Maternal and child health journal. 2016:1-9.

24. Siddiqi A, Jones MK, Bruce DJ, Erwin PC. Do racial inequities in infant mortality correspond to variations in societal conditions? A study of state-level income inequality in the US, 1992–2007. Social Science & Medicine. 2016;164:49-58.

25. Hosseinichimeh N, MacDonald R, Hyder A, Ebrahimvandi A, Porter L, Reno R, et al. Group model building techniques for rapid elicitation of parameter values, effect sizes, and data sources. System Dynamics Review. 2017;33(1):71-84.

26. Alleman BW, Smith AR, Byers HM, Bedell B, Ryckman KK, Murray JC, et al. A proposed method to predict preterm birth using clinical data, standard maternal serum screening, and cholesterol. American Journal of Obstetrics & Gynecology. 2013;208(6):472. e1-. e11.

27. Weber A, Darmstadt GL, Gruber S, Foeller ME, Carmichael SL, Stevenson DK, et al. Application of machine-learning to predict early spontaneous preterm birth among nulliparous non-Hispanic black and white women. Annals of epidemiology. 2018;28(11):783-9. e1.

28. Robinson JN, Norwitz E. Preterm birth: Risk factors, interventions for risk reduction, and maternal prognosis. https://wwwuptodatecom/contents/preterm-birth-risk-factors-interventions-for-risk-reduction-and-maternal-prognosis. 2019.

29. Martin JA, Hamilton BE, Osterman MJ, Driscoll AK, Drake P. Births: Final data for 2016. 2018.

30. He J-R, Ramakrishnan R, Lai Y-M, Li W-D, Zhao X, Hu Y, et al. Predictions of preterm birth from early pregnancy characteristics: born in guangzhou cohort study. Journal of clinical medicine. 2018;7(8):185.

31. Davey MA, Watson L, Rayner JA, Rowlands S. Risk-scoring systems for predicting preterm birth with the aim of reducing associated adverse outcomes. Cochrane Database of Systematic Reviews. 2015(10).

32. American College of Obstetrics and Gynecology. Practice Bulletin No. 130: Prediction and Prevention of Preterm Birth. Obstetrics & Gynecology. 2012;120(4):964-73.

33. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. The lancet. 2008;371(9606):75-84.

34. Berwick DM, Hackbarth AD. Eliminating waste in US health care. Jama. 2012;307(14):1513-6.

35. Schroeder SA. We can do better—improving the health of the American people. New England Journal of Medicine. 2007;357(12):1221-8.

36. Wilkinson RG, Pickett KE. The problems of relative deprivation: why some societies do better than others. Social science & medicine. 2007;65(9):1965-78.

37. Conrad P, Barker KK. The social construction of illness: Key insights and policy implications. Journal of health and social behavior. 2010;51(1_suppl):S67-S79.

38. Woolhandler S, Campbell T, Himmelstein DU. Costs of health care administration in the United States and Canada. New England Journal of Medicine. 2003;349(8):768-75.

39. Lantz PM, Lichtenstein RL, Pollack HA. Health policy approaches to population health: the limits of medicalization. Health Affairs. 2007;26(5):1253-7.

40. Wilson N, Couper I, De Vries E, Reid S, Fish T, Marais B. Inequitable distribution of healthcare professionals to rural and remote areas. Rural Remote Health. 2009;9(1060):1-21.

41. Swift EK. Guidance for the National Healthcare Disparities Report: National Academies Press; 2002.

42. Lantz PM, House JS, Lepkowski JM, Williams DR, Mero RP, Chen J. Socioeconomic factors, health behaviors, and mortality: results from a nationally representative prospective study of US adults. Jama. 1998;279(21):1703-8.

43. Szanton SL, Gill JM, Allen JK. Allostatic load: a mechanism of socioeconomic health disparities? Biological Research for Nursing. 2005;7(1):7-15.

44. Grech V. Evidence of socio-economic stress and female foeticide in racial disparities in the gender ratio at birth in the United States (1995–2014). Early human development. 2017;106:63-5.

45. Homer JB, Hirsch GBJAjoph. System dynamics modeling for public health: background and opportunities. 2006;96(3):452-8.

46. Milstein B, Homer J, Hirsch G. Analyzing national health reform strategies with a dynamic simulation model. American journal of public health. 2010;100(5):811-9.

47. Mahamoud A, Roche B, Homer J. Modelling the social determinants of health and simulating short-term and long-term intervention impacts for the city of Toronto, Canada. Social science & medicine. 2013;93:247-55.

48. Homer J, Milstein B, Hirsch GB, Fisher ES. Combined Regional Investments Could Substantially Enhance Health System Performance And Be Financially Affordable. Health Affairs. 2016;35(8):1435-43.

49. Sterman JD. Learning from evidence in a complex world. American journal of public health. 2006;96(3):505-14.

50. Brownson RC, Chriqui JF, Stamatakis KA. Understanding evidence-based public health policy. American journal of public health. 2009;99(9):1576-83.

51. Luke DA, Stamatakis KA. Systems science methods in public health: dynamics, networks, and agents. Annual review of public health. 2012;33:357-76.

Chapter 2. “Understanding State level Variations in U.S. Infant Mortality: 2000–2015

Authors: Alireza Ebrahimvandi, Niyousha Hosseinichimeh, Jay Iams

***Objective**:* To exploit state variations in infant mortality over time, identify diagnoses that contributed to reduction of the infant mortality rate (IMR) between 2000 and 2015, and examine factors associated with preterm related deaths at the state level.

Methods: Using linked birth-infant deaths period data files from 2000 to 2015, we examined patterns in the leading causes of infant deaths including (1) the preterm-related mortality rate (PMR), (2) the congenital malformation-related mortality rate (CMR), (3) the sudden infant death syndrome mortality rate (SMR), and (4) all other 130 major mortality causes (OMR). We compared these rates at both national and state levels in 2000 and 2015 to find reduction trends. Creating a cross-sectional time series of states' PMR data and some explanatory variables, we implemented a fixed-effect regression model controlling for infant, maternal, and institutional characteristics.

Results: We found substantial state-level variations in changes of the IMR (range= -2.87 to 2.08), PMR (range= -1.77 to 0.67) and the CMR (range= -1.23 to 0.59) between 2000 and 2015. Twenty-one states in which the IMR declined more than the national average of 0.99 (from 6.89 to 5.90) were labeled as *successful*. We also labeled 20 states that saw a decline in their IMR less than the national average as *unsuccessful*. In the successful states, we found a reduction in the PMR accounted for the largest decline in the IMR—0.90 fewer deaths per 1,000 live births—or six times more than the PMR decline (0.14) in unsuccessful states. Changes in the other subgroups did not differ significantly in successful and unsuccessful states. Regression results showed that the PMR is positively associated with inadequate care (P-Value <0.05). A one-percentage-point decline in the share of pregnant women with inadequate care is significantly associated with 0.011 fewer preterm-related deaths per 1,000 live births. The magnitude of this variable is small relative to the PMR mean. The percentage of teen pregnancies, multiple births, and pregnant women that smoke was not significantly associated with the state-level PMR.

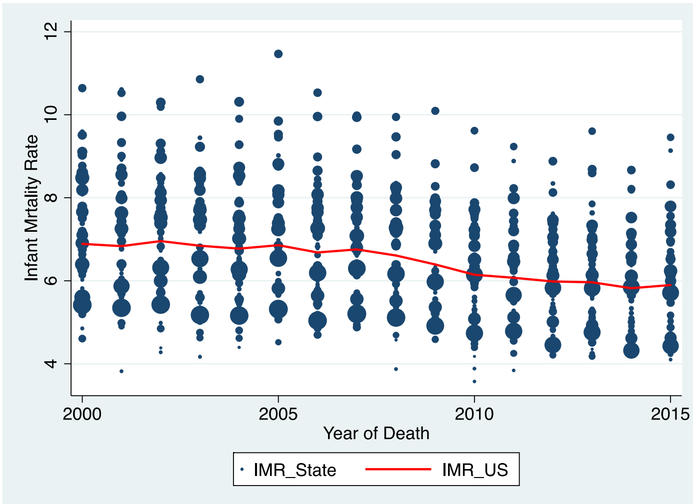
Policy Implications: Trends in the leading causes of mortality reduction are heterogeneous across states. States with high infant mortality need to focus on preterm-related deaths, as they are the largest contributor to the success of states with a high infant-death reduction. Although its

impact is not large, reducing the percentage of pregnant women with inadequate care is one of the mechanisms through which preterm-related deaths might decrease.

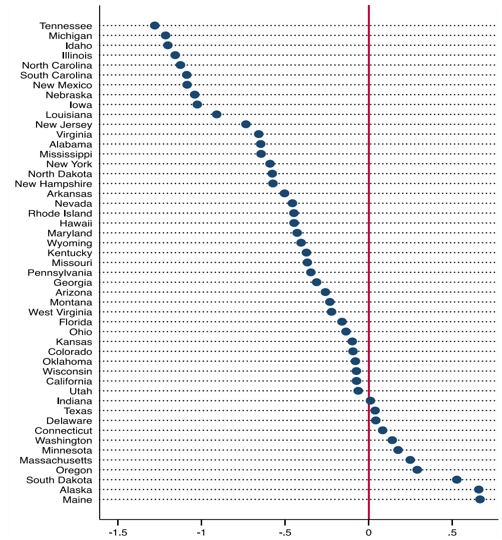
2.1 Background

The U.S. infant mortality rate (IMR) increased in 2015 after a decade of decline—from 2005 to 2014 (1). The U.S. IMR is 71% higher than the average rate for comparable countries in the Organization for Economic Co-operation and Development (OECD) (2). Although the U.S. rate declined by 14%, from 6.89 in 2000 to 5.90 in 2015, infant mortality has declined more slowly than in comparable countries. The IMR is defined as the number of infant deaths per 1,000 live births before their first birthday, and it is a representation of population health, quality of health care, and societal well-being (3).

Preterm birth, one of the most complicated factors associated with infant mortality (4, 5), has become a focus of research. Despite the reduction in recent decades, preterm birth in the U.S. (9.8%) remains higher than in European countries such as France (7.5%), Germany (8.3%), and Sweden (5.8%) (6). States vary substantially in terms of IMR size, trends, and preterm-related mortality rate (PMR). Figure 2-1.a shows the rate of infant mortality for 50 states from 2000 to 2015. The size of each dot is proportionate to the number of births in corresponding states and years. The continuous line shows the national average at each year. Four observations can be inferred from Figure 2-1.a: (1) states differ considerably in their IMR each year, (2) the IMR in larger states tend to be closer to the average, (3) states experienced a diverse change in their IMR from 2000 to 2015, and (4) on average, states experienced a decline in their IMR. However, the rates in Maine, South Dakota, and Texas are all higher in 2015 than in 2000. States also vary in mortality rates for babies born preterm. Figure 2-1.b shows the changes in PMR between 2000 and 2015 for each state. States such as Tennessee and Michigan reduced the PMR by more than 1.2 deaths per 1,000 live births, while the PMR increased by more than 0.5 deaths per 1,000 live births in Alaska, Maine, and South Dakota.



a. Infant mortality for each state from 2000 to 2015.



b. PMR reductions: per 1000 live births, between 2000 and 2015.

Figure 2-1. Trends of states' infant mortality and preterm-related mortality rates from 2000-2015.[1]

Note: The District of Columbia was dropped since the large change is attributed to demographic changes over the past decade.

The Centers for Disease Control and Prevention (CDC) reports national trends of infant mortality based on five leading causes that include: congenital malformations and chromosomal abnormalities, disorders related to short gestation and low birth weight, newborns affected by maternal complications of pregnancy, sudden infant death syndrome, and accidents. Change in state IMRs and variation across states have been reported from 2005 to 2014 (7), and infant mortality trends were also investigated in limited geographical locations. In five Southeastern states with the highest IMRs in the U.S., descriptive IMR statistics based on location, causes of death, infant and maternal characteristics were reported from 2005 to 2009 (8). Finally, a study of perinatal mortality in 50 states between 1989 and 2000 found that perinatal mortality among whites increased because of an increase in medically indicated preterm births (9). However, there are no studies comparing PMR trends by states over time. Many studies have focused on one state or a limited number of modifiable risk factors across states to investigate determinants of preterm births (9-14).

[1] Vermont is not shown because of data confidentiality.

Understanding drivers of IMR reduction in states that have achieved substantial rate reductions may help to improve high IMRs. In this study, we explore state variations of the IMR and categorize them as either successful—with an IMR reduction above the national average—or unsuccessful. We next examine the trends of four IMR subgroups—i.e., the PMR (preterm mortality rate): the CMR (congenital malformation-related mortality rate), the SMR (sudden mortality rate), and the OMR (other mortality rate)—in successful and unsuccessful states. Finally, we construct a cross-sectional time series data of the 50 states and run a fixed-effect model to examine the association between the PMR and explanatory variables. Specifically, our study aims to answer the following questions: (1) Which of the four IMR subgroup(s) are most responsible for the reduction of IMR at the national level from 2000 to 2015? (question 1); (2) Which subgroup(s) are most responsible for the IMR reduction in states that have successfully reduced their IMRs from 2000 to 2015? (question 2); and (3) Are variations in teen pregnancy, prenatal care, and multiple births associated with state variations in the PMR? (question 3).

2.2 Method

We obtained individual-level data from the linked birth-infant deaths period data files for the periods 2000 to 2015 (15) to compare states' performance in reducing the infant mortality rate and find the factors that might be associated with the PMR. We divided the IMR into four main subgroups: the preterm-related mortality rate (PMR), the congenital malformations-related mortality rate (CMR), the sudden mortality rate (SMR), and other mortality rates (OMR) that includes all other causes of infant mortality. The first subgroup, PMR, is created based on Callaghan and colleagues' criteria (16) and includes those with an underlying cause of death assigned to one of the following ICD-10 categories: K55.0, P010.0, P01.1, P01.5, P020, P02.0, P02.1, P02.7, P07, P10.2, P22, P27, P28.0, P28.1, P36, P52.0, P52.1, P52.2, P52.3, and P77.

The ICD-10 code only considers "Disorders related to short gestation and low birth weight, not elsewhere classified (P07)" as PMR. Callaghan and colleagues showed that the ICD-10 criteria for assigning causes of death do not capture the true number of prematurity-related infant deaths. Callaghan and colleagues categorized preterm-related deaths as those that met two criteria: (1) a biological connection with preterm birth, and (2) more than 75% of infants for such a given cause are born preterm (16).

Congenital malformations, deformations, and chromosomal abnormalities (ICD10: Q00–Q99) are classified under the CMR as the second subgroup. The third subgroup, the SMR, includes infants that died due to sudden infant death (SID) syndrome (ICD10: R95). All other causes are classified as OMR (Appendix A provides a detailed list of the causes and their related ICD10 code). These four categories—the PMR, CMR, SMR, and OMR—are mutually exclusive, which means that they should add up to the total IMR for each state and year (see Appendix B).

To identify the subgroup most responsible for the reduction of the IMR at the national level (question 1), we compared changes in the PMR, CMR, SMR, and OMR between 2000 and 2015. We calculated the percentage decline between 2000 and 2015 in each subgroup and compared the rates using the method described by Mathews and colleagues (17).

To investigate the subgroup most responsible for the IMR reduction in states (question 2), we assigned 41 states to one of two groups: successful and unsuccessful. We chose the national IMR reduction size between 2000 and 2015 (i.e., 0.99) as the cutoff point. A state with a reduction size of 0.99 or more was defined as "successful," while all others were "unsuccessful." We calculated the reduction in the PMR, CMR, SMR, and OMR for successful and unsuccessful states. We used a t-test to discern the statistical differences of mean-comparisons among states. To assess the appropriateness of using the pair-wised Student-t test, we used Bartlett's test for homogeneity of variances to compare mean the mortality rates of successful and unsuccessful states. We dropped nine states for the following reasons. First, some states, such as Massachusetts, had a low IMR in 2000, and a minimal opportunity to improve such rate. However, other states, such as Mississippi, experienced a large reduction in their IMR (1.19, a drop from 10.64 to 9.45), but still remained the worst state in the U.S. in the overall rate of infant death in 2015. Second, states in the two groups were not comparable because the mean IMR in 2000 was higher in successful states—7.53 versus 6.63. Nevertheless, we conducted the same analysis using the full sample of states and report the results in Appendix E .

In addition, we constructed a state-level, cross-sectional time series data of 50 states from 2000 to 2015 and used a fixed-effect, multiple linear regression analysis to investigate the association between the PMR and state-level explanatory variables (question 3). We accounted for serial correlation within states by clustering standard errors at the state level. The advantage of the fixed-effect model is that it controls for unobserved factors that are constant over time. Some of the important predictors of birth outcome were unchanged between 2000 and 2015. For example, although the IMR is higher among African-American women, their birth rates did not change in every state from 2000 to 2015. Our model controlled for such factors that remained the same during the study period. We selected our explanatory variables based on evidence in the literature and availability of data in the linked birth-infant deaths period data files. These variables include teen pregnancy, multiple births at infant level, prenatal care, and tobacco use during pregnancy.

Teen pregnancy is defined in each state as the percentage of mothers who were 19 or younger at the time of birth. Multiple births rate is a representation of live births for cases in which a mother delivered more than one baby in the same pregnancy. Tobacco use is the percentage of mothers who reported smoking at any time during pregnancy.

Prenatal care variables indicates the percentage of women in each state who received care based on the Kessner Index (18). This method defines three levels for prenatal care: adequate, intermediate, and inadequate. Since these three variables are perfectly correlated, we dropped the adequate category, and we interpreted the results with respect to it. The linked birth-infant deaths period data reported on these variables until 2002. For the period after 2003, we constructed variables based on Kessner's criteria presented in Appendix C . The Kessner Index has high predictive value for infant mortality and preterm birth (19, 20).

The PMR is reported as deaths per 1,000 live births, and all explanatory variables are in percentages. Thus, to interpret the regression results, we examined how many deaths per 1,000 could be avoided if an explanatory variable is changed by one percentage point (e.g., how many deaths per 1,000 live births would be avoided if we reduced teen pregnancy by one percentage point). Tobacco use and prenatal care were not available for some states between 2011 to 2015. We therefore ran two regressions. First, we regressed the PMR against teen pregnancy and multiple births using a cross-sectional time series of 50 states from 2000 to 2015 that included 800 data points (i.e., 50 states over 16 years). Second, the PMR and all explanatory variables were regressed in a data set of 46 states between 2000 to 2010 (505 data points). The number of observations in the second regression analysis was 505, because some explanatory variables, such as tobacco use in California, is not available (see Appendix D for a complete list of missing values). We also controlled for the effect of year by adding a continuous year variable. Calculations were performed on a Windows OS, using STATA/MP 14.0 software.

2.3 Results

Table 2-1 shows the national trends for the IMR values and its subgroups in the U.S. between 2000 and 2015. The IMR declined from 6.89 to 5.90 between 2000 and 2015, representing about one fewer death per 1,000 live births. The preterm-related mortality rate (PMR) was the largest subgroup of IMR in 2000, accounting for 2.46 deaths per 1,000 live births. The PMR subgroup was reduced to 2.07 deaths per 1,000 in 2015. In 2000, 1.41 deaths per 1,000 were caused by congenital malformations, which declined to 1.22 deaths per 1000 in 2015. The sudden mortality rate (SMR), the smallest subgroup of the IMR in 2000 and 2015, also declined from 0.62 to 0.39 deaths per 1,000, respectively. Other causes of deaths (OMR), the second largest subgroup, declined from 2.40 to 2.21 deaths per 1,000. The highest reduction, 0.39 deaths per 1,000 live births, occurred in the PMR subgroup. The percentage change in the SMR exceeded the other subgroups.

Table 2-1 Summary of infant mortality rates

Variable	2000	2015	Reduction Size	Reduction Rate	P-Value
IMR	6.89	5.90	1.00	14%	<0.001
PMR	2.46	2.07	0.39	16%	<0.001
CMR	1.41	1.22	0.19	14%	<0.001
SMR	0.62	0.39	0.23	36%	<0.001
OMR	2.40	2.21	0.19	8%	<0.001

Figure 2-2 depicts the subgroups' trends in 2000 and 2015 for the two groups of successful and unsuccessful states. Successful states reduced their rates significantly in each of the subgroups (P-value<0.05) while unsuccessful states only reduced their CMR significantly. When it came to comparing the *difference* in reduction sizes, the PMR was the only subgroup with a significantly different reduction in successful states versus unsuccessful. In other words, changes in the CMR, SMR, and OMR in successful states were not statistically different from changes in these subgroups in unsuccessful states.

In 2000, the mean of the IMR in successful states was 7.08, and it significantly declined to 5.47 (P-Value<0.001) in 2015, representing about 1.61 fewer deaths per 1,000 live births. The largest reduction is in the PMR in the successful states, a 0.90 reduction (P-Value <0.001), from 2.63 to 1.73. The reduction size in the CMR, SMR, and OMR are all significant and equal to 0.22 (from 1.44 to 1.22 with P-Value<0.001), 0.30 (from 0.63 to 0.33 with P-Value<0.001), and 0.18 (2.37 to 2.19 with P-Value=0.022), respectively. Fifty-six percent of the IMR reduction (0.90 out of 1.61) in successful states was due to PMR reduction.

In unsuccessful states, the IMR mean was 7.14 in 2000, which then declined by 0.65 (P-Value<0.001) per 1,000 live births and reached 6.49. A decrease in PMR from 2.48 deaths per 1,000 to 2.34—equivalent to a reduction size of 0.14—was not significant (P-Value=0.076). The CMR declined significantly by 0.27 from 1.53 to 1.26 (P-Value=0.003), the SMR reduced by 0.09

(P-Value=0.080) from 0.65 to 0.56, and the OMR declined by 0.16 (P-Value=0.352) from 2.49 to 2.33 (Figure 2-2).

The change in the PMR differed significantly between successful and unsuccessful states (P-Value<0.001). Successful states reduced their PMRs by 0.90, while the other group experienced a slight decrease of 0.14 in their PMRs over the period of 2000 to 2015. The reduction for the CMR, SMR, and OMR was not significantly different (P-Values are 0.115, 0.137, 0.157 respectively) between the two groups of states. Successful states reduced their CMRs by 0.22, which is not significantly different (P-Value= 0.115) than the reduction size in unsuccessful states—0.27. The SMR reduction size for successful states was 0.30, which is not significantly larger (P-Value=0.137) than the reduction size of unsuccessful states—0.09. The OMR reduction size was 0.18 in successful versus 0.16 in unsuccessful states, which was not significantly different (P-Value=0.157). Appendix F depicts the successful and unsuccessful states on the U.S. map.

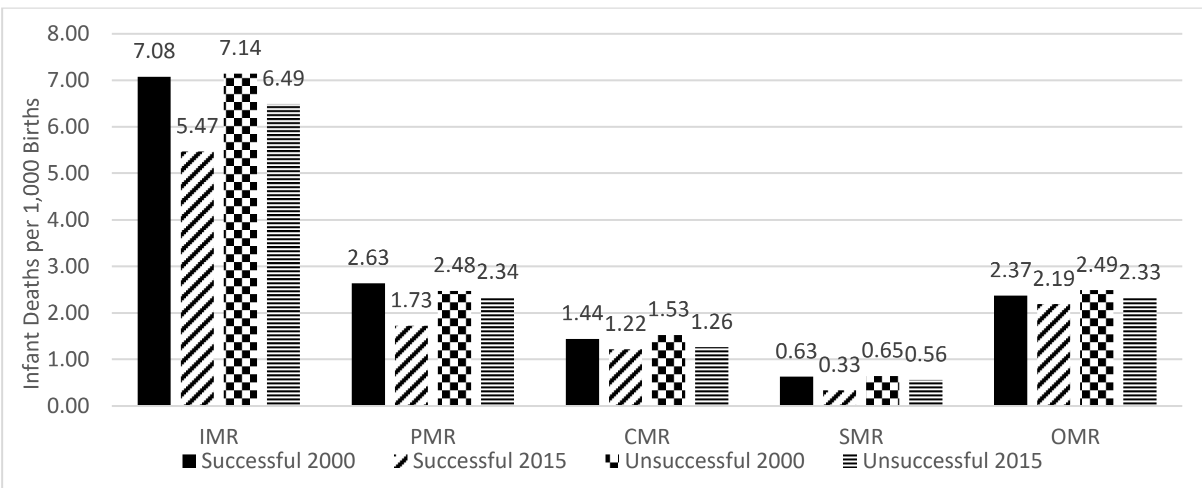


Figure 2-2. Performances of successful and unsuccessful states in reduction of IMR and its subgroups in 2000 and 2015 in selected states.

In summary, our analyses showed that the largest reduction size occurred in the PMR subgroup at the national level. Further, states with an IMR reduction above the national average—successful states—experienced a much larger decline in the PMR subgroups during the period of 2000 to 2015. Successful states had a higher PMR in 2000 compared to unsuccessful states, and reduced it substantially by 2015, while other subgroup reductions were not significantly different between the two categories of states. In the next step, we examine the association between the PMR and some of the determinants of the PMR reported in the literature in a data set of 50 states from 2000 to 2015.

2.3.1 Regression results for factors associated with PMR

Table 2-2 shows the summary statistics of each PMR and explanatory variables for selected years of the study period. These covariates are not weighted based on population of states and cannot be interpreted as national estimates. Few states do not report tobacco use, adequate, intermediate, and inadequate care in some years (see Appendix D). The first column shows the mean of each variable across the entire time frame. On average, 9.34% of pregnant women delivered their infant when they were teenagers, 3.35% of births were multiple births, 12.63% of pregnant women were tobacco users, 68.95% received adequate care, 20.72% received intermediate care, and 9.39% received inadequate care. The percentages of residents with different levels of care did not sum to 100% because of missing values between 2011 to 2015 in some states, such as Connecticut and New Jersey.

Table 2-2 Summary Data on Preterm-Related Mortality and State-Level Factors for 50 U.S. States from 2000 to 2015

Variable	Mean (SD)	Range	2000	2003	2006	2009	2012	2015
PMR per 1000 Live Birth	2.31 (0.74)	0.52-5.02	2.46	2.50	2.40	2.25	2.16	2.06
% of Teen Pregnancy	9.34 (2.98)	3.02-18.75	11.80	10.27	10.20	9.92	7.76	5.86
% of Multiple Births	3.35 (0.44)	2.08-4.90	3.07	3.34	3.37	3.42	3.39	3.43
% of Tobacco User	12.63 (5.04)	3.84-27.26	14.00	12.88	12.48	12.62	12.10	10.69
% of Adequate Care	68.95 (8.76)	39.35-87.38	72.86	73.31	68.93	66.86	64.92	67.76
% of Intermediate Care	20.72 (5.12)	9.13-46.95	18.08	17.98	18.31	22.11	23.76	21.50
% of Inadequate Care	9.39 (5.62)	1.39-41.00	5.07	8.71	8.93	11.04	11.32	10.73

Source: Linked Birth/ Infant Death Period Data, 2000-2015

Note 1: Data represented unweighted averages across states, and should not be interpreted as national estimates.

Note 2: Tobacco Use, Adequate, Intermediate, and Inadequate Care have missing data points (see Appendix D)

Table 2-3 shows the results for the two fixed-effect models. The first regression investigates the association between the PMR and two explanatory variables—teen pregnancy and multiple births—from 2000 to 2015 for 50 states, controlling for year effect and using the state-level fixed effect to control for differences between states. A one-percentage-point increase in the multiple birth rate is significantly associated with 0.18% more deaths per 1,000 live births. Teen pregnancy is not significant. The second regression reports the result when an institutional factor—prenatal care—and a maternal behavior characteristic—tobacco use—are added to the fixed-effect model. Since some states do not report tobacco use and prenatal care after 2010, the data set of the second regression include 46 states from 2000 to 2010. In this regression, the only significant explanatory variable is "inadequate care"—the percentage of pregnant women who do not receive adequate care in a state. A one-percentage-point decline in "inadequate care," from 9.39% (mean reported in Table 2-2) to 8.39%, is associated with 0.011% less preterm-related deaths per 1,000 live births. The magnitude of this variable is not large relative to the state-level PMR mean (2.31).

Table 2-3 Parameter Estimates for Fixed-Effect Regression Models of the Preterm-Related Mortality Rate

Variables	Regression 1 *(n=800)*		Regression 2 (*n=505*)	
PMR	Estimate	P-Value	Estimate	P-Value
% of Teen Pregnancy	0.009	0.653	0.002	0.966
% of Multiple Births	0.182	0.021**	0.117	0.429
% of Tobacco			-0.035	0.127
% of Intermediate Care			0.007	0.315
% of Inadequate Care			0.011	0.027**

**' Sig at 0.05

Note: Data represented unweighted averages across states, and should not be interpreted as national estimates.

2.4 Discussion

Despite the historic declines in the national infant mortality rate (IMR) during the past 16 years, significant variation in the reduction of infant mortality across states still exists. Some states have experienced an increase in their IMRs from 2000 to 2015. We categorized states into successful and unsuccessful groups based on their performance in reducing their IMRs and examined patterns in the leading causes of infant mortality across them. We found that decline in PMR was the major source of success in reducing the IMR during 2000-2015. A decline in the PMR accounts for 0.90 fewer deaths per 1,000—an approximate 56% reduction in the IMR—in successful states.

In addition, we found that preterm-related deaths fell more than other subgroups at the national level. The CDC listed congenital malformation as the first leading cause of infant death in the U.S. from 2005 to 2014 (7) and in 2013 (17). The CDC report, as well as other reports on the IMR (10, 21, 22), only considered short gestation and low birth weight in calculating the PMR. Callaghan and colleagues demonstrated that this classification of infant death does not capture the true magnitude of preterm-related deaths (16). We therefore used the criteria reported by Callaghan and colleagues to identify preterm-related deaths.

Despite statewide variations in the reduction of IMR and PMR, we found a homogenous improvement in the sudden mortality rate (SMR) across states. National-level initiatives such as the "Safe Sleep Campaign," which started in 1994, increased public awareness and led to lower rates of prone infant sleeping (23, 24). The rate of mortality related to sleeping dropped 36% on average in the U.S. between 2000 and 2015. Reduction of the other-related mortality rate (OMR) also did not differ in successful and unsuccessful states from 2000 to 2015. However, the decline in the OMR (8%) was substantially less than the 36% reduction seen in the SMR.

Prenatal care may improve birth outcomes even in low income and resource settings (25-27). Prior studies have found that prenatal care is associated with reduced PMRs. Mothers who did not receive any prenatal care between 2005 and 2009 in Southeastern states, which have the highest rates of infant mortality and lowest average incomes, had the highest IMRs (8). Although we found a significant association between access to prenatal care and the PMR, the magnitude of the coefficient was small. We used a fixed-effect regression analysis to identify variables that might be associated with PMR variations across states. Our analysis revealed that a one-percentage-point decline in mothers that received inadequate care led to a decrease in preterm-related deaths of 0.011 per 1,000 live births. The magnitude of this variable relative to the mean PMR is not large, suggesting that an increase in prenatal care as currently practiced would not lead to a state-level reduction in the PMR. The small-effect size of prenatal care might also be due to a limitation of data collection or lack of control for other key explanatory variables such as adhering to prevention programs.

Despite the emphasis on prenatal care, access to obstetric services is decreasing in the U.S. More than 9% of rural counties lost their obstetric services from 2004 to 2014, which made access to care harder for 28 million women of reproductive age (28). Our findings highlight the importance of eliminating an inadequacy of care. Models such as Coordinated Care Organizations significantly increase access to care with broad financing and delivery reforms that can reduce disparities in prenatal care and improve birth outcomes through a reduction in inadequate care (29). The adoption of expanded access to Medicaid has also been an effective policy. States that adopted Medicaid expansion have observed a reduction in their IMRs (30).

The risk factors, corresponding interventions, and causes of preterm mortality vary by target population. Iams and colleagues classified all possible interventions for a reduction of PMR risk factors into three main categories: primary (directed to all women), secondary (focusing on reducing existing risk), and tertiary (for improving preterm-infant outcomes) (31). Of these, the most improvement in the survival of preterm babies can be attributed to better neonatal care access (4). Obstetric interventions, such as antenatal corticosteroid treatment for women delivering preterm, also contribute to the decline in neonatal mortality.

Another preconceptional strategy to decrease infant mortality is the adoption of steps to reduce the risk of higher-order gestation (4). We found that the percentage of women with multifetal pregnancies is significantly and positively associated with the PMR in the first regression in which we controlled for teen pregnancy. After adding prenatal care variables and tobacco use to our model, multifetal births still correlate positively with the PMR, but statistical significance is lost, perhaps for two reasons. First, the second regression has fewer observations (for 46 states from 2000 to 2010) than the first regression (50 states from 2000 to 2015). Second, the first regression might have omitted the variable bias so that the P-value of multifetal pregnancy changed when we controlled for prenatal care and tobacco use. Multifetal pregnancy substantially increases the risk of preterm birth and infant death (17, 32, 33). In 2013, the IMR for twins (24.37) was more than 4 times the rate for single births (5.25) (17). The IMR for triplets (61.08) was nearly 12 times, and the rate for quadruplets (137.04) was 26 times the rate for single births. However, previous studies (17, 32, 33) did not use a multivariate analysis to disentangle the effects of multiple risk factors over time. Our analysis investigates the percentage impact of multifetal birth at the state level to examine the factors that might be associated with state variations in infant mortality while previous studies compared the IMR between singleton and multifetal births.

Cessation of cigarette smoking during pregnancy is recommended because it is a prevalent and preventable cause of infant mortality (34-36). It is estimated that 5% of infant mortality and 5.0%–7.3% of preterm-related deaths are attributable to maternal smoking (37, 38). Our findings are not consistent with these studies, which may be attributed to our different design. Traditional studies of smoking in pregnancy have examined cross-sectional data. We assessed the association of the PMR and smoking using cross-sectional time series data. Notably, between 2000 and 2004 in the U.S., smoking among childbearing age women decreased from 25.5% to 21.7%, while preterm birth rates increased from 11.6% to 12.5% (39). One study reported a model for the most important interventions for 39 countries and suggests that smoking cessation had the lowest contribution in reducing the rate of preterm deaths (40).

Our study was limited by the quality of CDC data that may vary by state and hospital over time (41). In addition, due to a lack of information, we could not control for some behavioral risk factors (e.g., drug abuse prevalence and behavioral factors, such as pregnant women's adherence to prevention programs), which may bias our regression results.

Overall, some states performed better in terms of reducing their IMRs over the past decade. Reducing preterm-related death was the biggest factor in states that have improved their infant deaths. It appears that state variations in reducing preterm-related deaths can be partially explained by better access to prenatal care, although the impact size is not large. More qualitative and in-depth analyses are needed to understand why some states have successfully reduced their preterm-related deaths better than others."

References

1. Xu J, Murphy SL, Kochanek KD, Arias E. Mortality in the United States, 2015. 2016.

2. How infant mortality rates in the United States compare to rates in other countries [Internet]. Peterson-Kaiser Health System Tracker. 2017 [cited Nov 12, 2017]. Available from: https://www.healthsystemtracker.org/chart-collection/infant-mortality-u-s-compare-countries/#item-infant-mortality-declined-u-s-comparable-countries.

3. Koh HK, Blakey CR, Roper AY. Healthy People 2020: a report card on the health of the nation. Jama. 2014;311(24):2475-6.

4. Iams J, Romero R, Culhane JF, Goldenberg RL. Primary, secondary, and tertiary interventions to reduce the morbidity and mortality of preterm birth. The lancet. 2008;371(9607):164-75.

5. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. The Lancet. 2008;371(9608):261-9.

6. MacDorman MF, Mathews T, Mohangoo AD, Zeitlin J. International comparisons of infant mortality and related factors: United States and Europe, 2010. 2014.

7. Mathews T, Driscoll AK. Trends in infant mortality in the United States, 2005-2014. Centers for Disease Control and Prevention, National Center for Health Statistics.(Hyattsville, Maryland: 2017); 2016.

8. He X, Akil L, Aker WG, Hwang H-M, Ahmad HA. Trends in infant mortality in United States: a brief study of the southeastern States from 2005–2009. International journal of environmental research and public health. 2015;12(5):4908-20.

9. Ananth CV, Joseph KS, Oyelese Y, Demissie K, Vintzileos AM. Trends in preterm birth and perinatal mortality among singletons: United States, 1989 through 2000. Obstetrics & Gynecology. 2005;105(5, Part 1):1084-91.

10. Paul DA, Mackley A, Locke RG, Stefano JL, Kroelinger C. State infant mortality: an ecologic study to determine modifiable risks and adjusted infant mortality rates. Maternal and child health journal. 2009;13(3):343-8.

11. Hibbs S, Rankin KM, David RJ, Collins Jr JW. The Relation of Neighborhood Income to the Age-Related Patterns of Preterm Birth Among White and African-American Women: The Effect of Cigarette Smoking. Maternal and child health journal. 2016:1-9.

12. Siddiqi A, Jones MK, Bruce DJ, Erwin PC. Do racial inequities in infant mortality correspond to variations in societal conditions? A study of state-level income inequality in the US, 1992–2007. Social Science & Medicine. 2016;164:49-58.

13. Hosseinichimeh N, MacDonald R, Hyder A, Ebrahimvandi A, Porter L, Reno R, et al. Group model building techniques for rapid elicitation of parameter values, effect sizes, and data sources. System Dynamics Review. 2017;33(1):71-84.

14. Hosseinichimeh N, Kim H, Ebrahimvandi A, Iams J, Andersen DJSR, Science B. Using a Stakeholder Analysis to Improve Systems Modelling of Health Issues: The Impact of Progesterone Therapy on Infant Mortality in Ohio. Systems Research and Behavioral Science. 2019.

15. Linked Birth / Infant Death Records 2000-2014, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program [Internet]. United States Department of Health and Human Services (US DHHS), National Center for Health Statistics (NCHS), Division of Vital Statistics (DVS). [cited Oct 2, 2017 5:01:19 PM]. Available from: http://wonder.cdc.gov/lbd-current.html

16. Callaghan WM, MacDorman MF, Rasmussen SA, Qin C, Lackritz EM. The contribution of preterm birth to infant mortality rates in the United States. Pediatrics. 2006;118(4):1566-73.

17. Mathews TJ, MacDorman MF, ME T. Infant mortality statistics from the 2013 period linked birth/infant death data set. National vital statistics reports. Hyattsville, MD: National Center for Health Statistics2015.

18. Kessner DM. Infant death: an analysis by maternal risk and health care: Institute of Medicine; 1973.

19. VanderWeele TJ, Lantos JD, Siddique J, Lauderdale DS. A comparison of four prenatal care indices in birth outcome models: comparable results for predicting small-for-gestational-age outcome but different results for preterm birth or infant mortality. Journal of clinical epidemiology. 2009;62(4):438-45.

20. Sonchak L. Medicaid reimbursement, prenatal care and infant health. Journal of health economics. 2015;44:10-24.

21. Brown Speights JS, Goldfarb SS, Wells BA, Beitsch L, Levine RS, Rust G. State-Level Progress in Reducing the Black–White Infant Mortality Gap, United States, 1999–2013. American journal of public health. 2017;107(5):775-82.

22. Tomashek KM, Qin C, Hsia J, Iyasu S, Barfield WD, Flowers LM. Infant mortality trends and differences between American Indian/Alaska Native infants and white infants in the United States, 1989–1991 and 1998–2000. American journal of public health. 2006;96(12):2222-7.

23. Moon RY. SIDS and other sleep-related infant deaths: expansion of recommendations for a safe infant sleeping environment. Pediatrics. 2011;128(5):e1341-e67.

24. Y Moon R, R Hauck F, R Colson E. Safe infant sleep interventions: What is the evidence for successful behavior change? Current pediatric reviews. 2016;12(1):67-75.

25. Jimenez J, Romero MI. Reducing infant mortality in Chile: success in two phases. Health affairs (Project Hope). 2007;26(2):458-65.

26. Quinlivan JA, Evans SF. Teenage antenatal clinics may reduce the rate of preterm birth: a prospective study. BJOG: An International Journal of Obstetrics & Gynaecology. 2004;111(6):571-8.

27. Hirai AH, Sappenfield WM, Ghandour RM, Donahue S, Lee V, Lu MC. The Collaborative Improvement and Innovation Network (CoIIN) to Reduce Infant Mortality: An Outcome Evaluation From the US South, 2011 to 2014. American journal of public health. 2018(0):e1-e7.

28. Hung P, Henning-Smith CE, Casey MM, Kozhimannil KB. Access To Obstetric Services In Rural Counties Still Declining, With 9 Percent Losing Services, 2004-14. Health affairs (Project Hope). 2017;36(9):1663-71.

29. Muoto I, Luck J, Yoon J, Bernell S, Snowden JM. Oregon's Coordinated Care Organizations Increased Timely Prenatal Care Initiation And Decreased Disparities. Health affairs (Project Hope). 2016;35(9):1625-32.

30. Bhatt CB, Beck-Sagué CM. Medicaid Expansion and Infant Mortality in the United States. American journal of public health. 2018;0(0):e1-e3.

31. Iams JD, Applegate MS, Marcotte MP, Rome M, Krew MA, Bailit JL, et al. A Statewide Progestogen Promotion Program in Ohio. Obstetrics & Gynecology. 2017;129(2):337-46.

32. Glinianaia SV, Rankin J, Sturgiss SN, Platt MPW, Crowder D, Bell R. The north of England survey of twin and multiple pregnancy. Twin Research and Human Genetics. 2013;16(1):112-6.

33. Martin JA, Hamilton BE, Osterman MJ. Three decades of twin births in the United States, 1980-2009. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics; 2012.

34. Burguet A, Kaminski M, Abraham-Lerat L, Schaal JP, Cambonie G, Fresson J, et al. The complex relationship between smoking in pregnancy and very preterm delivery. BJOG: an International Journal of Obstetrics & Gynaecology. 2004;111(3):258-65.

35. Phelan S. Smoking cessation in pregnancy. Obstetrics and gynecology clinics of North America. 2014;41(2):255-66.

36. Leung LW, Davies GA. Smoking Cessation Strategies in Pregnancy. Journal of obstetrics and gynaecology Canada : JOGC = Journal d'obstetrique et gynecologie du Canada : JOGC. 2015;37(9):791-7.

37. Dietz PM, England LJ, Shapiro-Mendoza CK, Tong VT, Farr SL, Callaghan WM. Infant Morbidity and Mortality Attributable to Prenatal Smoking in the U.S. American Journal of Preventive Medicine. 2010;39(1):45-52.

38. Salihu HM, Aliyu MH, Pierre-Louis BJ, Alexander GR. Levels of excess infant deaths attributable to maternal smoking during pregnancy in the United States. Maternal and child health journal. 2003;7(4):219-27.

39. Martin JA, Hamilton BE, Sutton PD, Ventura SJ, Menacker F, Kirmeyer S. Births: final data for 2004. National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System. 2006;55(1):1-101.

40. Chang HH, Larson J, Blencowe H, Spong CY, Howson CP, Cairns-Smith S, et al. Preventing preterm births: analysis of trends and potential reductions with interventions in 39 countries with very high human development index. Lancet (London, England). 2013;381(9862):223-34.

41. Martin JA, Wilson EC, Osterman MJ, Saadi EW, Sutton SR, Hamilton BE. Assessing the quality of medical and health data from the 2003 birth certificate revision: results from two states. National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System. 2013;62(2):1-19.

Chapter 3. Identifying the Early Signs of a Preterm Birth from U.S. Birth Records using Machine Learning Techniques

Alireza Ebrahimvandi, Niyousha Hosseinichimeh, Zhenyu James Kong

Objective—To increase understanding of preterm birth (PTB) risk factors that are present early in pregnancy by building a predictive model using 3.4 million U.S. birth records.

Study Design—The 2016 U.S. linked birth dataset is obtained and combined with two other area-level datasets, Area Health Resources File and County Health Ranking. Then, state-of-the-art machine learning (ML) techniques were applied to a cohort of 3.4 million singleton deliveries. The response variable is preterm birth, which includes both spontaneous and indicated PTB. Both women without a prior birth—nulliparous—and women with previous pregnancy—multiparous are included in the data to increase the generalizability of the predictive model. The results are compared using the area under the receiver operator characteristic curves (AUC) and interpreted using variable importance and partial dependence plots.

Results—The gradient boosting machines outperformed other ML techniques with an AUC of 75.58 percent for the validation dataset. The most important predictors of preterm birth are gestational and chronic hypertension, interval since last live birth, history of a previous preterm birth, and mother's prepregnancy body mass index. Parents education is the next influential variable in prediction of PTB. Race is still an important predictor of PTB, however, its relative importance decreases after considering the other risk factors and their interactions with race.

Conclusions—Application of ML techniques improved the performance measures in prediction of preterm birth. The results emphasize on the importance of socioeconomic factors like parent's education in a large cohort study. Therefore, more research is needed on the mechanisms in which the socioeconomic factors affect the biological responses.

3.1 Introduction

Preterm birth (PTB), which is defined as a birth before 37 weeks of pregnancy, is the leading cause of infant mortality in the U.S. and globally (1). In 2013, PTB accounted for 36% of U.S. infant deaths in their first year of life (2). In addition to the monetary cost of PTB, which exceeds 25 billion dollars annually, these babies may suffer from life-long deficiencies (3, 4). Many of the current interventions for reducing the likelihood of a preterm delivery like progesterone therapy are effective only if administered early—between 16 and 24 weeks of gestation—in the pregnancy (5). Therefore, it is critical to have an early assessment on the likelihood of a preterm delivery.

A noninvasive screening framework to quantify the risk among pregnant women in early pregnancy—before the second trimester—is valuable for several reasons. First, this allows for identification of at-risk women and the initiation of risk-specific treatments. Many of these treatments like vaginal progesterone supplementation should be administered before 24 weeks of gestation (6). Second, identifying the risk factors might help define a population useful for studying specific interventions. Third, the identification of risk factors might provide insight into the mechanisms of preterm birth which is still largely unknown (7).

A large and growing body of literature have focused on finding the individual risk factors of preterm birth (7-9). The most important individual risk factor for predicting a preterm delivery is a history of a previous PTB (both indicated and spontaneous) (10, 11). Race is another major predictor for a PTB. The preterm birth rate (PBR) among non-Hispanic (NH) Black is 52% more than NH White—13.77 vs. 9.04 respectively (12). Other significant risk factors of preterm birth include age (13), short cervix between 16 to 28 weeks of pregnancy (14), and chronic medical disorders like hypertension (15) or diabetes (16). Some studies attempted to increase generalizability of the risk factors by including large cohorts in their studies (17). Machine learning techniques are recently used in advancing the understanding of spontaneous PTB risk factors (18).

Despite the vast body of literature on the risk factors of PTB, very few interventions have been proven to effectively prolong gestational age in at-risk women (19). This is partly because two-thirds of preterm deliveries happen to the women with no risk factors (20). The current risk assessment in the obstetrical population shows limitation because of the low prevalence of individual risk factors in the general obstetric population (21). For example, the most important risk factor for preterm birth in singleton pregnancies is the history of a previous PTB (12, 21). However, the history of a previous PTB is not applicable to the women without a prior birth (nulliparous) which includes more than a third of the total births. Many of the proposed studies consider only the main effect of the individual risk factor of PTB while controlling for a limited number of confounding variables and interactions that were selected manually (9, 18, 20, 22). In sum, previous studies have not examined PTB risk factors that are present in early pregnancy for an all-embracing dataset controlling for diverse confounding factors and their interactions. To address this issue, we use the proper machine learning (ML) methods that has the capability of checking high-order interactions with minimal supervision. The importance of considering interactions is that it enhances the capability of the model to capture complex relationships. We also use a comprehensive dataset that can increase our understanding of preterm birth risk factors, as it enables us to check interactions of risk factors in the general obstetrical population.

In this study, we develop a predictive model with preterm birth as the response variable on the U.S. birth records. The dataset includes all deliveries which happened in the U.S. in 2016. We focus on identifying the risk factors of preterm birth (for both indicated and spontaneous), which is present early in pregnancy. We perform the analyses in four main steps. First, we apply several data preprocessing steps including auto-encoder deep neural network to eliminate anomalous observations and random errors. Second, we fit several machine learning techniques including logistic regression with elastic net regularization, random forest, and gradient boosting machines. Third, we apply Bayesian optimization to find the optimal hyperparameters for the ML methods more efficiently. Fourth, we use variable importance plot and partial dependence plot to interpret and explain the results.

3.2 Data

3.2.1 Data sources

We use the 2016 birth records provided by the CDC (23) and combine it with other data sources including the County Health Rankings, and the Area Health Resources File (see Table 3-1). All datasets were linked using a common geographical identifier, the FIPS county codes. This allows us to integrate and examine multiple influences on preterm birth. We performed the data cleaning and preparation in *STATA 14.0* and the processing has been coded in *R 3.6.0.*

Table 3-1 Datasets used

Dataset	Data Provider	Years
Linked Period Birth/Infant Death Records	CDC	2016
County Health Rankings	The County Health Rankings & Roadmaps (CHR)	2016
Area Health Resources File	U.S. Area Health Resources and Service Administration (AHRF)	2013-2016

3.2.2 Data preparation

The CDC reports about 380 variables for every birth that occurs in the U.S. Some of these variables are recoded a few times for the same feature. For example, mother's age is reported in four different types of recoding. We, therefore, only include the main variables. Appendix G shows the data preparation steps. The final set of variables that we use in the models are outlined in Appendix H . This set represents the variables of which we found at least one study in the literature that explained their importance in the incident of preterm birth. The merged dataset includes 3,664,509 observations with 77 variables.

The CDC dataset contains variables that are collected through self-reported survey at the time of birth from both practitioners and parents. Therefore, we need to ensure that the quality of an observation is good before using it in the model. To detect the data points that contains random errors or have significant difference from the other data points, we train an unsupervised autoencoder deep neural network (DNN) and then use it in an anomaly detection model (24). The autoencoder based anomaly detector (AD) outperforms the other methods that are based on distance, density, clustering, or the support vector machine (SVM) (25, 26). We also use grid search to find the parameters of the autoencoder that minimizes the mean square error (MSE). The final neural net that minimizes the MSE has three hidden layers with 20, 10, and 20 neurons, respectively. We also use hyperbolic tangent function ("Tanh") as the activation function for the neurons in the hidden layers. In addition, the model passes the training dataset in batches through the neural network, the number of epochs, 30 times.

Figure 3-1 shows the reconstruction mean square error (Reconstruction.MSE) for all of the observations in the dataset calculated by the anomaly detection method. The figure shows that the majority of observations has an MSE of 0.035 or less; therefore, we remove the observations with a greater value than this threshold. This will remove 5.07% of the records and at the same time, it keeps the proportion of singleton preterm birth at 7.73%, which is close to the initial distribution at 8.02%. The final dataset includes 3,610,827 observations with 77 variables (see Appendix H for more details).

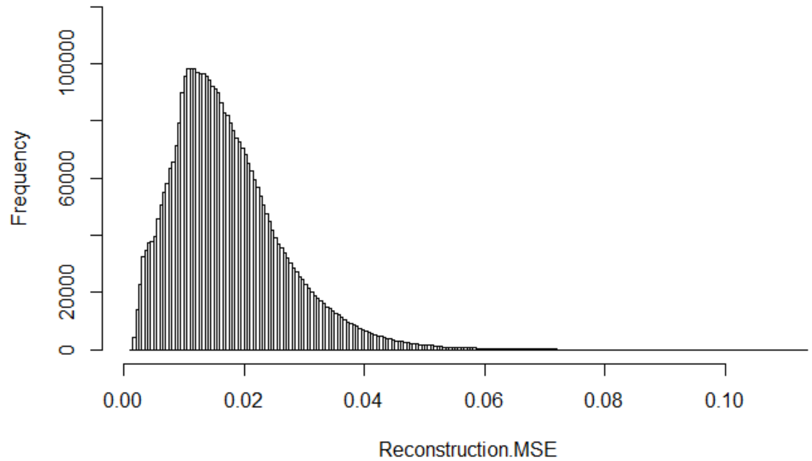


Figure 3-1 Histogram of Reconstruction.MSE for birth records after using autoencoders

3.2.3 Visualization

Data visualization is a challenging task in this study due to the large number of observations. To address this challenge, we used *Violin* graphs from the *ggplot2* package in *R* to plot the data and gain more information about the features and their relationship with preterm birth. Figure 3-2 is a sample of these graphs that shows kernel distribution of "History of Preterm" vs. "Gestational Age". This graph shows that mothers with a history of preterm delivery (represented as 1 in the graph) have a lower average gestational age (represented with a red dot) compared to those without such a history. Appendix I shows visualization of each variable.

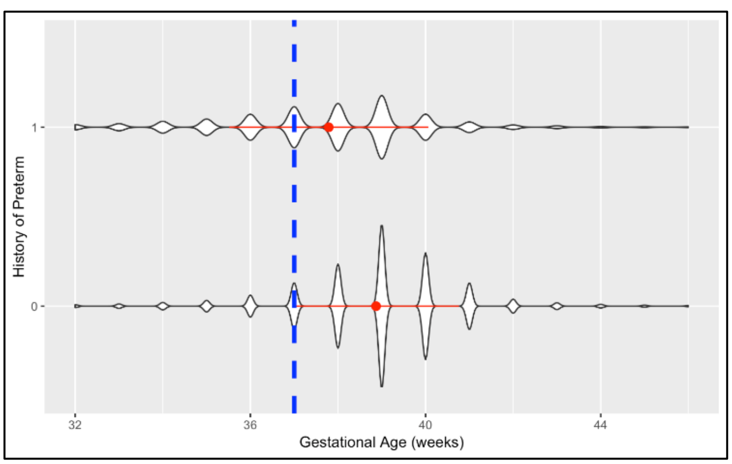


Figure 3-2 Data visualization; Gestational age distribution based on the history of preterm birth. The blue line shows the cutoff value for preterm birth.

3.3 Methods

There are many statistical and machine learning (ML) methods for classification purposes and choosing the right method is an important step of this process. One of the major criteria for choosing the machine learning techniques is their performance on the validation dataset. Our dataset has five characteristics that guide us in the selection of the methods.

First, the distribution of the response variable is imbalanced. Preterm birth in singleton pregnancies occurs only in eight percent of the deliveries and the remaining are full-term. One of the techniques to deal with this phenomenon is under or over sampling, or selecting the decision tree methods that can perform well in the prediction of the minority class, which is preterm birth. Second, many of the features like age and education have collinearity (Pearson's correlation coefficient= 0.41). This will limit the use of methods like logistic regression which has the assumption of little or no multicollinearity between independent features. Third, we are interested in finding the significant interactions among the variables. One of the best methods for learning the interactions with minimal supervision is decision trees (27). Fourth, our dataset has 3.6 million records with 77 variables, which limits the use of methods that are memory intensive like support vector machines. Fifth, the dataset has 20 categorical variables. This will limit the application of distance-based methods like K-Nearest Neighbor. Based on these five characteristics, we apply regularized logistic regression, random forest, gradient boosting machines (GBM), and LightGBM on our dataset.

3.3.1 Regularized Logistic Regression

The logistic regression method requires the observations to be independent of each other. Therefore, we need to remove "multiple births" from the dataset. Another assumption of the logistic regression is little to no collinearity among the features. To deal with this assumption and also avoid overfitting, we use regularization.

Least Absolute Shrinkage and Selection Operator (LASSO) and ridge regression are the two most common models for feature selection and handling the correlated features, respectively (28). These methods can also be implemented to reduce the prediction error variance. The LASSO regression penalizes the objective function by adding the l_1 penalty, which is the sum of the absolute coefficients $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$. Consider that we have data $(X^i, y_i),\ i = 1, 2, \ldots, N$, where

$X^i = \left(x_{i1}, ..., x_{ip}\right)^T$ and y_i are the independent and response for the i th observation, respectively. The objective function is to minimize

$$\sum_{i=1}^{N}\left(y_i - \sum_{j} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right| \qquad (1)$$

In equation (1), λ is the parameter that can control the sparsity—or complexity— of the model. A larger λ results in the greater the amount of shrinkage. The ridge regression has the same structure, except that it penalizes the second norm (l_2) in the objective function, which is he sum of the squares of the model coefficients: $\left\|\beta\right\|_2^2 = \sum_{j=1}^{p}\beta_j^2$.

The two penalties have different effects in the presence of correlated variables. The l_2 penalty shrinks coefficients toward each other simultaneously. On the other hand, the l_1 penalty tends to select only one of the correlated coefficients and sets the other one to zero. To keep the balance between these two behaviors, we use elastic net regularization. The elastic net combine the two penalties with argument α, which is a value between [0,1], in the same objective function (28). This will result in a balanced sparsity and stability in the final model. To tune the hyperparameters and select the best α and λ, we use a grid search.

3.3.2 Decision Trees

Decision tree (DT) is another classification technique for building predictive models. This machine learning technique involves dividing the feature space into sub-regions $R_j, j = 1, 2, ..., J$, where it starts with a single node, which branches into J terminal nodes of the trees (27). Then, it fits a simple model, like constant γ_j in each region j as

$$x \in R_j \Rightarrow f\left(x\right) = \gamma_j$$

Decision tree methods are simple and useful for interpretation (27). In addition, they are non-parametric methods that can handle both categorical and non-categorical variables. We use three types of decision trees in this paper for building the predictive models, random forest (RF), gradient boosting machines (GBM), and light GBM. We select these methods because they have been widely used in the classification of healthcare related problems and they outperform the other DT methods.

3.3.2.1 Random forest

The main idea of random forest (RF) is bagging. Random forest, which is also called wisdom of crowds, is the procedure of first building a large number of trees $\{T(x,\Theta_k), k=1,\ldots\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and then aggregating their votes for the popular class at input x (29). This method has two important parameters for tuning, the number of trees and the depth of each tree that govern the model complexity and accuracy.

Random forest has three characteristics that are particularly useful in the classification of preterm birth in this paper. First, random forest generates variable importance plots that are useful in exploring the independent effect of predictors and making inferences. Second, it can work with multicollinearity in the data and explore the effect of many different interactions depending on the depth of trees. Third, the method does not overfit and the predictions become more stable if we grow sufficiently large number of trees and perform cross-validation (27, 29). Despite the useful features of random forest, gradient boosting machines (GBM) performs better in the presence of imbalanced data.

3.3.2.2 Gradient boosting machines

Boosting is the algorithm of combining many simpler models to fit a highly accurate model (16, 30). This method is similar to random forest or bagging methods in being a committee method. However, these committee of weak learners improve in each iteration consecutively, and the members cast a weighted vote. These weights are adjusted in a way that the errors in the previous iterations get larger weights.

To illustrate the mathematical steps in a simple way, consider the current estimate of GBM for input x as a sum of trees, or weak learners

$$f_M(x) = \sum_{m=1}^{M} T(x;\Theta_m),$$

At each step in a stage-wise procedure, the purpose is to find the best parameters, $\hat{\Theta}_m$, by solving an optimization problem to minimize the loss function L

$$\hat{\Theta}_m = \arg\min_{\Theta_m} \sum_{i=1}^{N} L\left(y_i, f_{m-1}\left(x_i\right) + T\left(x_i;\Theta_m\right)\right)$$

This optimization is for the region set and constants $\Theta_m = \left\{R_{jm}, \gamma_{jm}\right\}_1^{Jm}$ of the next tree, considering the current model as $f_{m-1}(x)$. Defining the regions is difficult, but if we consider a given region R_{jm}, finding the optimal constants in each region, γ_{jm} is straightforward. We also grow multiple parallel classification trees sequentially and at the end, take the popular vote as the estimate for each input x.

In this paper, we use two different methods of gradient boosting machines for finding the best regions and constants, gradient descent and light GBM method (16, 31). The gradient descent has long been used in predictive modeling in the field of healthcare (16). In 2014, XGBoost was introduced as a new GBM that was one of the fastest algorithms among gradient boosted trees with a run time ten times faster than previous popular methods on a single machine (32). In 2017, Ke, Meng (31) introduced the light GBM, which returns more accurate outcomes in a shorter training time.

The iterative nature of the gradient boosting machines (GBM) makes it a better classifier for the purpose of this paper, which is increasing the performance of classifier in prediction of the preterm birth as the minority class while keeping the accuracy in prediction of full-term deliveries high. In addition to the good performance, the GBM provides both variable importance plot (VIP) and partial dependence plots (PDP) that are useful tools for interpretation of the results.

Two of the major problems in the training process of the decision trees are overfitting and long training times. As the dimension of the problem increases, the computation resources needed to train the model increases and finding the optimal hyperparameters become harder. Using a subset of observations or features is one of the techniques for reducing the training time, which also reduce the overfitting to the noise and improve generalization. The other common approaches to reduce the training time and prevent overfitting are random and grid search. However, using these methods are not efficient enough. Instead, we couple Bayesian optimization with the ML performance measures to reduce training time by only running the model sequentially on a set of hyperparameters that have the potential to improve the outcomes. It is shown that Bayesian optimization can reach to the same level of accuracy obtained by random or grid search in less iterations (33-35).

3.3.3 Handling missing observations and categorical variables

Handling missing data points and categorical variables in both methods are based on the capabilities of algorithms. Logistic regression needs to have a value for all of the data points or it will drop any observation with a missing value. Therefore, we use mean of each variable for imputing the missing values. For the decision trees, however, the missing data points carries some type of information and the method will keep such observations and consider them as one category.

Many of the variables in our dataset are categorical. The typical method to handle categorical variables is to create *N* new columns for categorical features with *N* levels. This method is called creating dummy variables or hot-encoding. However, we use a more efficient and interpretable method for handling the categorical variables. Instead of hot-encoding, we use a method in which strings are internally mapped to integers and splits are done over these integers.

3.4 Model comparisons and interpretations

3.4.1 Performance metrics

The area Under the receiver operating characteristic Curve (AUC) is a measure to assess the ability of a classifier to distinguish between true positives and negatives in the binary classifications. To calculate AUC, one needs to plot the True Positive Rate (TPR) against the False Positive Rate (FPR) where TPR is on y-axis and FPR is on the x-axis for all of the cutoff thresholds between 0 and 1. The area under this curve is AUC as a number between 0.5 and 1, where 1 is a perfect classifier and 0.5 as a classifier with no better prediction power than random guessing.

Our dataset is imbalanced with a majority of full-term deliveries; therefore, a model with high true negative rate (Specificity) may result in high AUC. However, the more important measure for this study is the true positive rate (Sensitivity or Recall), which is represented as the imbalanced response of preterm versus full-term in this study. Therefore, we use AUC under the precision and recall graph, which puts more emphasis on the positive class—preterm birth, in the model development step (36).

3.4.2 Variable importance and Partial Dependence plot

Variable importance is defined as the sum of all the improvement in the split-criterion that one variable creates in the forest. For example, in the split-criteria of "AUC", the importance of a variable is the percentage point lost if the variable is dropped. To simplify the interpretation, we use a scale of zero to one for showing the scaled importance of predictor variables.

To get the 'effect size' of each variable on the response, we use partial dependence plots (PDP). This is a useful tool in our study, particularly because we consider high-order interactions between our independent variables. Partial dependence plot returns the marginal 'effect size' of each variable on the response after accounting for the effect (average) of other responses: $\overline{f_s}\left(X_s\right)=\frac{1}{N}\sum_{i=1}^{N} f\left(X_s, x_{ic}\right)$.

Where X_c and X_s complement the set of X, and $\left\{x_{1c}, x_{2c}, ..., x_{Nc}\right\}$ are the values of X_c occurring in the training dataset of X.

It is important to note that the PDP does not ignore the effect X_c. The latter case can be estimated by $\tilde{f}_s\left(X_s\right)=\frac{1}{N}\sum_{i=1}^{N} f\left(X_s, x_{ic} \middle| X_s\right)$. The quantities $\overline{f_s}$ and $\tilde{f}_s$ will be the same only if the two events of c and s are independent, which is an unlikely situation.

3.5 Results

We randomly separated 75% of the data for training set and the remaining 25% for validation purposes. The performance metrics are reported for the test set that are not part of the training process. The number of cross-validations for the methods are five-fold.

3.5.1 Study design

In order to assess the performance of different machine learning algorithms and have a fair comparison, the hyperparameters of each method needs to be at their optimal point. All of the hyperparameter are optimized and evaluated using *H2O 3.26.0.2* package, an open source ML package. The *H2O* package runs the algorithms multi-threaded that reduces the total run time significantly. To prevent overfitting and reducing run-time, we also use early stopping methods. The model stops if it cannot improve the performance metric more than 1e-4 after 5 rounds. It takes about an hour to train a GBM model on a system equipped with a Core i7 2.50 GHz processor, and a 32.0 GB memory, which has an *Ubuntu 18.04.3* operating system.

We used grid search to find the best hyperparameters of logistic regression and random forest. However, we used Bayesian optimization (BO) to find the hyperparameters of the gradient boosting machines (GBM) and the lightGBM. The BO reduced the total required runs to reach these results in a way that a random grid search with 200 total runs returns the same performance as a BO with 40 initial runs and 67 subsequent iterations.

The parameters for Logistic Regression with Elastic Net regularization (LR-EN) are set as $\alpha = 0.25$, and $\lambda = 2.125E-4$ after performing a grid search. The results of Bayesian optimization for tuning the parameters of Gradient Boosting Machines return 480 decision trees (*ntrees*) with a learning rate of $\eta = 0.04$ and an annealing rate of 0.99. The maximum depth is 13 for each tree. This means that each tree checks up to 13 interactions among variables. Each tree is trained on a random sample of observations, n = 0.55* N, and each split of the tree is performed on a random sample of features, p = 0.80*M. The optimization result for LightGBM returns *ntrees* =280, $\eta = 0.008$, and maximum depth of 14. We also used “Lossguide” for the grow policy, “dart” for booster type, and “histogram” for tree method in the LightGBM method. For a detailed list of the hyperparameters, see Appendix J .

3.5.2 Results of the machine learning algorithms

Table 3-2 provides the performance metrics for each method. Recall, specificity, and accuracy are a function of the cut-off threshold. Therefore, we report these metrics corresponding to the threshold that returns the highest mean per class accuracy for all of the methods. Logistic regression with elastic net regularization (LR-EN) and random forest return very close testing and training AUC, which shows that they do not overfit to the noise. However, their AUC metrics are less than the gradient boosting machines (GBM) and the LightGBM on both testing and training datasets. The LightGBM returns the highest testing AUC at 75.91%. We pick the GBM as the best model for the prediction of preterm birth for two reasons. First, it returns a slightly higher recall (TPR) at 64.82% while maintaining the specificity at a comparable rate (73.01%) with LightGBM (73.93). Second, the GBM is better for interpretations, because it generates the variable importance plots that shows the contribution of each variable to the performance metric.

Table 3-2 Performance metrics for machine learning models

Method	Train AUC (%)	Test AUC (%)	Recall (%)	Specificity (%)	Accuracy (%)
LR-EN	66.59	66.61	51.98	71.68	70.22
RF	70.24	70.78	57.36	73.01	71.78
GBM	**77.94**	**75.58**	**64.82**	**73.01**	**72.37**
LightGBM	78.34	75.91	62.24	73.93	72.99

3.5.3 Comparison with other studies

Our study is unique because of the high-dimensional dataset that we use; therefore, there are few similar studies for comparison purposes. Weber, Darmstadt (18) developed their model on a high dimensional dataset with 1000 initial features and 2.7 million observations. However, they developed their predictive model for the early spontaneous preterm birth, which happens at a much lower rate of 1.02% compared to the singleton preterm deliveries at 7.63% in our study. Another study by Alleman, Smith (17) has the closest setup in terms of developing the predictive model for singleton pregnancies, but has a smaller dataset compared to our study.

Table 3-3 shows the comparison between the performance of our best GBM with the most relevant preterm birth studies. The criteria for including a paper in the comparison table is that it has to either use a data with large sample size that includes demographical information as predictors or use machine learning techniques for the model development. We report the sample size, prevalence of the positive class, test AUC, recall, and specificity for each study. As can be seen in Table 3-3, our best GBM model outperforms the frameworks in these studies by improving the AUC more than 5%, 9%, and 13% compared to the work of Goodwin, Iannacchione (37), Alleman, Smith (17) and Weber, Darmstadt (18), respectively. The improvement in the combined AUC, recall, and accuracy stems from pre-processing steps that remove anomaly and noise removal, regularization methods, an optimized set of hyperparameters, and the superior ability of the GBM algorithms in extraction of high-level features in the data.

Table 3-3 Performance comparison with the most related studies.

Model	Method	Sample size (n)	Prevalence of Positive Class (%)	Test AUC (%)	Recall (%)	Specificity (%)
Goodwin et al., 2002	Neural nets, Stepwise LR	19970	22.20	72.00	NR	NR
Vovsha et al., 2014	SVM with Radial Basis kernel	3002	NR	NR	57.60	62.10
Alleman et al., 2014	LR	2509	7.50	69.50**	31.20	90.60
Weber et al., 2018*	Super learner (Combination of RF, lasso, ridge)	336,214	1.02*	67.00	62.00	65.00
Best model in this study	GBM	3,610,827	7.73	75.58	64.82	73.01

*Early (before 32 weeks) spontaneous preterm

** Training AUC

NR= Not reported, LR= Logistic regression, RF= Random forest, SVM= Support vector machine

3.5.4 Interpretations

Figure 3-3 shows the scaled importance of the top 15 variables in the prediction of preterm birth in the obstetric population (See Appendix K for more details). Hypertension ("hyper"), interval since last live birth ("interval"), and a history of PTB ("Previous_preterm") are the most important predictors of a preterm birth. Mothers' prepregnancy BMI is also an important predictor of preterm birth. Figure 3-3 shows this interesting result that race has less relative importance when we consider other factors like parent's education, age, and adequacy of care during pregnancy.

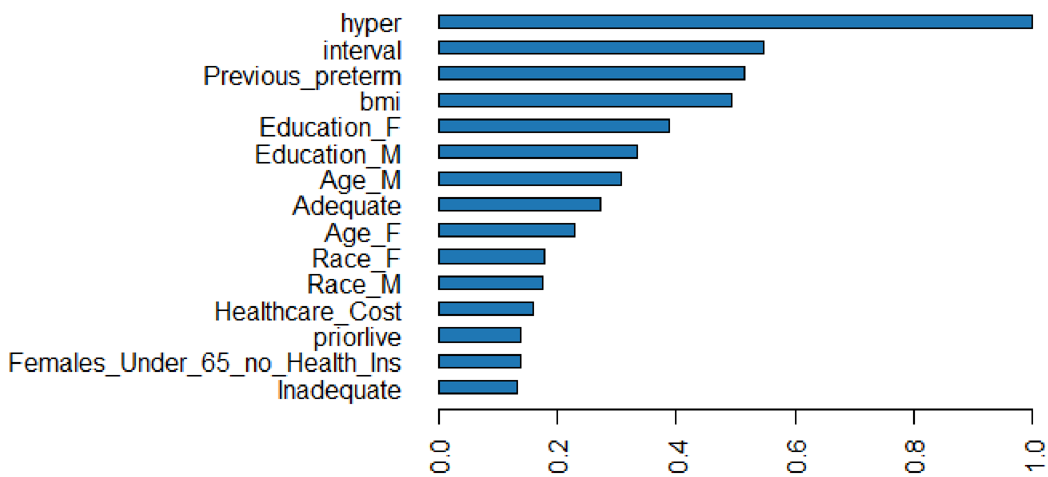


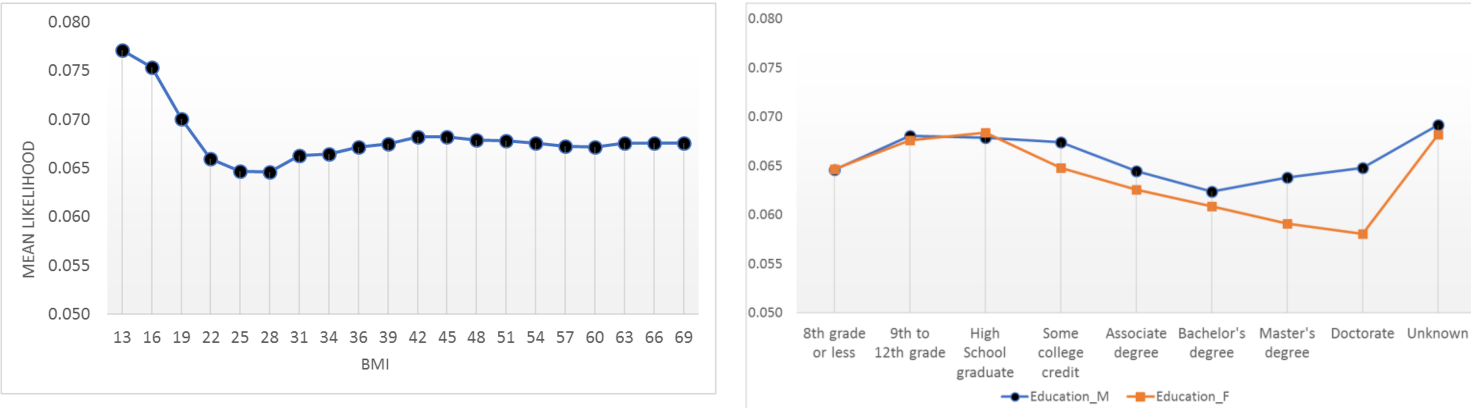
Figure 3-3 Variable importance plot

When we build the model on a high dimensional dataset that is representative of almost all the deliveries in the U.S., results show that the level of parent education is a more important predictor relative to demographic characteristics like race. Race is a significant predictor of both preterm birth and infant mortality where African American mothers has consistently been at a higher risk of a preterm delivery (12, 38). In 2016, 10.88% of Black singleton pregnancies resulted in a preterm baby versus 7.11% for White mothers. The results of partial dependence plots show that this likelihood is 7.02% (P-Value<0.001) for Black versus the 6.32% (P-Value<0.001) for White mothers when we account for the (average) effect of all factors like education and age of parents, and adequacy of care during pregnancy in each class.

Partial dependence plot shows the ‘effect’ of a variable on the response—likelihood of preterm birth—while accounting for the effect of other variables. The PDP is a great tool for interpretation because it calculates the \`effect size\` of each level given other conditions being equal. For example, mothers with a doctorate’s degree might have a higher age compared to those with a bachelor’s degree at the time of delivery. The PDP cancels the effect of age on the response and shows an ‘effect size’ that is attributable to the degree level. This is done by passing over the same dataset to the predictive model and fixing the *Education* variable at a certain level like doctorate. The average of these predictions for all of the observations is the marginal effect of the doctorate degree. The same procedure is repeated to obtain the marginal effects of the other levels of *Education.*

Figure 3-4 shows two examples of partial dependence plots (PDP). Figure 3-4.a shows the relationship between mother’s BMI and the likelihood of a preterm delivery. The PDP shows that mother with very low BMI—less than 22—are at higher risk of delivering a preterm baby. The important point about the PDP is that it considers the (average) effect of all other variables, which are called confounding variables in the statistical literature, in calculation of the likelihoods (27).

Figure 3-4.b shows the relationship of parent’s education and the likelihood of preterm birth. The likelihood of having a preterm infant for fathers decreases as their level of education increases. However, mothers with a Bachelor’s degree are the least likely group to have a preterm baby (6.24% with P-Value<0.001), and the likelihood increases for any degree more or less than that. The graph also shows that the missing values in education of father or mothers carries an important and insightful information showing that the likelihood of a preterm delivery for these types of observations are the highest (6.92% with P-Value<0.001) compared to other groups.



a. Preterm delivery likelihood for different values of BMI

b. Preterm delivery likelihood for different levels of education

Figure 3-4 Partial dependence plot for BMI and parent's education

3.6 Discussion

In this study, we deployed state-of-the-art statistical and machine learning techniques to first build a predictive model and then extract the risk factors of preterm birth (PTB) that are present during the early stages of pregnancy. The novelty of this study lies in exploring high-order interactions of risk factors by using a large dataset. We included both nulliparous and multiparous mothers, spontaneous and indicated preterm birth, but excluded multifetal pregnancies to increase the generalizability of our PTB prediction model. We reported the variable importance and partial dependence plots for the first time in the study of PTB.

The reported metrics indicate that our best GBM model improves the performance of preterm prediction compared to the similar works that combined maternal characteristics with important biological markers like serum analytes (17, 37). One of the major findings of this study is that race becomes a less important predictor for prediction of preterm birth—with relative importance of 4.13 percent—in a large cohort study when we add both individual risk factors (e.g., interval since live birth, education of parents, and whether the person received adequate care during pregnancy) and their interactions. This analytical finding is consistent with the theory of Lifecourse solutions for addressing the racial disparities in the preterm birth outcomes (39, 40). Th theory of Lifecourse emphasizes on the socioeconomic factors as the main determinants of health that can result in a positive shift in the long-term individual's health trajectory.

Hypertension is the most important predictor of preterm birth in a large cohort study, where 14.91 percent of the AUC improvement is attributed to this variable. The relative importance of hypertension is partly because of the deliveries that are scheduled preterm to prevent further complications in the pregnancy, especially when placenta is not providing enough nutrients and oxygen to the baby (41). The other important finding of this paper is that history of a previous PTB is not the most important variable in the prediction of PTB and it can only explain 5.63 percent of the AUC. This finding can be explained in two ways. First, history of preterm is useful only when the mother had a previous pregnancy. Second, the frequency of hypertension among the preterm population is almost two times the population of those with a history of a PTB. In 2016, the number of singleton pregnancies that resulted in preterm birth was 290,584. Among this population, 56,768 were hypertension positive, while a much smaller group—28,501—had a history of PTB.

The results of our GBM model agree with the findings of previous studies. The variables like hypertension ("hyper"), interval since last live birth ("interval"), and a history of PTB ("Previous_preterm") are among the most important predictors of a preterm birth, which is consistent with the findings of previous studies (7, 21). Also, our model shows that the 'effect' size of a history of PTB is the largest in the model (14.98 versus 6.51 with P-Value<0.001). However, the variable importance plot (VIP) reveals a novel and insightful finding compared to the previous studies. While the plot shows that the variables like previous preterm is an important predictor for preterm birth, it attributes a larger relative importance to factors like hypertension or interval since last live birth in the prediction of preterm in the general obstetric population. This new finding can be explained by the limitation of traditional studies. Methods like logistic regression only report the adjusted odd ratios for certain levels of the predictors, which is somewhat similar to the 'effect sizes' reported by the GBM. However, the GBM reports the collective importance of a variable including all of its levels in the predictive model. This capability of the DTs provides insights about the variables that can explain risk factors for larger groups of population. The new hierarchy of the important variables in prediction of PTB can address a gap in literature where already known risk factors cannot predict many actual preterm deliveries.

This study contributes to the literature in several ways. First, the results are generalizable to the US population. Past studies lacked generalizability for different reasons (17, 18). For example, some studies used a majority White population or their sample was from one geographical location to assess the PTB risk factors (17). A major strength of this study was application of data science on a population-based linked singleton births in the U.S. to address this gap. However, using the U.S. birth dataset had its own challenges like existence of anomalous observations and random errors. To mitigate this problem, we applied one of the advanced machine learning techniques, auto-encoders with deep neural nets, to perform data cleaning and preparation. Using such big dataset also entails long run times. The other strength of this study was implementation of Bayesian optimization in order to reach the optimal hyperparameters in a more efficient way. This study also contributes to the literature of preterm birth study by providing important insights by using advanced visualization techniques. The initial visualization of variables like mother's age versus gestational age (see Appendix I) shows a clear relationship between these two variables in which the risk of a preterm delivery is the highest at the extremes of maternal age. These findings match the results of multiple other in-depth analyses (13, 42). Partial dependence plots (PDP) are the other insightful tool that we used in this analysis. The PDPs like mother's BMI in Figure 3-4 shows that the extremes of pre-pregnancy BMI is associated with increased rates of PTB, which is compatible with the finding of other studies (21, 43). The PDP provides a better estimation of this association compared to previous studies (44), because it takes the (average) interdependent effect of other variables into account.

There is still significant room for improving the precision of preterm birth in large cohort studies. Positive predictive value (precision) of the past studies varied between 17 to 30 percent depending on the sample used in the analysis (20, 45). Our model shows a maximum precision of 28.13% in a national-level data, which approaches the best practices of similar studies. However, this metric is still relatively low. This low precision is due to the lack of knowledge regarding the cause(s) of PTB and absence of important predictors of preterm birth (e.g., cervical length) in the CDC dataset (20). Our study was subject to other limitations. Despite using the obstetric estimation for categorization of the PTB, there remains potential for errors (46). However, we used large samples and multifold cross-validations that minimize the effect of the incorrect categorization. Also, some of the biomarkers like cervical length or fetal fibronectin that are routinely measured in the obstetrical screenings were unavailable in the U.S. linked birth datasets. The association of these biomarkers and their interactions on the likelihood of a PTB can be assessed in future research.

References

1. Blencowe H, Cousens S, Chou D, Oestergaard M, Say L, Moller A-B, et al. Born too soon: the global epidemiology of 15 million preterm births. 2013;10(1):S2.

2. Mathews TJ, MacDorman MF, ME T. Infant mortality statistics from the 2013 period linked birth/infant death data set. National vital statistics reports. Hyattsville, MD: National Center for Health Statistics2015.

3. Butler AS, Behrman RE. Preterm birth: causes, consequences, and prevention: National Academies Press; 2007.

4. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. The Lancet. 2008;371(9608):261-9.

5. Iams JD. Identification of candidates for progesterone: why, who, how, and when? Obstetrics and gynecology. 2014;123(6):1317-26.

6. American College of Obstetrics and Gynecology. Practice Bulletin No. 130: Prediction and Prevention of Preterm Birth. Obstetrics & Gynecology. 2012;120(4):964-73.

7. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. The lancet. 2008;371(9606):75-84.

8. Boots AB, Sanchez-Ramos L, Bowers DM, Kaunitz AM, Zamora J, Schlattmann P. The short-term prediction of preterm birth: a systematic review and diagnostic metaanalysis. American journal of obstetrics and gynecology. 2014;210(1):54. e1-. e10.

9. Davey MA, Watson L, Rayner JA, Rowlands S. Risk-scoring systems for predicting preterm birth with the aim of reducing associated adverse outcomes. Cochrane Database of Systematic Reviews. 2015(10).

10. Bhattacharya S, Raja EA, Mirazo ER, Campbell DM, Lee AJ, Norman JE, et al. Inherited predisposition to spontaneous preterm delivery. Obstetrics & Gynecology. 2010;115(6):1125-33.

11. Laughon SK, Albert PS, Leishear K, Mendola P. The NICHD Consecutive Pregnancies Study: recurrent preterm delivery by subtype. American journal of obstetrics and gynecology. 2014;210(2):131. e1-. e8.

12. Martin JA, Hamilton BE, Osterman MJ, Driscoll AK, Drake P. Births: Final data for 2016. 2018.

13. Fuchs F, Monet B, Ducruet T, Chaillet N, Audibert F. Effect of maternal age on the risk of preterm birth: A large cohort study. PloS one. 2018;13(1):e0191002.

14. Newman R, Goldenberg R, Iams J, Meis P, Mercer B, Moawad A, et al. Preterm prediction study: comparison of the cervical score and Bishop score for prediction of spontaneous preterm delivery. Obstetrics and gynecology. 2008;112(3):508.

15. Magee L, Von Dadelszen P, Chan S, Gafni A, Gruslin A, Helewa M, et al. The control of hypertension in pregnancy study pilot trial. BJOG: An International Journal of Obstetrics & Gynaecology. 2007;114(6):770-e20.

16. Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics. 2001:1189-232.

17. Alleman BW, Smith AR, Byers HM, Bedell B, Ryckman KK, Murray JC, et al. A proposed method to predict preterm birth using clinical data, standard maternal serum screening, and cholesterol. American Journal of Obstetrics & Gynecology. 2013;208(6):472. e1-. e11.

18. Weber A, Darmstadt GL, Gruber S, Foeller ME, Carmichael SL, Stevenson DK, et al. Application of machine-learning to predict early spontaneous preterm birth among nulliparous non-Hispanic black and white women. Annals of epidemiology. 2018;28(11):783-9. e1.

19. Nelson DB, McIntire DD, McDonald J, Gard J, Turrichi P, Leveno KJ. 17-alpha Hydroxyprogesterone caproate did not reduce the rate of recurrent preterm birth in a prospective cohort study. American journal of obstetrics and gynecology. 2017;216(6):600. e1-. e9.

20. Robinson JN, Norwitz E. Preterm birth: Risk factors, interventions for risk reduction, and maternal prognosis. https://wwwuptodatecom/contents/preterm-birth-risk-factors-interventions-for-risk-reduction-and-maternal-prognosis. 2019.

21. Iams JD. Prevention of preterm parturition. New England Journal of Medicine. 2014;370(3):254-61.

22. He J-R, Ramakrishnan R, Lai Y-M, Li W-D, Zhao X, Hu Y, et al. Predictions of preterm birth from early pregnancy characteristics: born in guangzhou cohort study. Journal of clinical medicine. 2018;7(8):185.

23. Linked Birth / Infant Death Records 2007-20017 [Internet]. United States Department of Health and Human Services (US DHHS), National Center for Health Statistics (NCHS), Division of Vital Statistics (DVS). [cited Oct 2, 2017 5:01:19 PM]. Available from: http://wonder.cdc.gov/lbd-current.html

24. Bengio Y, editor Deep learning of representations: Looking forward. International Conference on Statistical Language and Speech Processing; 2013: Springer.

25. Goldstein M, Uchida S. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PloS one. 2016;11(4):e0152173.

26. Leng Q, Qi H, Miao J, Zhu W, Su G. One-class classification with extreme learning machine. Mathematical problems in engineering. 2015;2015.

27. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning: Springer series in statistics New York, NY, USA:; 2001.

28. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological). 1996;58(1):267-88.

29. Breiman L. Random forests. Machine learning. 2001;45(1):5-32.

30. Schapire RE, Freund Y. Boosting: Foundations and algorithms. Kybernetes. 2013.

31. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al., editors. Lightgbm: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems; 2017.

32. Chen T, Guestrin C, editors. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016: ACM.

33. Bergstra J, Yamins D, Cox DD. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. 2013.

34. Mendoza H, Klein A, Feurer M, Springenberg JT, Hutter F, editors. Towards automatically-tuned neural networks. Workshop on Automatic Machine Learning; 2016.

35. Feurer M, Hutter F. Hyperparameter optimization. Automated Machine Learning: Springer; 2019. p. 3-33.

36. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PloS one. 2015;10(3):e0118432.

37. Goodwin LK, Iannacchione MA, Hammond WE, Crockett P, Maher S, Schlitz K. Data mining methods find demographic predictors of preterm birth. Nursing research. 2001;50(6):340-5.

38. Meis PJ, Goldenberg RL, Mercer BM, Iams JD, Moawad AH, Miodovnik M, et al. The preterm prediction study: risk factors for indicated preterm births. American journal of obstetrics and gynecology. 1998;178(3):562-7.

39. Lu MC, Kotelchuck M, Hogan V, Jones L, Wright K, Halfon N. Closing the Black-White gap in birth outcomes: a life-course approach. Ethnicity & disease. 2010;20(1 Suppl 2):S2-62-76.

40. Wadhwa PD, Entringer S, Buss C, Lu MC. The contribution of maternal stress to preterm birth: issues and considerations. Clinics in perinatology. 2011;38(3):351-84.

41. Krishna U, Bhalerao S. Placental insufficiency and fetal growth restriction. J Obstet Gynaecol India. 2011;61(5):505-11.

42. Fraser AM, Brockert JE, Ward RH. Association of young maternal age with adverse reproductive outcomes. New England journal of medicine. 1995;332(17):1113-8.

43. Hendler I, Goldenberg RL, Mercer BM, Iams JD, Meis PJ, Moawad AH, et al. The Preterm Prediction Study: association between maternal body mass index and spontaneous and indicated preterm birth. American journal of obstetrics and gynecology. 2005;192(3):882-6.

44. Honest H, Bachmann LM, Ngai C, Gupta JK, Kleijnen J, Khan KS. The accuracy of maternal anthropometry measurements as predictor for spontaneous preterm birth—a systematic review. European Journal of Obstetrics & Gynecology and reproductive biology. 2005;119(1):11-20.

45. Meertens LJ, van Montfort P, Scheepers HC, van Kuijk SM, Aardenburg R, Langenveld J, et al. Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation. Acta obstetricia et gynecologica Scandinavica. 2018;97(8):907-20.

46. Martin JA, Osterman M, Kirmeyer S, Gregory E. Measuring gestational age in vital statistics data: transitioning to the obstetric estimate. National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System. 2015;64(5):1-20.

Chapter 4. The Capability Trap on a Macro-societal Level: An Explanation for Government Failure to Improve Healthcare System

Alireza Ebrahimvandi, Niyousha Hosseinichimeh, Navid Ghaffarzadegan

Abstract

Healthcare problems and expenditure in the United States have been both increasing over the past decade. The rise in expenditure has also shifted national resources from other sectors such as education, infrastructure, and environment. We argue that the macro-level resource allocation decision can result in a capability-trap-like phenomenon leading to deterioration of socio-economic status of the population, and worsening health outcomes. In a multi-method study, we model and simulate the mechanism, and provide statistical evidence for simulation outcomes. Simulation results show that health problems can exacerbate in the long-term due to a lack of investment in fundamental dynamics that influence socio-economic resources, capital, and status of the society. The result is growth in both healthcare expenditure and problems. We postulate that the system has tipping point values, and too much or too little healthcare spending result in vicious cycles that deteriorate healthcare conditions. In a supplementary step, we conduct statistical analysis of health indicators across different US states over the past 23 years. The results confirm our simulation outcomes pointing to the negative consequences of increasing healthcare expenditure in the expense of shifting resources from other sectors such as education.

Keyword: Capability tap, healthcare system, healthcare expenditure, socio-economic system, systems thinking, system dynamics, simulation, tipping point

4.1 Background

The U.S. health system is in a poor and costly condition. It lags behind many of the advanced Organization for Economic Cooperation and Development (OECD) nations in several key health indicators including infant mortality, prevalence of chronic conditions, and access to care (1). While the outcomes are undesirable, the expenditure is the highest both per capita and in total (2). In addition, there are significant state variations in health outcomes (3).

Studies offer mixed explanations on why the U.S. health is in a poor and costly condition (4, 5). While many blame over-medicalization or private insurers for creating administrative burdens for providers (6, 7), others point to the unequal distribution of primary care providers that limits their ability to offer the highest-quality care (8). Some point to social and income inequality as the source of poor outcomes (9, 10). Many economists would argue that health drives socioeconomic (SE) status—because poor health negatively affects job opportunities and socioeconomic position that creates social drift (11, 12). However, most social scientists look at the adverse health outcomes like overall mortality as a function of socioeconomic and behavioral factors that trigger physiological responses, resulting in poor functioning of the immune system in the long-run (13-15). Before the mid-1980s, healthcare researchers largely ignored the SE status as a major factor in their studies except as a control variable (12). Since 1985, social scientists increasingly conducted research on the nature of the relationships between socioeconomic factors and health (16).

Systems thinking approaches have contributed to better understanding of healthcare systems and poor health outcomes. These frameworks are powerful decision support systems for explaining health outcomes such as chronic or acute diseases over time and making better decisions for improving adverse results (17-23). However, most of systems models still have boundaries that mainly revolve around the health system, and does not include interdependencies with other social and economic sectors. In addition, in most models, since the economy is treated as an exogenous factor, resources that finance interventions are not generated endogenously within the models. The problem is hard to model and explore as there are various dynamic complexities and interactions between different economic sectors that hinder our ability to reach correct conclusions about the impacts of our decisions on the health system (24-26). Reaching an optimal answer becomes particularly harder when there are various interacting agencies and actors involved who might have multiple and sometimes conflicting interests.

In this study, we conduct a multi-method approach that includes system dynamics and statistical analysis. Our purpose is to bridge the gap in the literature and provide policymakers a comprehensive framework that can help them in making better and more robust decisions for advancing the health and well-being of their communities. In this novel approach, we consider the effect of health-related decisions on both the health outcomes and the social status of the population in the long-run. We also include effects of socio-economic status on shaping health outcomes. In the first step, we develop a system dynamics simulation model. We use the State of Ohio as our case study to ground our modeling framework and develop our model structure. This improves our understanding of why such a state with one of the best healthcare systems in the U.S. struggles to improve the health outcomes despite high expenditures in medical care. The result is an illustrative system dynamics model structure that is different from past studies as it explicitly includes interactions between social status to the health outcomes, and resource generation/depletion in a community. It also considers multiple feedbacks of decisions on resources allocated to health and their impacts on shaping socio-economic status of people. Our simulation results point to non-linear dynamics of healthcare spending, and potential consequences of too much or too little spending in healthcare. The proposed framework supports critical thinking and more robust conclusions about how resource allocations are likely to affect the performance of the health system in the short and long term. After analyzing simulation results, we provide empirical evidence for the arguments based on statistical examination of US states over a period of 23 years. Overall, the study shows that solutions for improvement in population health outcomes do not necessarily lie in the healthcare system, and healthcare outcomes can deteriorate, not just *despite* increasing health expenditure, but *because* of increasing healthcare expenditure.

In the next section, we first review the theory of Lifecourse that explains the relationship between socio-economic status and health. This theory offers strategies aimed at optimization of health trajectories both at the individual and population level (27). In analysis of health outcomes, this theory allowed us to expand our scope of examination from health care system to a broader context in which an individual becomes vulnerable to pathogens. Based on this perspective, we then applied the theory of capability trap to explain why many states cannot overcome the spiral of health deterioration. The theory of capability trap offers insight into the performance of the system and how it depends on the resource allocation decisions that lead to building capabilities. Then, we evaluate the state of Ohio to assess their resource allocation trends over time to different categories such as health and socio-economic capital and then evaluate their health trends under an economic shock. We combine these two theories and evidence from the case of Ohio to introduce a model structure followed by simulation results and more empirical evidence. We conclude with a discussion with respect to the implications of a comprehensive framework of resource allocations and how it might affect health outcomes. Finally, we discuss the optimal amount of resource that should be allocated to each category under different conditions.

4.2 Theoretical Background

4.2.1 Why Socioeconomic Status Matters in Health Outcomes?

Theory of Lifecourse shifts the focus from health services to when the biological and physiological characteristics of the immune system are shaped. A vast body of health research has focused on finding and explaining the causes of diseases and contributors to health. The biomedical models drove the first era of healthcare. They initially provided simplified causes for different acute diseases and illnesses (28, 29). The next era of healthcare was driven by a more comprehensive view. The newer bio-psychosocial models accounted for a wider range of factors influencing health. These models included the role of social and behavioral factors in illnesses and designed programs to manage chronic illnesses in addition to acute illnesses by changing unhealthy lifestyle choices in adulthood (27). More recently, models of Lifecourse health development (LCHD) proposed a new explanation for individual and population health trajectories. The LCHD models showed the major impacts of socioeconomic and psychological factors operating early in life on long-term health outcomes (27, 30, 31). These models synthesize research from a wide range of biological, behavioral and social sciences and provide an integrated, dynamic definition for health development. The time horizon of LCHD models begins before conception and continues throughout the lifespan. The LHCD models paved the way for the creation of new approaches for improvements of the individual and population health.

Theory of Lifecourse suggests that exposure to acute and chronic stress over time erodes the ability of the immune system to work effectively (32). People who are exposed to stress during their lifetime are more vulnerable because they have higher levels of allostatic load. Allostatic load is defined as the cumulative physiological toll that accumulates through financial distress, illness or injury, exposure to environmental hazards, or risky behaviors throughout life (33). Furthermore, living in an economically strained household has a great impact on the stress level of its members. This portion of the population is at a higher risk of developing a vulnerability to adverse health outcomes or biological predispositions that can be passed to their next generations in many ways such as delivering a premature or low-birth-weight baby (34-36). In addition, vulnerable people have a higher propensity for engaging in unhealthy behaviors like bad diet, smoking, and drinking that are also risk factors for many chronic diseases (33, 37, 38).

Positive socioeconomic factors such as higher levels of education can result in a positive shift in an individual’s health trajectory (27). Education has a unique dimension in forming social and health status. Educated people are healthier not because they can afford better or more health services, but because they can buy themselves out of privation (39). Education increases the sense of control, which eventually leads to a healthy and less stressful lifestyle (40). Educational attainment is the main bridge between the status of one generation and the next. It also functions as the "main avenue of upward mobility" (39). Therefore, investing in education can serve as one of the major tools for decision-makers to create positive and sustainable changes in the social status of people in vulnerable communities. However, resources are not necessarily allocated with a concerted approach in a community and major categories like education and social investments may lose their priorities to other urgent issues. In the next section, we introduce a problem that has the same structure, but has been studied in a different context. The theory of capability trap has been discussed in the context of organizational and business management, however, we will argue how we can apply the same theory to the domain of health.

4.2.2 The Theory of Capability Trap and How it Can Explain the Spiral Health Deterioration

The theory of capability trap was introduced by Repenning and Sterman in early 2000s to explain the failure of process improvement programs in organizational settings (41, 42). This theory was first developed based on an inductive study of an organization's attempt to improve its core businesses. This theory explains the delays between the process of investing in improvement activities and recognizing the rewards. Many organizations choose temporary gains by "working harder" and even taking shortcuts. However, working harder comes at the expense of investing less time and resources for improving capabilities. Therefore, these gains often lead to a "better before worse" trade-offs that managers unwittingly choose over "working smarter." This situation demands an ever-increasing level of effort to maintain performance. Realizing the fact that the system is in the trap is very hard, and it is even harder to overcome the trap. The outcome is a vicious cycle where the organization spends on fixing short term problems while capabilities are deteriorating and leading to more problems.

The capability trap phenomenon is not limited to large organizations. The examples of this dynamic can be found in a wide range of activities and entities from personal life to large organizations to societal level. At the personal level, we often fail to commit to a regular exercise program that will improve our long-run well-being, because of the short-term stresses of other tasks. At the industry level, Sterman (43) discusses why organizations' efforts for sustainable and pro-social improvements often lead to "sizzle and fizzle". Lyneis and Sterman (44) look into the problem of maintenance in a large scale university, and expand the theory to discuss a "win-win" investment strategy that improves both socially desirable outcomes in energy efficiency and organizational outcomes.

A recent literature review by Landry and Sterman (45) provides a comprehensive application of the capability trap in different contexts including the domain of health. Discussing within health system capability trap and beyond health system trap they point to an idea that is similar to our line of thinking in this paper. Landry and Sterman write:

> ".. [H]ealthcare is not the only determinant of health. By expanding the scope of examination, we discover that the United States has consistently ranked among the lowest in public social service expenditure as a percentage of GDP, which is an important determinant of its long-term demographic health, compared to its OECD counterparts… This gives rise to yet another, even larger capability trap,… in which 'downstream' systems like healthcare are heavily funded at the expense of investment in 'upstream' social capabilities. … Ultimately, there is an understanding that healthcare sector specific interventions are 'insufficient to address population-level health disparities' due to its position within a larger system with more momentum." (p. 26-27).

Landry and Sterman provide a casual loop diagram in an aggregate level, and consistent with the generic representation of capability trap dynamics, calling it an expanded health system. The diagram represents resources and capabilities that are not directly health-related and call it "community capabilities."

In this paper, we offer a formal simulation model grounded in the Lifecourse theory to systematically analyze healthcare system problems. We also provide a supplementary empirical analysis. To be precise, it might be better to refer to our model as a capability-trap-like mechanism, instead of capabilities, as we look into the distribution of the population in terms of their socio-economic status. Since our work also point to the dominance of a vicious cycle leading to shifting resources from long-term solutions to seemingly urgent problems, it is in fact like a capability trap model.

4.3 Data and Method

Our approach in this multi-method project is depicted in Table 4-1. Our model structure is grounded in qualitative data from the case of Ohio as it experiences a downward spiral of health deterioration despite having one of the best healthcare systems in the nation and increasing resource allocations for healthcare both at urgent services and prevention. We perform this study in four steps, each provides a supporting information to a final theory. First, we review a case study and use it to conduct an inductive reasoning. The case is inferred in the model building process, helping verify model structure. We then further develop the structure of our model to include the theoretical grounds of Lifecourse perspective and specifically capture the socio-economic status as a key stock in shaping the population health (27). Next, we simulate model for a wide range of parameters to move toward a more generalizable and comprehensive theory. Finally, we use the insights from the model to offer a set of testable strategic policies toward the allocation of resources in order to reach better long-term health outcomes, and test the results in an empirical setting. This final step includes statistical analysis of healthcare spending and status data of all US states.

Table 4-1 Steps and triangulation in our multi-method approach in this study

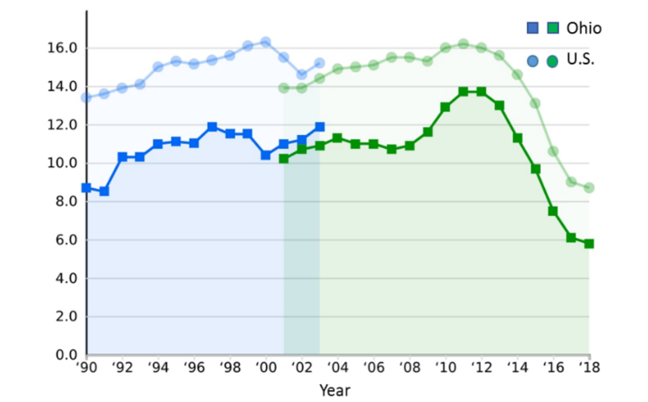
	Step	**Outcome**
1	Qualitative and quantitative study of the case of healthcare in Ohio	• Recognition of interactions between health and socioeconomic status and spending • Input to system dynamics model (step 2)
2	Develop a system dynamics model	• A Capability-tap-like theory of healthcare system problems and spending in Ohio
3	Further analysis of the simulation model for a wide range of parameters	• A more general capability-tap-like theory of interdependencies between healthcare system problems and spending • Emergence of tipping points and bifurcation in outcomes
4	Statistical analysis based on data from US state over 23 years.	• Empirical evidence for outcomes of step 3.

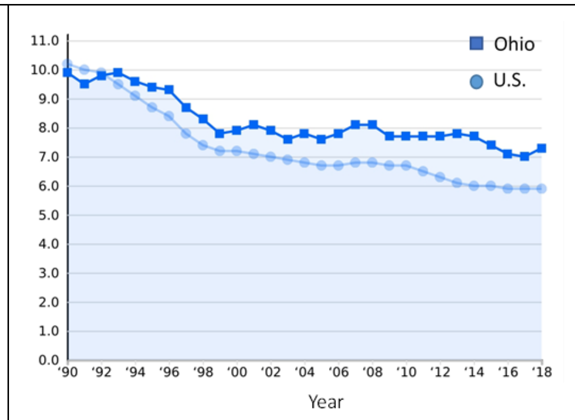
4.3.1 An overview of the case that we used as a ground for building the model

The state of Ohio has experienced one of the most interesting trends in healthcare outcomes. Ohio's healthcare spending is consistently more than other states and yet it gets less for it even compared to the neighboring states. The classical justifications that attribute some of the high healthcare spendings to the high investment in research and tech do not fit to the case of states' spending (46). This leaves a big confusion about the reasons for such poor health outcomes, because Ohio also has expanded preventive care like immunization and provided health insurance to a larger percentage of the population compared to the U.S. average.

Figure 4-1.a shows the uninsured rate from 1990 to 2018 based on two different surveys. Ohio consistently had a lower (better) uninsured rate compared to the U.S. average. Despite good healthcare access, two examples of worsening health outcomes in Ohio is infant mortality (Figure 4-1.b), and premature deaths (Figure 4-1.c) as population health indicators. These measures have all been better than the U.S. average in 1990 and they became worse than the U.S. national average over time. Many other health indicators show the same troubling trend. The rank of Ohio in overall health, which is defined by the World Health Organization as “a state of complete physical, mental, and social well-being”, has dropped from 27 to 40 in a comparison with other states (Figure 4-1.d). Many behavioral indicators in Ohio like obesity (Figure 4-1.e) or the rate of drug overdose deaths (Figure 4-1.f) has followed the same deteriorating trend.

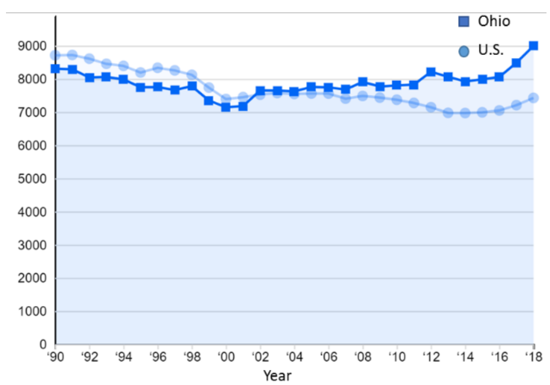
The correlation between the pattern of behavioral indicators and health outcomes in Ohio is clear. However, it is not clear why all of these indicators are experiencing the spiral downward trend. Policymakers in Ohio have tried to improve major health issues by leveraging behavioral factors. For example, Governor Kasich mentioned that "Ohio has one of the worst infant mortality rates in the nation and that is simply unacceptable" in a speech in 2014. He also suggested that "Initiatives like today's summit and our efforts to reduce drug addiction are good first steps, but we must work together to focus support and resources to those mothers and babies most at-risk" (47). These legislation and interventions, however, had limited short-term impacts as can be seen in Figure 4-1.b. On the other side, the theory of Lifecourse suggests that improving health outcomes require a more comprehensive approach to health problems. Therefore, it is important to build a case to show the impacts of social determinants on these indicators and show how the resources are managed in Ohio during the past two decades.

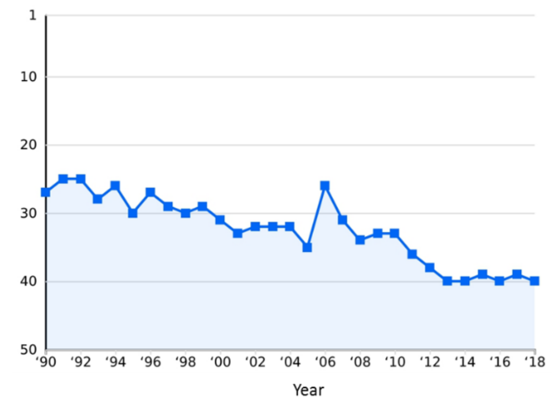




(a) Ohio uninsured rate compared to the U.S.*

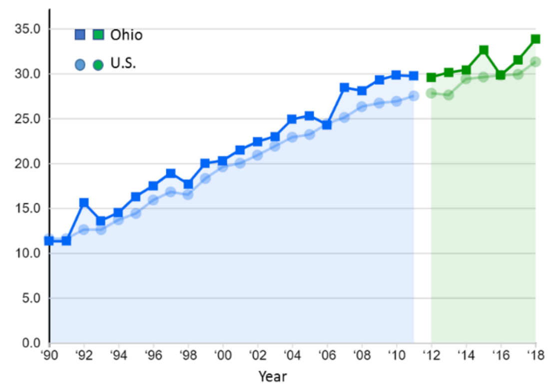
(b) Infant mortality per 1,000 live births

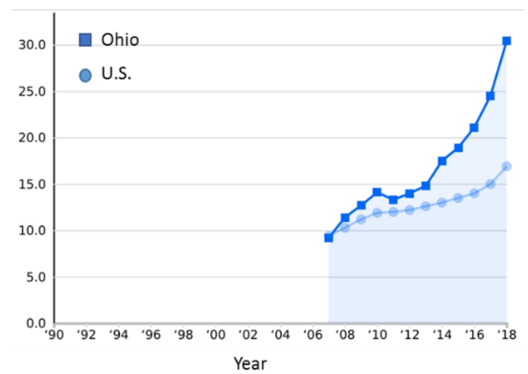




(c) Years lost before age 75 per 100,000 population

(d) Ohio overall health ranking in the U.S.





*(e) Percentage of adults with a body mass index of 30.0 or higher***

(f) Drug overdose deaths per 100,000 population

Figure 4-1 Comparison of trends in Ohio and U.S. national average in key behavioral and health determinants

Source: American Health Rankings

* Percentage of the population that does not have health insurance privately, through their employer or through the government (2-year average) from two different surveys conducted by the US Census Bureau: current population survey is used for pre-2002 and American community survey is used for post-2000.
** The blue part is based on the reported height and weight pre-2011 Behavioral Risk Factor Surveillance System methodology that is different compared to the green part for post-2012.

Table 4-2 shows how Ohio's budget is allocated between different categories from 2000 to 2015. These statistics show the percentage of the budget that is spent both from the General Revenue Fund (GRF) and the total spending (including capital and non-capital). The GRF funding is financed mainly from the resident's income and business sales taxes. However, the total spending includes both the GRF and the federal funds that are allocated to the state. To stay relevant, we report the numbers and percentages for the GRF, because federal funds are mainly a function of the decisions that are made about the GRF. For example, if a state decides to expand Medicaid, it has to spend more of the GRF for providing insurance coverages and then the federal government will accordingly contribute a certain percentage to the state budget.

As suggested in Table 4-2, resources are increasingly driven away from public social services (social capital) to healthcare. The resources that are allocated to K-12 education decreased from 26.2% to 24.7% of the GRF from 2000 to 2015. The same trend in the decreased allocation of resources also occurred in higher education, correction, and transportation. On the other side, the share of Medicaid funds increased from 30.6% to about half of the GRF—48.2%—in the same time period.

Table 4-2 Ohio's spending from 2010 to 2015 in each category by a percentage of the GRF fund (and total budget). The percentages from the total fund are shown in parentheses.

Year	Education	Higher Education	Corrections	Transportation	Other	Medicaid
2000	26.2 (18.2)	12.6 (7.0)	7.9 (4.8)	0.2 (8.5)	22.4 (42.5)	30.6 (19.1)
2003	27.6 (19.0)	10.6 (5.7)	7.1 (3.9)	0.2 (8.1)	17.5 (40.1)	37.0 (23.1)
2006	27.6 (18.9)	9.9 (5.2)	6.9 (3.7)	0.1 (7.7)	15.3 (39.4)	40.2 (25.1)
2009	24.7 (19.5)	9.9 (5.4)	7.2 (3.6)	0.1 (4.9)	21.8 (46.1)	36.4 (20.6)
2012	24.7 (17.4)	8.3 (4.2)	6.3 (3.1)	0.0 (5.1)	16.4 (41.4)	44.3 (28.9)
2015	24.7 (16.8)	7.7 (4.1)	5.6 (2.9)	0.0 (5.1)	13.7 (33.7)	48.2 (37.4)

Source: National Association of State Budget Officers
Note: "Other" expenditure includes public assistance, Children's Health Insurance Program (CHIP), state police, economic development, employer contributions to pensions and health programs.

The evidence from neighboring states supports this assumption that Ohio is missing investment opportunities in socio-economic capital, specifically in education. States like Michigan or

Pennsylvania spent a higher percentage of their budget on K-12 education and a lower percentage on Medicaid (Figure 4-2). In 2015, the share of budget expenditure of Ohio, Michigan, and Pennsylvania on K-12 education was 24.7%, 36.5%, and 35.3% and the share of spending on Medicaid was 48.2%, 26.9%, and 29.4%, respectively. Although some of these variations are due to differences in accounting practices, these statistics are representative of the priorities in resource allocations of these state.

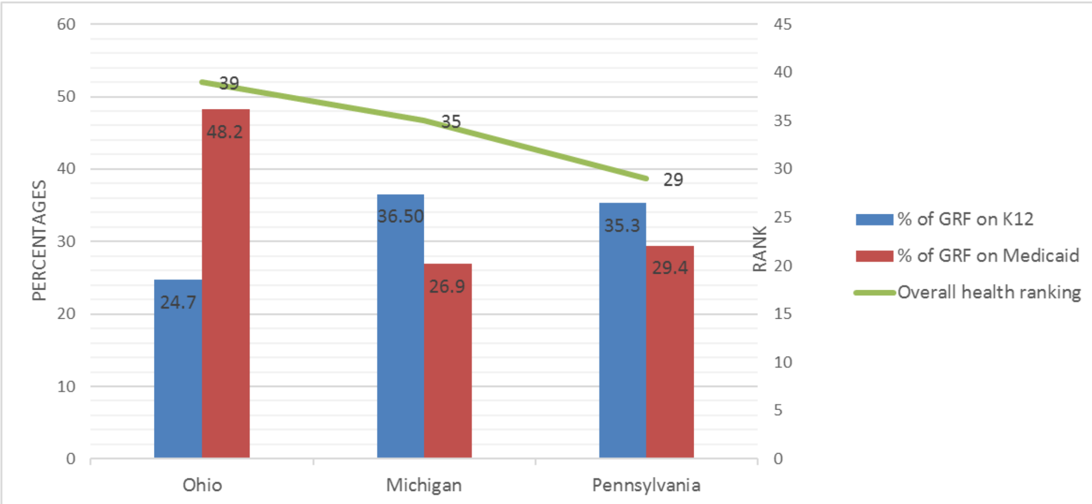


Figure 4-2 Comparison of Ohio with neighboring states in healthcare spending and overall health ranking in 2015

Interestingly, both Michigan and Pennsylvania had a better ranking in overall health compared to Ohio in 2015. The ranking of Ohio, Michigan, and Pennsylvania among 50 states were 39, 35, and 29, respectively in 2015. This is despite the better historical ranking of Ohio compared to these states in the 1990s as suggested by Figure 4-3.

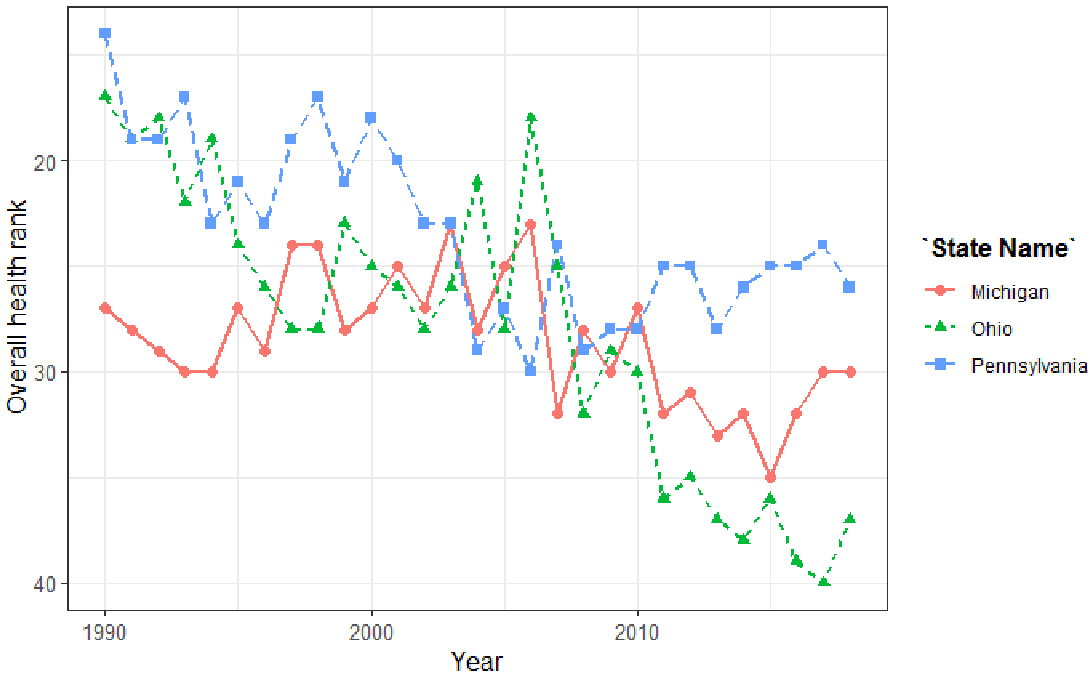


Figure 4-3 Historical ranking of Ohio, Michigan, and Pennsylvania in overall health outcomes among US states

4.4 Modeling Approach

Our modeling approach is system dynamics, for several reasons. First the question is a dynamic question and deals with resource allocation that changes over time. Second, decisions about resource allocation are arguably endogenous variables that affect the status of the system which will in turn influence future decision. In such context a feedback-rich approach helps examine complex counter-intuitive outcomes. Finally, our aim is to help impact decision makers mental models and system dynamics model have proved to be effective tools for communication across organizational boundaries (48). The model uses the case to verify causal relations, but it is still an illustrative model and is formulated based on hypothetical values. That is to say, our focus is not on the point prediction but more on understanding the structural features of health dynamics and their interaction with socio-economic status.

4.4.1 Population Flow

Previous system dynamics health models mainly include compartmental models that divide the population into different groups based on their vulnerability to different diseases (17). We split the population into two main groups based on their socio-economic (SE) status. Figure 4-4 shows the two socio-economic groups in our model: *SE vulnerable* (V) and *SE stable* (S). People move from V to S when there is a *Change in SE status* (C_s) [see *(2)* and *(3)*]. A negative value of C_s would mean a flow from S to V. To simplify, we assume net changes in total population (total of birth, death, in- and out-migration) is zero for the period of analysis.

$$\frac{d}{dt}S = C_S \tag{2}$$

$$\frac{d}{dt}V = -C_S \tag{3}$$

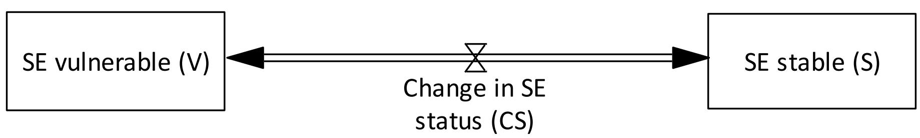


Figure 4-4 Two modes of Socioeconomics (SE) status and population flow between them

Note: The rectangular boxes show the stock variables, and the flow symbols are inflows and/or outflows from the stocks.

4.4.2 Health Issues

Based on the literature, the average number of health issues per person is different for the two groups of V and S population (38). The SE vulnerable are more prone to risky behaviors like smoking, drinking, and drug abuse that are risk factors for a wide range of health issues (33). This group is also less likely to have self-payed or employer-sponsored insurance to cover their healthcare costs. Therefore, most of their health issues need to be addressed by public funds. In our model, the parameters *Health issues per SE vulnerable* and *SE stable* are introduced at two different values of h_1 and h_2, respectively. The total number of *New health issues of SE vulnerable* (I_{H1}) and *New health issues of SE stable* (I_{H2}) is formulated proportional to the per capita health issues, h_i, and their corresponding population, V and S (equations *(4)* and *(5)*).

$$I_{H1} = h_1 * V \tag{4}$$

$$I_{H2} = h_2 * S \tag{5}$$

As depicted in Figure 4-5, the rate of *New health issues* (I_H) per unit of time is represented by a simple summation of I_{H1} and I_{H2}*(6)*. The *Health issues* (H) is a stock variable, that increases when new health issues is more than health addressing issues (O_H).

$$I_H = I_{H1} + I_{H2} \quad (6)$$

$$H = \int (I_H - O_H)\, dt + H_0 \quad (7)$$

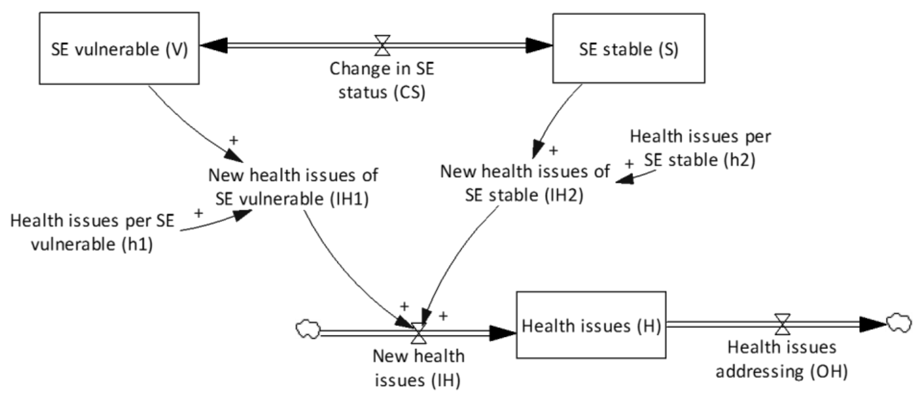


Figure 4-5 Population flow and health issues

The model assumes that the major part of the tax dollar resources is allocated between two categories of spending, resources to address health issues, and resources that affect socio-economic status of individuals, represented by SE capability. Examples of healthcare spending includes investment in healthcare technology, hospitals, medicine, and emergency room, and investment in SE capital includes education, job creation, transportation and infrastructure development.

The model makes an assumption about priority of resource allocation on urgent health matters, and the assumption is consistent with other capability trap models: The decision makers first make resource allocation decisions for addressing health issues based on cases of health issues, and then the rest of the resource is allocated on building SE capital. To help generalizability of the model we introduce a parameter, (μ), which is a cap on how much of total resources (*r*) can be spent on health. This is represented in equation *(8)*. Health issues are generally addressed at the rate of $\frac{\gamma}{\tau_1}$, where τ_1 represents *Average time to address health issues* and γ represents *Health resources adequacy (9).* Required resources is simply proportional to total health issues.

$$r_H = Min(r_H{}', r\mu) \quad (8)$$

$$O_H = \frac{\gamma}{\tau_1} H \quad (9)$$

$$\gamma = Min\left(1, \frac{r_H}{r_H{}'}\right) \quad (10)$$

$$r_H{}' = \lambda_1 H \quad (11)$$

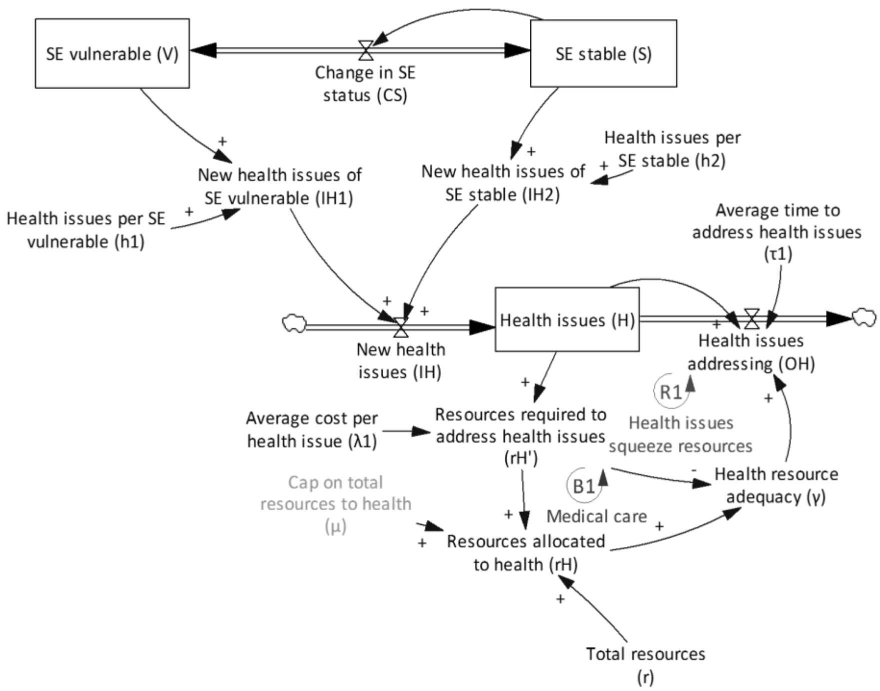


Figure 4-6 Addressing health issues

Note: The notations "B" and "R" under the loop shapes stands for "balancing" and "reinforcing", respectively.

Figure 4-6 also shows two feedback mechanisms in the model; ***B1*** and ***R1***. The balancing loop, ***B1***, represent effects of resource allocation on addressing health issues, and loop ***R1*** represents the need for more resource as health issues increase.

4.4.3 Investment in Socio-Economic Capability

As shown in Figure 4-7, the *socio-economic (SE) capability* (C) is a stock variable that accumulates by *Investment in SE capital* (I_C) and declines by *SE capital depletion* (O_C) rate *(12)*. The depletion rate, O_C, follows Little's law (49) with *Life expectancy of SE capita* of (τ_3) *(13)*. We assumed that government is aware of desired level of SE, thus can determine *Desired investment* *in SE* (d) based on desired per capita SE (γ_2) and replacing the depletion rate. This will give them a steady state level of SE, as represented in the following:

$$\frac{d}{dt}C = I_C - O_C \qquad (12)$$

$$O_C = \frac{C}{\tau_3} \qquad (13)$$

$$I_C = Min(d, r_S) \qquad (14)$$

The rate of desired investment in SE capital, d, closes a fraction of the gap between the current C level and the *Desired SE capital*(D_S), where the fraction is given by 1/Time to cover gap (τ_2), plus the rate of *SE capital depletion* (O_C). The D_S is simply the total *Desired SE capital per capita*(λ_2) of population (P) *(16)*.

$$d = \frac{\Delta}{\tau_2} + O_C \qquad (15)$$

$$D_S = \lambda_2 \times P \qquad (16)$$

However, building SE capital requires resources. As stated, the budgeting priorities may change every year and affect percentage of the General Revenue Fund (GRF) allocated to categories like health or education. Our case study shows that states are likely to allocate a higher percentage of their GRF to health as the number of health issues increases. Therefore, we make this simplifying assumption that a community that has to spend less on healthcare will have more flexibility in allocating resources to build social capabilities *(17)*.

$$r_S = r - r_H \qquad (17)$$

Figure 4-7 also shows Loop ***B2***. This loop balances the *Capital allocation* to keep it at a steady level. It makes sure that if *Investment in SE capital* (I_C) suddenly increases, it returns back to the normal level after a while. In this mechanism, an increase in I_C increases the *SE Capital* (C). A larger C then decreases the *Gap in SE Capital* (Δ), which subsequently decreases the *Desired investment in SE* (d). A lower d then reduces I_C and this will close the balancing loop.

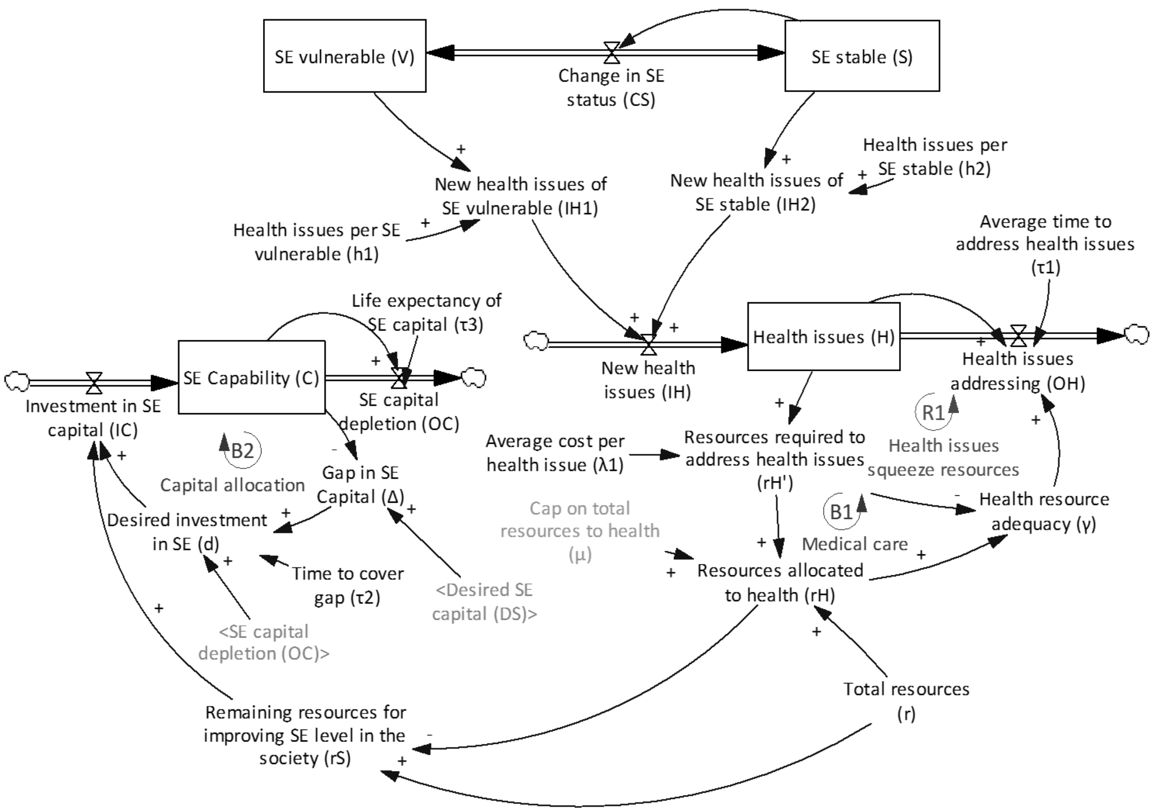


Figure 4-7 The model with the socio-economic capital

4.4.4 Dynamics of Social Status Mobility

SE status is a complex multi-dimensional concept that can include three major components of human capital (such as education, knowledge, and skills), material capital (such as household income and equity), and social capital (such as family, friends, and membership in a community or an organization) (50). To keep our model simple, and without formulating the details of those three domains, we assume that as SE capital increases, it creates an upward mobility in the SE status (Figure 4-8). The process of change in SE status happens with a delay, (τ_4). We represent the indicated population of SE stable as S_P and the current population in the stable status as S, where S is a smooth function approaching S_P after the delay. S_P assumed to be proportional to the investment in SE capital.

$$C_S = \frac{(S_P - S)}{\tau_4} \quad (18)$$

$$S_P = I_{f_S} \times P \quad (19)$$

S_p is the *Indicated population of SE stable* and P is the *Total population*. Also, I_{f_S} shows the *Indicated fraction of SE stable* (20). This fraction applies two conditions at the same time. First, it enforces an upper bound, f_S, for the maximum number of people who can move to the S stock. We add this limit to the model because not all of the people who receive education and social capital will move to the stable status. In reality, many complex factors like residential segregations, income inequality, and inefficiencies of social programs like education systems reduce the chances of empowerment for many people (51, 52). Second, the I_{f_S} fraction shows the percentage of population that can receive SE resources in the model. To calculate this fraction, we divide the *SE capital per capita* (c_p) by the *Desired SE capital per capita*(λ_2).

$$I_{fS} = Min\left(1, \frac{c_p}{\lambda_2}\right) \times f_S \quad (20)$$

$$c_p = \frac{C}{P} \quad (21)$$

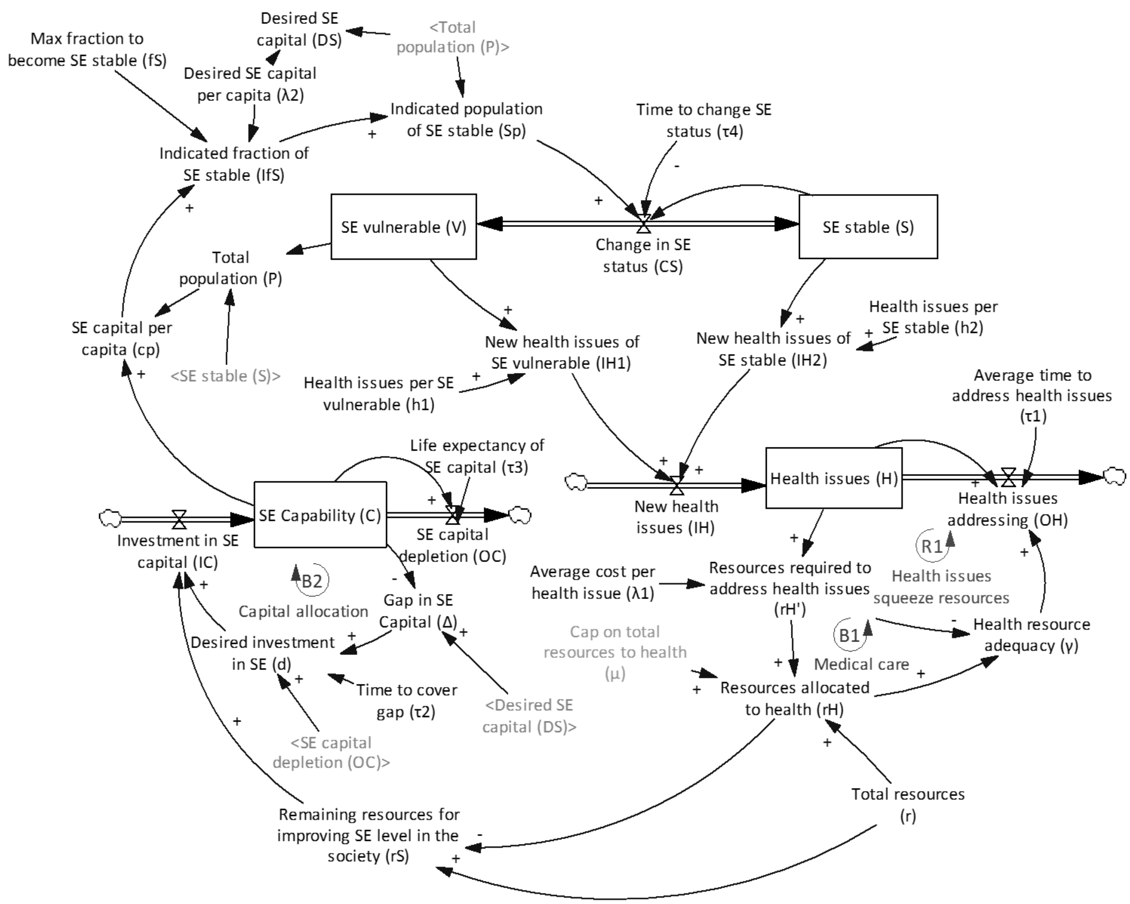


Figure 4-8 The model with the socio-economic capital

4.4.5 Resource Generation Mechanisms

One of the major assumptions of this model is that resources are generated endogenously, which means that the total amount of economic resources in a community is dependent on well-being of its residents. This is a simple representation of government General Revenue Fund (GRF) based on tax-payers' contributions. The normal contribution of each group is different for SE vulnerable and stable at τ_1 and τ_2, respectively. Since a large-scale economic crisis can affect individuals' revenues, they can reduce people's tax payments and the government's revenue too. For example, some states in the Rust Belt region like Ohio have never fully recovered from the 2000 and 2007 financial crises (53). Therefore, we design a mechanism in our model to include an *External shock* (Ω) that affect the economy $[\mathrm{t_i} = \beta_\mathrm{i} \times (1 - \Omega) \quad \text{for i} = 1,2]$ (Figure 4-9). This shock will be proportionate to the financial shock size (ϕ) for a duration of d years. The tax revenues will be calculated as follows:

$$T = T_1 + T_2 \tag{22}$$

$$T_1 = t_1 \times V \tag{23}$$

$$T_2 = t_2 \times S \tag{24}$$

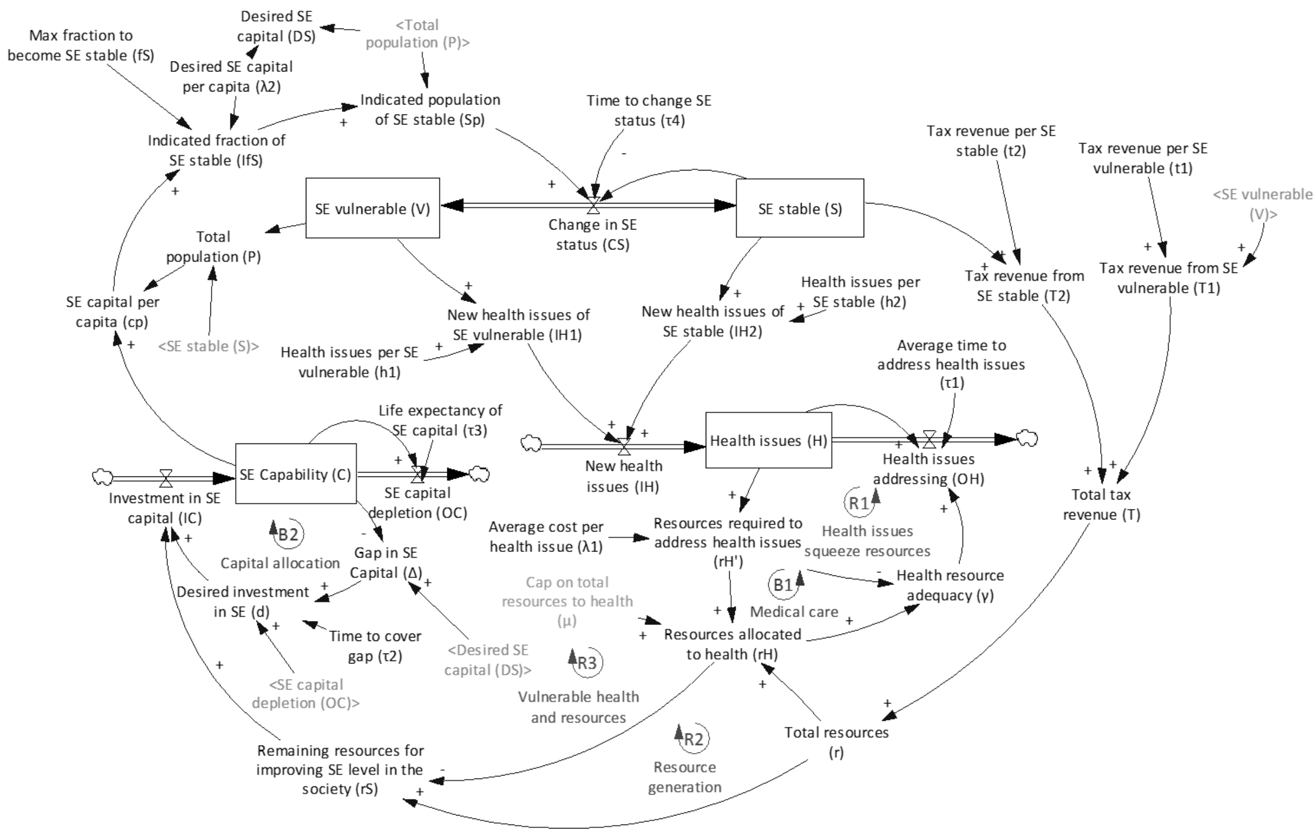


Figure 4-9 Resource generation scenarios

Figure 4-9 also shows the two biggest loops in the model, **R2** and **R3**. Loop **R2** indicates the reinforcing effect of *Investment in SE capital* (I_C) in the creation of more resources in the long-term. A society with more resources, r, has more flexibility to increase *the rate of* I_C. More I_C raises the C, which will lead to a larger *Indicated population of SE stable* (S_P). As the rate of S_P increases, more of the community will transit to the SE stable status. A community with a relatively large SE stable population has more tax revenues that can later invest in I_C. Loop **R2** can function as a virtuous cycle that propels a community further ahead or as a vicious cycle that pushes it back. Loop **R3** is another reinforcing loop that has a similar structure to loop **R2**. It shows that a community with a relatively higher percentage of SE Vulnerable (V) has to spend more of its resources on addressing health issues. The V may not have more health issues, but they are more likely to depend on the public resources for addressing them. Therefore, the community has less flexibility on allocating resources to I_C and the subsequent behavior is similar to loop **R2**.

In the next section, we analyze the model to explore different modes of outcomes over time and find the best set of parameters that result in desirable long-term outcomes in key indicators of well-being of society. These indicators include the number of health issues (H), fraction of SE vulnerable (f_V), and total resources (r). We evaluate these indicators under a 35% decline in the financial resources($\phi = 0.35$), and then a wide range of values for ϕ. We then test the model for different values of *Cap on total resources to health* (μ). Table 1 presents parameter values for the base-run simulation that generates a steady state behavior in the model.

Table 4-3 Parameter values for the base-run simulation

Parameters	Value	Dimension
V_0: Initial value of SE vulnerable population	50	person
S_0: Initial value of SE stable population	50	person
h_1: Health issues per SE vulnerable	5	Disease/Year/person
h_2: Health issues per SE stable	1	Disease/Year/person
H_0: Initial number of health issues	300	Disease
τ_1: Average time to address health issues	1	Year
τ_2: Time to cover gap	3	Year
τ_3: Life expectancy of SE capital	10	Year
τ_4: Time to change SE status	3	Year
f_S: Max fraction to become SE stable	0.5	Dimensionless
β: Constant Tax revenue	500	$/Year
β_1: Normal tax SEV	1	$/Year/person
β_2: Normal tax SES	9	$/Year/person
μ: Cap on total resources to health	1	Dimensionless
d: External shock duration	5	Year

4.5 Simulation Results

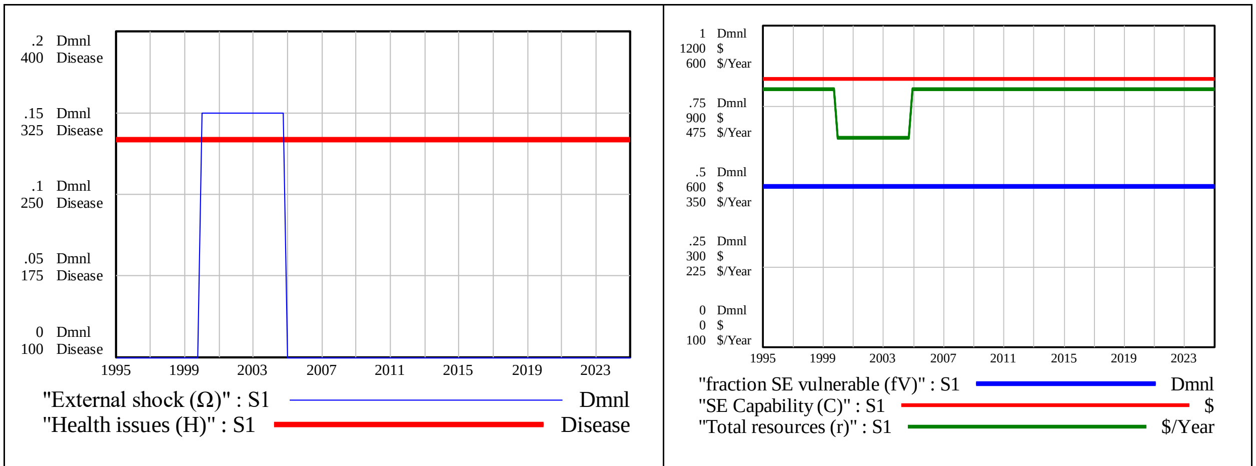
Our initial results, using the parameter values in Table 4-3, show that the outcomes can be significantly different depending on how the resources are managed in the community, especially after a considerable economic shock. We specifically look at effects of short term exogenous economic shocks, which can represent effects of changes in the global or national political system or in more intense conditions, effects of short-term economic recessions. Given the local governments' common decision model of securing resources for high priority healthcare issues, one can analyze long term resiliency of the community to such external shocks, and analyze if the system can endogenously react in a way that can quickly recover after those external events. To operationalize different conditions, we design different sets of simulation experiments (scenarios).

In short, we have three different scenarios. The first scenario (**S1**) is a small shock that drops the economic resources by 15% ($\phi = 0.15$). In our model, this condition represents an external shock in a level that slack resources of local governments are enough to absorb the negative impacts. Simulation results from the first scenario depict a behavior that is close to the basic steady-state pattern of the system. In the second scenario (**S2**), we consider a sizable economic shock, where the community loses 30% ($\phi = 0.30$) of their total resources, for a period of five years. In the third scenario (**S3**), we increase the economic shock by only five more percentage point, to 35% ($\phi = 0.35$) and report the outcomes. In this scenario, similar to **S2**, the external shock ends after five years. These two scenarios depict different conditions where the local government may or may not recover in long-run. Later, we use the last scenario, and examine policies to improve government response. To keep the analysis consistent and comparable across different scenarios, we keep the rest of the parameters equal to the values introduced in Table *4-3*.

4.5.1 S1: The Case of a Small Economic Shock

We define the first scenario to capture the behavior of the model in cases where the system undergoes a small shock. To better illustrate this condition, we assume that the community experiences a relatively small drop in economic resources. The range and size of a small shock can be different for each community, depending on their preparedness in dealing with the fluctuation in their income. The size of a small shock for our hypothetical community has a value between [0, 20] percent of the economic resources, where the shock will create similar outcomes as long as it falls in this range.

Figure 4-10 shows the simulation results for a 30-year timeline between 1995 and 2025, where an exogenous shock is introduced at year 2000 for a period of five years. The shock is formulated by a 15% drop in economic resources ($\phi = 0.15$) during the same five-year time period. As Figure 4-10.a shows, the shock will not have a significant impact on the number of *Health issues* (H) in the long-run. The system is able to sustain the previous level of performance, and does not adversely fluctuate due to this shock. Figure 4-10.b represents the socio-economic status of the community, and shows that the community has enough slack resources to address health issues without compromising the *SE Capital* (C). The continued investment in C helps keep the *fraction of SE Vulnerable* (f_V) at a steady-state level. As a result, the system does not experience any fluctuation in health outcomes that need to be addressed through public resources.



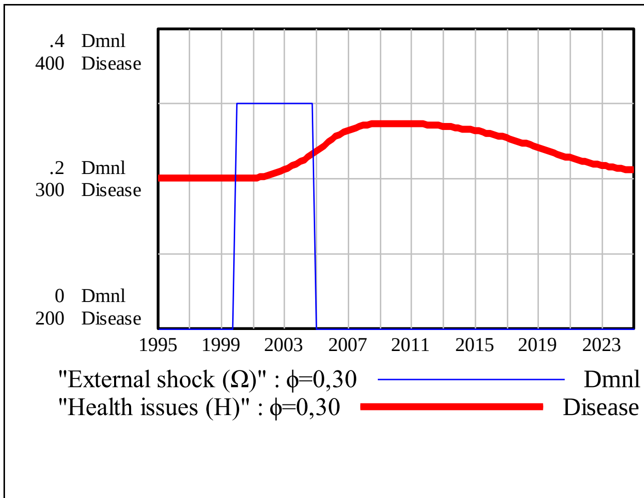
(a) (b)

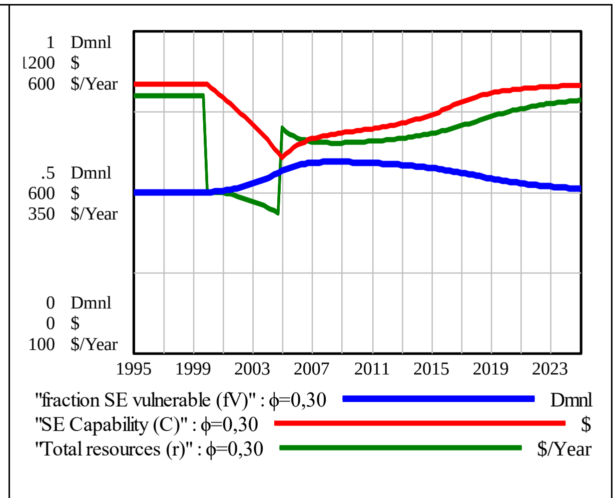
Figure 4-10 Trends of (a) health outputs and (b) SE status with a 15% economic shock

4.5.2 S2: Large Economic Shock

In the second scenario, we test the behavior of the model under a rather sizable drop in economic resources of 30% for five years ($\phi = 0.30$). Contrary to ***S1***, Figure 4-11.a shows that H is negatively impacted in a way that it deteriorates both during and well after the shock. A few years after the shock is introduced, we see an increasing trend in the number health issues, and the trend lasts long after the economic shock has ended. In this case, and in the long-run, the community can address all health issues with available resources; therefore, the *Health resource* *adequacy* (γ) does not change, and remains sufficient for addressing H. The main reason behind this increase in the stock of H is the increase in the *fraction of SE vulnerable* (f_V).

As can be seen in Figure 4-11.b, the economic shock decreases total resources, which are limited and mainly spent on addressing health issues. This will drive the resources further away from the SE Capital (C). Consequently, more people will be dependent on public resources for addressing their health issues, and H increases accordingly. This increase, however, is only temporary—the system slowly regenerates its resources, which can be allocated to C after the shock. As a result, the system goes back to its normal state after f_V drops to the levels it had before the shock.





(a)	*(b)*

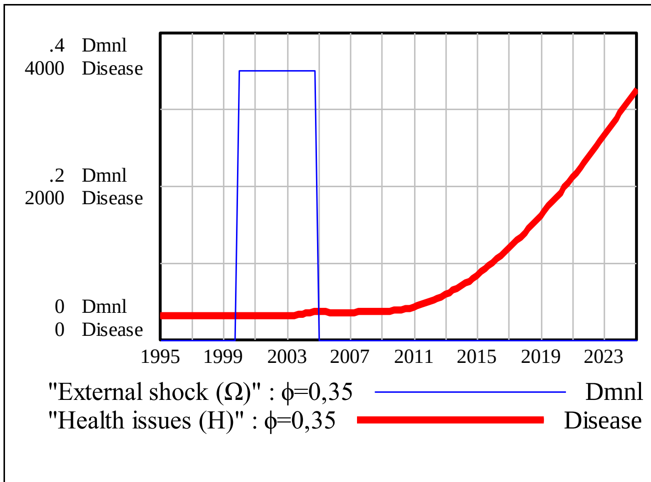
Figure 4-11 Trends of (a) health outputs; (b) SE status with a 30% economic shock

4.5.3 S3: Just a Little Larger Economic Shock than S2

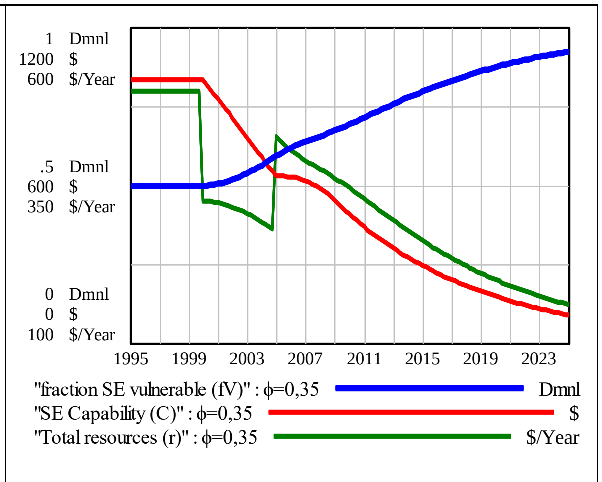
In the previous cases, the government's decision rule of how to allocate resources had relatively little to no adverse long-term consequences on the health outcomes, or at most it had short term effects. However, this is not always the case, and if the resources are not managed properly in certain conditions with larger external shocks, the system's dynamics can create vicious cycles with adverse consequences. In the third scenario, we consider a case where the economic shock is only 5% more than the second scenario. This is relatively a small step in increasing the economic shock, however, it creates significantly different outcomes. Figure 4-12.a depicts the long-term effects of a 35% decline in economic resources $(\phi = 0.35)$ on the number of health issues (H). The graph shows that H is increasing with an exponential trend, and in two phases.

In the first few years after the shock, there is a small jump in the H as a result of a sharp decline in *health resource adequacy* (γ). Figure 4-12.a indicates that the investment in *SE capital* (C) declines during and after the shock, which leads to an increase of the *fraction of SE Vulnerable* (f_V) in the community.

The main consequences of the decline in f_V is that it turns the reinforcing loop of investment in SE to a vicious cycle in the community in which the resources are generated less and are spent more on addressing immediate health issues—loops **R2** and **R3,** respectively. The result of these interacting forces creates the second phase of increase in H after a longer period of time. In this phase, the system goes through a downward spiral in health status. This pattern emerges because the reinforcing loops **R2** and **R3** become dominant in the form of vicious cycles, which leads to significant changes in the long-term outcomes. As the f_V increases, *Total resources* (r) declines continuously. In addition, the community faces a larger number of new health issues that has to address through public funds. A smaller r creates a backlog in H and further reduces the adequacy of health resources for addressing the health issues. The community gets locked into shifting more and more resources away from the SE capital as the number of H grows. The simulation model in Figure 4-12.b clearly shows that the overshoot of H happens when the community underinvests in the socioeconomic capital. Consequently, these dynamics create unexpected outcomes for policymakers where they spend even more on addressing health issues, but the system ends up with more health problems than before.



(a)



(b)

Figure 4-12 Trends of (a) health outputs; (b) SE status with a sizable economic shock in the endogenous scenario

4.5.4 Broader Range for the Effect of Economic Shocks

Economic shocks have different durations and effect sizes. To have a comprehensive view of the model behavior in all these situations, we run 1000 simulations for different values of external shocks (ϕ) with different length of times. The economic shocks are all introduced at year 2000 and we would like to assess their effect 25 years later. We show the final number of *Health issues* (H), which happens in 2025, under three different shock durations in Figure 4-13. To stay consistent with previous cases, we assume that the system has the same resource allocation mechanism, similar to sections 4.5.1 to 4.5.3, in which the resources are allocated to health with higher priority.

Figure 4-13 depicts effects of a wide range of economic shocks on healthcare with differing size and period of shock. The figure shows that tipping point exists for a wide range of conditions, and consistent with the intuition, larger and longer economic downturns lead to more long-term problems in health, and potentially a hard recovery for the system.

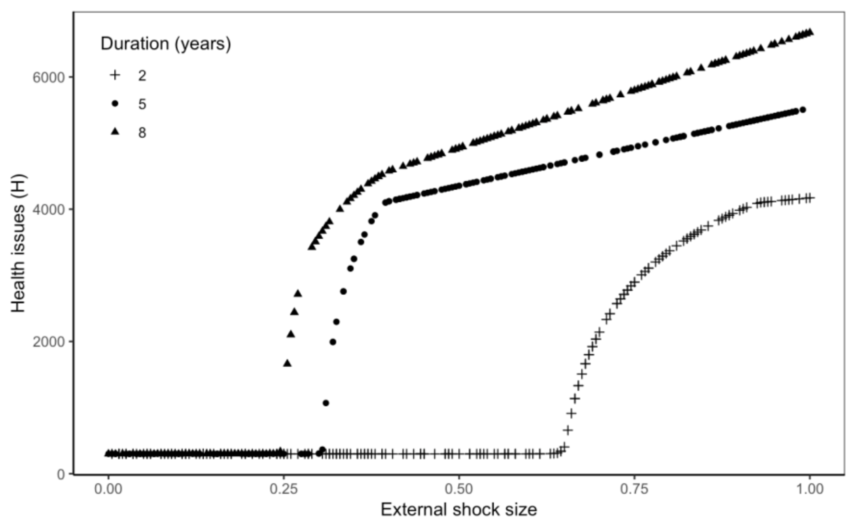


Figure 4-13 Effect of short-term economic shocks with differing sizes and durations, all started in 2000, on health issues long after the shock ended, in 2025.

Of course, such behavior in the real world is uncommon, and it happens in the model due to simplifying assumptions. However, these trends show the potential of falling into the vicious cycles that throw the community in a downward spiral of deterioration in health outcomes. Adopting a long-term horizon for evaluating the impacts of decisions plays an important role in making the proper choices that shape health outcomes. Thus far, we had an assumption that the community allocates the resources to the health issues and then it invests the remaining for building SE capital. Next, we relax this assumption and discuss how different allocation policies can shape the outcomes in the long-run.

4.5.5 Assessing different decisions on allocation of resources

In this part, we evaluate the effect of different allocation policies on health outcomes after a significant economic shock ($\phi = 0.35$). The shock starts in the year 2000 and lasts for five years. The parameters and the size of the shock remain fixed for all cases and the resources are generated endogenously. We change the size of *Cap on total resources to health* (μ) and evaluate the effects of these different allocation decisions on outcomes.

4.5.5.1 Optimal range for allocation decisions

Testing different scenarios highlights three major dominant modes of behavior in the final outcomes of the model after a major decline in the economic resources. The first mode entails a condition that the outcomes become better before worse. This mode represents a condition with a deteriorating trend that has a high and growing number of H at the conclusion of the simulation period. The second mode shows a permanent shift in the status of the system, where the outcomes get worse but remain constant toward the end of the simulation time. The third mode is the return to normal, in which the community first experiences a rather difficult period of time in terms of undesirable health outcomes and then goes back to the steady-state comparable to the initial phases of the simulation. Figure 4-14 compares the number of *Health issues* (H) for a set of experiments where each line represents the result of a simulation with a different *Cap on total resources to health* (μ).

First mode: Heath problem escalation?

Too much priority on health issues: Two of the simulations in which there is no limit for the resources to health and an 85% cap on the total resources to health—$\mu = 100$ and $\mu = 85$, respectively—fall into this category (Figure 4-14). Both of cases show that H first has an improvement, but only at the expense of deteriorating the SE status of the population that sends the community towards a spiral of deterioration in the *fraction of SE vulnerable* (f_V). The number of H declines for a short period of time after the shock. In this stage, the government has health problems under control, but only at the expense of deteriorating the SE status of the population that sends the community towards a downward spiral. During this period, the community has to overspend in addressing health issues that creates a problematic condition. Resources are allocated to addressing health issues with a higher priority, which drives the necessary resources away from the *SE Capital* (C). The squeezed C leads to an increase in the *fraction of SE vulnerable* (f_V). Consequently, health issues increase, and because a larger fraction of population is dependent on public funds, it further shrinks government resources. In the same time, government's earning from tax payers declines as the socio-economic status of citizens decline.

Too little priority on health issues: The 50 percent cap $(\mu = 0.50)$ shows a similar pattern in increasing H, but with a different underlying reason. This time, underspending for health creates the growing backlog of H. In this scenario, the SE status of the population does not change, but the community experiences a shortage of resources that are necessary for addressing health issues. The community experiences a rise in H even before an economic shock, and the decline in the resources only exacerbates the situation. Appendix M shows the graph of health issues and SE status for each simulation.

Second mode: shift to a new (worse) steady state

The second mode of the simulation results includes a state where the stock of health issues (H) increases to a new steady state level (Figure 4-14). In this case $(\mu = 60)$, f_V stays the same (see Appendix M) and the community does not allocate enough resources for addressing health issues leading to accumulation of H. The system reaches a new equilibrium in this condition. The community can mitigate the problem by allocating more resources to health without jeopardizing the SE Capital (C).

Third mode: full recovery

A cap on health resources that is close to the optimal value (e.g. $\mu = 0.70$ or $\mu = 0.80$) creates a two-phased feedback in the system. First, the pattern of the outcomes changes to a worse before better. The first phase includes a sharp increase in the number of H during the shock. This jump stems from the lack of health resources to address the issues for the duration of the economic shock, which creates the backlog in H. This shock is bigger than what we saw in cases with larger allocations to health (e.g. $\mu = 1.00$). However, this larger jump is followed by an improvement to the normal levels after the shock.

The underlying reason for the return of the outcomes to the normal level is that the community has invested enough SE Capital (C) to keep the fraction of SE vulnerable (f_V) in balance. Therefore, after the effects of economic shock disappears, the community can replenish the resources and address all of the backlogged H. The choice of the cap size is dependent on the current state of the system.

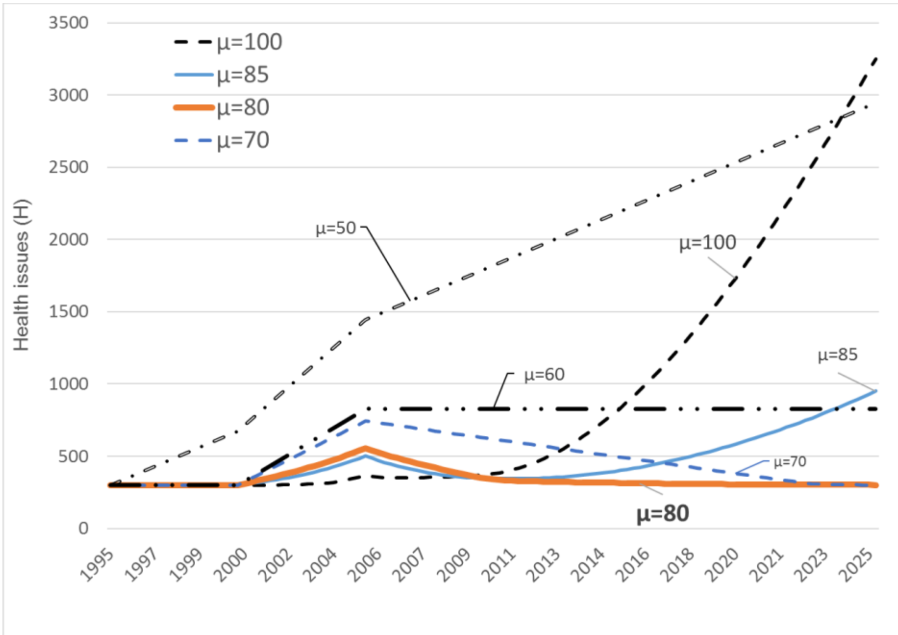


Figure 4-14 Simulation results showing the number of Health issues (H) under different allocation policies

The set of simulation experiments shown in Figure 4-14 highlights three key features of the system's dynamics. First, allocation of resources to health has a tipping point. Keeping the cap of resources to health around the tipping point results in the long-run desirable outcomes (e.g. $\mu^* = 0.70$ or $\mu^* = 0.80$). However, attempts to further improve the outcomes by moderate amounts can result in better before worse patterns (e.g. $\mu = 0.85$). Health outcomes improve at first, but only at the expense of deteriorating SE status that drives the community toward a deteriorating condition even though resources to health are higher than before. The decision to increase resources just a little more to mitigate the negative health outcomes may unexpectedly trigger a vicious cycle sending the system in a state where it is difficult to recover from ($\mu = 1.00$). Second, once the deterioration of the SE status for a large fraction of the population begins, it is self-sustaining. Once past the tipping point, the increasing *fraction of SE vulnerable* (f_V) depend more on the public funds, shifting more resources away from *SE Capital* (C), and the vicious cycle that deteriorates the outcomes gains more momentum. Third, the health spending below the tipping point also returns obnoxious outcomes. An underinvestment can send the health issues to a new worse state ($\mu = 0.60$). The number of health issues are more than before, because they are not addressed at the rate that they are introduced. If the health resources get squeezed further, it creates a backlog of health issues—resembling to the chronic diseases—that is difficult, if not impossible, to address (e.g. $\mu = 0.50$). A large underinvestment in health such as $\mu = 0.50$ or less creates an unsteady state before even any economic shock occurs. The community experiences a growing number of health problems that are not addressed. These issues arise from varying sources such as long wait times in hospitals due to the inadequate number of physicians or hospital beds. Also, many people develop disease complications as a result of lack of routine visits, immunization, and other inefficiencies in the health system that prevent people from receiving a proper health care.

Figure 4-15 illustrates the effect of different allocation decisions on the long-run health outcomes. The number of health issues are shown in the conclusion of simulation time, i.e. year 2025. These outcomes are sensitive both to under and over investment. Allocation of resources with values in less than 65% of all economic resources ($\mu < 65$) creates a backlog of health issues. On the other hand, allocating more than 85% of economic resources on health will create even more health issues at the end. The underinvestment is clearly more troubling than overinvestment, however, neither are optimal. Optimal allocations to health ($\mu \in [65\%, 85\%]$) create a sustainable long-term balance in the outcomes, because it leaves enough resources for building SE capital.

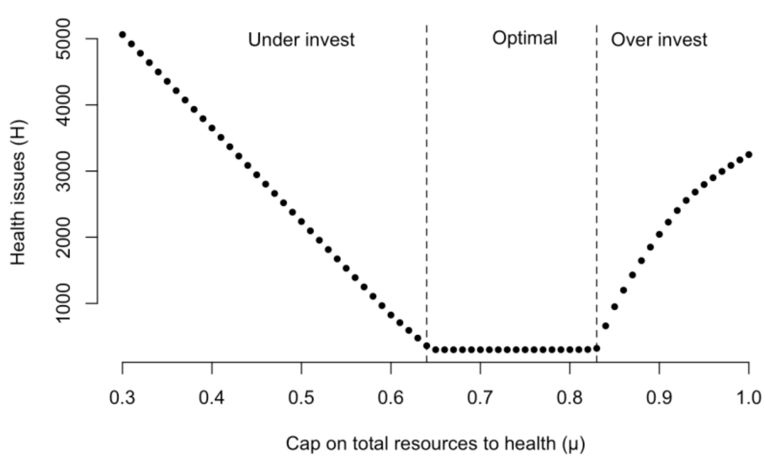
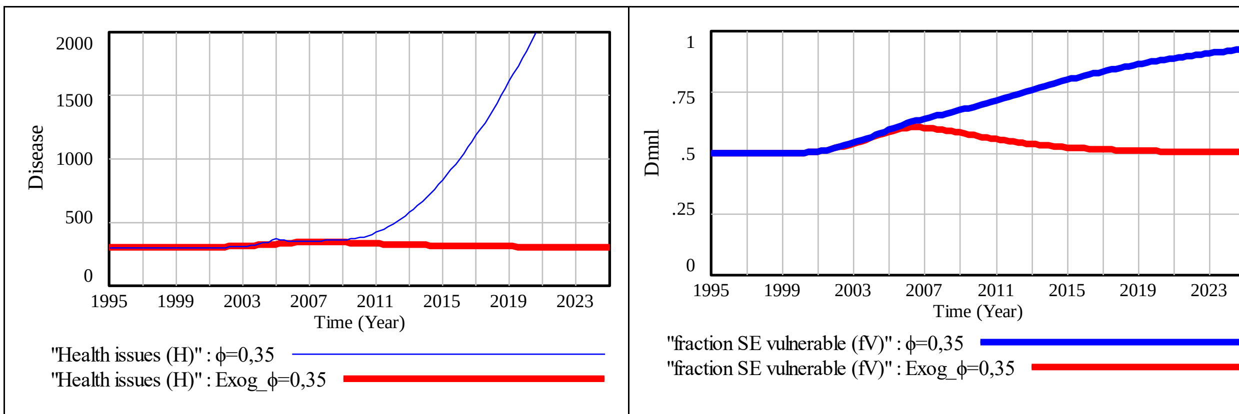


Figure 4-15 Relationship between health outcomes (H) in 2025 and allocation decisions (μ)

4.5.6 Model analysis: The role of endogenous resources

In this section we would like to shed more lights into the outcomes of the model. To that end, there are two major reinforcing loops in the system, and it might be hard to intuit which one is more dominant in the vicious cycle model driving healthcare outcomes down. We test the effects by switching them off and running the model for conditions where a part of the model is deactivated.

In this simulation, we consider the economic resources as an exogenous variable that are not generated by the community and it has a steady stream that is supplied at a constant rate of τ per year and independent of the residents. To have a fair comparison and stay consistent with the endogenous approach, we assume that an *External shock* (Ω) will also affect the *Total resources* (r). As Figure 4-16.a suggests, in contrast to the endogenous approach, the number of H does not increase exponentially after a 35% decline in the resources if the system was financed by exogenous resources. In this case, the number of H increases for a shorter period time and it returns to normal levels faster compared to smaller shocks in the endogenous approach. Figure 4-16. compares the resources after an economic shock in a system with both the endogenous and exogenous resources. This increase happens because the shock decreases the *SE Capital* (C), as suggested by Figure 4-16.b, and it subsequently increases the *fraction of SE vulnerable* (f_V) in the community. However, since the resource levels, r, quickly go back to the normal levels after the shock, the f_V subsequently returns to the normal level and the number of H drops. This test explains the effect of resources as an endogenous variable. It showed that it is important to include the endogeneity of resources when they are indeed endogenous.



(a)	(b)

Figure 4-16 Comparison of (a) fraction of SE vulnerable; (b) resources in both endogenous and exogenous approach after a similar economic shock

4.6 Empirical Evidence

While the main objective of this manuscript is to develop a new explanation, which is a capability-trap-like theory for the healthcare system, we would like to take another step to examine the consistency of our claims and simulation results with data. In this section, we check the implications of our study and whether they are consistent with the data. We perform this analysis in two steps. First, we run several regression models to explore associations between lagged healthcare ranking and two major spending categories of K-12 education and Medicaid. The hypothesis is that more K-12 spending should result in higher healthcare ranking, potentially with a few years of delay. Second, we take a closer look at the trend of healthcare spending in each state over time and analyze the health ranking accordingly. The objective of this second analysis is to see if there is actually variation of responses and bifurcation of outcomes after the US economic recession.

4.6.1 Fixed-effect Regression Model

We use a fixed-effect, multiple linear regression analysis to check the association between the health ranking and state-level explanatory variables from 1991 to 2013. The serial correlation within states is controlled by clustering standard errors for each state. In addition, we control for year effect. The dependent variable is state health ranking and the independent variables are government spending in K-12 Education and Medicaid as a percentage of the GRF fund with a one-year lag. Note that a lower health ranking shows a better status across 50 states.

Table 4-4 shows the results of the fixed-effect model. These results are in strong agreement with observations from the trend of Ohio budget allocation and health ranking. Spending in education improves the health ranking for next year, and healthcare spending paradoxically has an inverse effect on health ranking. The regression results show that a one-percentage-point increase in Education spending is significantly ($p<.05$) associated with a 15% decline (improvement) in the next year's health ranking. On the other hand, a one-percentage-point increase in Medicaid spending is significantly ($p<.05$) associated with a 13% increase (deterioration) in the next year's health ranking. We also checked lags of two, three, and four years and reported the results Table 4-4. These results are similar to one-year lag.

Table 4-4 Parameter Estimates for Fixed-Effect Regression Models of the Preterm-Related Mortality Rate

Variables	**Regression *(n=1093)***	
*Health Ranking **(year= t+1)***	Estimate	*P-Value*
% of Education Spending	-0.15	0.042**
% of Medicaid Spending	0.133	0.049**
*Health Ranking (**year= t+2**)*		
% of Education Spending (year= t)	-0.13	0.06*
% of Medicaid Spending (year= t)	0.11	0.07*
*Health Ranking (**year= t+3**)*		
% of Education Spending (year= t)	-0.10	0.07*
% of Medicaid Spending (year= t)	0.11	0.06*
*Health Ranking (**year= t+4**)*		
% of Education Spending (year= t)	-0.12	0.02**
% of Medicaid Spending (year= t)	0.12	0.04**

'**' Sig at 0.05 '*' Sig at 0.10

Note: Data represented unweighted averages across states, and should not be interpreted as national estimates.

4.6.2 Evidence for variation in health outcomes

In another step of the empirical analysis, this section aims to explore the relationship between government healthcare spending and state overall health ranking. To perform this analysis, we divide states into different groups. The grouping is based on the overall health rank of the state in 1991 and 2013. Figure 4-17 shows the rank of each state in 1991 and 2013.

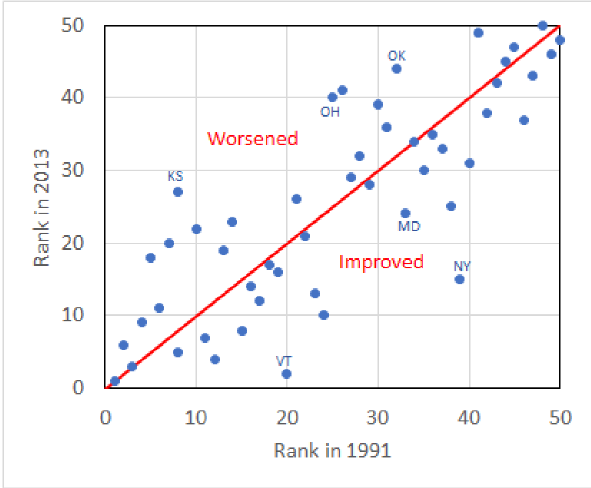
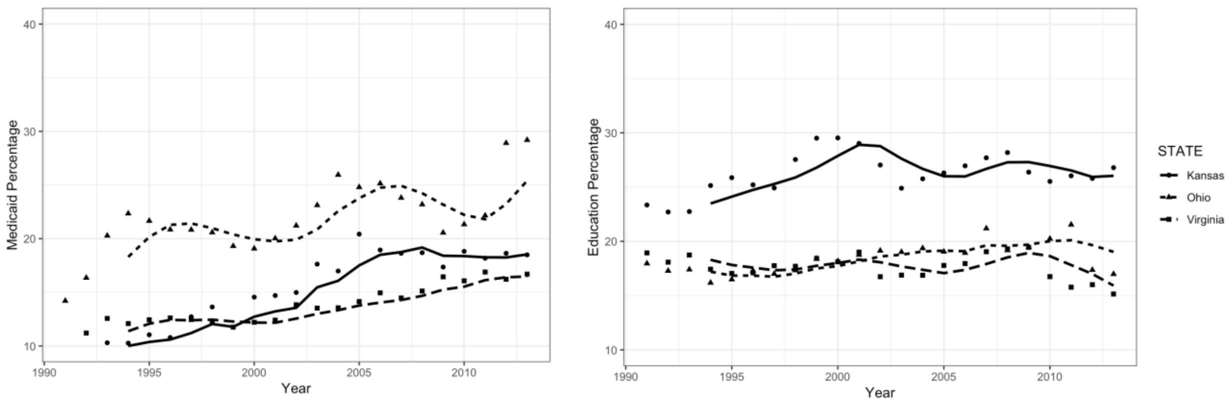


Figure 4-17 Comparison of state's health ranking in 1991 and 2013

In this analysis, we focus on the states that moved from a ranking better than average to worse or vice versa. These states either experienced a significant deterioration (G1) or an improvement (G2) in their health ranking over 23 years. Therefore, we provide scatterplots for the percentage that each state spends on education and Medicaid for these two groups—see Appendix N for the trends of groups G1 and G3. To better track the trend of spending over time, we also add a four-year moving average line for each state.

G1: Good start, but a poor end

Only three states experienced a shift from better than average to worse in their health ranking. Figure 4-18.a shows that all of these three states have spent more and more on health over time, and in return, their health ranking deteriorated even further. On the other hand, these three states spent either the same or less on education (Figure 4-18.b).

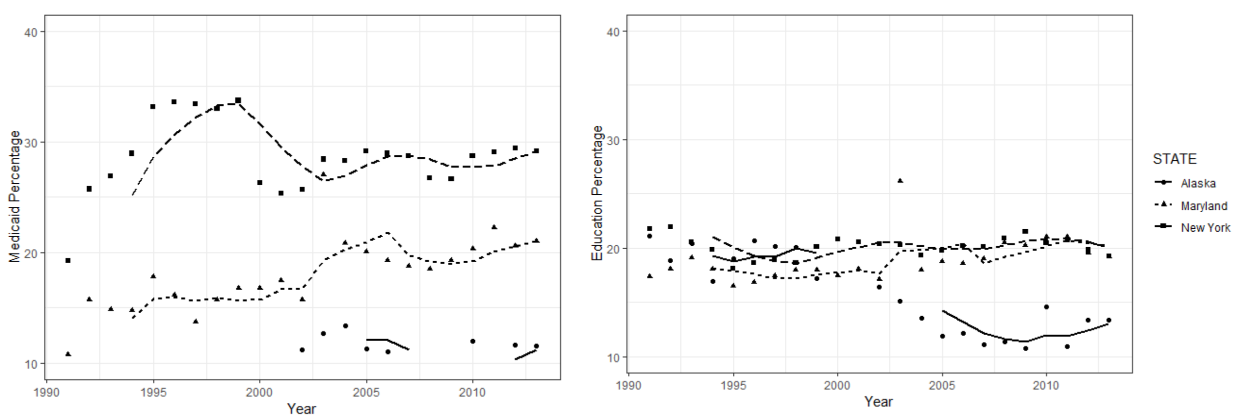


a) *b)*

Figure 4-18 Trends in a) Medicaid and b) Education spending for states that experienced a drop in health ranking

Group 2: Poor start, but a good end

Two states plus Alaska have managed to improve their health ranking (Figure 4-19). Both Maryland and New York have not increased their health spending substantially. On the other hand, they had a slightly positive slope in their spending on education. Note that we do not analyze Alaska, because it has a different financing system due to different geography and abundant natural resources.



a) *b)*

Figure 4-19 Trends in a) Medicaid and b) Education spending for states that experienced drop in health ranking

4.7 Discussion and Conclusion

In this study, we examined why some states like Ohio suffer from poor health outcomes, despite massive investments in their healthcare sector. By using the evidence from Ohio and the two comprehensive theories from the literature, we offer a generalizable theory along with an analytical framework to explain why governments struggle to fix health problems. For the first time, we integrate the theory of Lifecourse and Capability Trap to propose a new systematic framework to evaluate the decisions that aim to address major health issues. We provide a simulation-based model that can be used as a framework to support the process of strategic decision making on how resources should be allocated to a healthcare sector and also towards building SE capabilities. We then tested the model under different conditions and scenarios. The empirical analysis of 50 states over the past 23 years supports the modeling results.

Our simulation results reveal several non-linear behaviors with a tipping point around which small changes in resource allocations to healthcare can lead to significantly different outcomes. These results suggest that there is an optimal range for these allocations. Proper management of resources become more important, especially when there is a major economic downturn. This is mainly because economic shocks significantly reduce public funds and increase the number of people who are dependent on such resources.

Attempts to further increase healthcare resources outside of the optimal range can create short-term improvements in health outcomes, but can send the system into a spiral of deterioration that is difficult, if not impossible, to recover from. Balancing these trade-offs become complex and often vexing for policymakers, as the success in improving health outcomes sends a salient confirmation that their interventions are paying off. However, these temporary improvements often mask the underlying, less obvious degradation of SE capabilities. Excessive allocation of limited resources to the healthcare sector can improve health outcomes, especially after an economic downturn. Despite this fact, outcomes may appear better before worse patterns. The consequences of shifting resources to the healthcare sector manifest with a delay, and will come at the expense of developments in socioeconomic (SE) capabilities. Investments in categories like education can build SE capabilities that contribute to the developmental scaffolding of health, which ultimately improves health trajectories, particularly for children (27, 35). Protective factors such as high-quality preschools or access to appropriate nutritional supports are health determinants that play a crucial role in shaping individual health outcomes in adulthood (27).

These results are similar to the underlying findings of different capability trap studies conducted on sociotechnical systems and service industries (42, 44, 45, 54).

On the other hand, our simulations show that underinvestment in the healthcare sector can result in even more severe consequences than overinvestment. Failure to allocate necessary resources to healthcare, especially during sensitive periods, can result in compounded losses in health potential and reserves (27). Failing to provide enough access to free immunization or basic healthcare coverage that can help in early detection of diseases creates cascading costs and problems for a healthcare system. These results are also in line with the theory of Lifecourse. This theory considers a big part of the healthcare expenditure as an investment in the individual's health capital, which adds to their long-term reserves. More particularly, a system with underinvestment in health misses the opportunity of primary care and prevention, which drags the community into expensive treatments. Health investments are more important for infants and children, as early childhood is the time for intensive health developments and screening, as well as detecting disorders (55, 56).

Our empirical analysis provides evidence for the validity of the simulation results. In this analysis, we used a fixed-effect regression model to check the association between healthcare ranking and healthcare and education spendings for each state. The regression results show that allocating a higher percentage of the state's General Revenue Fund (GRF) to education is significantly associated with improvements in the healthcare ranking in the following years. On the other hand, allocation of a higher percentage to the healthcare sector in a given year is significantly associated with a deterioration in the healthcare ranking for subsequent years. These empirical analyses echo the findings of other theories, which suggest that the main determinants of health trajectories are the social scaffolding that addresses upstream roots of health (27, 57).

Our model's implication can also explain the trends of health outcomes in many states. For example, the state of Ohio with one of the most interesting patterns in health outcomes in the past three decades suffers from overinvestment in the healthcare sector at the expense of underinvestment in SE capabilities. Their health ranking was better than the national average in 1990s, but they suffer from poor outcomes in many key health and behavioral indicators like infant mortality and drug overdose deaths in 2018 (58-60).

Healthcare spending of Ohio as a percentage of the GRF is one of the highest across all 50 states (61). Many counties in Ohio like Cuyahoga (greater Cleveland area) experienced an economic shock in 2001. The economic shock was caused by the layoffs and wage-cuts of auto manufacturing companies like Ford and General Motors. Other national shocks like the dotcom crash of 2000-2002 escalated the problem even further, and the per capita income of Ohio residents dropped from 1999 to 2008 (62). On the other hand, policymakers in Ohio have squeezed the overall resources by cutting income-tax rates by a third since 2005 and gave tax deductions for income from limited liability companies and other businesses. These cuts and deductions costs Ohio more than one billion dollars a year—5% of their GRF—annually (63).

All of these events have resulted in a status where both families and government have to spend a larger portion of their resources for treating health problems and less than before on education and building social capital. Ohio residents have long suffered from insufficient and unequal distributions of educational funding despite the many court battles over the years that started in 1990s (64, 65). Missing the opportunities to invest in SE capabilities affected the low-income population even harder in Ohio. The performance index score of schools in districts with the highest property value is 33 percent higher than districts with the lowest property value in 2018 (66). The children living in poor districts are less likely to attend college, which will likely lead to periods of unemployment and financial stress in their adult life. Unless the community provides occupational opportunities and supports for the most at-risk youth, risks will continue to multiply over their lifespan and further declines their health trajectory (27).

We are not the first to point out this systemic problem. Some studies attempted to understand why the US is the sickest of wealthy nations by using the Lifecourse theory approach (67). Fine and Kotelchuck (68) discuss a strategic approach on timing and planning for healthcare with respect to the theory of Lifecourse. Others attempted to tackle the American healthcare paradox by taking a systematic approach—the Capability Trap— in a qualitative model and explain why spending more is getting us less (45, 69). Homer, Milstein (70) assessed a regional health system and suggested that the best mix of intervention is the one that improves the socio-economic status. Our study provided a more comprehensive perspective that connects all these theories and then we presented detailed empirical evidence to support these theories and the simulation results. In addition, our model includes the endogeneity of resource generation in the community for the first time in a dynamic study.

This study was subject to some limitations. We considered universal education as the only mechanism for improving the SE status of individuals or families in the long run. Despite the undeniable effect of education on the prosperity of families and society as a whole, there are other ways to create sustainable improvements. The measures for creating a healthy and sustainable ecosystem also include policies and regulations that increase gender equity, long-term food security, access to clean water, and basic sanitation (71). The other limitation of this study was the level of aggregation in our model. We categorized all of the SE capability and healthcare spending in two sectors. Although we considered an average effect for each sector, the model can be further expanded to include detailed policies. For example, a future research can focus on expanding the SE capability section. Comparing the trend of each state in allocation of resources to each subcategory can improve our understanding about the effect of different interventions on improving health of the community. Another potential research area is investigating the effect of health expenditure based on preventive and treatment categories. The allocation of resources to preventive services has been optional for states even for the low-income families under Medicaid insurance (72). Therefore, studying state's policies and adding such details can further increase our understanding about interventions and why states have performed significantly different over time.

The purpose of this study was to provide an explanation of why some states struggle to improve their health outcomes. Our simulation results point to the importance of keeping the balance between healthcare and social investments. Overall, our analytical framework can support the robust decision-making process of policymakers, and contribute to the national debate on the optimal allocation of resources to the healthcare sector.

References

1. Squires D, Anderson C. U.S. health care from a global perspective: spending, use of services, prices, and health in 13 countries. Issue Brief (Commonw Fund). 2015;15:1-15.

2. How does health spending in the U.S. compare to other countries? [Internet]. Peterson-Kaiser Health System Tracker. 2018 [cited Dec 7, 2018]. Available from: https://www.healthsystemtracker.org/chart-collection/health-spending-u-s-compare-countries/#item-start.

3. Rosenberg BL, Kellar JA, Labno A, Matheson DH, Ringel M, VonAchen P, et al. Quantifying geographic variation in health care outcomes in the United States before and after risk-adjustment. PLoS One. 2016;11(12):e0166762.

4. Berwick DM, Hackbarth AD. Eliminating waste in US health care. Jama. 2012;307(14):1513-6.

5. Schroeder SA. We can do better—improving the health of the American people. New England Journal of Medicine. 2007;357(12):1221-8.

6. Woolhandler S, Campbell T, Himmelstein DU. Costs of health care administration in the United States and Canada. New England Journal of Medicine. 2003;349(8):768-75.

7. Lantz PM, Lichtenstein RL, Pollack HA. Health policy approaches to population health: the limits of medicalization. Health Affairs. 2007;26(5):1253-7.

8. Wilson N, Couper I, De Vries E, Reid S, Fish T, Marais B. Inequitable distribution of healthcare professionals to rural and remote areas. Rural Remote Health. 2009;9(1060):1-21.

9. Wilkinson RG, Pickett KE. The problems of relative deprivation: why some societies do better than others. Social science & medicine. 2007;65(9):1965-78.

10. Conrad P, Barker KK. The social construction of illness: Key insights and policy implications. Journal of health and social behavior. 2010;51(1_suppl):S67-S79.

11. Swift EK. Guidance for the National Healthcare Disparities Report: National Academies Press; 2002.

12. Gornick ME. Measuring the effects of socioeconomic status on health care. Guidance for the National Healthcare Disparities Report. 2002:45-73.

13. Lantz PM, House JS, Lepkowski JM, Williams DR, Mero RP, Chen J. Socioeconomic factors, health behaviors, and mortality: results from a nationally representative prospective study of US adults. Jama. 1998;279(21):1703-8.

14. Szanton SL, Gill JM, Allen JK. Allostatic load: a mechanism of socioeconomic health disparities? Biological Research for Nursing. 2005;7(1):7-15.

15. Grech V. Evidence of socio-economic stress and female foeticide in racial disparities in the gender ratio at birth in the United States (1995–2014). Early human development. 2017;106:63-5.

16. Adler NE, Ostrove JM. Socioeconomic status and health: what we know and what we don't. Ann N Y Acad Sci. 1999;896:3-15.

17. Homer JB, Hirsch GBJAjoph. System dynamics modeling for public health: background and opportunities. 2006;96(3):452-8.

18. Milstein B, Homer J, Hirsch G. Analyzing national health reform strategies with a dynamic simulation model. American journal of public health. 2010;100(5):811-9.

19. Mahamoud A, Roche B, Homer J. Modelling the social determinants of health and simulating short-term and long-term intervention impacts for the city of Toronto, Canada. Social science & medicine. 2013;93:247-55.

20. Hosseinichimeh N, Kim H, Ebrahimvandi A, Iams J, Andersen DJSR, Science B. Using a Stakeholder Analysis to Improve Systems Modelling of Health Issues: The Impact of Progesterone Therapy on Infant Mortality in Ohio. Systems Research and Behavioral Science. 2019.

21. Ghaffarzadegan N, Larson RC, Fingerhut H, Jalali MS, Ebrahimvandi A, Quaadgras A, et al. Model-Based Policy Analysis to Mitigate Post-Traumatic Stress Disorder. Policy Analytics, Modelling, and Informatics: Springer; 2018. p. 387-406.

22. Hosseinichimeh N, MacDonald R, Hyder A, Ebrahimvandi A, Porter L, Reno R, et al. Group model building techniques for rapid elicitation of parameter values, effect sizes, and data sources. System Dynamics Review. 2017;33(1):71-84.

23. Ghaffarzadegan N, Ebrahimvandi A, Jalali MS. A Dynamic Model of Post-Traumatic Stress Disorder for Military Personnel and Veterans. PLOS ONE. 2016;11(10):e0161405.

24. Sterman JD. Learning from evidence in a complex world. American journal of public health. 2006;96(3):505-14.

25. Brownson RC, Chriqui JF, Stamatakis KA. Understanding evidence-based public health policy. American journal of public health. 2009;99(9):1576-83.

26. Luke DA, Stamatakis KA. Systems science methods in public health: dynamics, networks, and agents. Annual review of public health. 2012;33:357-76.

27. Halfon N, Larson K, Lu M, Tullis E, Russ S. Lifecourse health development: past, present and future. Maternal and child health journal. 2014;18(2):344-65.

28. Porter R. The Greatest Benefit to Mankind: A Medical History of Humanity (The Norton History of Science): WW Norton & Company; 1999.

29. Breslow L. Health measurement in the third era of health. American journal of public health. 2006;96(1):17-9.

30. Kingston D, Sword W, Krueger P, Hanna S, Markle-Reid M. Life course pathways to prenatal maternal stress. 2012;41(5):609-26.

31. Kotelchuck MJM, journal ch. Evaluating the healthy start program: A life course perspective. 2010;14(5):649-53.

32. McEwen BS. Protective and damaging effects of stress mediators. 1998;338(3):171-9.

33. Wadhwa PD, Entringer S, Buss C, Lu MC. The contribution of maternal stress to preterm birth: issues and considerations. Clinics in perinatology. 2011;38(3):351-84.

34. Messer LC, Kaufman JS, Dole N, Savitz DA, Laraia BA. Neighborhood Crime, Deprivation, and Preterm Birth. Annals of Epidemiology. 2006;16(6):455-62.

35. Lu MC, Halfon N. Racial and ethnic disparities in birth outcomes: a life-course perspective. Maternal and child health journal. 2003;7(1):13-30.

36. Hogue CJR, Bremner JD. Stress model for research into preterm delivery among black women. American journal of obstetrics and gynecology. 2005;192(5):S47-S55.

37. Wellen KE, Hotamisligil GS. Inflammation, stress, and diabetes. The Journal of clinical investigation. 2005;115(5):1111-9.

38. Nandi A, Glymour MM, Subramanian S. Association among socioeconomic status, health behaviors, and all-cause mortality in the United States. Epidemiology (Cambridge, Mass). 2014;25(2):170-7.

39. Mirowsky J. Education, social status, and health: Routledge; 2017.

40. Mirowsky J, Ross CE. Education, learned effectiveness and health. London Review of Education. 2005;3(3):205-20.

41. Repenning NP, Sterman JD. Nobody ever gets credit for fixing problems that never happened: creating and sustaining process improvement. California management review. 2001;43(4):64-88.

42. Repenning NP, Sterman JD. Capability traps and self-confirming attribution errors in the dynamics of process improvement. Administrative Science Quarterly. 2002;47(2):265-95.

43. Sterman J. Stumbling towards sustainability. Leading Sustainable Change. 2015:50-80.

44. Lyneis J, Sterman J. How to save a leaky ship: Capability traps and the failure of win-win investments in sustainability and social responsibility. Academy of Management Discoveries. 2016;2(1):7-32.

45. Landry E, Sterman J, editors. The capability trap: prevalence in human systems. 35th International Conference of the System Dynamics Society; 2017.

46. Cutler DM, Rosen AB, Vijan S. The value of medical spending in the United States, 1960–2000. New England journal of medicine. 2006;355(9):920-7.

47. Higgs R. Gov. John Kasich announces efforts to reduce infant mortality rate in Ohio 2014 [updated Jan 12, 2019. Available from: https://www.cleveland.com/open/2014/12/gov_john_kasich_announces_effo.html.

48. Black LJ. When visuals are boundary objects in system dynamics work. System Dynamics Review. 2013;29(2):70-86.

49. Allen AO. Probability, statistics, and queueing theory: Academic press; 2014.

50. Oakes JM, Rossi PH. The measurement of SES in health research: current practice and steps toward a new approach. Social science & medicine. 2003;56(4):769-84.

51. Rothstein J. Inequality of educational opportunity? Schools as mediators of the intergenerational transmission of income. Journal of Labor Economics. 2019;37(S1):S85-S123.

52. Chetty R, Hendren N, Kline P, Saez E. Where is the land of opportunity? The geography of intergenerational mobility in the United States. The Quarterly Journal of Economics. 2014;129(4):1553-623.

53. Hanauer A. Ohio's economy no longer fully recovers after recessions 2019 [Available from: https://www.epi.org/blog/ohios-economy-no-longer-fully-recovers-after-recessions/.

54. Oliva R, Sterman JD. Cutting corners and working overtime: Quality erosion in the service industry. Management Science. 2001;47(7):894-914.

55. Halfon N, Hochstein M. Life course health development: an integrated framework for developing health, policy, and research. The Milbank Quarterly. 2002;80(3):433-79.

56. Conti G, Heckman JJ. Economics of child well-being: Springer; 2014.

57. Foster KR. The sociobiology of molecular systems. Nature Reviews Genetics. 2011;12(3):193.

58. Mathews T, Driscoll AK. Trends in infant mortality in the United States, 2005-2014. Centers for Disease Control and Prevention, National Center for Health Statistics.(Hyattsville, Maryland: 2017); 2016.

59. Ebrahimvandi A, Hosseinichimeh N, Iams J. Understanding State-Level Variations in the US Infant Mortality: 2000 to 2015. Am J Perinatol. 2018.

60. Hedegaard H, Warner M, Miniño AM. Drug overdose deaths in the United States, 1999–2015. 2017.

61. National Association of State Budget Officers. State Expenditure Report, 1990-2018. 2019.

62. Schiller Z. Report to the Ohio budget planning and management commission. www.policymattersohio.com; 2010.

63. Patton W. Ohio Budget Basics. 2019 January 30, 2019

64. Fleeter H. 20 Years of School Funding Post-DeRolph. Ohio Education Policy Institute; 2018.

65. DeRolph v. State of Ohio. Ohio Supreme Court.

66. Institute OEP. OEPI Analysis of 2018 School District Report Card Data. 2018.

67. Bradley E, Taylor L. The American health care paradox: Why spending more is getting us less: Public Affairs; 2013.

68. Fine A, Kotelchuck M. Rethinking MCH: The life course model as an organizing framework. US Department of Health and Human Services, Health Resources and Services Administration, Maternal and Child Health Bureau. 2010:1-20.

69. Council NR, Population Co. US health in international perspective: Shorter lives, poorer health: National Academies Press; 2013.

70. Homer J, Milstein B, Hirsch GB, Fisher ES. Combined Regional Investments Could Substantially Enhance Health System Performance And Be Financially Affordable. Health Affairs. 2016;35(8):1435-43.

71. Griggs D, Stafford-Smith M, Gaffney O, Rockström J, Öhman MC, Shyamsundar P, et al. Policy: Sustainable development goals for people and planet. 2013;495(7441):305.

72. Ku L, Paradise J, Thompson V. Data Note: Medicaid's Role in Providing Access to Preventive Care for Adults. Kaiser Family Foundation; 2017.

Chapter 5. Conclusion

High infant mortality rates (IMR) in the U.S. have been a major public health concern for decades. Many studies have focused on understanding the causes, risk factors, and interventions that can reduce infant mortality (1-15). However, the death of an infant is the result of the interplay between many risk factors, which in some cases can be traced to the infancy of their parents. Consequently, these complex interactions challenge the effectiveness of many interventions (16, 17). The long-term goal of this study is to advance the common understanding of effective interventions for improving health outcomes and, in particular, infant mortality. To achieve these goals, I implemented systems and data science methods to contribute to the understanding of infant mortality causes and risk factors. Based on these new explanations, I offer a systematic approach that can help in addressing adverse birth outcomes—including high infant mortality and preterm birth (PTB) rates—which is the central contribution of this dissertation.

5.1 Understanding State-level Variations in U.S. Infant Mortality

In essay one (Chapter 2), I focused on the trends of infant mortality at the state-level between 2000 and 2015. Our goal was to identify patterns in the leading causes of infant mortality across those states that successfully reduced their infant mortality. I exploited variations in medical causes of infant mortality to identify the ones that contributed the most to IMR reduction. I answered three specific questions: (1) Which of the IMR subgroup(s) are most responsible for the reduction of IMR at the national level from 2000 to 2015?; (2) Which subgroup(s) are most responsible for the IMR reduction in states that have successfully reduced their IMRs from 2000 to 2015?; and (3) Are variations in teen pregnancy, multiple births, and prenatal care associated with state variations in the preterm-related mortality rate? I used statistical tools, including significant test and fixed-effect regression model to address these questions.

This study contributed to the research in infant mortality in several ways. First, it provided new insights into the impact of the different diagnoses that contributed to the reduction of infant mortality in successful states. Understanding drivers of IMR reduction in states that have achieved substantial rate reductions may help to improve high IMRs in other states. In addition, it contributed to the literature by investigating whether the variations in teen pregnancy, prenatal care, and multiple births are associated with state variations in the PMR. This study also provided recommendations to policymakers in states with high infant mortality rates to leverage interventions targeting preterm-related mortalities such as reducing the percentage of pregnant women with inadequate care. However, better access to prenatal care could only partially explain the preterm-related mortality variations because of its small impact size. Therefore, I conducted the second study to investigate the risk factors of PTB in the largest obstetric population that has ever been studied in this field. The objective in the second study was to increase the understanding of the PTB risk factors that are both generalizable and identifiable during the early stages of pregnancy.

5.2 Identifying the Early Signs of a Preterm Birth

In essay two (Chapter 3), I used statistical and machine learning techniques to enhance our understanding of the influential risk factors of preterm birth (PTB). In this study, I used the high-dimensional dataset of U.S. birth records in 2016 and combine it with two other major area-level datasets to increase the number of potential predictors of preterm birth that are present in early pregnancy—before the second trimester. Analyzing a national-level data with machine learning techniques increases the generalizability of the risk factors in the obstetric population, which has received little attention in the literature (18, 19).

This study contributes to the overall literature of preterm birth in many ways. First, the dataset that we used in our analysis, the U.S. linked birth datasets, is unique and representative of the national obstetric population. Second, I included the factors so that the study design become more generalizable. Third, this study provides the level of importance for each variable, which is the first time this comparison has been reported in a PTB study. Fourth, I use proper machine learning (ML) methods that have the capability of checking high-order interactions with minimal supervision. It is important to consider the interactions, because it enhances the capability of the model in capturing complex relationships. Combining ML with big data increased our understanding of preterm birth risk factors, as it helped me to check high-order interactions between risk factors in the national obstetrical population. A major finding of this study is that socioeconomic factors such as parent education are more important than generally known factors such as race in the prediction of PTB. However, much remains to be understood about the preterm birth risk factors and the relative impact of interventions that can control these factors.

5.3 The Capability Trap on a Macro-societal Level

As suggested by the second study, there is an important relationship between socioeconomic factors such as education level and major health outcomes like preterm birth. No intervention can improve the socioeconomic status of a mother during pregnancy. These results point to the need for more comprehensive approaches that change the focus from medical interventions during pregnancy to the time where mothers become vulnerable to the risk factors of PTB. In the third essay (Chapter 4), I took a more aggregate perspective to study the dynamics of population health that results in undesirable outcomes in major areas like infant mortality. I was particularly interested in understanding why some states are performing poorly despite having the best healthcare systems in the nation and allocating most of their resources to health. To do this, I used a system dynamics approach to capture interconnections between health and socioeconomic (SE) factors in a community in the long-term. I also considered the effects of resource allocation to health on SE status in the community.

The novelty of this study lies in exploring the dynamic interactions of socioeconomic status, resource generation-depletion, and aggregate health outcomes in a community over time. We postulate that the system has tipping point values, and too much or too little healthcare spending result in vicious cycles that deteriorate healthcare conditions. In a supplementary step, we conducted statistical analysis of health indicators across different US states over the past 23 years. The results confirm our simulation outcomes pointing to the negative consequences of increasing healthcare expenditure at the expense of shifting resources from other sectors such as education.

5.3.1 Broader Impacts

Societal: According to the U.S. government’s Healthy People 2020 guidelines, one of the high-priority health issues is reducing infant mortality rates (IMR). The national goal of the U.S. is to reduce IMR with a focus on reducing disparities and variations. This study contributes to an accomplishment of these goals by providing insights to both individuals and policymakers on causes, risk factors and interventions that targets infant mortality. We also provide insight to the effect of interventions that target preterm-related mortalities, such as reducing the percentage of pregnant women who received inadequate care. These findings are specifically helpful for states with high infant mortality rates. The second study builds on these findings to discover the more generalizable risk factors of PTB that can impact a larger portion of the population. A major finding of the second study is that socioeconomic factors like parent education are more important than generally known factors like race in the prediction of birth outcomes such as PTB. This finding is significant evidence for theories like Lifecourse, which postulate that the main determinants of a health trajectory are the social scaffolding that addresses the upstream roots of health. Considering these evidences, we propose a comprehensive approach for addressing key health outcomes, such as PTB or infant mortality. This framework can support the robust decision-making process of policymakers and families for their long-term health planning. ***Health and healthcare system:*** We also provide explanations for the poor health outcomes of different communities, and we then built an analytical framework based on these explanations. The final analytical framework can contribute to the heated national debate on optimal allocation of resources to the healthcare sector. ***Outreach:*** We published and presented our findings in different conferences and journals in both management systems and health domains. I plan to continue this research and present it to different professional societies and communicate our findings with the purpose of increasing chances of implementation.

In conclusion, this dissertation contributes to a better understanding of complexities in infant mortality and health-related policies. Moreover, we proposed important research questions for future studies in the previous chapters of this dissertation. This work contributes to the body of health literature both in terms of the application of statistical and machine learning techniques, as well as advancing health-related theories.

References

1. Alio AP, Richman AR, Clayton HB, Jeffers DF, Wathington DJ, Salihu HM. An Ecological Approach to Understanding Black–White Disparities in Perinatal Mortality. Maternal and Child Health Journal. 2010;14(4):557-66.

2. McLeroy KR, Bibeau D, Steckler A, Glanz K. An Ecological Perspective on Health Promotion Programs. Health Education Quarterly. 1988;15(4):351-77.

3. Elder TE, Goddeeris JH, Haider SJ. Racial and ethnic infant mortality gaps and the role of socio-economic status. Labour economics. 2016;43:42-54.

4. Iams JD, Applegate MS, Marcotte MP, Rome M, Krew MA, Bailit JL, et al. A Statewide Progestogen Promotion Program in Ohio. Obstetrics & Gynecology. 2017;129(2):337-46.

5. Latendresse G. The interaction between chronic stress and pregnancy: preterm birth from a biobehavioral perspective. Journal of midwifery & women's health. 2009;54(1):8-17.

6. Lefmann T, Combs-Orme T. Prenatal stress, poverty, and child outcomes. Child and Adolescent Social Work Journal. 2014;31(6):577-90.

7. Class QA, Khashan AS, Lichtenstein P, Langstrom N, D'Onofrio BM. Maternal stress and infant mortality: the importance of the preconception period. Psychological science. 2013;24(7):1309-16.

8. Burguet A, Kaminski M, Abraham-Lerat L, Schaal JP, Cambonie G, Fresson J, et al. The complex relationship between smoking in pregnancy and very preterm delivery. BJOG: an International Journal of Obstetrics & Gynaecology. 2004;111(3):258-65.

9. Gage TB, Fang F, O'Neill E, Dirienzo G. Maternal education, birth weight, and infant mortality in the United States. Demography. 2013;50(2):615-35.

10. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. The Lancet. 2008;371(9608):261-9.

11. Katz J, Lee ACC, Kozuki N, Lawn JE, Cousens S, Blencowe H, et al. Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis. The Lancet. 2013;382(9890):417-25.

12. Mathews TJ, MacDorman MF, ME T. Infant mortality statistics from the 2013 period linked birth/infant death data set. National vital statistics reports. Hyattsville, MD: National Center for Health Statistics2015.

13. Iams JD. Identification of candidates for progesterone: why, who, how, and when? Obstetrics and gynecology. 2014;123(6):1317-26.

14. MacDorman MF, Mathews T. Recent trends in infant mortality in the United States. 2008.

15. Sommers BD, Baicker K, Epstein AM. Mortality and access to care among adults after state Medicaid expansions. The New England journal of medicine. 2012;367(11):1025-34.

16. Iams JD. Prevention of preterm parturition. New England Journal of Medicine. 2014;370(3):254-61.

17. Chen A, Oster E, Williams H. Why Is Infant Mortality Higher in the United States Than in Europe? American economic journal Economic policy. 2016;8(2):89-124.

18. Alleman BW, Smith AR, Byers HM, Bedell B, Ryckman KK, Murray JC, et al. A proposed method to predict preterm birth using clinical data, standard maternal serum screening, and cholesterol. American Journal of Obstetrics & Gynecology. 2013;208(6):472. e1-. e11.

19. Weber A, Darmstadt GL, Gruber S, Foeller ME, Carmichael SL, Stevenson DK, et al. Application of machine-learning to predict early spontaneous preterm birth among nulliparous non-Hispanic black and white women. Annals of epidemiology. 2018;28(11):783-9. e1.

Appendices

Appendix A List of OMR Causes

From wonder.cdc.gov:

1) Location: The state, region and division data are derived from the "STRESFIPB" variable in the public use files for years 1999-2002, and from "MRSTATEFIPS" in the public use files for years 2003-2004, and from "MRTERR" for years 2005-2015. The county data are derived from the combined values in variables "STRESFIPB" + "CNTYRFPB" in the public use files for years 1999-2002, and from "MRSTATEFIPS" + "MRCNTYFIPS" in the public use files for years 2003-2004, and from "MRTERR"+ "MRCNTY" for years 2005-2015.

2) Gestational age at birth based on Last Menstrual Period (LMP).
3) The "Gestational Age at Birth" data for years 1995-1998 and the "Gestational Age 10" (formerly named "Gestational Age Group2") data are derived from the "GESTAT10" variable in the public use data for years 1995-2002, and derived from the "GESTREC10" variable in the public use data for years 2003 and later.

"A00-B99 (Certain infectious and parasitic diseases)
C00-D48 (Neoplasms)
D50-D89 (Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism)
E00-E88 (Endocrine, nutritional and metabolic diseases)
F01-F99 (Mental and behavioural disorders)
G00-G98 (Diseases of the nervous system)
H00-H57 (Diseases of the eye and adnexa)
H60-H93 (Diseases of the ear and mastoid process)
I00-I99 (Diseases of the circulatory system)
J00-J98 (Diseases of the respiratory system)
K00-K14 (Diseases of oral cavity, salivary glands and jaws)
K20-K31 (Diseases of oesophagus, stomach and duodenum)
K35-K38 (Diseases of appendix)
K40-K46 (Hernia)
K50-K52 (Noninfective enteritis and colitis)
K55.1 (Chronic vascular disorders of intestine)
K55.2 (Angiodysplasia of colon)
K55.8 (Other vascular disorders of intestine)
K55.9 (Vascular disorder of intestine, unspecified)
K56 (Paralytic ileus and intestinal obstruction without hernia)
K57 (Diverticular disease of intestine)
K58 (Irritable bowel syndrome)
K59 (Other functional intestinal disorders)
K60 (Fissure and fistula of anal and rectal regions)

K61 (Abscess of anal and rectal regions)

K62 (Other diseases of anus and rectum)

K63 (Other diseases of intestine)

K65-K66 (Diseases of peritoneum)

K70-K76 (Diseases of liver)

K80-K86 (Disorders of gallbladder, biliary tract and pancreas)

K90-K92 (Other diseases of the digestive system)

L00-L98 (Diseases of the skin and subcutaneous tissue)

M00-M99 (Diseases of the musculoskeletal system and connective tissue)

N00-N98 (Diseases of the genitourinary system)

O00-O99 (Pregnancy, childbirth and the puerperium)

P00 (Newborn affected by maternal conditions that may be unrelated to present pregnancy)

P01.2 (Newborn affected by oligohydramnios)

P01.3 (Newborn affected by polyhydramnios)

P01.4 (Newborn affected by ectopic pregnancy)

P01.6 (Newborn affected by maternal death)

P01.7 (Newborn affected by malpresentation before labour)

P01.8 (Newborn affected by other maternal complications of pregnancy)

P01.9 (Newborn affected by maternal complication of pregnancy, unspecified)

P02.2 (Newborn affected by other and unspecified morphological and functional abnormalities of placenta)

P02.3 (Newborn affected by placental transfusion syndromes)

P02.4 (Newborn affected by prolapsed cord)

P02.5 (Newborn affected by other compression of umbilical cord)

P02.6 (Newborn affected by other and unspecified conditions of umbilical cord)

P02.8 (Newborn affected by other abnormalities of membranes)

P02.9 (Newborn affected by abnormality of membranes, unspecified)

P03 (Newborn affected by other complications of labour and delivery)

P04 (Newborn affected by noxious influences transmitted via placenta or breast milk)

P05 (Slow fetal growth and fetal malnutrition)

P08 (Disorders related to long gestation and high birth weight)

P10.0 (Subdural haemorrhage due to birth injury)

P10.1 (Cerebral haemorrhage due to birth injury)

P10.3 (Subarachnoid haemorrhage due to birth injury)

P10.4 (Tentorial tear due to birth injury)

P10.8 (Other intracranial lacerations and haemorrhages due to birth injury)

P10.9 (Unspecified intracranial laceration and haemorrhage due to birth injury)

P11 (Other birth injuries to central nervous system)

P12 (Birth injury to scalp)

P13 (Birth injury to skeleton)

P14 (Birth injury to peripheral nervous system)

P15 (Other birth injuries)

P20 (Intrauterine hypoxia)

P21 (Birth asphyxia)

P23 (Congenital pneumonia)

P24 (Neonatal aspiration syndromes)

P25 (Interstitial emphysema and related conditions originating in the perinatal period)

P26 (Pulmonary haemorrhage originating in the perinatal period)

P28.2 (Cyanotic attacks of newborn)

P28.3 (Primary sleep apnoea of newborn)

P28.4 (Other apnoea of newborn)

P28.5 (Respiratory failure of newborn)

P28.8 (Other specified respiratory conditions of newborn)

P28.9 (Respiratory condition of newborn, unspecified)

P29 (Cardiovascular disorders originating in the perinatal period)

P35 (Congenital viral diseases)

P37 (Other congenital infectious and parasitic diseases)

P38 (Omphalitis of newborn with or without mild haemorrhage)

P39 (Other infections specific to the perinatal period)

P50 (Fetal blood loss)

P51 (Umbilical haemorrhage of newborn)

P52.4 (Intracerebral (nontraumatic) haemorrhage of newborn)

P52.5 (Subarachnoid (nontraumatic) haemorrhage of newborn)

P52.6 (Cerebellar (nontraumatic) and posterior fossa haemorrhage of newborn)

P52.8 (Other intracranial (nontraumatic) haemorrhages of newborn)

P52.9 (Intracranial (nontraumatic) haemorrhage of newborn, unspecified)

P53 (Haemorrhagic disease of newborn)

P54 (Other neonatal haemorrhages)

P55 (Haemolytic disease of newborn)

P56 (Hydrops fetalis due to haemolytic disease)

P57 (Kernicterus)

P58 (Neonatal jaundice due to other excessive haemolysis)

P59 (Neonatal jaundice from other and unspecified causes)

P60 (Disseminated intravascular coagulation of newborn)

P61 (Other perinatal haematological disorders)

P70-P74 (Transitory endocrine and metabolic disorders specific to newborn)

P76 (Other intestinal obstruction of newborn)

P78 (Other perinatal digestive system disorders)

P80-P83 (Conditions involving the integument and temperature regulation of newborn)

P90-P96 (Other disorders originating in the perinatal period)

R00-R09 (Symptoms and signs involving the circulatory and respiratory systems)

R10-R19 (Symptoms and signs involving the digestive system and abdomen)

R20-R23 (Symptoms and signs involving the skin and subcutaneous tissue)

R25-R29 (Symptoms and signs involving the nervous and musculoskeletal systems)

R30-R39 (Symptoms and signs involving the urinary system)

R40-R46 (Symptoms and signs involving cognition, perception, emotional State and behaviour)

R47-R49 (Symptoms and signs involving speech and voice)

R50-R68 (General symptoms and signs)

R70-R79 (Abnormal findings on examination of blood, without diagnosis)

R80-R82 (Abnormal findings on examination of urine, without diagnosis)

R83-R89 (Abnormal findings on examination of other body fluids, substances and tissues, without diagnosis)

R90-R94 (Abnormal findings on diagnostic imaging and in function studies, without diagnosis)
R95 (Sudden infant death syndrome)
R99 (Other ill-defined and unspecified causes of mortality)

U00-U99 (Codes for special purposes)

V01-Y89 (External causes of morbidity and mortality)

R96 (Other sudden death, cause unknown)

R98 (Unattended death)”

Appendix B Causes of Mortality per state

Table B-1 represents the data for IMR and all four subgroups for each state in 2015.

Table B-1 IMR and Subgroups per State

No	State	IMR	PMR	CMR	SMR	OMR
1	Alabama	8.31	2.77	1.46	1.07	3.02
2	Alaska	6.91	1.86	1.33	0.89	2.84
3	Arizona	5.47	1.63	1.34	0.16	2.34
4	Arkansas	7.53	1.85	1.59	1.57	2.49
5	California	4.43	1.57	1.12	0.24	1.50
6	Colorado	4.66	1.91	1.01		1.71
7	Connecticut	5.65	2.83	0.56		2.04
8	Delaware	9.22	4.66	1.07		3.40
9	Florida	6.24	2.37	1.20	0.28	2.39
10	Georgia	7.79	2.91	1.45	0.84	2.58
11	Hawaii	5.70	2.12	0.98		2.33
12	Idaho	4.69	1.01	1.49	0.44	1.75
13	Illinois	6.00	2.44	1.11	0.09	2.35
14	Indiana	7.31	2.43	1.69	0.40	2.77
15	Iowa	4.23	1.04	1.09	0.48	1.62
16	Kansas	5.95	2.12	1.46	0.46	1.92
17	Kentucky	6.68	1.79	1.38	1.04	2.50
18	Louisiana	7.56	2.57	1.21	0.70	3.09
19	Maine	6.58	2.06	1.27		2.70
20	Maryland	6.59	3.10	1.17	0.83	1.49
21	Massachusetts	4.32	2.20	0.53	0.15	1.43
22	Michigan	6.53	2.38	1.34	0.29	2.51
23	Minnesota	5.17	1.86	1.25	0.16	1.90
24	Mississippi	9.45	3.10	1.51	0.47	4.38
25	Missouri	6.49	2.08	1.43	0.24	2.73
26	Montana	5.80	0.95	1.59	0.95	2.30
27	Nebraska	5.70	1.31	1.57	0.82	1.95
28	Nevada	5.18	1.29	1.35		2.42
29	New Hampshire	4.10	1.21			2.41
30	New Jersey	4.68	1.60	0.85	0.44	1.80
31	New Mexico	5.07	1.01	1.39		2.60
32	New York	4.63	1.64	0.84	0.09	2.06
33	North Carolina	7.35	2.92	1.22	0.15	3.05
34	North Dakota	7.16	1.77	1.77	1.24	2.39

No	State	IMR	PMR	CMR	SMR	OMR
35	Ohio	7.18	2.76	1.32	0.60	2.49
36	Oklahoma	7.30	2.39	1.54	0.94	2.43
37	Oregon	5.15	1.86	0.99	0.50	1.80
38	Pennsylvania	6.15	2.55	1.00	0.47	2.14
39	Rhode Island	5.91	2.91			2.00
40	South Carolina	6.91	2.53	1.24	0.43	2.72
41	South Dakota	7.30	1.78	1.54	0.81	3.16
42	Tennessee	6.94	1.90	1.44	0.24	3.37
43	Texas	5.71	1.75	1.40	0.42	2.15
44	Utah	5.02	1.46	1.28	0.45	1.83
45	Vermont	4.57				2.37
46	Virginia	5.90	2.09	1.25	0.57	1.99
47	Washington	4.88	1.58	1.24	0.61	1.45
48	West Virginia	7.07	2.22	1.57	0.71	2.52
49	Wisconsin	5.80	2.13	1.21	0.18	2.28
50	Wyoming	4.89	1.67	1.29		1.80

Note that any states with less than 10 deaths in each subgroup is eliminated due to CDC data user agreement.

Appendix C Adequacy of Prenatal Care: KESSNER Index

ADEQUATE*

Gestation (weeks)****	Number of Prenatal Visits
13 or less AND	1 or more or not stated
14-17 AND	2 or more
18-21 AND	3 or more
22-25 AND	4 or more
26-29 AND	5 or more
30-31 AND	6 or more
32-33 AND	7 or more
34-35 AND	8 or more
36 or more AND	9 or more

INADEQUATE**

Gestation (weeks)****	Number of Prenatal Visits
14-21*** AND	0 or not stated
22-29 AND	1 or less or not stated
30-31 AND	2 or less or not stated
32-33 AND	3 or less or not stated
34 or more AND	4 or less or not stated

INTERMEDIATE: All combinations other than specified above

* In addition to the specified number of visits indicated for adequate care, the Interval to the first prenatal visit has to be 13 weeks or less (first trimester).
** In addition to the specified number of visits indicated for inadequate care, all Women who started their prenatal care during the third trimester (28 weeks or later) are considered inadequate.
*** For this gestation group, care is considered inadequate if the time of the first visit is not Stated.
**** When month and year are specified but day is missing, input 15 for day. Adequacy categories are in accord with recommendations of American College of Obstetrics and Gynecology and the World Health Organization.

Appendix D List of missing values for each value

Tobacco use and prenatal care have many missing values between 2011-2014. California did not report tobacco use for any year. Therefore, we will remove California in the second regression. Additionally, Florida, Georgia, and Michigan have incomplete or missing values for prenatal care and smoking. Therefore, we removed these States as well.

In 2011, 15 areas with “Not Available” data for tobacco use:

Alabama", "Alaska", "Arizona", "Arkansas", "Connecticut", "Hawaii", "Maine", "Massachusetts", "Michigan", "Minnesota", "Mississippi", "New Jersey", "Rhode Island", "Virginia", "West Virginia"

In 2012, 13 areas with “Not Available” data for tobacco use: "Alabama", "Alaska", "Arizona", "Arkansas", "Connecticut", "Hawaii", "Maine", "Michigan", "Mississippi", "New Jersey", "Rhode Island", "Virginia", "West Virginia"

In 2013, 10 areas with “Not Available” data for tobacco use: "Alabama", "Arizona", "Arkansas", "Connecticut", "Hawaii", "Maine", "Michigan", "New Jersey", "Rhode Island", "West Virginia"

In 2014, 4 areas with “Not Available” data for tobacco use: "Connecticut", "Hawaii", "New Jersey", "Rhode Island"

In 2015, 2 areas with “Not Available” data for tobacco use: "Connecticut", "New Jersey"

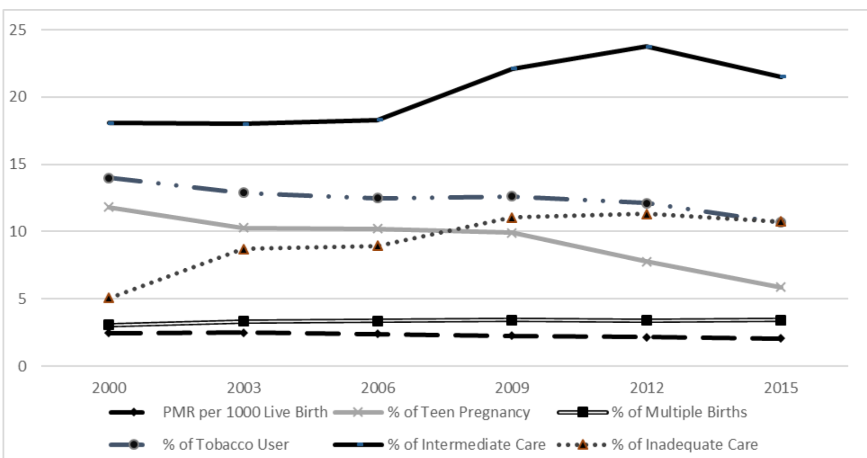


Figure D-1 Data on Preterm-Related Mortality and State-Level Factors for 50 U.S. States

Appendix E Performance Analysis of full list of States

Without dropping any states from categories, the mean IMR in successful states was 7.53 in 2000 and declined to 5.90 in 2015 (P-Value<0.001), representing 1.63 less deaths per 1,000 live births (Figure E-1). In successful states, reduction in the PMR accounted for the largest reduction in IMR—a 0.76 reduction (P-Value <0.001) from 2.63 to 1.87 (Figure E-1). Forty-seven percent of the IMR reduction (0.76 out of 1.63) in successful states was due to the PMR reduction. The reduction size in the CMR, SMR, and OMR were 0.32 (from 1.57 to 1.25 with P-Value=0.001), 0.34 (from 0.72 to 0.38 with P-Value<0.001), and 0.20 (2.60 to 2.40 with P-Value=0.086), respectively.

The IMR mean in unsuccessful states was 6.63 in 2000, which then declined by 0.29 per 1,000 live births to 6.34 deaths per 1,000 in 2015 (P-Value=0.051). In unsuccessful states, the PMR slightly decreased by 0.04 from 2.28 deaths per 1,000 to 2.24 (P-Value=0.60), the CMR declined by 0.12 from 1.37 to 1.25 (P-Value=0.017), the SMR reduced by 0.13 from 0.68 to 0.55 (P-Value=0.069), and the OMR stayed almost the same by 0.01 decline from 2.30 to 2.29 (P-Value=0.967) (Figure E-1).

The changes in the PMR differed significantly between successful and unsuccessful states (P-Value<0.001). Successful states reduced their PMRs by 0.76, while the other group experienced a slight decrease in the PMR—0.04—over the period of 2000 to 2015. The reduction for the CMR and SMR was also significantly different between the two groups of states (P-Values of 0.028 and 0.036, respectively). Successful states reduced their CMRs by 0.32, which is more than 2.5 times of the reduction size in unsuccessful states—0.12. The SMR reduction size for successful states was 0.34, which is significantly larger than the reduction size of unsuccessful states—0.13. Although the OMR size was the largest after the PMR, its reduction size was not large in any categories of states (0.20 in successful versus 0.01 in unsuccessful states), this difference was not significant (P-Value=0.172).

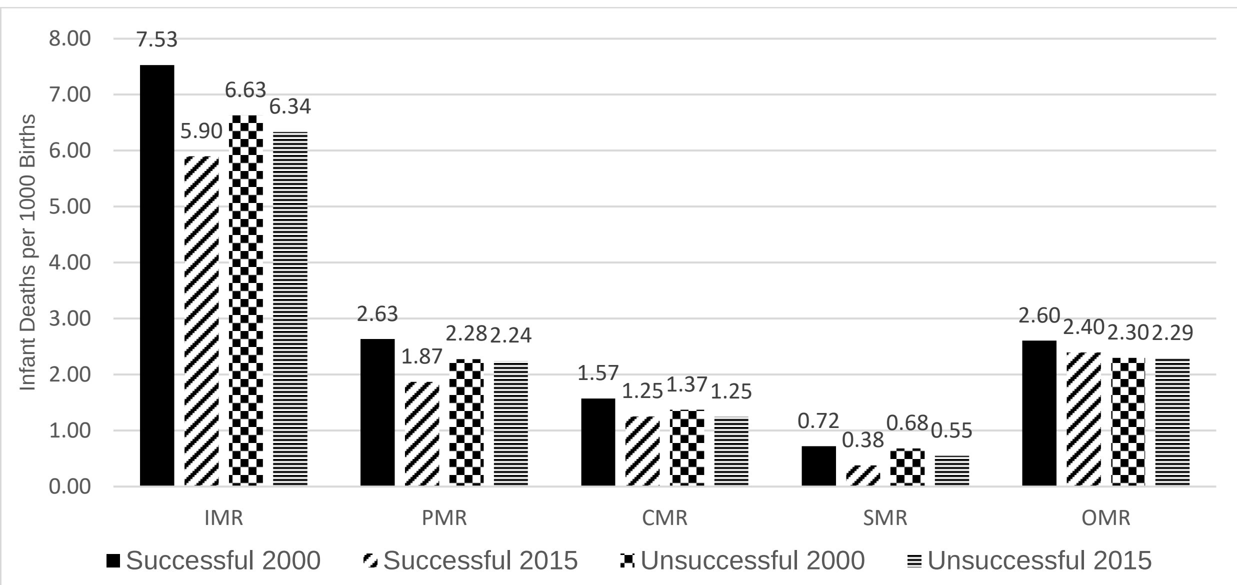


Figure E-1 Performances of successful and unsuccessful states in reduction of IMR and its subgroups in 2000 and 2015

Appendix F Category of states

Table F-1 List of States Based on Performance in Reduction of IMR between 2000 to 2015

Successful States		Unsuccessful States		States dropped from the sample	
state	Reduction Size	state	Reduction Size	state	Reduction Size
Arizona	-1.28	Alaska	0.00	Alabama	-1.20
California	-0.99	Arkansas	-0.62	Louisiana	-1.47
Colorado	-1.49	Connecticut	-0.86	Maine	1.73
Hawaii	-2.39	Delaware	-0.37	Massachusetts	-0.27
Idaho	-2.87	Florida	-0.67	Mississippi	-1.19
Illinois	-2.49	Georgia	-0.66	South Dakota	2.08
Iowa	-2.20	Indiana	-0.46	Tennessee	-2.13
Michigan	-1.67	Kansas	-0.68	Utah	-0.24
Nebraska	-1.44	Kentucky	-0.42	Washington	-0.32
Nevada	-1.34	Maryland	-0.93		
New Hampshire	-1.72	Minnesota	-0.45		
New Jersey	-1.59	Missouri	-0.70		
New Mexico	-1.54	Montana	-0.13		
New York	-1.77	Ohio	-0.50		
North Carolina	-1.27	Oregon	-0.42		
North Dakota	-1.18	Pennsylvania	-0.96		
Oklahoma	-1.17	Rhode Island	-0.40		
South Carolina	-1.85	Texas	0.09		
Vermont	-1.73	West Virginia	-0.36		
Virginia	-1.01	Wisconsin	-0.83		
Wyoming	-1.82				

To make initial average rates of IMR comparable among two groups of successful and unsuccessful states, we removed states with high initial IMR (IMR at 2000 > 9) from “Successful” and states with low initial IMR (IMR at 2000 < 5.5) from “Unsuccessful” and repeated the analysis. Therefore, we removed Alabama, Louisiana, Mississippi, and Tennessee from successful states because of high initial IMR in 2000—9.51, 9.03, 10.64, and 9.07 respectively. Also, we dropped Massachusetts, Maine, South Dakota, Utah, and Washington with low initial IMR from unsuccessful states—4.59, 4.85, 5.22, 5.26, and 5.20 respectively (The third column in Table F-1).

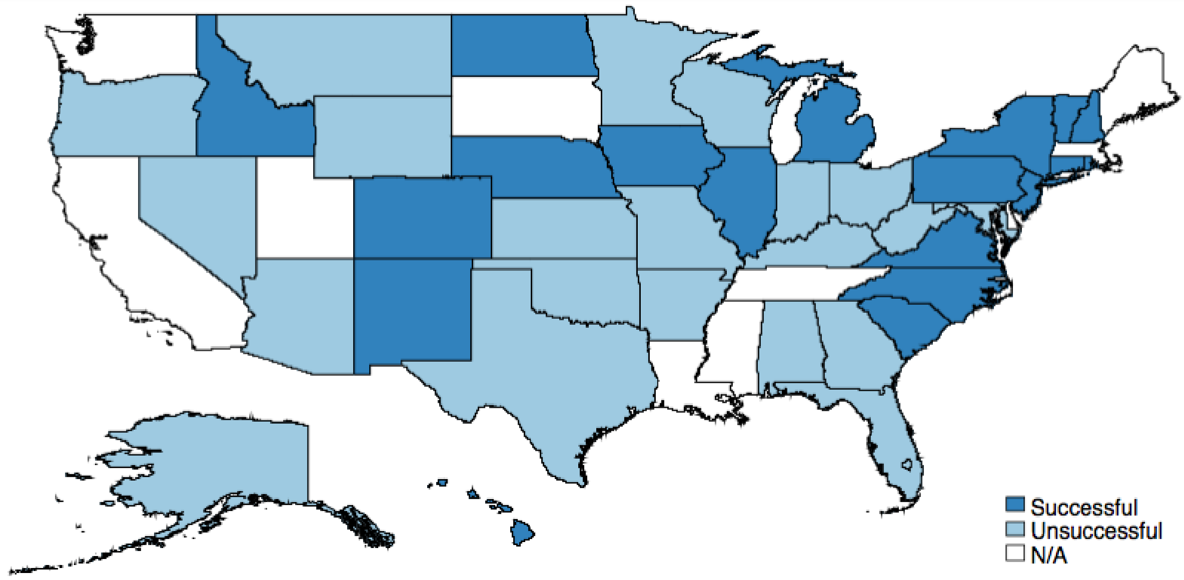


Figure F-1 Category of state Based on Performance in Reduction of IMR between 2000 and 2015

Appendix G Data preparation steps

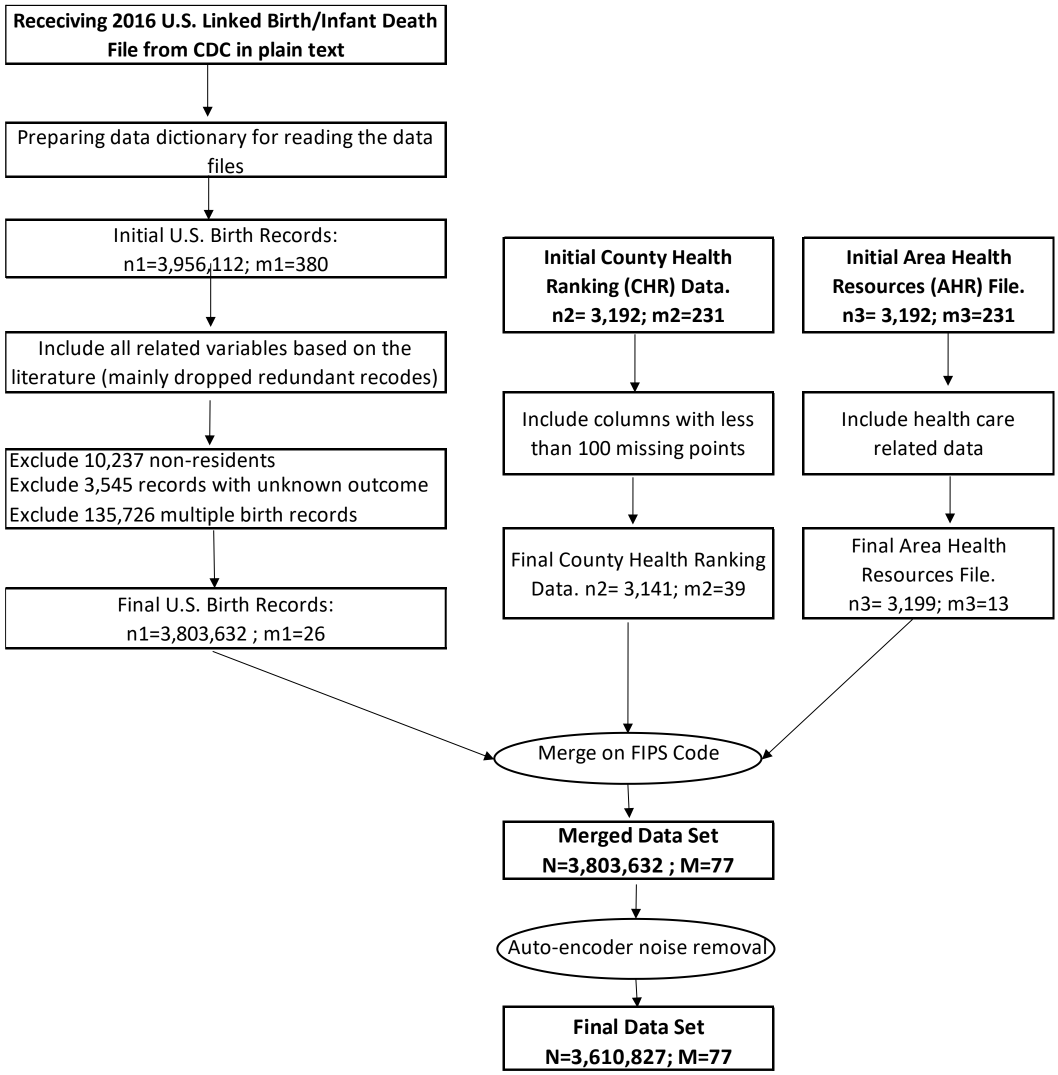


Figure G-1 Steps of data preparation.

Note that “n1” is the number of observations, and “m1” is the number of variables in the CDC dataset; “n2” and “n3” shows the number of counties in CHR and AHRF, respectively; and “m2” and “m3” shows the number of variables in CHR and AHRF, respectively.

Appendix H List of variables and their description

Table H-1 Variables included in analysis- CDC variables

Variable	Type	Scale	Definition
Individual Level data from CDC			
preterm	Binary	0/1	Preterm [17-37)
Age_M	Integer	12-50	Age of mother
Age_F	Integer	12-80	Age of father
Married	Binary	0/1	Marital status
bmi	Continuous	13-68.9	Body mass index
sex	Binary	0/1	Sex of infant; 1 if male
Priorlive	Integer	0-16	Number of prior live children
Priordead	Integer	0-12	Number of prior dead children
Race (Race_M or Race_F)	Categorical	White Black Indian Asian Hispanic Unknown	Maternal and paternal race. Race is stored in two different variables, one for maternal and the other for paternal.
Education (Education_M or Education_F)	Categorical	"1" 8th grade or less "2" 9th through 12th grade with no diploma "3" High school graduate or GED completed "4" Some college credit, but not a degree "5" Associate degree "6" Bachelor’s degree "7" Master’s degree "8" Doctorate or Professional Degree "9" Unknown	Maternal and paternal education level. Education is stored in two different variables, one for maternal and the other for paternal.
Sexually Transmitted Disease (STD)	Binary	0/1	1 if mother diagnosed with infections including Gonorrhea, Syphilis, or Chlamydia
StCnty	Categorical	"AK013" to "WY045"	The first two digits show the mother's residence state abbreviation. The last three digits show's county code.
Hypertension/ Eclampsia (hyper)	Binary	0/1	1 if any hypertension or pre-pregnancy hypertension eclampsia indicators in CDC dataset are positive for mother

Diabetes (diab)	Binary	0/1	1 if any pre-pregnancy diabetes, or gestational diabetes in CDC dataset are positive for mother
Adequacy of healthcare-KESSNER index	Coded as three binary variables.	Adequate (1) Intermediate (1) Inadequate (1)	Prenatal Care Utilization Index (Kessner et al, 1973). See Appendix for full definition.
cig_before	Integer	0-98	Number of cigarettes daily before pregnancy. 98 shows that mother smoked 98 or more cigarettes daily.
cig_1	Integer	0-98	Number of cigarettes daily in the first trimester.
cig_2	Integer	0-98	Number of cigarettes daily in the second trimester.
Pay	Categorical	Medicaid Private Self Other Unknown	Source of payment in the pregnancy and delivery.
Interval	Categorical	Plural delivery 1st child <11 month 12-17 18-23 months 24-35 months 36-47 months 49-59 months 60-71 months 72 months and over	Interval of Last Live Birth
WIC	Binary	0/1	Special Supplemental Nutrition Program for Women, Infants, and Children (WIC)
Previous_preterm	Binary	0/1	Previous preterm birth
Infertility_treatment	Binary	0/1	Infertility treatment
Previous_cesareans	Binary	0/1	Previous cesareans

Table H-2 Variables included in analysis- County Health Rankings variables

Area Level data from County Health Rankings			
Poor_Health	Continuous	0-100%	"Percentage of adults in a county who consider themselves to be in poor or fair health in past 30 days"
Poor_Physical_Days	Continuous	2.2- 6.5	"Average number of physically unhealthy days reported in past 30 days"
Poor_Mental_Days	Continuous	2.1- 5.6	"Average number of mentally unhealthy days reported in past 30 days"
Adult_Smoking	Continuous	0-100%	"Percentage of the adult population in a county who both report that they currently smoke every day or most days and have smoked at least 100 cigarettes in their lifetime."
Adult_Obesity	Continuous	0-100%	"Percentage of the adult population (age 20 and older) that reports a body mass index (BMI) greater than or equal to 30 kg/m2"
Physical_Inacitvity	Continuous	0-100%	"Percentage of adults ages 20 and over reporting no leisure-time physical activity in the past month"
Access_to_Exercise	Continuous	0-100%	"Percentage of individuals in a county who live reasonably close to a location for physical activity. Locations for physical activity are defined as parks or recreational facilities. Individuals are considered to have access to exercise opportunities if they: 1. Reside in a census block that is within a half mile of a park, or 2. Reside in an urban census block that is within one mile of a recreational facility, or 3. Reside in a rural census block that is within three miles of a recreational facility."

Exessive_Drinking	Continuous	0-100%	"Percentage of a county's adult population that reports binge or heavy drinking in the past 30 days"
Alcohol_Driving_Death	Continuous	0-100%	"Percentage of motor vehicle crash deaths with alcohol involvement"
Uninsured_Total	Continuous	0-100%	"Percentage of the population under age 65 without health insurance coverage."
Prev_Hospital_Stays	Continuous	0-100%	"Age-adjusted hospital discharge rate for ambulatory care-sensitive conditions per 1,000 fee-for-service Medicare enrollees."
College_Degree	Continuous	0-100%	"Percentage of the population ages 25-44 with some post-secondary education, such as enrollment in vocational/technical schools, junior colleges, or four-year colleges"
Unemployment	Continuous	0-100%	"Percentage of the county's civilian labor force, age 16 and older, that is unemployed but seeking work."
Children_In_Pov	Continuous	0-100%	"Percentage of people under age 18 living in poverty"
Income_Inequality	Continuous	2.6- 8.7	"Ratio of household income at the 80th percentile to that at the 20th percentile. A higher inequality ratio indicates greater division between the top and bottom ends of the income spectrum."
Chil_in_Single_Parent	Continuous	0-100%	"Percentage of children (less than 18 years of age) living in family households that are headed by a single parent"
Social_Asoc	Continuous	0-81	"Number of membership associations per 10,000 population. Rates measure the number of events in a given time period (generally one or more years) divided by the average number of people at risk during that period."

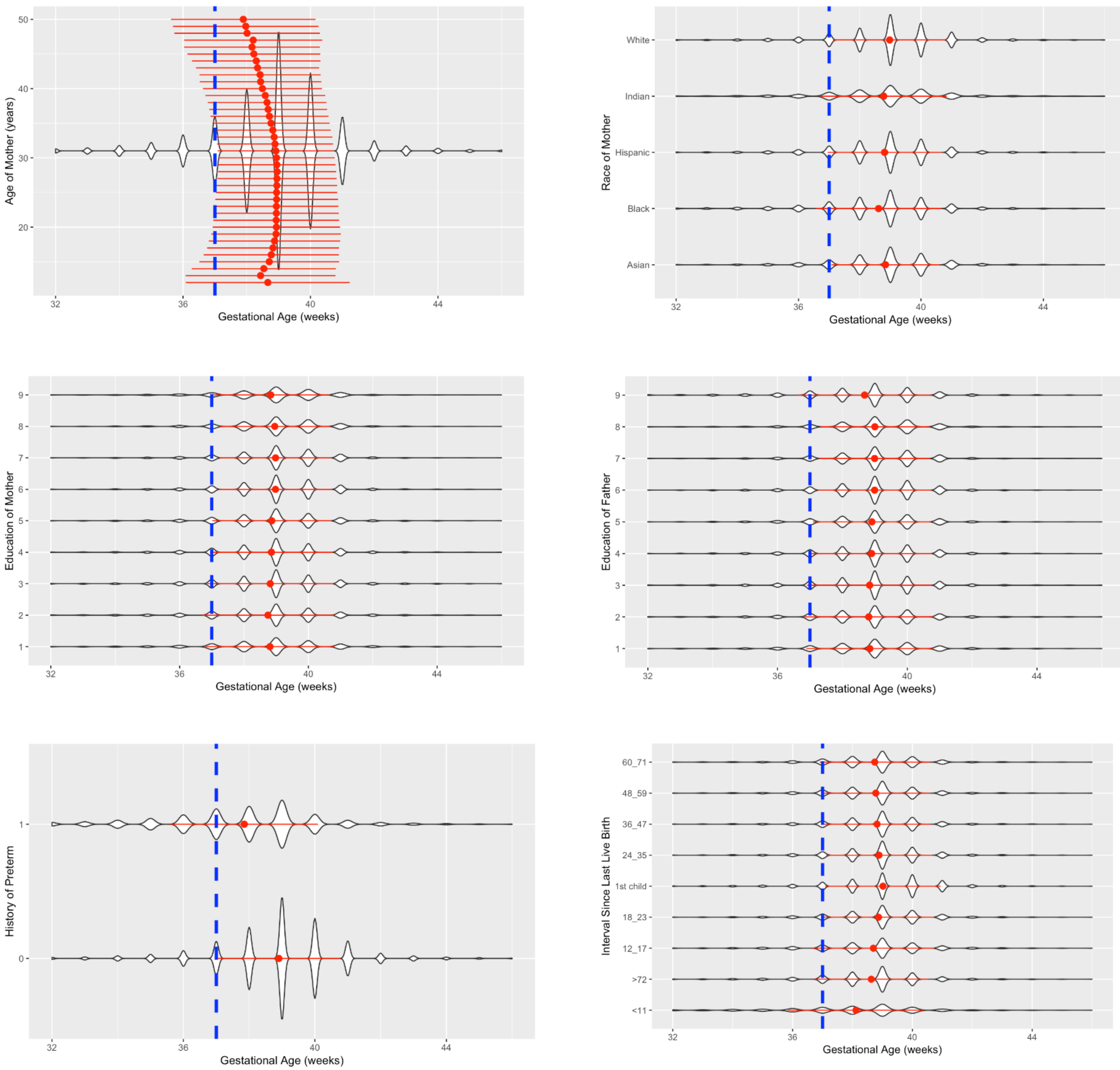
Air_Particulate_Matter	Continuous	0-100%	"Average daily density of fine particulate matter in micrograms per cubic meter. Fine particulate matter is defined as particles of air pollutants with an aerodynamic diameter less than 2.5 micrometers."
Drinking_Water_Viol	Continuous	0-100%	"Annual average percentage of the population served by community water systems who receive drinking water that does not meet all applicable health-based drinking water standards"
Severe_Housing_Prob	Continuous	0-100%	"Percentage of households with one or more of the following housing problems: 1. Housing unit lacks complete kitchen facilities; 2. Housing unit lacks complete plumbing facilities; 3. Household is overcrowded; or 4. Household is severely cost burdened."
Driving_Alone_Work	Continuous	0-100%	"Percentage of the workforce that usually drives alone to work"
Long_Drive_Work	Continuous	0-100%	"Percentage of workers who drive alone (via car, truck, or van) with a commute longer than 30 minutes"
Diabetes	Continuous	0-100%	"Percentage of adults aged 20 and above with diagnosed diabetes in a given county"
Food_Insecurity	Continuous	0-100%	"Percentage of the population that did not have access to a reliable source of food during the past year"
Lim_Healthy_Food	Continuous	0-100%	"Percentage of the population that is low income and does not live close to a grocery store. "Low income" is defined as having an annual family income of less than or equal to 200 percent of the FPL. Living close to a grocery store in rural areas means living less than 10 miles from a grocery store whereas in nonrural areas, it means less than 1 mile."

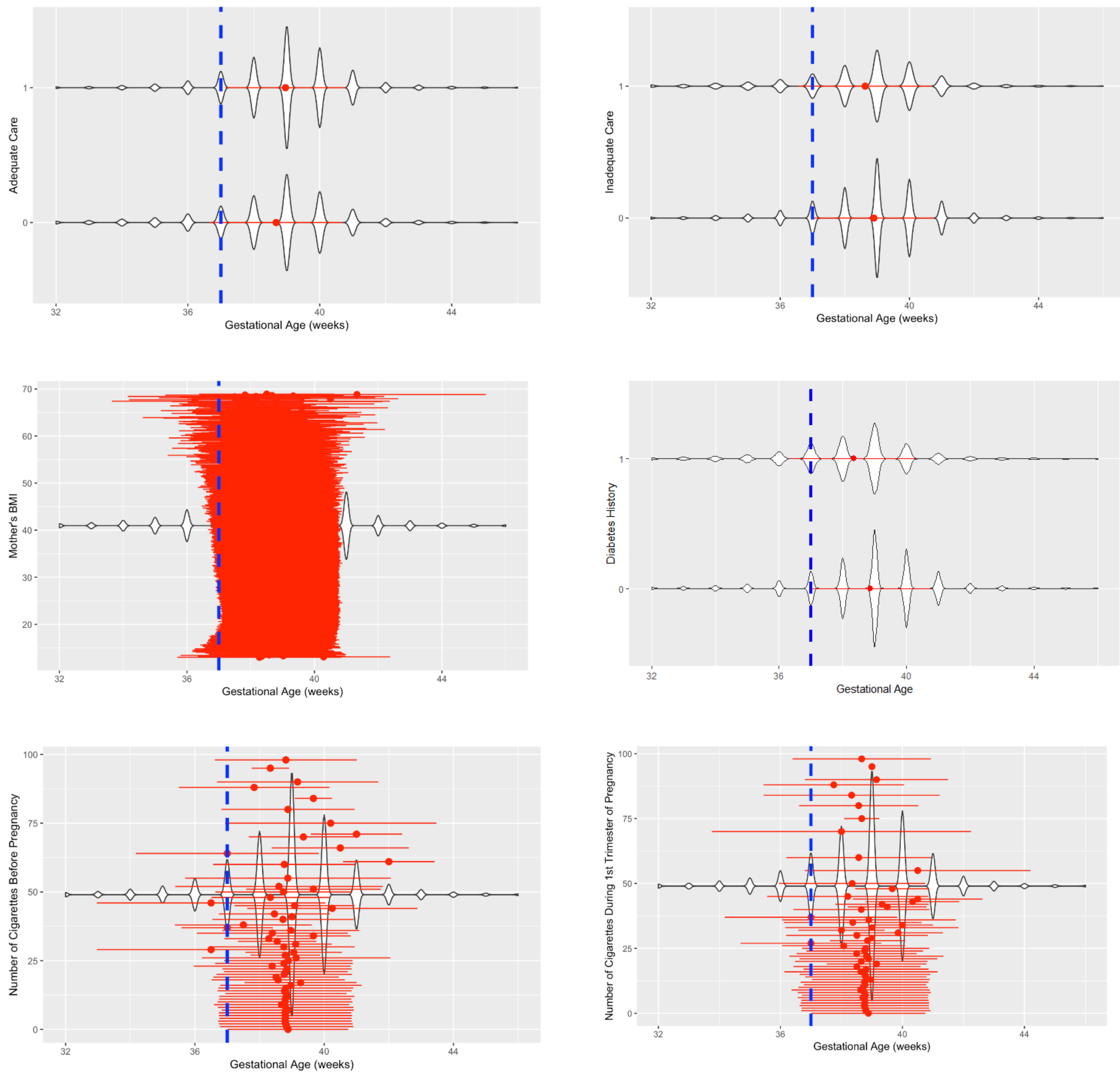
Insuf_Sleep	Continuous	0-100%	"Percentage of adults who report that they sleep less than 7 hours per night on average"
Uninsured_Adults	Continuous	0-100%	"Percentage of the population ages 18 to 64 that have no health insurance coverage in a given county"
Uninsured_Children	Continuous	0-100%	"Percentage of the population under age 19 that has no health insurance coverage in a given county."
Healthcare_cost	Continuous	4149- 18543	"Price-adjusted Medicare reimbursements (Parts A and B) per enrollee"
Median_Household_Income	Continuous	0-100%	"Median Household Income is based on one year of survey data and is created using complex statistical modeling. Modeling generates more stable estimates for places with small numbers of residents or survey responses. "
Perc_Und_Eighteen	Continuous	0-100%	"Percentage of the population below 18 years of age"
Perc_Ove_SixtyFive	Continuous	0-100%	"Percentage of the population over 65 years of age"
Black_Perc	Continuous	0-100%	"Percentage of the population that is non-Hispanic African American"
Hisp_Perc	Continuous	0-100%	"Percentage of the population that is Hispanic"
Rural_Perc	Continuous	0-100%	"Percentage of the population that lives in a rural area"
Dentists_Ratio	Continuous	0-100	"Ratio of dentists to 10,000 population"
Other_PC_Physicians_Ratio	Continuous	0-100	"Ratio of other primary care providers to 10,000 people"

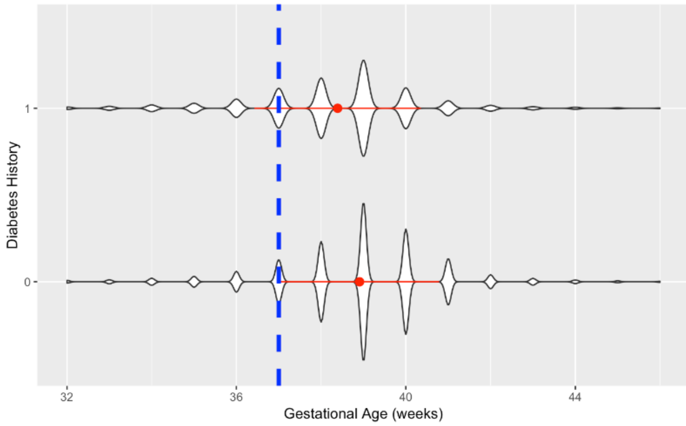
Area Level data from Area Health Resources File			
Perc_in_Poverty	Continuous	0-100%	Percentage of persons in poverty
Females_Under_65_no_Health_Ins	Continuous	0-100%	Percentage of females in poverty
FemalesDivorced	Continuous	0-100%	Percentage of divorced females
Number_of_Hospitals_Per_10k	Continuous	0- 10	Ratio of hospitals to 10,000 population
Hospital_Total_Prsnl_Per_10k	Continuous	0- 3385	"Ratio of hospital personnel (including full-time (35 hours or more) and part-time (less than 35 hours) personnel who were on the hospital/facility payroll at the end of the hospital's reporting period) to 10,000 population"
Hospital_Beds_Per_10k	Continuous	0- 777	"Ratio of hospital beds to 10,000 population"
Hosp_w_Obstetric_Care_Per_10k	Continuous	0- 5	"Ratio of hospitals with obstetric care to 10,000 population"
Population_Density	Continuous	0- 70,000	"Population density per square mile"
Number_of_MDs_Per_10k	Continuous	0- 339	"Ratio of total active MDs (federal and non-federal) to 10,000 population"
Number_of_ObGyn_Per_10k	Continuous	0- 7.4	"Ratio of total active ObGyns to 10,000 population"
Inpatient_days_in_cnty	Continuous	0- 28 (Mean= 0.7)	"Annual average of inpatient days per person in county"
Outpatient_days_in_cnty	Continuous	0- 53 (Mean= 2.5)	"Annual average of outpatient days per person in county"

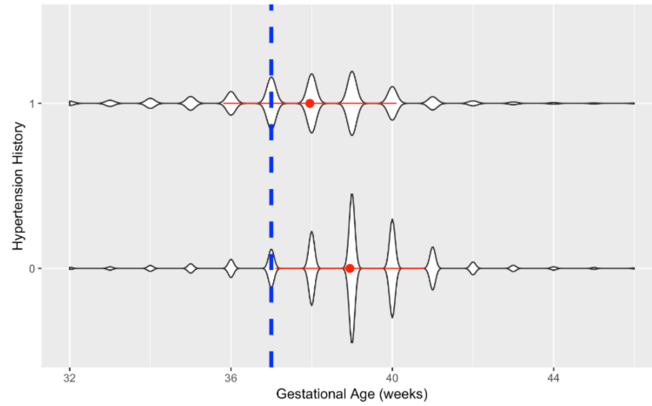
Appendix I Visualization of gestational age (gestational age) vs. selected covariate

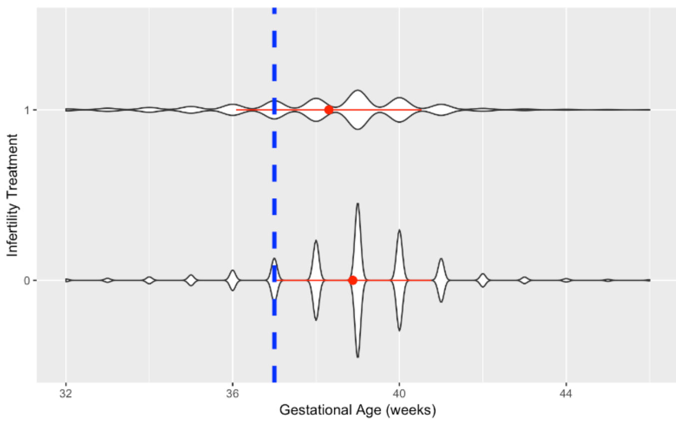
Note that we used *violin* graphs to plot the density of data points at each level. Also, the dotted vertical line separates the preterm and full-term births. Preterm births are at the left side of the vertical line, and the full-terms are on the right.

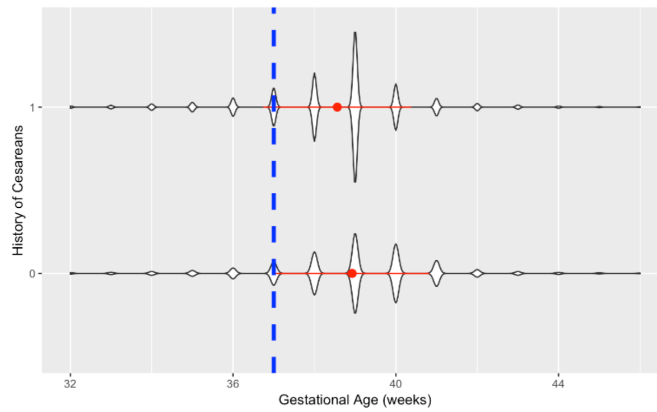


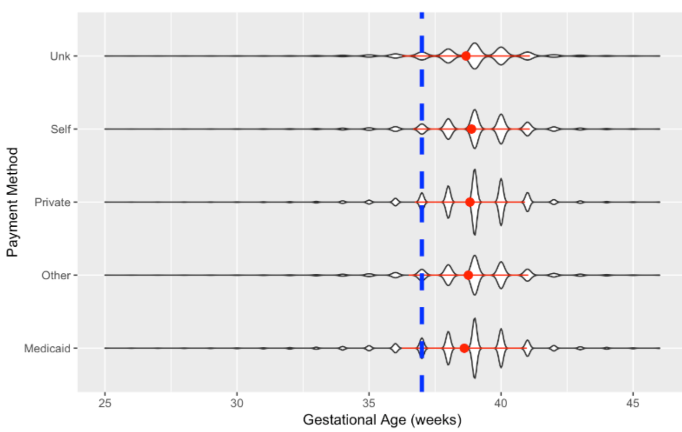

Adequate Care
Gestational Age (weeks)
Inadequate Care
Gestational Age (weeks)
Mother's BMI
Gestational Age (weeks)
Diabetes History
Gestational Age
Number of Cigarettes Before Pregnancy
Gestational Age (weeks)
Number of Cigarettes During 1st Trimester of Pregnancy
Gestational Age (weeks)

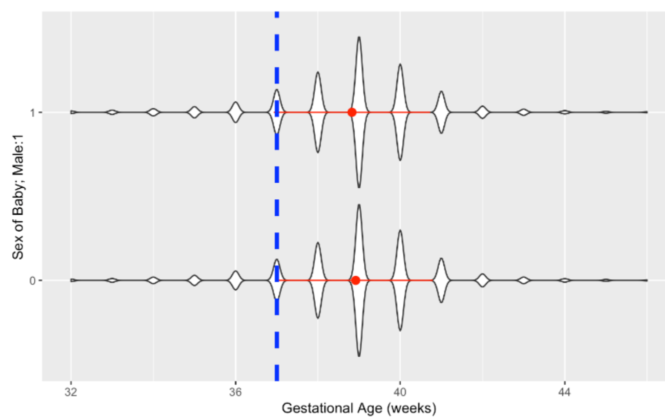

Diabetes History
1
0
32
36
40
44
Gestational Age (weeks)

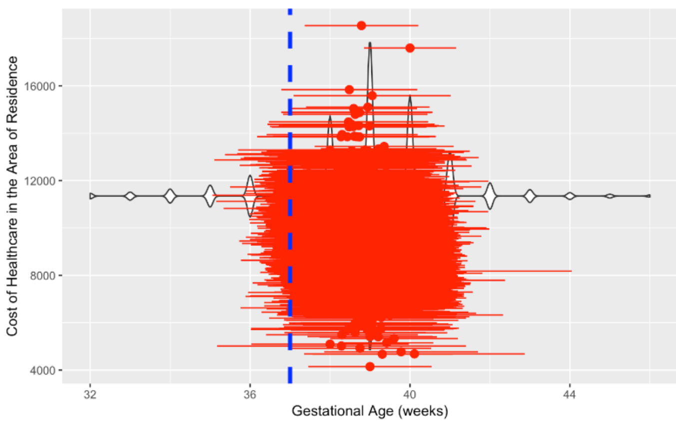

Hypertension History
1
0
32
36
40
44
Gestational Age (weeks)

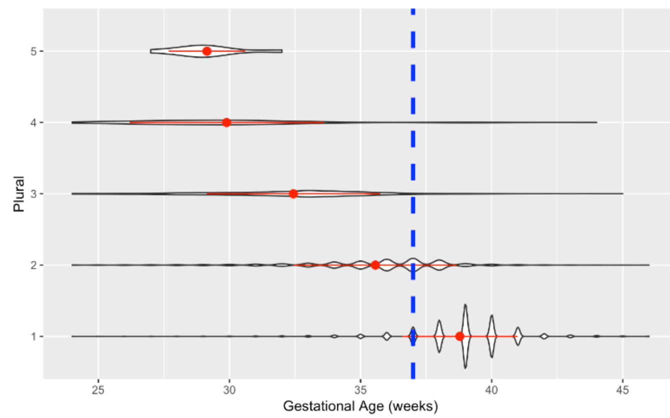

Infertility Treatment
1
0
32
36
40
44
Gestational Age (weeks)


History of Cesareans
1
0
32
36
40
44
Gestational Age (weeks)


Payment Method
Unk
Self
Private
Other
Medicaid
25
30
35
40
45
Gestational Age (weeks)


Sex of Baby; Male:1
1
0
32
36
40
44
Gestational Age (weeks)


Cost of Healthcare in the Area of Residence
16000
12000
8000
4000
32
36
40
44
Gestational Age (weeks)


Plural
5
4
3
2
1
25
30
35
40
45
Gestational Age (weeks)

Appendix J Hyperparameters of RF, GBM, and lightGBM

Table J-1 Random forest hyperparameters

Parameter	Value	Range
Number of trees	600	Grid search in the [20, 500]
Max depth	13	Grid search in the [5, 20]
Cross validation	5	5 to 10 is good for preventing overfitting.
Histogram type	"Round Robin"	The algorithm will cycle through all histogram types (UniformAdaptive, Random, and QuantilesGlobal) one per tree.
Min split improvement	0.0001	To balance training AUC and overfitting.
Sample rate	0.65	Grid search test
The number of columns to randomly select at each level	0.8	Grid search test
Categorical encoding	Enum_Limited	Enum creates one column per category feature. Enum_limited maps strings to integers.

Table J-2 The GBM hyperparameters

Parameter	Optimal Value	Range
Learning Rate	0.04	Feeding the grid search results into surrogate modeling
Learning Rate Annealing	0.99	Recommended by best practices
Number of trees	480	Feeding the grid search in the [50, 1000] to the Bayesian optimization
Max depth	14	Feeding the grid search in the [5, 15] to the Bayesian optimization
Sample rate	0.55	Feeding the grid search in the [0.1, 1] to the Bayesian optimization
Column sample rate	0.79	Feeding the grid search in the [0.1, 1] to the Bayesian optimization

Table J-3 The LightGBM hyperparameters

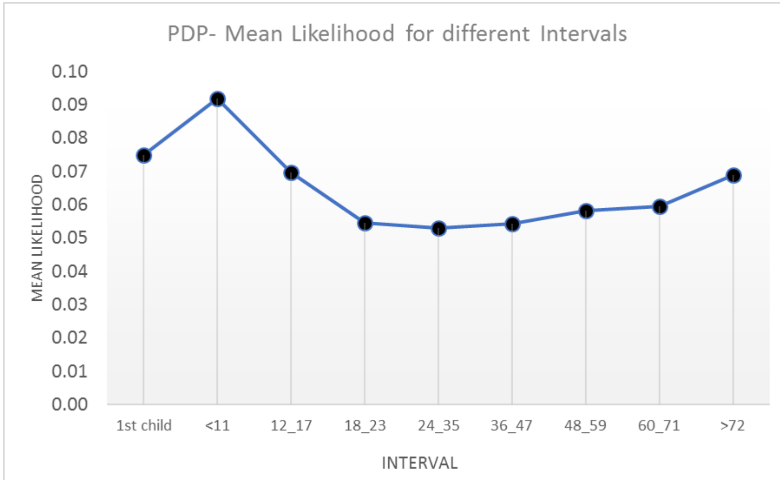
Parameter	Optimal Value	Range
Learning Rate	0.008	Feeding the grid search results into surrogate modeling
Number of trees	250	Feeding the grid search in the [50, 1000] to the Bayesian optimization
Max depth	13	Feeding the grid search in the [5, 15] to the Bayesian optimization
Sample rate	0.55	Feeding the grid search in the [0.2, 0.8] to the Bayesian optimization
Min child row	20	Grid search between [10, 100]
Column sample rate	0.79	Feeding the grid search in the [0.1, 1] to the Bayesian optimization

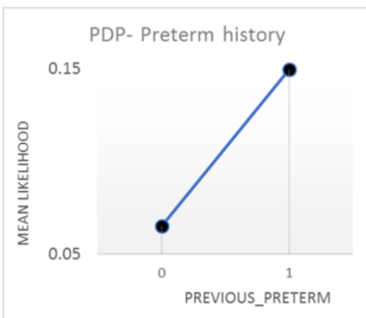
Appendix K The importance of variables in the predictive model

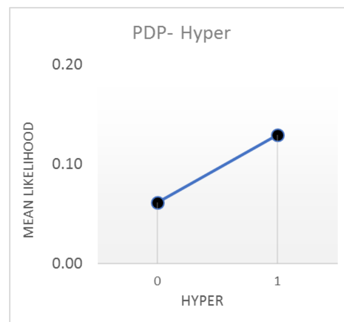
variable	*relative_importance*	*scaled_importance*	*percentage*
hyper	312937.2813	1	0.1092
interval	171301.625	0.5474	0.0598
Previous_preterm	161489.9844	0.516	0.0563
bmi	154535.0781	0.4938	0.0539
Education_F	121954.1797	0.3897	0.0425
Education_M	104375.5859	0.3335	0.0364
Age_M	96351.9531	0.3079	0.0336
Adequate	84977.4063	0.2715	0.0296
Age_F	72292.6953	0.231	0.0252
Race_F	56246.9297	0.1797	0.0196
Race_M	54843.5078	0.1753	0.0191
Healthcare_Cost	49813.7344	0.1592	0.0174
priorlive	43385.2227	0.1386	0.0151
Females_Under_65_no_Health_Ins	42884.2578	0.137	0.015
Inadequate	41785.1211	0.1335	0.0146
FemalesDivorced	40594.8438	0.1297	0.0142
pay	38487.6563	0.123	0.0134
Prev_Hospital_Stays	37273.3906	0.1191	0.013
diab	36240.5117	0.1158	0.0126
Long_Drive_Work	32736.0488	0.1046	0.0114
Hospital_Total_Prsnl_Per_10k	32257.8086	0.1031	0.0113
Alchohol_Driving_Death	31987.0273	0.1022	0.0112
Median_Household_Income	31884.6074	0.1019	0.0111
Number_of_MDs_Per_10k	31324	0.1001	0.0109
Income_Inequality	28918.209	0.0924	0.0101
Perc_in_Poverty	28736.9961	0.0918	0.01
Access_to_Exercise	27700.8594	0.0885	0.0097
Population_Density	27560.0508	0.0881	0.0096
Black_Perc	27256.7754	0.0871	0.0095
Outpatient_days_in_cnty	26760.5508	0.0855	0.0093
Hospital_Beds_Per_10k	26310.9961	0.0841	0.0092
College_Degree	25634.748	0.0819	0.0089
sex_M	24694.5098	0.0789	0.0086
Number_of_ObGyn_Per_10k	24203.2891	0.0773	0.0084
Insuf_Sleep	24158.4629	0.0772	0.0084
Driving_Alone_Work	23626.7188	0.0755	0.0082
Hisp_Perc	23499.4375	0.0751	0.0082
Exessive_Drinking	23287.0313	0.0744	0.0081
Perc_Ove_SixtyFive	23045.1094	0.0736	0.008
Chil_in_Single_Parent	22389.7988	0.0715	0.0078
Previous_cesareans	21795.127	0.0696	0.0076
Physical_Inacitvity	21707.918	0.0694	0.0076
cig_2	21690.3262	0.0693	0.0076
Married	21663.6855	0.0692	0.0076
Severe_Housing_Prob	21630.8984	0.0691	0.0075
Rural_Perc	21055.4102	0.0673	0.0073
Poor_Physical_Days	20756.043	0.0663	0.0072
WIC	20681.3223	0.0661	0.0072
Adult_Obesity	20585.418	0.0658	0.0072
Perc_Und_Eighteen	20537.6309	0.0656	0.0072
Poor_Mental_Days	20207.4531	0.0646	0.007
Lim_Healthy_Food	20075.7383	0.0642	0.007
Social_Asoc	19916.6934	0.0636	0.0069
Food_Insecurity	19213.4922	0.0614	0.0067
Children_In_Pov	17782.291	0.0568	0.0062
Uninsured_Adults	17158.373	0.0548	0.006
cig_1	17148.5605	0.0548	0.006
Poor_Health	16795.875	0.0537	0.0059
Uninsured_Children	16560.6484	0.0529	0.0058
Inpatient_days_in_cnty	16391.0723	0.0524	0.0057
Nursing_Home_Beds_Per_10k	16185.1221	0.0517	0.0056

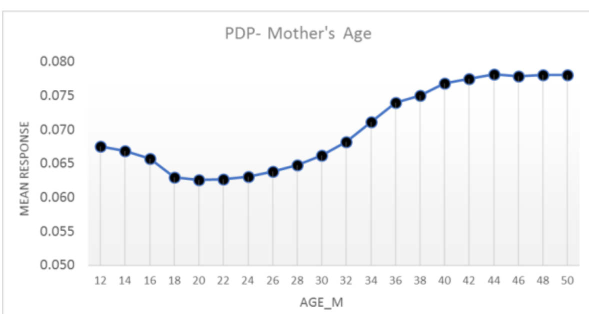
Intermediate	15514.2002	0.0496	0.0054
Other_PC_Physicians_Ratio	15313.3584	0.0489	0.0053
Uninsured_Total	15272.0635	0.0488	0.0053
Air_Particulate_Matter	15162.8574	0.0485	0.0053
Adult_Smoking	14632.1748	0.0468	0.0051
Unemployment	13191.8438	0.0422	0.0046
Dentists_Ratio	12383.541	0.0396	0.0043
priordead	10832.5498	0.0346	0.0038
cig_before	10333.5176	0.033	0.0036
Number_of_Hospitals_Per_10k	9831.4629	0.0314	0.0034
Diabetes	9789.0313	0.0313	0.0034
Hosp_w_Obstetric_Care_Per_10k	8908.4805	0.0285	0.0031
Drinking_Water_Viol	5445.4966	0.0174	0.0019
Infertility_treatment	5433.4639	0.0174	0.0019
STD	5334.0708	0.017	0.0019

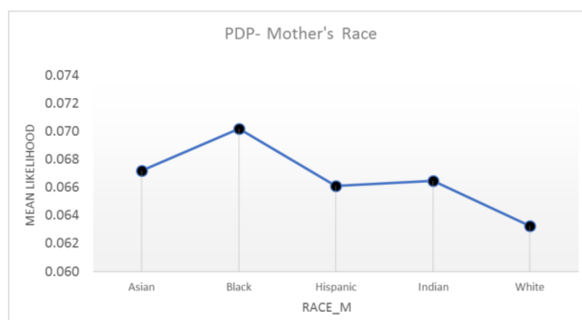
Appendix L Partial dependence plots





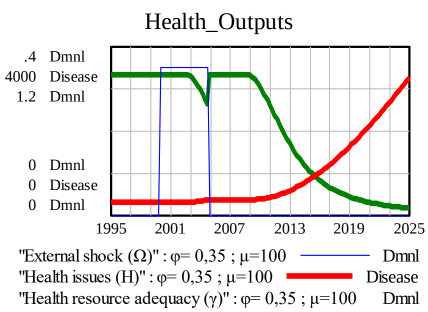




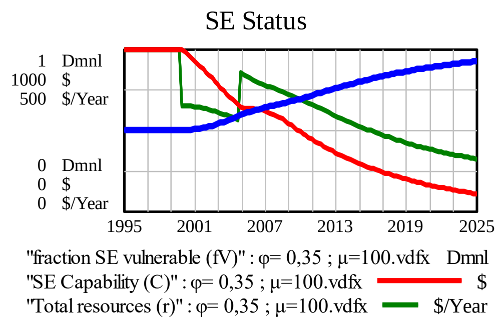


Appendix M Experimental setup

Phi=35, mu=100

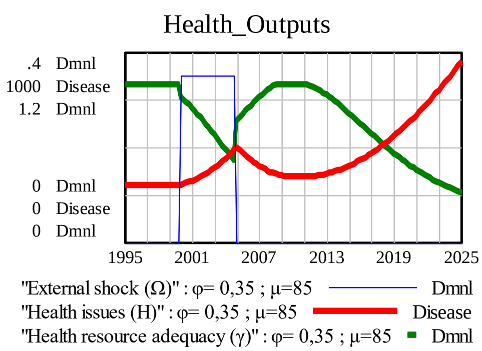


(a)

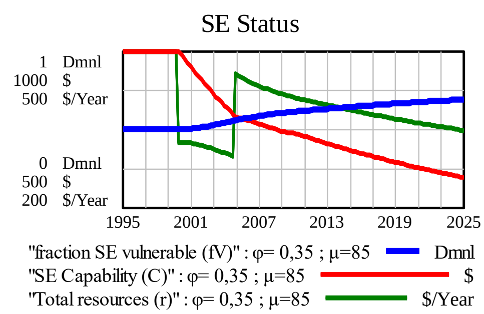


(b)

Phi=35, mu=85

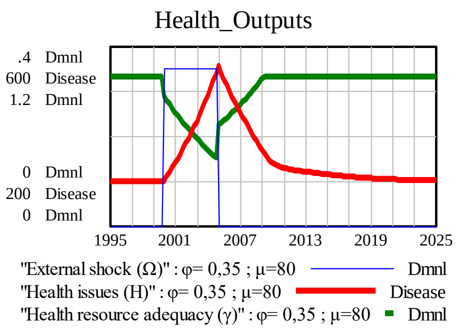


(a)

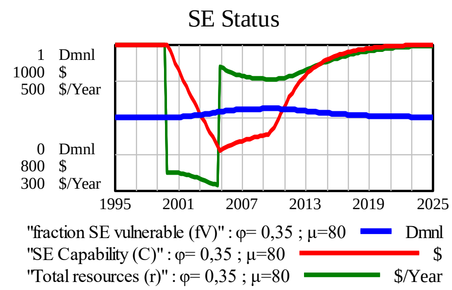


(b)

Phi=35, mu=80

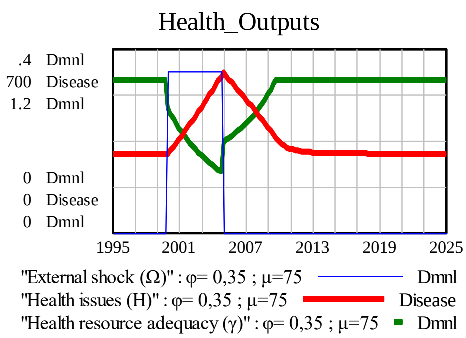


(a)

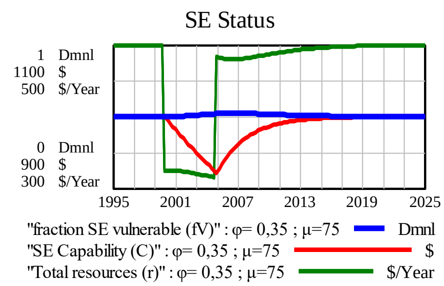


(b)

Phi=35, mu=75

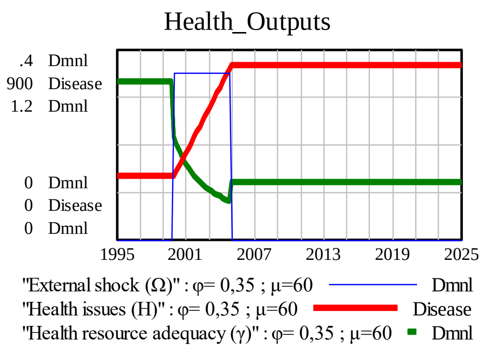


(a)

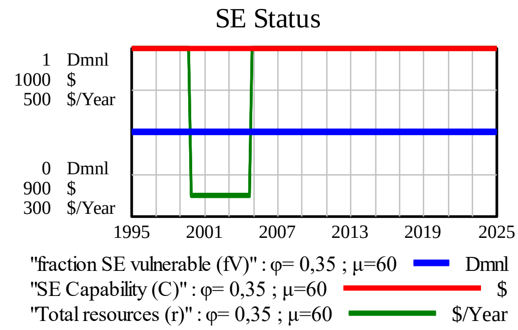


(b)

Phi=35, mu=60

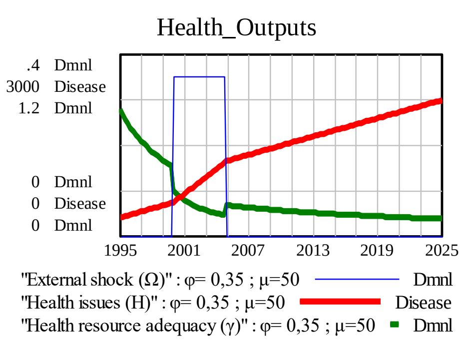


(a)

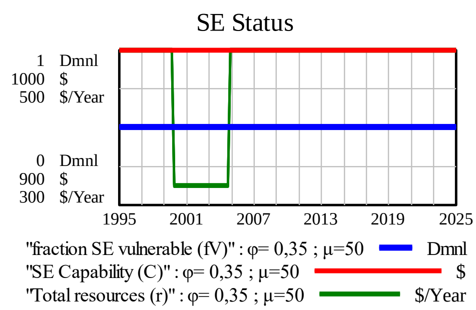


(b)

Phi=35, mu=50



(a)

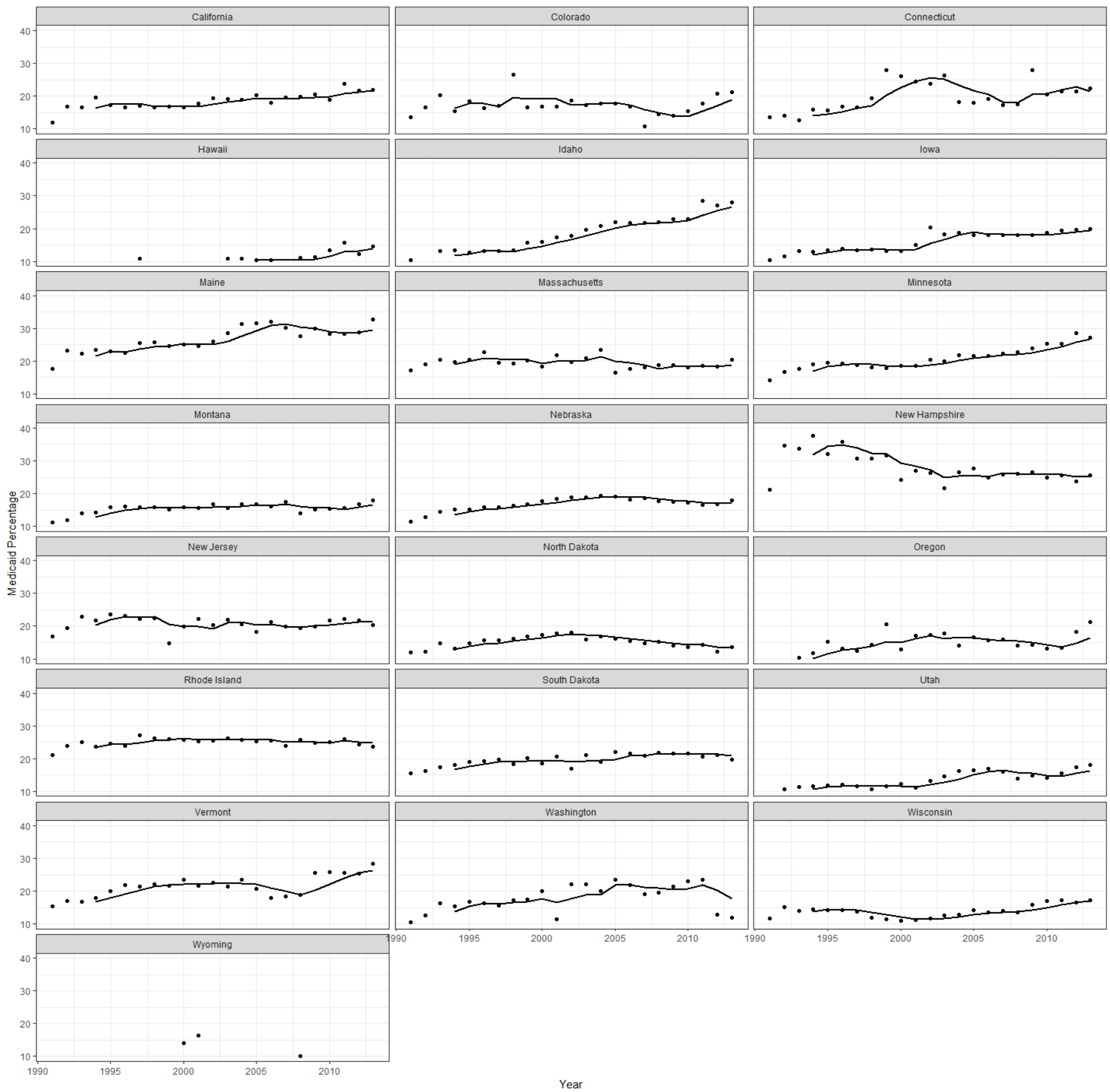


(b)

Appendix N Trend of health spending for G1 and G3

Health spending for G3: Good start, good end

Most of the states in this category has spent fairly the same percentage of their GRF budget on health. Although there are states like Minnesota that has a positive slope in the percentage of spending on health. A closer look at Minnesota health ranking shows that a higher spending did not help them. During this time, they lost their second rank and dropped to 7th among all states. Another state like Idaho experienced the same outcome, but with less severity. Increased health spending in Idaho only helped them to keep their 16th rank over time in health ranking.



Health spending for G4: Poor start, poor end

Most of the states in this category has spent more on health in 2013 compared to 1991. The majority of the trend lines are also positive. However, their health ranking is still relatively poor.