A COMPUTATIONAL SOCIAL SCIENCE APPROACH TO THE SOCIAL DETERMINANTS OF CANCER

by

Cristina Metgher
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Computational Social Science

Committee:

_____ Dr. Robert Axtell, Dissertation Director

_____ Dr. Andrew Crooks, Committee Member

_____ Dr. Naoru Koizumi, Committee Member

_____ Dr. Maksim Tsvetovat, Committee Member

_____ Dr. Kevin Curtin, Department Chairperson

_____ Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science

_____ Dr. Peggy Agouris, Dean, College of Science

Date: _____ Spring Semester 2016
George Mason University
Fairfax, VA

A Computational Social Science Approach to the Social Determinants of Cancer

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Cristina Metgher
Master of Science
Technical University of Moldova, 2010
Bachelor of Science
Technical University of Moldova, 2008

Computational Social Science Program
Department of Computational Data Sciences

Spring Semester 2016
George Mason University
Fairfax, VA

This work is licensed under a creative commons attribution-noderivs 3.0 unported license.

DEDICATION

I dedicate this dissertation to my mother, Valentina and to my brothers Pavel and Sergiu. Also, to my maternal grandparents, who both died prematurely of cancer.

ACKNOWLEDGEMENTS

It is thanks to many people that I was able to complete this work, and I am extremely grateful to each of you.

Firstly, I would like to extend my sincere gratitude to my advisor, Dr. Robert Axtell, for his constant availability, wide-ranging expertise, and support, and to my committee – Dr. Andrew Crooks, Dr. Naoru Koizumi and Dr. Maksim Tsvetovat for their guidance, insightful discussions, and helpful comments. Also, I thank all of you for being understanding and flexible regarding my immigration situation. Your support means a lot to me.

I am grateful to Open Society Institute for providing me with a four-year fellowship and to George Mason University for providing me with a Dissertation Finishing Scholarship. I thank the Santa Fe Institute, University of Washington and Johns Hopkins University for providing me with the financial support to attend summer schools and workshops that helped shape this research. I am also thankful to Dr. Karel Pacak of NIH and Dr. Carlos Sluzki of GMU for sharing with me their expertise on cancer and social epidemiology.

I cannot thank Karen Underwood (of the CSS program) enough. Your continued administrative guidance since my first weeks in the U. S. regarding coursework requirements, qualifying exams, proposal defense until the end – it means a lot to me. Thank you for bearing with my strange Eastern European accent of 6 years ago and with my endless questions. You are amazing. I also thank my CSS friends and colleagues: Ernesto Carella, Omar Guerrero, Ates Hailegiorgis, Sarah Wise, Chenna Cotla, Maciek Latek, Mohammad Rizi, Bianica Pint and Anamaria Berea.

Thank you, Kirill Cherkashin for material and emotional support and for giving me space to be miserable when I needed to. I also thank my half-sister, Maria Barbu, for being there for me throughout this process.

Last, but certainly not least, I thank Cristian Ziliberberg.

TABLE OF CONTENTS

Page

List of Tables xii

List of Figures xiv

List of Equations xix

List of Abbreviations xx

Abstract xxii

Chapter 1: Introduction 1

1.1 Context and Motivation 1

1.2 Terminology 5

1.3 Research Goals 6

1.4 Aims and Objectives 7

1.5 Dissertation Organization 8

Chapter 2: Social Determinants of Cancer 12

2.1 Introduction 13

2.2 Social Determinants of Cancer 13

2.3 Social Epidemiology Theories 17

2.3.1 Psychosocial Theory 18

2.3.2 The Social Production of Disease Theory 19

2.3.3 Ecosocial Theory 20

2.3.4 Other Social Theories 21

2.4 Social Networks 26
2.5 Current Methodologies in Social Epidemiology 29
2.6 Complex Systems and Cancer 30
2.7 Summary 33
Chapter 3: Colorectal Cancer 35
3.1 Introduction 36
3.2 CRC Overview and Staging 37
3.3 CRC Risk Factors 41
3.4 CRC Prevention 45
3.4.1 CRC Screening Methods 47
3.4.2 Sensitivity and Specificity 51
3.4.3 CRC Screening Barriers and Facilitators 53
3.5 CRC Treatment and Survival 56
3.6 CRC Recurrence and Quality of Life 60
3.7 CRC Disparities 61
3.8 Delay in Diagnosis and Treatment 65
3.9 Summary 71
Chapter 4: Methodology 72
4.1 Introduction 73
4.2 Computational Social Science Modeling Approaches 74
4.2.1 Cellular Automata (CA) 74
4.2.2 System Dynamics (SD) 75
4.2.3 Microsimulation Models (MSM) 75

4.2.4 Agent Based Models (ABM) 76
4.2.5 Geographic Information Systems (GIS) 77
4.2.5 Social Network Analysis (SNA) 83
4.3 Data-driven ABM – a "bottom-up" Approach to SDC 83
4.4 Calibration 87
4.4.1 Input vs. Output Calibration 88
4.5 Verification and Validation of ABMs 89
4.6 "Understanding our Creations." 93
4.6.1 UML vs. Pseudocode 93
4.6.2 The ODD Protocol 95
4.7 Summary 96
Chapter 5: Study Area 98
5.1 Introduction 99
5.2 The Geography of the District 100
5.3 Socio-Demographic Characteristics of DC Residents 103
5.3.1 Age distribution 104
5.3.2 Race/Ethnicity 107
5.3.3 Marital Status 110
5.3.4 Education 111
5.3.5 Income and Poverty 114
5.3.6 Employment 118
5.3.6 Occupations 120
5.4 Health Characteristics of DC Residents 121

5.4.1 Life Expectancy 123
5.4.2 Overweight and Obesity 124
5.4.3 Physical Activity 125
5.4.4 Smoking/Tobacco Use 127
5.4.5 Alcohol Use 128
5.4.6 Health Insurance/ Access to Heath Care 130
5.5 Health Disparities 133
5.6 Colorectal Cancer in DC 135
5.6.1 CRC Status in DC 135
5.6.2 CRC Incidence in DC 138
5.6.3 CRC Mortality and Survival in DC 141
5.6.4 CRC Screening Status in DC 147
5.6.5 CRC Disparities Summary 151
5.7 Summary 154
Chapter 6: Cancerscape 156
6.1 Introduction 157
6.2 Data 159
6.2.1 Census Data 160
6.2.2 GIS Data 161
6.2.3 Joining Data to Shapefiles 162
6.3 State Machines for Specifying Behavior 163
6.4 Follow the Average (FTA) Mechanism 167
6.5 Model Overview 169

6.5.1 Purpose 174
6.5.3 Process Overview and Scheduling 180
6.6 Design Concepts 184
6.6.1 Basic Principles 184
6.6.2 Interaction 185
6.6.3 Emergence 185
6.6.4 Adaptation 185
6.6.5 Stochasticity 186
6.6.6 Collectives 186
6.6.7 Observation 186
6.7 Details 187
6.7.1 Initialization 188
6.7.2 Input 189
6.7.3 Submodels 190
6.8 Summary 220
Chapter 7: Verification, Validation & Results 221
7.1 Introduction 222
7.2 Model Verification and Validation (V&V) 224
7.2.1 Documentation 227
7.2.2 Debugging 228
7.2.3 Code Walkthroughs 229
7.2.4 Unit Testing (Iterative Programmatic Testing) & Mocking 229
7.2.5 Model Profiling 230
7.3 Model Calibration and Sensitivity Analysis 231

7.4 Model Validation 235
7.4.1 Model Validation Considerations 235
7.4.2 Matching Spatial Data 239
7.4.3 General Considerations 242
7.4.4 ABM and GIS Integration within Cancerscape 242
7.4.5 Model Scenario Analysis Overview 248
7.5 Theme I: Exploring CRC Incidence 249
7.5.1 Overall CRC Incidence Results 250
7.5.2 Disparities in CRC Incidence 253
7.6 Theme II: CRC Screening 256
7.7 Theme III: CRC Survival 259
7.8 Theme IV: Delay in CRC Diagnosis and Treatment 266
7.9 Scenarios and Discussion of Results 269
7.9.1 Scenario 1: Reducing CRC Screening Age for Average Population 269
7.9.2 Scenario 2: Increasing CRC Knowledge for General Population 274
7.9.3 Scenario 3: Increasing the Health Insurance Coverage for Colonoscopies 280
7.10 Summary 285
Chapter 8: Policy Implications 287
8.1 Introduction 287
8.2 CRC Legislation 288
8.3 Overall Implications of the Cancerscape Framework 289
8.4 Implications of Scenario Analyses Results 290
8.5 Summary 300

Chapter 9: Conclusions 304
9.1 Introduction 304
9.2 Summary of Research Findings 306
9.3 Scientific Contributions 313
9.4 Limitations 316
9.5 Future Research 317
9.6 Concluding Remarks 319
Appendix A: Cancerscape Results Presentation Dashboard 321
Appendix B: Algorithm for Generating the Results Dashboard 322
Appendix C: Spatial distribution of CRC incidence for African-Americans, 2000-2004 324
Appendix D: High Level Model Overview 326
Appendix E: Overview of Cancerscape Modules Interaction 327
Appendix F: Summary of Default Input Parameters 328
Appendix G: Summary of Main Scenario Results 334
References 338

LIST OF TABLES

Table Page
Table 1: The structure of the dissertation 9
Table 2: The cancer control continuum (Source: NCI, 2015) 14
Table 3: Characteristics of screening tests for colorectal cancer (Source: Pignone et al., 2002) 50
Table 4: An example of a 2x2 (two by two) table 52
Table 5: Sensitivity and specificity of screening tests 52
Table 6: The ODD protocol overview 95
Table 7: Employment status of residents of age 16 and over, DC 2000 (Source: USCB, 2000) 118
Table 8: DC CRC stage at diagnosis by sex and race, 2000-2004 (Source: Altekruse et al., 2010) 152
Table 9: Data types required by Cancerscape 160
Table 10: Periodicity in Cancerscape 163
Table 11: The 'follow the average' (FTA) rule 168
Table 12: Percentage incidence of polyp≤5mm at age [a, b], given R = r and F = f, for polyp pathway of CRC development (Source: Gopalappa et al., 2011) 188
Table 13: Mean times to progress from event *i* to event *j*, given R = r and F = $f1\lambda i|rfj$ (Source: Gopalappa et al., 2011) 196
Table 14: CRC stage distribution (%) by race/ethnicity, 1999-2006 (Source: Altekruse et al., 2007) 197
Table 15: Relation between stage and the most frequently observed symptoms (Source: Gonzalez-Hermoso et al., 2004) 198
Table 16: Pseudocode showing the process of assigning education based on race age and gender 206
Table 17: Average number of people and relation types (Source: McCarty et al., 2001) 209
Table 18: Properties of a human social network (Source: Wise, 2014) 210
Table 19: The pros and cons of selected CRC screening tests (Source: ACS, 2014) 215
Table 20: The four levels of assessing ABM performance and analysis (Source: Axtell & Epstein, 1994) 223
Table 21: The change in survival rates by recommended screening age 274
Table 22: The change in incidence rates by health insurance coverage for colonoscopies 280
Table 23: Relative change in CRC incidence by year for Scenario 1 291
Table 24: Relative change in CRC survival by year for Scenario 1 292

Table 25: Relative change in CRC incidence by year for Scenario 2 293
Table 26: Relative change in CRC survival by year for Scenario 2 293
Table 27: Relative change in CRC incidence by year for Scenario 3 296
Table 28: Relative change in polyp removal rates by year for Scenario 3 297
Table 29: Relative change in CRC mortality by year for Scenario 3 298

LIST OF FIGURES

Figure Page
Figure 1: Social determinants of cancer framework by Hiatt & Breen (2008)..... 16
Figure 2: Distribution of health practices and risk factors (such as cigarette smoking, alcohol consumption). Percent of men and women who have several high-risk behaviors (3-5) by level of social network index (Alameda County study, n = 2229 men; n = 2496 women, ages 30-69 years) (Source: Berkman & Glass, 2000)..... 27
Figure 3: Conceptual models of how social networks affect health (Source: Berkman & Glass 2000) 28
Figure 4: Comparative data showing the change in death rate for major diseases. The death rate for cancer has declined at a much slower pace over the past 60 years (Source: Agus, 2012)..... 31
Figure 5: Anatomy of the lower digestive system (Source: Blausen Medical Communications, Inc., 2016)..... 37
Figure 6: Colorectal cancer growth (Source: Mayo Clinic, 2016)..... 40
Figure 7: Continuum of cancer prevention, treatment, and outcomes (Source: Price et al., 2012, adapted from Zapka et al., 2003) 46
Figure 8: Algorithm for CRC screening and surveillance (Source: Winawer et al., 1997) 51
Figure 9: Relative survival for all races, both sexes by SEER summary stage 2000 (Source: SEER, 2016)..... 59
Figure 10: CRC incidence trends by age and sex. Rates are age adjusted to the 2000 United States standard population (Source: American Cancer Society, 2014) 63
Figure 11: CRC stage distribution (%) by race/ethnicity, 2003-2009 (Source: NCI, 2013) 64
Figure 12: Categorization of delay (Source: Hansen et al., 2008)..... 66
Figure 13: Decision to seek delay care in three stages of an illness episode (Source: Safer et al., 1979) 67
Figure 14: Andersen model of total patient delay (Source: Andersen et al., 1995)..... 68
Figure 15: Pathways to treatment model (Source: Walter et al., 2012)..... 70
Figure 16: Power vs. difficulty continuum of model development (Source: Crooks & Castle, 2012) 82
Figure 17: The logic of simulation as a method (Source: Gilbert & Troitzsch, 2005)..... 85
Figure 18: A general validation process of ABMs (Source: Ngo and See, 2012)..... 92
Figure 19: The study area (Source: Google Maps, 2016)..... 100
Figure 20: Washington, DC 2000 census tracts (Source: NeighborhoodInfo DC, 2016)101
Figure 21: The 2000 Washington, DC Census profile (Source: USCB, 2000) 103

Figure 22: Age groups distribution for 2000 in DC (Source: USCB, 2000) 105
Figure 23: Population by 5-year age groups in 2000: Washington, DC (Source: Singer, 2004) 105
Figure 24: Share of population 65 and older, 2000: Washington, DC (Source: Berube et al., 2003) 106
Figure 25: Population proportion by race/ethnicity in Washington, DC, 2000 (Source: USCB, 2000)..... 108
Figure 26: Race and ethnicity Washington, DC profile, 2000 (Source: Eric Fisher, Data Source: Census 2000). Red represents the White population, Blue – the African Americans; Green – Asian; Orange represents Hispanics and Gray refers to the population ethnicity marked as Other in the census. Each dot marks 25 people; created by Eric Fischer 109
Figure 27: Population proportion by marital status in Washington, DC, 2000 (Source: USCB, 2000)..... 110
Figure 28: Proportion of population of age 25 and over by education level in DC, 2000 (Source: USCB, 2000) 113
Figure 29: Proportion of population of age 25 and over with at least a bachelor's degree by race/ethnicity, 2000 (Source: USCB, 2000) 114
Figure 30: Per-capita income for 1999 (Source: USCB, 2000)..... 115
Figure 31: Median household income by race/ethnicity in DC, 2000 (Source: USCB, 2000) 116
Figure 32: Proportion of individuals living in poverty in DC, 2000 (Source: Berube et al., 2003) 117
Figure 33: Unemployment rates by wards in DC, 2000 (Source: USCB, 2000)..... 119
Figure 34: Proportion of employed civilian population 16 years and over by occupation in DC, 2000 (Source: USCB, 2000)..... 121
Figure 35: Top ten leading causes of death in DC, 2000 (Source: DC DOH, 2002)..... 122
Figure 36: Life expectancy at birth, by race and sex in DC between 1999–2001 123
Figure 37: Life expectancy at birth, by race and sex in U.S. between 1999-2000 123
Figure 38: Overweight prevalence by race, DC 2000..... 125
Figure 39: Obesity prevalence by race, DC 2000 125
Figure 40: Physical activity prevalence by race in DC, 2000..... 126
Figure 41: Smoking frequency by race/ethnicity for adults 18 and over in DC, 2000 ... 128
Figure 42: Alcohol use by age group, Washington DC, 2000 (Source: DC Department of Health, 2001)..... 129
Figure 43: Health insurance coverage status and type of coverage, DC 2000 131
Figure 44: Proportion of individuals with any type of health care access/coverage, DC 2000 (Source: CDC BRFSS, 2000) 132
Figure 45: National Vital Statistics System Data on age-adjusted CRC death rates (Source: U.S. Cancer Statistics Working Group, 2015) *Rates are per 100,000 persons and are age-adjusted to the 2000 U.S. standard population..... 136
Figure 46: CRC incidence by age groups for all races, both sexes (Source: SEER, 2016) 137

Figure 47: Age-adjusted overall CRC incidence and mortality rates in Washington, DC between 2000-2004 (Source: DC Department of Health, 2000b)..... 138
Figure 48: CRC incidence by gender in DC for the years 2000 and 2005 (Source: NCHS, 2015) 139
Figure 49: CRC incidence by race in DC for the years 2000 and 2005 (Source: NCHS, 2015) 140
Figure 50: Age-adjusted cancer death rates for the 4 primary sites with the highest rates within DC, for men and women in 2000, DC (Source: USCS, 2015) 142
Figure 51: Overall CRC mortality rates between 2000 and 2004, DC (Source: DC Department of Health, 2000b) 142
Figure 52: Overall CRC mortality rates for White and African American population groups in Washington, DC between 2000-2004 (Source: DC DOH, 2000b)..... 143
Figure 53: Overall CRC incidence rates for males and females population groups in Washington, DC between 2000-2004 (Source: DC DOH, 2000b)..... 144
Figure 54: Trends in CRC incidence rates by race/ethnicity and sex, 1975-2010 (Source: SEER, 2013) 145
Figure 55: Five-year CRC-specific Survival (%) by race/ethnicity, 2003-2009 (Source: SEER, 2013) * probability of not dying from colorectal cancer within 5 years of diagnosis 146
Figure 56: CRC covariate adjusted survival by stage (Source: ACS, 2008) 147
Figure 57: Adults aged 50+ who have had a blood stool test within the past two years in DC by age group, 2002 and 2004 (Source: CRC BRFSS, 2002) 148
Figure 58: Adults aged 50+ who have ever had a sigmoidoscopy or colonoscopy in DC by age group, in 2002 and 2004 (Source: CDC BRFSS, 2002)..... 149
Figure 59: Adults aged 50+ who have ever had a sigmoidoscopy or colonoscopy in DC by gender, in 2002 and 2004 (Source: CDC BRFSS, 2002)..... 150
Figure 60: Proportion of people getting CRC screenings by race (Source: NCHS, 2015) 152
Figure 61: CRC mortality rates, per 100,000 people (Source: NCHS, 2015) 153
Figure 62: Cancer mortality rates in DC from 2001-2004, per 100,000 (Source: DC DOH, 2000b) 154
Figure 63: The tools used for building Cancerscape: Java 7, IntelliJ IDEA 15, MASON, JavaScript, R and Github 158
Figure 64: QGIS interface showing prepared shapefiles to be included in the model ... 161
Figure 65: Sample of the attribute table of the DC 2000 demographic shapefile..... 162
Figure 66: Agent daily activities FSM diagram..... 164
Figure 67: Time used on an average workday for employed persons ages 25 to 54 with children (Source: ATUS, 2015) 165
Figure 68: Hours per day that individuals age 55 and over spent doing selected activities (Source: ATUS, 2015) 166
Figure 69: High-level model diagram..... 170
Figure 70: A framework addressing the influence of social factors on cancer..... 171
Figure 71: Section of class diagram for person 175

Figure 72: UML diagram showing the relation between the *Person*, *Cancer* and *CancerProgression* 177
Figure 73: Model screenshot with generated tracts and roads 178
Figure 74: Cancerscape model close-up including primary care units (hospital icon), houses (dark cyan) and workplaces (light cyan) 179
Figure 75: Person schedules in Cancerscape 180
Figure 76: UML sequence diagram presenting the interaction between the Person, Primary Care Physician and the Specialist 182
Figure 77: Close up of Cancerscape GUI showing the simulation layers available for display 187
Figure 78: Cancer surveillance in Cancerscape 194
Figure 79: The Vogelstein model for multistep carcinogenesis (Source: Vogelstein & Kinzler, 1993) 195
Figure 80: Natural history of CRC 196
Figure 81: The risk of contracting cancer increases with age 199
Figure 82: The cancer index diagram 200
Figure 83: High level diagram of population generation within the Cancerscape world 202
Figure 84: The overall distribution of houses across Cancerscape environment 204
Figure 85: A close-up of houses along the road networks in Cancerscape 205
Figure 86: Agents color-coded by their cancer status 207
Figure 87: The generated agent social networks in Cancerscape 212
Figure 88: The degree distribution of the generated social network 213
Figure 89: Overview of the screening and treatment processes in Cancerscape 217
Figure 90: The initial UI console window 218
Figure 91: Real world and simulation world relationships with verification and validation (Source: Sargent, 2000) 226
Figure 92: Example of code documentation 227
Figure 93: Cancerscape profiling with JProfiler 8.1.2 231
Figure 94: Washington DC 2000 population groups by age 238
Figure 95: Modeled population groups in Cancerscape 239
Figure 96: A close-up of the Washington DC map (Source: OpenStreetMap, 2015) 240
Figure 97: A close-up view of the Cancerscape equivalent of Figure 96 241
Figure 98: Cancerscape model runs with all GIS layers for five consecutive years 245
Figure 99: Close up of agents, businesses, and hospitals 247
Figure 100: Modeled CRC incidence by gender in DC, 2000-2004 250
Figure 101: Observed (A) and modeled (B) incidence rates for CRC in DC, 2000-2004 252
Figure 102: Modeled overall CRC incidence rates by race in DC, 2000-2004 254
Figure 103: Modeled CRC screening rates by gender, between 2000-2004 257
Figure 104: Adults of age 50 years and over who have ever had a sigmoidoscopy or colonoscopy by gender, in 2002 and 2004 (Source: BRFSS, 2015) 258
Figure 105: Modeled screening rates for Whites and African Americans by health insurance coverage in DC, 2000-2004 259
Figure 106: A: Observed (A) and modeled (B) CRC mortality rates in DC, 2000-2004 260

Figure 107: Modeled CRC mortality rates by gender in DC, 2000-2004 262
Figure 108: Modeled CRC mortality rates by race in DC, 2000-2004 263
Figure 109: Modeled 5-year CRC survival rate by stage at diagnosis, 2000-2004 264
Figure 110: Modeled mortality rates due to CRC and failed treatment, 2000-2004 265
Figure 111: Modeled delay in seeking medical attention and beginning treatment by health insurance (in days) 267
Figure 112: Modeled delay in seeking medical attention and beginning treatment by race (in days) 268
Figure 113: Modeled CRC incidence by recommended screening age starting at 30, 40 and 50 years 271
Figure 114: Modeled polyp removal rates by recommended screening age starting at 30, 40 and 50 years 272
Figure 115: Hours Modeled CRC survival by recommended screening age starting at 30, 40 and 50 years 273
Figure 116: Modeled CRC incidence by CRC knowledge levels at 30%, 60% and 90% 275
Figure 117: Modeled delays by health insurance coverage for CRC knowledge levels at 30%, 60% and 90% 277
Figure 118: Modeled CRC polyp removal rate by CRC knowledge levels at 30%, 60% and 90% 278
Figure 119: Modeled CRC survival rate by CRC knowledge levels at 30%, 60% and 90% 279
Figure 120: Modeled CRC incidence rate by colonoscopy coverage levels at 30%, 60% and 90% 281
Figure 121: Modeled CRC polyp removal rates by colonoscopy coverage levels at 30%, 60% and 90% 282
Figure 122: Modeled CRC incidence rate by colonoscopy coverage levels at 30%, 60% and 90% 283
Figure 123: Modeled CRC survival rate by colonoscopy coverage levels at 30%, 60% and 90% 285
Figure 124: Colossal Colon founder Molly McMaster crawls through the 40-foot-long model of a human colon. The exhibit is on a 20-city tour around the United States (Source: Reynolds, 2003) 295
Figure 125: In a public service announcement, two doctors apprehend the Polyp Man. The campaign debuted last year during Colon Cancer Awareness Month (Source: Reynolds, 2003) 295
Figure 126: Relative influence of the five major determinant categories of health: rough estimations (Source: Tarlov, 1999) 305

LIST OF EQUATIONS

Equation	Page
Equation 1	5
Equation 2	5
Equation 3	210
Equation 4	211

LIST OF ABBREVIATIONS

American Academy of Family Physicians ..... AAFP
Agent Based Modeling/Agent-based Model ..... ABM
American Time Use Survey ..... ATUS
Body Mass Index ..... BMI
Behavioral Risk Factor Surveillance System ..... BRFSS
Cellular Automata ..... CA
Center for Disease Control and Prevention ..... CDC
Colorectal Cancer ..... CRC
Computational Social Science ..... CSS
Comma-separated Values ..... CSV
Computed Tomography ..... CT
Computed Tomographic Colonography ..... CTC
District of Columbia ..... DC
Familial Adenomatous Polyposis ..... FAP
Fecal Occult Blood Test ..... FOBT
Finite State Machine ..... FSM
Follow the Average ..... FTA
Geographic Information System ..... GIS
Graphical User Interface ..... GUI
Health Care Provider/Practitioner ..... HCP
JavaScript Object Notation ..... JSON
Multi-Agent Simulator of Neighborhoods/Networks ..... MASON
Metastatic Colorectal Cancer ..... MCRC
Metropolitan Statistical Area ..... MSA
Microsimulation Model ..... MSM
National Cancer Institute ..... NCI
Non-governmental Organization ..... NGO
National Institutes of Health ..... NIH
Object-oriented Programming ..... OOP
Primary Care Physician ..... PCP
Primary Care Physician ..... PCP
Pattern Oriented Modeling ..... POM
Quality of Life ..... QoL
Quality-adjusted Life Year ..... QALY
System Dynamics ..... SD
Social Determinants of Cancer ..... SDC

Surveillance, Epidemiology, and End Results ..... SEER
Socioeconomic Status ..... SES
Social Network Analysis ..... SNA
Tumor Node Metastases ..... TNM
User Interface ..... UI
Unified Modeling Language ..... UML
United States Census Bureau ..... USCB
United States Preventive Services Task Force ..... USPSTF

ABSTRACT

A COMPUTATIONAL SOCIAL SCIENCE APPROACH TO THE SOCIAL DETERMINANTS OF CANCER

Cristina Metgher, Ph.D.

George Mason University, 2016

Dissertation Director: Dr. Robert Axtell

Cancer is a complex system of systems - in which heterogeneous actors interact dynamically with each other and with the environment across time. It is the second leading cause of death in the U. S. after cardiovascular disease, striking people from all occupations and backgrounds. The U. S. President Nixon declared the "War on Cancer" in 1971. Since then, much progress has been made towards understanding this disease. However, due to the reductionist thinking that has been predominant in cancer research for the last decades, the mortality rates due to cancer have hardly changed.

This dissertation uses a computational social science approach to address the complexity inherent into the social determinants of cancer. It examines, models and tests hypotheses that explore how the social determinants of cancer may influence cancer dynamics and outcomes. This work aims to move beyond the reductionist view predominant in cancer research by employing an agent-based computational approach, and documenting the process of building a data-driven model that simulates the dynamics

of cancer in a realistic population of boundedly rational agents. The results of this dissertation aim to advance the application of computational social science methodologies to the study of the social determinants of cancer and to assist policy makers in taking more informed decisions regarding cancer control interventions.

CHAPTER 1: INTRODUCTION

Contents

1.1 Context and Motivation 1

1.2 Terminology 5

1.3 Research Goals 6

1.4 Aims and Objectives 7

1.5 Dissertation Organization 8

1.1 Context and Motivation

Cancer poses an unprecedented challenge to scientific inquiry. A total of 1,685,210 new cancer cases and 595,690 cancer deaths are projected to occur in the United States in 2016 (Siegel et al., 2016). Approximately 1/4 to 1/3 of these cancer cases can be attributed to poor nutrition, physical inactivity, overweight and obesity (American Cancer Society, 2013). The number of Americans expected to die of cancer is expected to rise in the next years, as the nation's population ages. In 1971, U. S. President Nixon signed the National Cancer Act and declared the "War on Cancer", creating the National Cancer Program. While the progress in understanding cancer has been tremendous since then, the mortality rates due to cancer have hardly changed (Agus, 2012). This is, in part, due to

the reductionist thinking that has been predominant in cancer research for the last decades.

According to the Centers for Disease Control and Prevention (CDC), CRC is the second leading cause of cancer deaths in the United States (of the cancers affecting both genders). On January 1, 2010, in the United States there were approximately 1,154,481 men and women alive who had a history of cancer of the colon and rectum – 569,930 men and 584,551 women (Howlader et al., 2014). As with most cancers, colorectal cancer can be deadly if not detected early. Late stage 5-year survival rate for CRC can be as low as 6%.

Cancer embodies many of the characteristics of a complex system. A wide range of factors interacting across time at several levels influences the incidence, prevalence, dynamics and the mortality rates due to cancer (Hiatt & Breen, 2008). The current understanding of carcinogenesis is based on the conventional reductionist view of disease. In the past decades, the focus of cancer research has been on single individual genes gone wrong, single mutations, single associations. Cancer research largely neglected the importance of an emphasis on the individual and society from a systems science perspective (Agus & Gell-Mann, 2012) – a system that is composed of, and influenced by several other systems.

Among the main goals of this dissertation is to encourage cancer-related social science research using novel computational social science methodologies. Social determinants have been defined as “the circumstances in which people are born, grow up, live, work, and age, as well as the systems put in place to deal with illness” (WHO,

2008). They have often been referred to as the "fundamental cause of disease" (Link & Phelan, 1995) or "the causes of the causes" (Marmot & Wilkinson, 2005). These factors are particularly important for cancer research as they consequently maintain an association with diseases even when intervening mechanisms change. The social scattering, as well as the social determinants contributing to the population patterns of disease, is the focus of social epidemiology (Berkman et al., 2000; Krieger, 2001a; Krieger, 2005).

Social determinants influence health behaviors of individuals, groups, and populations (Kaplan, 1998; Marmot, 1995). They interact with genetic and biological factors (gene-environment interactions), shape individual and group behavior (either directly or indirectly), contextualize individual risk factors, embody the access to resources and offer a life-course perspective on disease development. The social determinants of health have traditionally been approached using reductionist, correlation-based analytic methods (e.g. regression techniques). Although useful, the epidemiological methods employed to address this multilevel complexity are not without limitations. These methods have been criticized (Oakes & Kaufman, 2006) because of their limited ability to incorporate and account for the vast variety of social causal relationships and interactions involved in disease causation and development. Traditional methods currently employed to explore the social determinants of disease are insufficient in addressing the complexity inherent in the social determinants of cancer (El Sayed et al., 2012; Hiatt & Breen 2008). To make further progress in cancer control research, an

interdisciplinary approach using new methods capable of accounting for the social influences is needed (Maglio & Mabry, 2011).

We are experiencing an unprecedented growth in computational power that makes computational social science modeling techniques an increasingly attractive methodology for understanding the causal pathways of disease. These techniques can be particularly useful for policymakers, by providing the opportunity to look into a wide range of possibilities of disease dynamics by simulating a multitude of "what-if" scenarios. During the last decades agent-based modeling (ABM) has become a reliable technique for building models to study macro level outcomes from the bottom up in several disciplines. ABM has been fruitfully applied to the explore the spread of several infectious diseases, such as H1N1 (Parker & Epstein, 2011; Brown et al., 2011), cholera (Hailegiorgis & Crooks, 2012), measles (Perez & Dragicevic, 2009) and it's becoming increasingly popular in studying chronic conditions, such as obesity (Hammond, 2009; Bourisly, 2013; Ejima et al., 2013), smoking (Kasaie et al., 2013) and drinking patterns (Gorman et al., 2006). ABM is not new to cancer research. This methodology has been widely applied to model tumor growth at the cellular level (Zhang et al., 2009; Sun et al., 2012; Dréau et al., 2009). Nevertheless, no approaches involving ABM have been proposed to address the social determinants of cancer at the individual and population level. This research aims to address this gap and the limitations of the reductionist thinking in cancer research by employing a "bottom-up" approach that integrates existing models, theories and knowledge from several disciplines involved in researching the social determinants

of cancer. Before proceeding with introducing the first chapter, and presenting the research aims, a few concepts need to be defined.

1.2 Terminology

A brief review of the terminology specifically related to the epidemiology of cancer that will be further used in this dissertation is presented below.

- Incidence. Cancer incidence rate is the number of new cancers of a specific site/type occurring in a particular population within a specified time period.

$$\text{Incidence rate} = \frac{\text{new cancer cases occuring during a given period}}{\text{total population at risk during the same period}} \qquad \text{Equation 1}$$

- Prevalence. Cancer prevalence is a measure of the total people with a previous cancer diagnosis alive at a particular point in time. It usually includes both the incidence as well as the pre-existing cases.
- Morbidity. Morbidity is another term for illness, referring to having a disease or a symptom of a specific disease.
- Recurrence. Recurrence refers to the return of cancer post treatment, usually after a period during which cancer cannot be detected.
- Mortality. Mortality statistics (in the case of cancer) refer to the number of deaths resulting from this particular disease within a certain population and during a certain period.

$$\text{Mortality rate} = \frac{\text{cancer deaths occuring during a given period}}{\text{total population at risk during the same period}} \qquad \text{Equation 2}$$

- Survival. Cancer survival statistics are typically expressed as the proportion of patients alive at some point after a diagnosis of cancer.

- Clinical trial. A scientific study that has the goal of looking for new treatments or improving the current ones.
- Screening. Screening for cancer refers to looking for cancer signs in (seemingly) asymptomatic individuals. If a screening test result is abnormal, more tests need to be done to find out if a person has cancer. These are called diagnostic tests.
- Prevention. The prevention of cancer most commonly refers to the activities known to decrease the risk of getting cancer.
- Etiology. Etiology relates to the cause, set of causes, or manner of causation of a particular disease.

1.3 Research Goals

Among the most significant goals of this work is to advance the understanding of social determinants of cancer using novel computational social science methodologies. This dissertation investigates several systems involved in shaping the social determinants of cancer using agent-based modeling, providing a "test-bed" for evaluating hypotheses on how social systems might affect cancer dynamics.

To illustrate the application of the proposed agent based framework, a study case is proposed. The study case will focus on *the social determinants of colorectal cancer in Washington, DC*. The agent-based model, coupled with geographic information systems (GIS), is driven by empirical data collected from multiple sources and simulates the multilevel influences of social systems on the dynamics of cancer in an artificial society of situated, boundedly-rational agents.

The main research question this dissertation addresses is: "*Can a bottom-up agent-based approach help to gain insight into the complexity inherent in the social determinants of cancer?*". The proposed ABM attempts to explore potential scenarios to answer questions about how society may affect the distribution of cancer at population and community levels. Through which mechanisms can social factors influence the initiation and development of cancer? Why and how does it affect only a small portion of the population exposed? The proposed ABM attempts to provide answers to substantial questions related to the social determinants of cancer, as well as to raise new issues and potentially drive new research directions and data collection efforts.

1.4 Aims and Objectives

The main goal of this dissertation research is to investigate the use of agent-based modeling in the context of social determinants of cancer. To accomplish this, the following formal list of aims and objectives was created:

1. Review and examine the literature on social determinants of cancer and distinguish the factors and behaviors that are most relevant to be included in the model.
2. Review and examine the literature on colorectal cancer and distinguish the factors and behaviors that are most relevant to be included in the model.
3. Synopsize the existing computational social science methodologies and motivate the use of agent-based modeling as the methodology of choice to address the social determinants of cancer.
4. Design and build an agent-based model that captures the complexity inherent in the social determinants of cancer.

5. Verify, validate and analyze the results of the agent-based model implemented during the previous step.
6. Apply the agent based model to different scenarios, to increase the validity of the model, and explore policies that might work best in certain situations.
7. Assess the potential of agent based modeling for addressing the social determinants of cancer.

1.5 Dissertation Organization

The structure of the dissertation, with the corresponding chapters as they relate to research aims and objectives, is presented in Table 1.

Table 1: The structure of the dissertation

Aim/Objective(s)	Chapter(s)
Review and examine the literature on the social determinants of cancer to distinguish the factors and behaviors that are most relevant to be included in the model.	*Chapter 2.* Social Determinants of Cancer
Review and examine the literature on colorectal cancer to distinguish the factors and behaviors that are most relevant to be included in the model.	*Chapter 3.* Colorectal Cancer
Synopsize the existing computational social science methodologies and motivate the choice of computational social science methods to study the social determinants of cancer.	*Chapter 4:* Methodology
Design and build an agent-based model that captures the complexity inherent in the social determinants of cancer.	*Chapter 5:* The Study Area *Chapter 6:* Cancerscape
Verify, validate and analyze the result of the agent-based model executed in the previous step.	*Chapter 7:* Verification, Validation & Results
Apply the model to different scenarios, to increase the validity of the model and explore various policy scenarios aiming to reduce the burden of CRC.	*Chapter 7:* Verification, Validation & Results *Chapter 8.* Policy Implications
Assess the results and comment on the potential of agent-based modeling to address the social determinants of cancer.	*Chapter 9.* Conclusions

A general overview of the literature on the social determinants of cancer that are included in the model is presented in Chapter 2. This chapter discusses the existing methodologies, as well as motivates the need for new approaches to address the multilevel complexity of the social determinants of cancer.

Chapter 3 extends the review of the social determinants of cancer and provides an overview of colorectal cancer, motivating the choice of this particular cancer site for the study case. Aspects such as colorectal cancer staging, symptoms, risk factors, screening, treatment and survival that are included in the model are discussed.

After the choice of the cancer site is introduced in Chapter 3, Chapter 4 proceeds by reviewing the existing computational social science methodologies and provides the motivation for a data-driven ABM to address the social determinants of cancer. In this chapter, I also describe the process of building, testing, verifying and validating ABMs, and explain some of the design aspects of building an agent-based model.

After setting the context by providing the background on social determinants of cancer and computational social science methodologies, Chapter 5 gives an overview of the study area and motivates the choice of Washington, DC as the study site for the model. It begins by providing an overview of the socio-demographic characteristics of the Washington, DC residents and proceeds by specifying their health-related characteristics. It also presents the geography of the district, health disparities, and the colorectal cancer statistics for Washington, DC. The data presented in this chapter is based on the year 2000, as this year will serve as the starting point of the model.

In Chapter 6, I describe the model using the ODD protocol (Overview, Design concepts, and Details), which is a commonly used protocol for standardizing published descriptions of agent-based models (Grimm et al., 2006). Verification and Validation (V&V) of the model are discussed in Chapter 7. The results after running the model under different scenarios are also presented in this chapter.

The practicality of using this particular methodological approach to inform policy and decision-making processes in the context of the social determinants of cancer is discussed in Chapter 8. This chapter presents the legislation related to CRC in Washington, DC and discusses the implications of both Cancerscape as a tool, as well as the implications of the model results regarding informing policy design of interventions aimed at decreasing the CRC incidence and mortality.

Finally, Chapter 9 draws overall conclusions and outlines the advantages and disadvantages of applying a computational social science approach to the social determinants of cancer. It examines the extent to which the aims and objectives of this dissertation have been accomplished and presents the broader scientific implications of this work. In the final part of this dissertation, I proceed with suggesting future directions for this research avenue. Ways of potentially extending this framework to include aspects such as modeling the costs and benefits of particular policy interventions and also expand the framework to other cancer types are discussed.

CHAPTER 2: SOCIAL DETERMINANTS OF CANCER

Contents

2.1 Introduction 13

2.2 Social Determinants of Cancer 13

2.3 Social Epidemiology Theories 17

2.3.1 Psychosocial Theory 18

2.3.2 The Social Production of Disease Theory 19

2.3.3 Ecosocial Theory 20

2.3.4 Other Social Theories 21

2.4 Social Networks 26

2.5 Current Methodologies in Social Epidemiology 29

2.6 Complex Systems and Cancer 30

2.7 Summary 33

> "Today the network of relationships linking the human race to itself and to the rest of the biosphere is so complex that all aspects affect all others to an extraordinary degree. Someone should be studying the whole system, however crudely that has to be done, because no gluing together of partial studies of a complex nonlinear system can give a good idea of the behavior of the whole."
>
> (Murray Gell-Mann, 1997)

2.1 Introduction

This chapter highlights the current knowledge on social epidemiology and provides a general overview of colorectal cancer. After laying out the historical background of social epidemiology and social epidemiology of cancer in particular, colorectal cancer risk factors, symptoms, progression, screening, and treatment are reviewed. Further, I discuss prior modeling efforts for social determinants of cancer, the gaps in these currents efforts of modeling colorectal cancer, and the potential of computational social science approaches (such as agent-based modeling, combined with GIS and social networks) to address the limitations of previous approaches.

2.2 Social Determinants of Cancer

Social determinants have been referred to as the "fundamental causes" of health and disease (Link et al., 1998), or upstream/distal cultural, social, environmental and economic factors, which predispose individuals and groups to engage in either protecting or health damaging behaviors (McKinlay, 1979; Kaplan 1996; Marmot 1998).

The social determinants of cancer refer to the social, economic, cultural and environmental factors that influence individual exposures, behaviors and access to resources, which can in turn trigger carcinogenic chains of events. Even considering for the multitude of definitions of the socioeconomic status (SES), people with a lower SES have higher chances to get cancer and die from it compared to individuals with a higher SES. The connection between SES and cancer outcomes has widely been studied (Ionescu et al., 1998; Ward et al., 2004; Clegg et al., 2009; Le et al., 2008).

The access to health care is affected by society, by the distributions of hospitals across the spatial environment, by norms and policies regarding health insurance

coverage among others. All these factors have been linked to cancer screening practices and outcomes (Optenberg et al., 1995; Wee et al., 2005; Smith et al., 2006). At the initiation stage of carcinogenesis, appropriate screening practices may contribute to early detection. Other important social factors that have shown to modulate cancer risk are diet (Cross et al., 2010; Norat et al., 2005), physical activity (Chao et al., 2004), alcohol consumption (Ferrari et al., 2007), tobacco smoke and cultural and societal values (Marmot et al., 2008). To better describe the different aspects of cancer prevention, detection, diagnosis, treatment, and survivorship, the National Cancer Institute established the "cancer control continuum", shown in Table 2.

Table 2: The cancer control continuum (Source: NCI, 2015)

CANCER CONTROL CONTINUUM				
PREVENTION	DETECTION	DIAGNOSIS	TREATMENT	SURVIVORSHIP
Tobacco control Diet Physical activity Sun exposure Virus exposure Alcohol use Chemoprevention	Pap test Mammography FOBT Sigmoidoscopy Prostate-Specific Antigen	Informed decision-making	Health services and outcomes research	Coping Health promotion for survivors

Introduced in the 1970s, the cancer control continuum has changed over time to incorporate the most recent research findings on cancer control. For example, because the number of cancer survivors is growing, survivorship was explicitly added to the framework (Rowland, 2008).

Hiatt & Breen (2008) proposed a conceptual framework that integrates social determinants of cancer into the cancer continuum, as shown in Figure 1. Drawing on the

grid introduced by Krieger (2005), the framework accounts for the dynamic nature of cancer control and the different levels of analysis, while also incorporating the impact of interventions across the continuum. For example smoking, obesity, cultural beliefs, dietary patterns and drinking habits play a role in cancer prevention, access and quality of healthcare, knowledge about personal risk factors and screening recommendations influence the early detection of cancer. Socioeconomic status and health literacy influence delay, knowledge, attitudes and cultural beliefs of the primary care physician in timely diagnosis. Among other factors, social support, socioeconomic status and stage at diagnosis influence both treatment and quality of life (Hiatt & Rimer, 1999). The framework in Figure 1 also illustrates how social determinants relate to other levels and the interventions types along the cancer care continuum. The healthcare systems are less likely to influence the incidence of cancer during the pre-disease states.

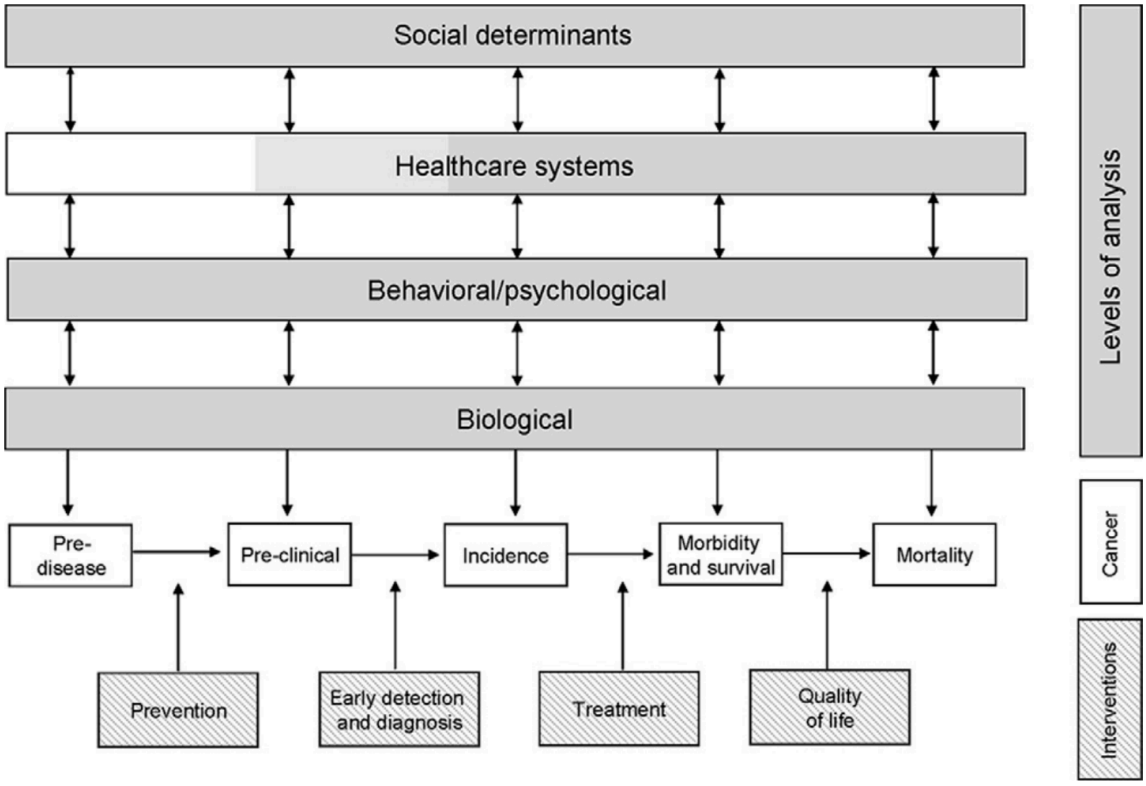


Figure 1: Social determinants of cancer framework by Hiatt & Breen (2008)

Population health at any point in time is influenced by the non-linear interaction among people and between people and their environment (both socially and spatially). Disparities have been observed across the cancer continuum concerning incidence, mortality, and survival. While the social determinants of cancer framework draws the relationships as linear, in reality, they are quite complex and multi-directional (Hiatt & Breen, 2008). The complex interaction between factors at different levels is nonlinear as well. The social determinants of cancer modulate cancer incidence, prevalence, and mortality via a wide range of mediating factors – such as individual and group behaviors (Doll & Peto, 1981).

2.3 Social Epidemiology Theories

Social epidemiology is a branch of epidemiology that deals with the social determinants of the distribution of health/disease (Berkman & Kawachi, 2000). It emerged as a consequence of the holistic models of health developed between World War I and World War II (Galdston, 1949) and after the emergence of the 'social medicine' framework during the 1940s (Ryle, 1994). Social epidemiology studies the social distribution and the social determinants of health (Berkman & Kawachi, 2000). Based on the definition provided by the WHO (2008), social determinants of health address the "the circumstances in which people are born, grow up, live, work and age, well as the systems put in place to deal with illness". The determinants mentioned above are in turn modulated by a broader range of influences: such as the economy, social and health policies, as well as politics. The projected mortality rates attributable to the social determinants in the United States are comparable to the rates attributed to pathophysiological and behavioral causes (Galea et al., 2011).

The theoretical basis of social epidemiology can be traced back to Émile Durkheim's work on suicide. Durkheim (1951) demonstrated that "social facts" could be used to explain changing patterns of the aggregate tendency towards suicide. Durkheim explored the difference in suicide rates between Protestants and Catholics, arguing that the suicide rates for Catholics were lower, mainly due to the stronger social control (Durkheim, 1951). Durkheim (1951) attributed the variations in suicide to societal phenomena such as the lack of connections between people (group attachment) and the lack of behavior regulation. Based on the ideas of Durkheim, the functionalism theory emerged, focusing on structures that create the society and on the way society can remain

stable. According to functionalism, society is striving to maintain a state of equilibrium. Society is made by groups of connected structures such as institutions (like education systems, businesses, non-governmental organizations (NGOs), medicine, marriage) and social facts (ways of thinking and acting formed by the society - such as the law, suicide, religion). Society is more than the sum of its parts. The institutions adapt and adjust to the evolving nature of social systems. Among the key theories invoked by contemporary social epidemiologists are: (1) the psychosocial theory, (2) the theory of the social production of disease and (3) ecosocial theory (Krieger, 2001b).

2.3.1 Psychosocial Theory

Psychosocial theory refers to both behavioral and endogenous biological responses to human interactions. This theory is based on the "health-damaging potential of psychological stress", that emerges as a result of hopeless situations, overwhelming responsibilities, or lack of social and emotional support (Elstad, 1998). Chronic stress can have a broad range of effects on the individual, such as altering host susceptibility, becoming directly pathogenic by affecting neuroendocrine function or inducing health-damaging behaviors (such as the use of psychoactive substances, diet, and sexual behaviors) (Berkman & Kawachi, 2000; Krieger, 2001a). Cassel (1976) argued that 'social environment' alters host susceptibility by affecting neuroendocrine function through the following psychosocial factors: dominance hierarchies, social disorganization and rapid social change, marginal status in society, including social isolation and bereavement. They may also act as a buffer to all of the above. The 'psychosocial asset' of social support states that the disease an individual gets depends on prior exposures and experiences (Krieger, 2012b). Following Cassel's work, McEwen & Seeman (1999)

described the ‘wear-and-tear from chronic over-activity or under-activity’ of the systems protecting the body, modulating stress, both internally and externally.

This brought a new perspective on psychosocial stressors, that not only do they alter susceptibility, but can also be directly pathogenic. The emerging life-course perspective states that the health status of an individual at any age is the result of prior conditions and living circumstances, starting from in utero onwards (Davey-Smith et al., 2001; Kuh & Shlomo, 1997). In summary, psychosocial theory focuses on responses to ‘stress’ and on stressed people in need of psychosocial resources.

This brought a new perspective on psychosocial stressors, emphasizing that not only do they alter susceptibility, but can also be directly pathogenic. The emerging life-course perspective states that the health status of an individual at any age is the result of prior conditions and living circumstances, starting from in utero onwards (Davey-Smith et al., 2001; Kuh & Shlomo, 1997). In summary, psychosocial theory focuses on responses to ‘stress’ and stressed people in need of psychosocial resources.

2.3.2 The Social Production of Disease Theory

The social production of disease and political economy of health address the economic causes, as well as the political factors that contribute the overall patterns of health and illness, including structural barriers to people living healthy lives (Conrad, 2005; Eyer & Sterling, 1977; Crawford, 1977). This theory highlights the capital accumulation and how the higher-level structures enforce it. According to this theory, the determinants of health are analyzed according to who benefits from specific policies and who does not. Occupational hazards, pollution, deleterious workplace organization, inadequate pay scales, tax and labor laws, spending on social programs versus prisons and the military

contribute to determining the health status of society. The theory of the social production of the disease also incorporates social inequalities related to race and ethnicity (Krieger, 1987), sexuality (Doyal, 1995) and gender (Davis et al., 1996) across various socioeconomic strata.

The health inequalities and disparities caused by economic and political decisions are considered to be the fundamental causes of the social inequalities in health. Another particularly challenging problem is economic inequality. A recent study indicates that human behavior under economic inequality further shapes inequality (Nishi & Christakis, 2015).

2.3.3 Ecosocial Theory

Ecosocial theory and the related multi-level dynamic perspectives incorporate social as well as clinical factors into a dynamic, historical and ecological outlook so that new insights into the social determinants of health can be developed (Krieger, 1994). The explanations must account for persisting and altering distributions of disease over time and space (Berkman & Kawachi, 2000).

Some core concepts incorporated in the ecosocial theory are:

- *Scale* – that refers to the dimensions of the observed spatiotemporal phenomenon.
- *Embodiment* – referring to incorporating social and biological factors from in utero to death, stating that no aspect of human life can be understood without knowledge of individual and societal history and ways of living.
- The additive *interaction* among exposure, susceptibility, and immunity – stating that each factor and its distribution is perceived at several levels (individual,

family, neighborhood, community, national) and in multiple perspectives (home, work, school) manifesting at many scales over time and space.

- *Dynamic states* – referring to the interplay of 'inputs' and 'outputs' that go into the framework.
- *Levels of organization* – referring to the nested hierarchies, from individual to population, to the ecosystem.

As the prefix 'eco' (ecological) implies, the goal is to employ a science devoted to the study of evolving interactions between living and inmate organisms over time and space. Recognizing the value of social, political and economic processes in shaping epidemiological profiles is necessary in attempting to create some integral principles to guide policymakers to specific actions for improving population health outcomes.

2.3.4 Other Social Theories

There are several other theories about social epidemiology. Among the most relevant ones are functionalism, conflict theory, symbolic interactionism, social constructionism, feminist theory and rational choice and exchange theory.

Functionalism. Functionalism states that different public institutions are necessary for the survival of the society, adjusting to minor changes to keep functioning in a stable way. For example, when people get sick, the medical system ensures they return to their functional state. When sick, the individual cannot contribute to society, therefore disturbing its stability on a small/minor scale. Illness is being perceived as developing out of social conditions (Frankenberg, 1974).

Conflict Theory. Conflict theory refers to inequalities (such as social, political) between population groups (Wallace & Wolf, 1999). Conflict theories address the disparities

regarding who gets access to medical care, hospitals, and health insurance. Wealthier people can afford to pay for high-quality medical care while individuals from lower income groups cannot afford health care without health insurance. This may result in skipping hospital visits and longer recovery times with more complications. Unequal access to education and housing opportunities or jobs affects individual health outcomes.

Another example showcasing conflict theory is air pollution (Rosenstock, 1974). Factories may prefer more tolerant rules to reduce costs. The people living in these regions may want rigorous regulations, to avoid health issues known to correlate with air pollution coming from factories. Therefore, stricter regulation might benefit the health of individuals, but at the detriment of the income of factories.

Symbolic Interactionism. Symbolic interactionism served as the basis for social constructionism. It is a perspective in sociology stating that individuals give the world meaning based on their interactions with it and based on the meaning they attributed to these interactions (Blumer, 1986). Blumer (1986) assumed that people interact and act on things based on the meaning derived from social interaction, with this meaning also being subject to individual interpretation.

More recently, new media became an increasingly popular social construct. Whether online or offline, people act according to the meaning they have derive as a result of the interaction with other people (Fernback, 2007). In March 2000, The TV personality Katie Couric agreed to undergo a colonoscopy procedure (that is a screening method for detecting colorectal cancer that involves searching for and removing polyps from the colon/rectum) on national television to raise awareness about CRC early

detection. The televised procedure has been associated with an increase in colonoscopy procedures (Cram et al., 2003). Symbolic interactionism, whether online or offline, is an essential social construct to be considered when planning cancer control interventions.

Social Constructionism. Social constructionism refers to the conceptual framework about the communally constructed understandings of the world (Shotter, 1993; Burr, 2015). For example, money is just a paper until the society agrees that it should have value. Also, society attaches different meanings to different behaviors. Individuals create stereotypes – preconceptions about other people or groups – and then relate to those individuals/groups based on the preconceptions they have created about them. Preconceptions are most commonly formed regarding gender, race, age, or subcultures. Perceptions can affect how the doctors diagnose and treat patients, or how the patient relates to his doctor (Bury, 1986). For instance, the patient might not mention a major symptom assuming that does not matter enough to mention, or the doctor might make false assumptions depending on the appearance of the patient. With so many things believed to be socially constructed (such as money, social class, race, illness, danger, emotion among others), the research based on this theory has been criticized for being applied to our beliefs about these phenomena/issues, as opposed to the facts themselves (Hacking, 1999).

Social constructionism addresses the way in which people and groups contribute to the creation of perceived social reality (Berger & Luckmann, 1991). Medical discourse, as well as social networks and cultural beliefs, may affect individual behaviors, influencing what people perceive as “normal” or “abnormal”, and what is

regarded as illness (Foucault, 2012). In the context of cancer, Sontag (2001) addresses the issue that a cancer diagnosis is stigmatizing and dehumanizing for individuals. The socially constructed connotation of cancer as a death sentence is rather common in current society, significantly inflicting damage on the selves of those suffering from cancer.

Feminist theory. Feminist theory focuses on the variations and disparities observed among women and men. Physicians' gender preference is a well-documented area of research (Kerssens et al., 1997). Taking medical school admission for example, even though the admittance of women into medical schools is increasing, medicine is still a male-dominated field.

Rational choice and exchange theory. The assumption behind the rational choice theory is that everything people do is fundamentally rational – that is assuming that the individual actions are based on weighing the costs and benefits of possible actions to maximize personal benefit (Blume & Easley, 2008). The rational theory assumes that complex social phenomena can be explained through rational individual actions.

Rooted in structural anthropology, exchange theory is an application of rational choice theory to social interactions, viewing society as a series of interactions between individuals (Cook et al., 2013). Homans (1958) borrowed concepts from animal behavioral studies to apply them to "human exchange" areas, such as influence, conformity, interactions and competition or esteem. It is commonly applied to study family and work relationships, partner selection, parenting, caregiving and other interpersonal connections by weighing the rewards of every interaction. If an interaction

results in approval, it is more likely to be repeated, as opposed to interactions resulting in social disapproval.

Both these theories assume that individuals have access to the information they need to make rational choices and that fulfillment comes from other people. It is assumed that people form relationships for their benefit. Choices of action can be limited by social factors such as socioeconomic status, gender or ethnicity. These factors may determine individuals to make choices that aren't in their best interest.

One criticism of the rational choice theory is that it is not always possible to explain every social structure by the actions of individuals. The rational and exchange theory reduces the formation of relations to a linear process, which is not always the case. People often follow social norms that determine them to act in the interest of other people (such as volunteering or paying taxes). These theories assume that people make rational choices based on evaluating the rewards and punishments of their interaction with things and other individuals and that behavior is determined by self-interest.

Referring to theory in the context of social epidemiology means referring simultaneously to both biology and society, or as Krieger (2001b) argues – referring to the embodiment. Humans are both biological and social entities. The social epidemiology theories aim to elucidate the processes at both biological and social level that contribute to a particular distribution of a disease within a population group. The three most important theories in social epidemiology are psychosocial, social production of illness and ecosocial theory. To gain better insight into the social epidemiology discipline and to

distinguish between trends and developments within the fields - more and improved theories, as well as the data needed to test these theories are required (Krieger, 2001b).

People are highly interconnected with each other, and so is their health (Christakis & Fowler, 2009). People having more social connections have better health compared to those who have fewer connections. There is strong empirical evidence linking social networks and health (Berkman et al., 1992; House et al., 1988). In the next section, I will provide an overview of social networks, describing the mechanisms by which they affect health.

2.4 Social Networks

Relationships matter. People's health is interconnected (Smith & Christakis, 2008). Empirical studies on the effect of social networks on health and mortality, in particular, emerged in the 1970s (Cassel, 1976; Cobb, 1976; Berkman & Syme, 1979; House et al., 1982). According to Berkman & Glass (2000), there are several mechanisms by which social networks affect health:

- Perceived and real social support,
- Social influence (such as norms),
- Social engagement,
- Face-to-face contact (physical exposure to secondhand cigarette smoke),
- Access to resources (such as jobs, money, knowledge).

Data from Alameda County, as shown in Figure 2, displays the social network size (connectedness) to be inversely related to risk behaviors, such as cigarette smoking, drinking alcohol, physical inactivity, obesity and dietary patterns.

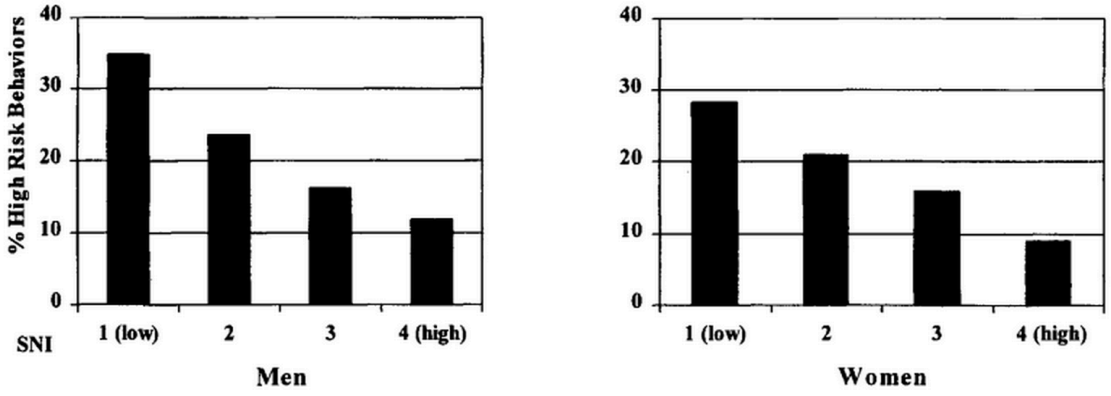


Figure 2: Distribution of health practices and risk factors (such as cigarette smoking, alcohol consumption). Percent of men and women who have several high-risk behaviors (3-5) by level of social network index (Alameda County study, n = 2229 men; n = 2496 women, ages 30-69 years) (Source: Berkman & Glass, 2000)

A conceptual framework on how social networks might affect health is presented in Figure 3. In the context of social networks, it is important to account for collateral effects (Christakis, 2004). For instance, the death of a spouse increases the risk of death in the other spouse (Lillard & Waite, 1995; Schaefer et al., 1995). The same phenomenon is observed regarding morbidity, or the caregiver burden (Schulz & Beach, 1999). A woman surviving breast cancer may motivate other women in her social network to undergo screening procedures for mammography (Murabito et al., 2001).

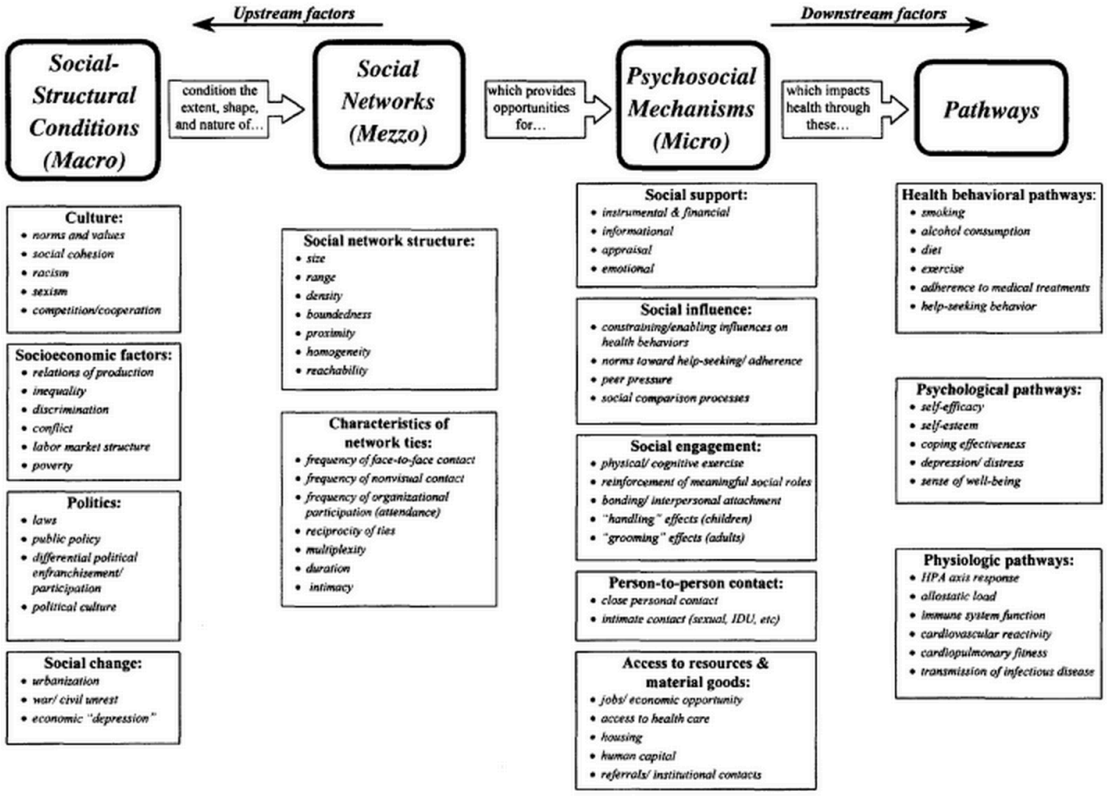


Figure 3: Conceptual models of how social networks affect health (Source: Berkman & Glass 2000)

Epidemics of conditions such as obesity, alcoholism or depression might also spread in an infectious manner (Andrews et al., 2002). Moreover, even loose social networks can serve as channels for such effects. The role of social connections to physical health (Gottlieb, 1981; Berkman & Glass, 2000; Cattell, 2001; Barger, 2013; Cohen, 2004), as well as the relationship between social networks and cancer mortality and screening in particular (Pinquart & Duberstein, 2010; Suarez et al., 2000; Peters-Golden, 1982; Sapp et al., 2003; Heaney & Israel, 2008) has widely been documented.

2.5 Current Methodologies in Social Epidemiology

The social determinants have traditionally been approached using correlation-based analytic methods (e.g. regression techniques). Although useful, the methods used to address this multilevel complexity are not without limitations. They have largely been criticized because of their limited ability to incorporate and account for the vast variety of casual social relationships and interactions (Maglio & Mabry, 2011). Traditional social epidemiology methods are insufficient to address the complexity inherent in the social determinants of cancer (El-Sayed et al., 2012; Hiatt & Breen, 2008). To make further progress in cancer control research, new methods, as well as an interdisciplinary approach that can account for nonlinearity of interactions and social influences, are needed (Parunak et al., 1998).

While equation-based models have their analytical and conceptual advantages, they work poorly for certain problems. These models often merge multiple mechanisms into a single parameter. Within an ABM, each process can be independently varied and tested. An individual agent can have its state and parameter vectors allowing incorporating easily empirically derived parameter estimates, like Type I survival curves for human populations (that are common to developed countries). Among other advantages of ABM over other techniques is the triviality to incorporate stochasticity. Nonetheless, exploiting the possibility of including per-agent stochasticity can make the model computationally expensive. The ABM methodology is discussed in more detail in Chapter 4.

To continue making meaningful advances in cancer research and accomplish more effective interventions for reducing the burden of cancer, a transdisciplinary

research framework is needed. The framework should take into account the broader scope of the social determinants of cancer and disentangle the interactions between the social, environmental, behavioral and biological factors that contribute to cancer development and progression (Hiatt & Breen, 2008).

The development of social epidemiology within the field of epidemiology made the necessity of a multilevel approach to cancer control research obvious. Many levels of analysis are necessary to shed light on the diverse mechanisms and pathways of cancer development and to determine how they are linked to the social environment. A complex systems approach is needed to address the complexity of the social determinants of cancer, and critical thinking is required to disentangle the complex interactions between the multiple actors interacting at many levels across time.

2.6 Complex Systems and Cancer

Despite the considerable molecular and cellular advancements in cancer research, the mortality rates due to cancer barely changed in the last decades comparing to the other major diseases, as shown in Figure 4. Therefore, addressing cancer in its complexity, accounting for its heterogeneity and nonlinearity is of ultimate significance. To get a better understanding of cancer and its determinants, it has been advocated to address it as a complex system, as it resembles the characteristics of one (Agus & Gell-Mann, 2012; Gell-Mann, 2002; Barabási et al., 2011; Galea et al., 2010; Moore et al., 2011; Grizzi & Chiriva-Internati, 2006).

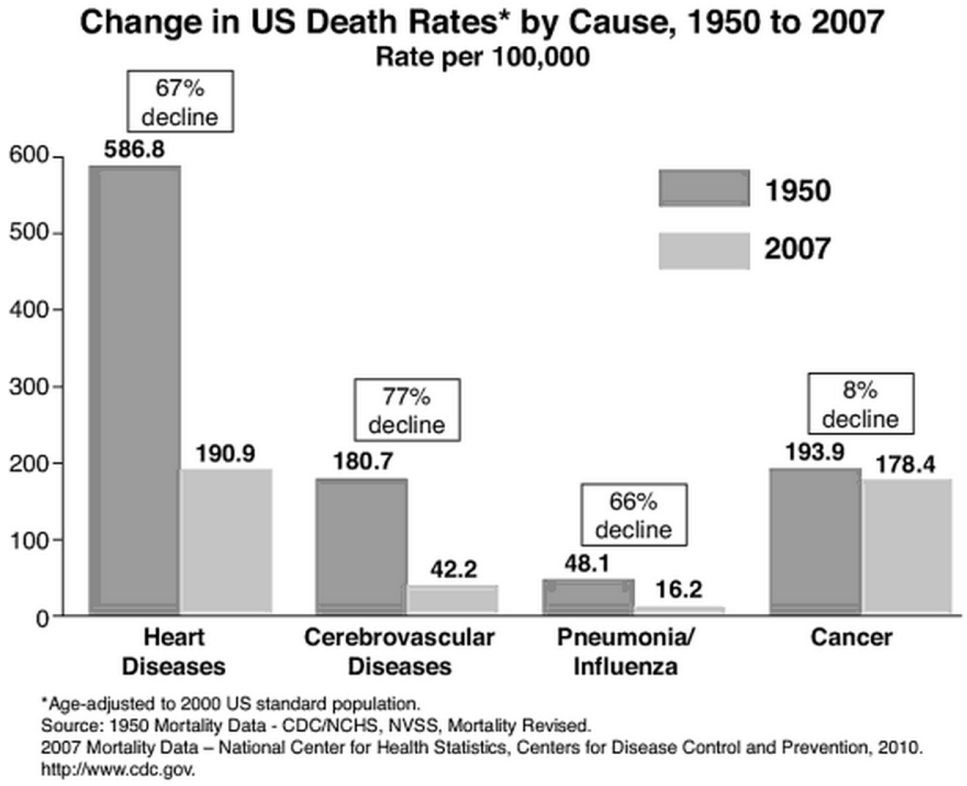


Figure 4: Comparative data showing the change in death rate for major diseases. The death rate for cancer has declined at a much slower pace over the past 60 years (Source: Agus, 2012)

Several properties make cancer an exceedingly complex system. Heterogeneity characterizes the process of cancer progression. It can be observed at the level of population, individual, tumor, genetic background and even response to treatment (Heng et al., 2009). The predominant strategies that address cancer by simplifying this heterogeneity are faulty, as the findings resulting from these homogeneous experimental studies are challenging to apply to the real world.

Another property of cancer that makes it resemble a complex system is interdependence. Biological, genetic, environmental, social, economic and several other risk factors interact with each other across multiple levels to influence cancer

progression. The interaction of a combination of these factors over time eventually leads to cancer development. In a complex system, the whole is more than the sum of the parts. Health itself can be seen as an emergent property of complex systems.

The development and progression of cancer may be based on signal and boundary interactions (Holland, 2012). The same way societies exist within boundaries – cancer can only exist within the limits of the body. No matter where they begin within the body, cancer cells cannot permeate and extend beyond the boundaries of the body. One cancer cell cannot cause death in isolation. Cancer cells communicate with other cells in their immediate proximity, converting them into cancerous cells and then spreading to other body sites within the human body. Underlying signals and mechanisms facilitate this interaction and the dynamics within such systems.

Another complex system characteristic that can be observed within cancer systems is path-dependence. In path-dependence - "history matters". History does matter in most complex systems, cancer development included. Processes and events with similar starting conditions that follow similar rules can lead to different outcomes. These results are usually sensitive to initial conditions and to other choices made along the way, where single events may ricochet through the system causing substantial perturbations in unexpected ways. By their nature, complex systems are not linear. In the case of cancer development, a significant number of endogenous and exogenous factors interact in non-linear, difficult to predict ways to influence both the initiation and the development of cancer.

Controlling a complex system need not mean the control of its individual constituent parts. Much effort is invested into understanding and treating cancer, rather than attempting to control it. In physics, as well as within other areas – scientists build models of phenomena we do not yet completely understand – to increase our understanding of them. This work attempts to accomplish a similar goal in the context of cancer while moving away from the reductionist view on cancer. Looking at cancer through the complex systems lens may lead to a deeper understanding of the disease and potentially result in innovative methods of addressing not only the understanding of cancer but its control as well.

2.7 Summary

In this chapter, I described the social determinants of cancer and set the context for understanding the social epidemiology of health while highlighting trends within the field. Because theory matters, I reviewed the theoretical grounds of this relatively new discipline, drawing from several contemporary theories and concepts that serve as the basis for modern social epidemiology. I also reviewed the ways social determinants of cancer link to cancer development, progression, and mortality and what mediating factors (such as individual and group behaviors) may modulate these influences. A quick introduction to social networks is provided, pinpointing their applications to health and cancer control in particular.

Existing methodologies predominant in social epidemiology are discussed in Section 2.5, and the need for new approaches to address the complexity inherent in the social determinants of cancer is examined. The chapter concludes with a subsection on complex systems and cancer, discussing the characteristics of this disease that make it a

complex system, and the need to use complex system methodologies to shed light on the social determinants of cancer and the mechanisms linking these with cancer incidence and mortality trends. In the next chapter, I discuss colorectal cancer – the cancer site selected for the study case of the model.

CHAPTER 3: COLORECTAL CANCER

Contents

3.1 Introduction 36

3.2 CRC Overview and Staging 37

3.3 CRC Risk Factors 41

3.4 CRC Prevention 45

3.4.1 CRC Screening Methods 47

3.4.2 Sensitivity and Specificity 51

3.4.3 CRC Screening Barriers and Facilitators 53

3.5 CRC Treatment and Survival 56

3.6 CRC Recurrence and Quality of Life 60

3.7 CRC Disparities 61

3.8 Delay in Diagnosis and Treatment 65

3.9 Summary 71

> "Even with our current understanding of how cancer develops and grows, this disease is mostly a mystery and not within human control"
>
> (American Cancer Society, 2013)

3.1 Introduction

A rapidly growing demand for cancer care has been observed in the United States, mainly due to three issues: an increasing burden of cancer (it is projected that the number of new cancer cases in the United States will increase by 45% by 2030), the growing number of cancer survivors (a 35% growth in cancer survivors is projected by 2022) and due to the large number of newly insured patients (an additional of 29 million Americans are expected to be insured by 2017) (ASCO, 2014). All these concerns together with the complexities involved in the cancer development processes make cancer an important national issue requiring new approaches to investigate its causation and dynamics.

Colorectal cancer (CRC) is the third leading type of cancer worldwide (Shike et al., 1990). Of the cancers affecting men and women, CRC is the second leading cancer killer in the United States, with the District of Columbia having the highest rates of CRC deaths in the country (CDC, 2011). Even though it is considered a type of cancer that is largely preventable, 134,490 new cases of CRC and about 49,190 estimated deaths are expected in 2016 in the United States (Siegel et al., 2016). In 2011, an estimated 1,162,426 people were living with CRC in the United States. Scientific evidence links CRC to several lifestyle factors, such as tobacco use, physical inactivity, and obesity among others. This chapter will provide an overview of CRC development, discussing the staging, risk factors, screening and testing methods available, treatment options, survival and the quality of life. The chapter concludes a discussion of CRC disparities and factors that contribute to delay in CRC diagnosis and treatment.

3.2 CRC Overview and Staging

Colorectal cancer (CRC) is a condition in which cancerous cells develop in the colon or the rectum, as shown in Figure 5. It begins when normal cells in the lining of the colon or rectum change and grow uncontrollably, forming a mass called a tumor. A neoplasm can either be benign (non-cancerous) or malignant (with the ability to spread to other parts of the body). The colon and the rectum make the large intestine, that is responsible for how the body processes waste.

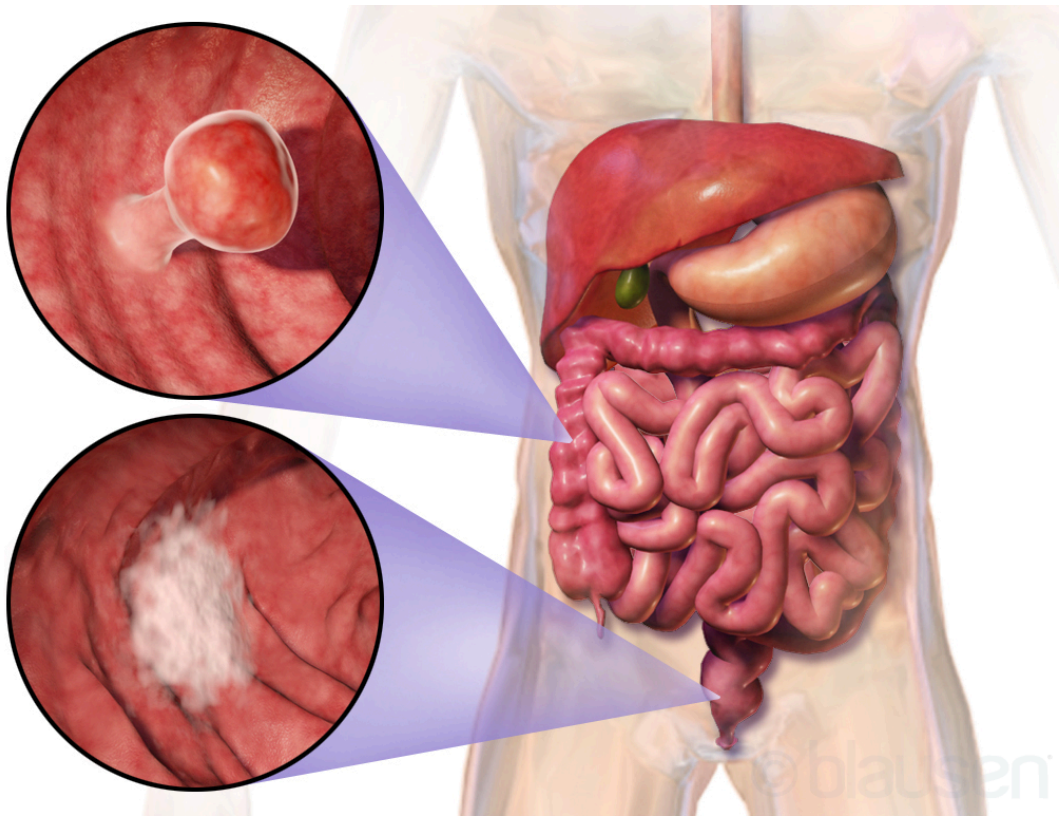


Figure 5: Anatomy of the lower digestive system (Source: Blausen Medical Communications, Inc., 2016)

Most colorectal cancers happen in individuals without a family history of cancer. The risk of contracting CRC begins to increase after the age of 40 years and rises suddenly at ages 50 to 55; the risk doubles with each succeeding decade and continues to grow exponentially (NCI, 2016). Even though CRC care improved considerably, the improvement in survival for individuals with late-stage CRC is insignificant (Lynch & Chapelle, 2003). Therefore, to reduce CRC morbidity and mortality, it is of utmost importance to focus on prevention.

The following signs and symptoms may accompany CRC:

- A change in bowel habits,
- Blood in the stool,
- Diarrhea and constipation that continues for more several days,
- An urge to have a bowel movement when the bowel is empty,
- A difference in the stool characteristics (e.g., more narrow than usual),
- Frequent gas pains, bloating, fullness, or cramps,
- Unintentional weight loss,
- Feeling exhausted,
- Decreased appetite,
- Cramping or discomfort in the lower abdomen.

In some case, the blood loss because of cancer can lead to anemia – therefore causing tiredness and weakness. Early CRC may not present with any symptoms. Therefore, undergoing timely screening is crucial.

Cancer may also metastasize, spreading from its original location to other parts of the body. After diagnosis, tests are done to find out if cancer cells have spread (via tissue, lymph system or blood) from the colon or rectum to other parts of the body, as shown in Figure 6. Staging measures the degree to which cancer has spread within the body.

According to the TNM (for tumors/nodes/metastases) staging system (Compton & Greene, 2004), CRC cases are grouped into the following stages:

- Stage 0: cancer cells are limited within the lining of the inner bowel with little cancer to having spread.
- Stage I: tumor is confirmed, cancer has grown from the inner lining into the muscle wall, and no cancer has spread to the lymph nodes.
- Stage IIA: cancer has advanced out of the bowel wall but hasn't reached the lymph nodes.
- Stage II-B: cancer has grown out the bowel wall and spread to the tissue or organ next to the bowel. No lymph nodes or other parts of the body are affected.
- Stage III-A: cancer has is still contained in the inner layer of the bowel wall and/or into the muscle layer. Between one and three lymph nodes contain cancer.
- Stage III-B: cancer has grown into the outer lining of the bowel wall or surroundings. Between one and three lymph nodes contain cancer.
- Stage III-C: cancer can be of any size and has spread to four or more lymph nodes. No cancer has spread to any part of the body.
- Stage IV: cancer has spread to other areas of the body (e.g. liver or lungs)

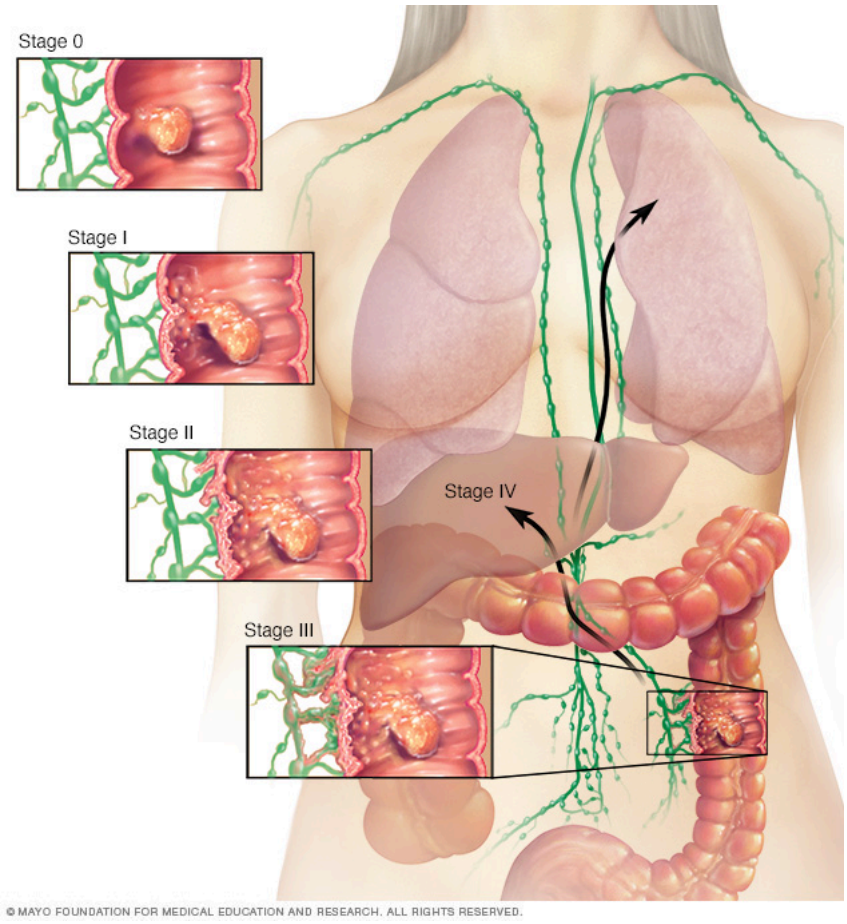


Figure 6: Colorectal cancer growth (Source: Mayo Clinic, 2016)

Another commonly used cancer staging system is the Surveillance, Epidemiology, and End Results (SEER) summary staging system (Howlader et al., 2014):

- **In situ**: cancer has not invaded the colon or rectum
- **Local:** cancer has grown into the wall of the colon or rectum, no invasion of the tissues outside the wall yet.
- **Regional**: cancer has developed into the wall of the colon or rectum and invaded nearby tissue or has spread to lymph nodes.

- **Distant**: cancer has advanced and spread to other parts of the body (e.g. liver or lungs)

3.3 CRC Risk Factors

A risk factor is anything that increases a person's chance of developing cancer. Some of these factors are modifiable while others are not. Among the non-modifiable CRC risk factors are a family history of CRC and a personal history of chronic inflammatory bowel disease.

Family history. Individuals with a first-degree family member (such as a parent, sibling, or offspring) previously diagnosed with CRC, are at increased risk of developing this disease, as opposed to individuals with no family history of CRC. The risk increases 3 to 6 times if the cancer diagnosis of the family member happened at an earlier age or if there is more than one affected relative (Butterworth et al., 2006; Johns & Houlston, 2001). Approximately 5% to 10% of the people annually diagnosed with CRC inherited their disease in an inherited autosomal way (Lynch & Chapelle, 2003). The most common ones are Lynch syndrome and the familial adenomatous polyposis (FAP). The latter is characterized by the development of hundreds to thousands of colorectal polyps in affected individuals, which have a lifetime risk of CRC approaching 100% by age 40 (Galiatsatos & Foulkes, 2006; Bernstein et al., 2001).

Personal history. Chronic inflammation of the bowel over extended periods of time (such as ulcerative colitis or Crohn's disease) increases the risk of developing CRC. It is estimated that about 18% of patients with a 30-year history of ulcerative colitis will get CRC (Bernstein et al., 2001). Diabetes is another significant risk factor for CRC, with

men having a higher risk compared to women (Campbell et al., 2010). A personal history of adenomatous polyps, especially having large or numerous polyps also increases the risk of CRC.

Among the most significant modifiable risk factors that increase the risk of CRC are excessive alcohol consumption, cigarette smoking, and obesity (Wu et al., 2009; Aune et al., 2011). Among the factors associated with decreased CRC rates are physical activity, diet modification (e.g. reducing red meat and increasing fruits and vegetable consumption), non-steroidal anti-inflammatory drugs, aspirin, polyp removal, postmenopausal female hormone supplements and vitamins. The modifiable factors that are most significant to CRC development and progression are presented below.

Physical activity. Physical activity is among the most critical modifiable risk factors for CRC. Studies have shown that people who are most physically active have a 25% lower risk of CRC compared to their less physically active counterparts (Boyle et al., 2012). The more physically active an individual is, the lesser his risk of developing CRC. This fact applies to both recreational and occupational physical activity (Samad et al., 2005). A large body of epidemiological evidence shows that physical activity is linked to lower risk of CRC for both men and women. It is possible for sedentary people who become active later in life to reduce their CRC risk (Chao et al., 2004).

Obesity. Obesity is associated with higher CRC risk for both men and women. Being overweight or obese increases the risk of developing CRC independent of physical activity (Larsson & Wolk, 2007). A meta-analysis of 31 studies with 70000 events found the relationship between CRC and obesity to be direct and independent (Moghaddam et

al., 2007). Even though obesity is considered among the most significant risk factors for CRC and a public health concern, the findings of the extent to which obesity contributes to the development of CRC have been inconsistent (Bianchini et al., 2002; Dai et al., 2007; Larsson & Wolk, 2007). This association has been observed more consistently in men compared to women (Dai et al., 2007; Renehan et al., 2008).

Diet. Recent studies on the effects of diet on CRC risk indicate the following dietary practices as high risk:

- Red and/or processed meat (Cross et al., 2010; Norat et al., 2005; Chao et al., 2005; Potter, 1996)
- Higher increase of dairy products (also milk and calcium, irrespective of milk fat content) decreases the risk of CRC (Aune et al., 2012)

Among the dietary practices shown to decrease CRC risk are the following:

- Dietary fiber, cereal fiber, and whole grains have been shown to reduce CRC risk (Fuchs et al., 1999; Aune et al., 2011)
- Moderate fruit and vegetable intake are associated with lower risk of CRC (Michels et al., 2000; Wu et al., 2009; Aune et al., 2011)
- Dietary folate intake was shown to decrease CRC risk (Sanjoaquin, 2005)

Smoking. In 2000, smoking represented an important cause of global mortality (Ezzati & Lopez, 2003). The International Agency for Research on Cancer reported in 2009 that there is sufficient evidence to conclude that there is a causal relation between smoking and CRC (Secretan et al., 2009). It may take about three or four decades between the exposure to smoking and a CRC diagnosis. A prospective study has shown that smoking

for at least 20 years in the past is associated with larger tumors (Giovannucci et al., 1994). In a meta-analysis, smoking was shown to double the risk of CRC polyps (Botteri et al., 2008). Because of the strong link between cigarette smoking and CRC risk, several studies looked into the effect of smoking cessation efforts on CRC (Chao et al., 2000). Early smoking cessation strategies were associated with evident CRC risk reductions.

Alcohol. Alcohol is a significant known risk factor for several human cancers (Glade, 2008). Moderate and heavy alcohol drinking has been linked to higher risk of developing CRC (Cho et al., 2004; Ferrari et al., 2007). Studies show that consuming two to four alcoholic drinks a day increases CRC risk by 23% compared to drinking just one alcoholic beverage a day (Ferrari et al., 2007). A review examining the role of alcohol consumption and CRC risk in 52 studies found evidence that alcohol consumption (particularly beer) is an important etiologic factor for CRC in both men and women (Kune & Vitetta, 1992). Another review including 27 cohort studies found moderate to no increase in CRC risk resulting from alcohol consumption (Franceschi & La Vecchia, 1994).

Individual behaviors affect health not just before, but also after a cancer diagnosis. Even though a cancer diagnosis can be devastating, evidence shows that a cancer diagnosis may often act as a catalyst, contributing to changing the patterns in a variety of health behaviors (Andrykowski et al., 2006). Cancer patients and survivors often report adopting healthier lifestyle behaviors such as reduced tobacco use, increased physical activity and exercise, adopting healthier diets subsequent a cancer diagnosis and treatment (Maskarinec et al., 2001; Patterson et al., 2003; Pinto et al., 2000; Satia et al.,

2004). The motivation behind this willingness of adopting health-promoting behaviors may be triggered by either an individual readiness to improve the overall health and decrease the risk of disease or enhance their response to cancer treatment and reduce the chances of cancer recurrence.

The interaction between these modifiable risk factors, with other non-modifiable factors such as age, a family history of CRC or race/ethnicity, modulates the risk of developing CRC.

3.4 CRC Prevention

CRC is a largely preventable cancer type, as it almost always grows from precancerous polyps. Three types of cancer prevention can be distinguished: primary, secondary and tertiary prevention. The goal of primary prevention is to prevent the onset of CRC. It can be accomplished through reducing the prevalence of risk factors and increasing protective factors and/or screening/testing for CRC. Secondary prevention is aimed at CRC mortality by early detection of CRC, mostly through screening tests. When primary and secondary prevention were omitted, tertiary prevention can be employed to improve the prognosis and the quality of life in individuals with CRC.

CRC is mostly treatable if found early. The prevention and treatment continuum for cancer, shown in Figure 7, highlights opportunities for prevention and treatment across the different stages of the cancer continuum.

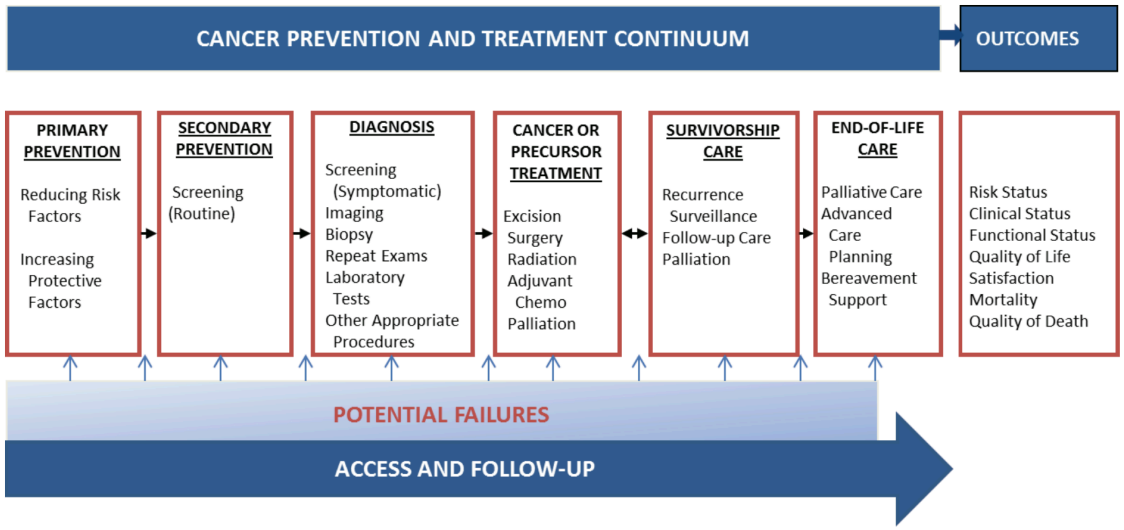


Figure 7: Continuum of cancer prevention, treatment, and outcomes (Source: Price et al., 2012, adapted from Zapka et al., 2003)

The goal of CRC screening is to look for signs of cancer in the seemingly asymptomatic population. With timely CRC screening practices, precancerous growths can be detected and removed – so that cancer can be prevented from occurring in the first place. The relatively long period of development from precancerous polyps to invasive CRC provides a unique opportunity for prevention for this specific cancer type (Ferrari et al., 2007).

Screening at regular intervals was shown to be highly effective in preventing CRC (Mandel et al., 1993). CRC screening is recommended for both men and women beginning at the age of 50. As part of the screening process, it is important to use tests that detect both cancer and precancerous growths, as opposed to testing primarily for the purpose of cancer detection.

3.4.1 CRC Screening Methods

The tests available for detecting precancerous growths (and cancer as well) in both men and women are the following:

- **Flexible sigmoidoscopy**: during which a specialist inserts a slender and flexible tube through the rectum into the colon. It provides a visual examination of the rectum and lower one-third of the colon (Levin et al., 2008). Some studies found sigmoidoscopy to be associated with a 21% reduction in CRC incidence and a 26% reduction in CRC mortality (Schoen et al., 2012).
- **Colonoscopy**: while similar to the sigmoidoscopy, this procedure implies the use of a colonoscope, which is a much longer and more complex instrument. With a colonoscope, the entire colon can be visualized, and if any polyps are identified, they can be removed during the same procedure (Levin et al., 2008). Before a colonoscopy procedure, preparation on the patient behalf (complete colon cleansing), as well as sedation, are required. This procedure is among the most sensitive for detecting polyps or colorectal cancer, and has longer rescreening interval of all forms of testing (about ten years in average-risk people). One of the limitations of this procedure compared to other tests is the higher risk of complications, such as bowel tears and bleeding (Waye et al., 1992).
- **Barium enema with air contrast** – is a radiological procedure that became less common with the increased use of colonoscopy. During this procedure,

the colon is partially filled with barium sulfate through the rectum. After that, air is introduced to expand the colon for better x-ray quality. The method is less invasive compared to colonoscopy and if an abnormality is noticed the patient is consequently referred for a colonoscopy for further examination (Toma et al., 2008).

- **Computed tomographic colonography** (CTC) is another radiological method that links a special x-ray machine to a computer for generating 2D or 3D views of the entire colon or rectum (Levin et al., 2008). Sedation is not required, but a bowel cleansing is helpful for a successful examination. CTC requires no recovery time, is less invasive and lasts approximately 10 to 15 minutes. The patients with abnormalities are further referred for colonoscopy, usually on the same day, as the bowel is already prepared.

Among the tests primarily effective at detecting cancer rather than just precancerous polyps are:

- **Fecal occult blood test** (FOBT) is a test to check stool (solid waste) for blood that can only be seen with a microscope. Polyps and colorectal cancers often bleed, and FOBT attempts to find these blood traces. Small samples of stool are placed on special cards and returned to the doctor or laboratory for testing. Blood in the stool may be a sign of polyps or cancer. Results from large clinical trials indicate that regular use of FOBT may decrease CRC mortality by 32% (after 30 years of follow-up), and decrease CRC incidence by 20%. FOBT, however, performs poorly in its ability to detect the disease. Even

though as many as one-third of primary care physicians carry out this test yearly (Nadel et al., 2005), an extensive study demonstrated that FOBT missed 95% of the precancerous polyps and cancers, which were revealed subsequently by colonoscopy (Collins et al., 2005).

- **Stool DNA** test was approved for screening in 2008 but wasn't commercially available until FDA approved a newer test in August 2014 (Imperiale et al., 2014) as a result of the advancements if the molecular properties of CRC. The adopted test was able to detect accurately CRC or polyps 92% of the time in a sample of 10000 adults at average CRC risk. Instead of looking for blood in the stool, stool DNA looks for specific DNA mutations associated with precancerous polyps or tumors. Subjects with a positive outcome should further be referred to colonoscopy.

The individual characteristics of screening tests, their ability to detect cancer, as well as their effectiveness in reducing the incidence and mortality due to CRC is summarized in Table 3.

Table 3: Characteristics of screening tests for colorectal cancer (Source: Pignone et al., 2002)

Screening Strategy for CRC	Effectiveness in Reducing Incidence of and Death from CRC	Evidence Grade†	Ability To Detect Cancer	Evidence Grade†
Digital rectal examination	Case–control study found no difference in mortality rates; OR, 0.96 [0.56–1.7] (8)	Level II—poor	Pathologic data suggest <10% of CRC is within reach of examining finger	
Office FOBT (one card)	Unknown		Only 58% of cancer cases are detected on the first of three cards, suggesting lower sensitivity than three-card testing (10)	Level III—fair
Home FOBT (three cards), unrehydrated	Biennial testing: 2 trials found mortality reductions of 15% [1%–26%] (13) and 18% [1%–32%] (14)	Level I—good	One-time sensitivity 30%–40% Unrehydrated FOBT finds about 25% of cancer cases (9)	
Home FOBT (three cards), rehydrated	Annual testing: 33% [13–50] reduction in mortality (2); 20% [10%–30%] reduction in cancer incidence (16) Biennial testing: 21% [3%–38%] reduction in	Level I—good	Single-test accuracy 50% [30%–70%] for cancer, 24% [19%–29%] for advanced neoplasms (17) Over 13 years, rehydrated FOBT finds	Level III—good
Sigmoidoscopy	Small RCT found decreased CRC mortality rates with screening; RR, 0.50 [0.10–2.72] (18) Case–control studies suggest a 59% [31%–75%] reduction in mortality rate within reach of scope (3)	Level I—fair Level II—good	One-time screening detects 68%–78% of advanced neoplasia (17, 20)	Level III—good
Combined FOBT and sigmoidoscopy	Nonrandomized trial found a 43% reduction in mortality rate when FOBT was added to rigid sigmoidoscopy; RR, 0.57 [0.56–1.19] (4)	Level I—fair to poor	One-time screening detects 76% of advanced neoplasia (17) Increased yield when sigmoidoscopy is added to FOBT (21–23)	Level III—good
Double-contrast barium enema	Unknown		Sensitivity of one-time test for cancer or large polyps 48% [24%–67%] (24)	Level III—fair
Colonoscopy	Case–control study found an OR of 0.43 [0.30–0.63] for death from CRC; CRC incidence decreased by 40%–60% (26)	Level II—fair	Sensitivity for large adenomas >90%; sensitivity for cancer probably higher (27)	Level III—good

* CRC = colorectal cancer; FOBT = fecal occult blood test; NA = not applicable (see text); OR = odds ratio; RCT = randomized, controlled trial; RR = relative risk. Numbers in square brackets are 95% CIs; numbers in parentheses are reference numbers.
† Level I = evidence from one or more controlled trials; level II = evidence from cohort or case–control studies; level III = evidence from diagnostic accuracy studies or case series. For each level, the investigators have assigned a quality score based on methods described in reference 7.

There are several algorithms for CRC screening that take into consideration individual risk profiles, family history, and personal preferences, as shown in Figure 8. Knowledge about the individual risk factors (and about this particular cancer type in general) among other factors, may influence timely seeking of medical care before symptoms occur. Further, the individual together with his/her physician determine the most suitable and appropriate method of screening.

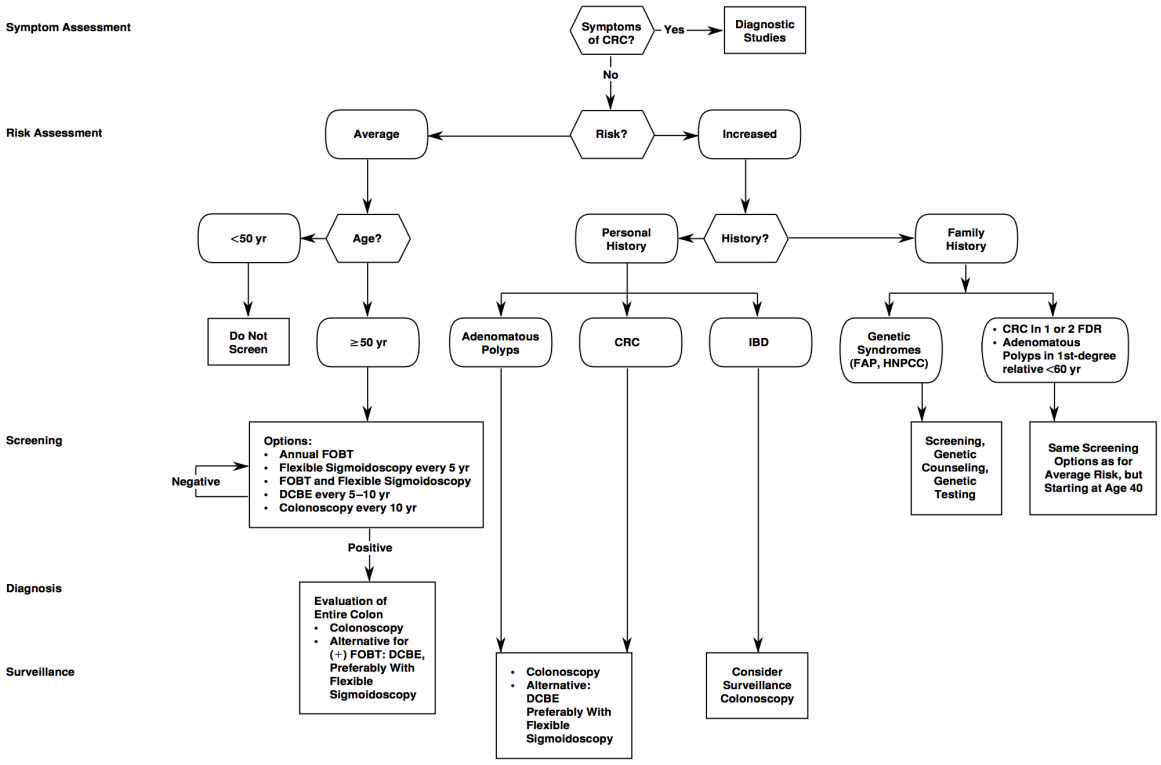


Figure 8: Algorithm for CRC screening and surveillance (Source: Winawer et al., 1997)

USPSTF officially recommends screening for CRC for most men and women starting the age of 50 (USPSTF, 2002). However, if the individual has a strong family history of CRC and is at increased risk of developing CRC, screening for CRC should start much sooner (with as much as ten years earlier than the first case of CRC in a first-degree relative).

3.4.2 Sensitivity and Specificity

CRC screening tests are evaluated by measures such as “Sensitivity” and “Specificity”. Sensitivity and Specificity are the most common measures used to assess the clinical performance of screening tests, and they are independent of the population subjected to

the test. Sensitivity measures the potency of a test to classify an individual accurately as "diseased". It refers to the probability that the screening test will be positive among individuals that are indeed suffering from the disease.

The capacity of a test to accurately distinguish a person as healthy (free-of-disease) is called specificity (Parikh et al., 2008). Specificity refers to the probability that a screening test will be negative for those individuals that are not suffering from the disease being screened. The outcomes are often organized in 2 by 2 contingency tables, as shown in Table 4 and Table 5.

Table 4: An example of a 2x2 (two by two) table

	Gold standard		
	Disease present	***Disease absent***	***Total***
Test positive	True positive (TP) a	False positive (FP) b	*Total test positives:* *a+b*
Test negative	False negative (FN) c	True negative (TN) d	*Total test negatives:* *c+d*
	Total diseased: *a+c*	*Total normal:* *b+d*	*Total population:* *a+b+c+d*

The sensitivity and specificity being calculated as:

Table 5: Sensitivity and specificity of screening tests

	Disease present	***Disease absent***
Test positive	a (TP)	b (FP)
Test negative	c (FN)	d (TN)
	Sensitivity: *a/(a+c)*	*Specificity:* *d/(b+d)*

A perfect case would be considered 100% sensitive and 100% specific if all diseased individuals are identified as diseased and all healthy individuals are identified as healthy. Theoretically, both these measures will have a minimum error rate.

Given that each of the tests mentioned in this section has benefits and harms, and different levels of sensitivity and specificity, each person should discuss with his primary care physician the risks and benefits of the recommended CRC screening methods to make an informed decision about what is appropriate for that specific individual.

While there are several effective prevention and detection tests for CRC available, not all population groups can benefit from them. Next, I will discuss the barriers to people getting screened as recommended.

3.4.3 CRC Screening Barriers and Facilitators

Cancer incidence and mortality could considerably be reduced by timely screening (USPSTF, 2002). According to CDC (2008), only 60.8% of the respondents (based on BRFSS data) 50 years of age or over reported having had a recent screening procedure in 2006, 56.8% in 2004 and 53.9% in 2002. While the overall trend of CRC screening test use increased (from 2002 to 2006), disparities in test use persist. Individuals at high risk for CRC (either because of family history of other medical conditions) should be screened before the age of 50. The initial age and frequency vary depending on individual circumstances.

Notwithstanding the availability of several screening tests and the evidence supporting the effectiveness of testing, only 65.1% of the United States population above age 50 was following the guidelines of recommended CRC testing in 2012 (CDC, 2013). The prevalence of screening is lower in individuals 65 years of age or younger who are

non-white, have less than 13 years of education, lack health insurance and are recent immigrants. Women are slightly more likely to follow screening recommendations than men. It is unclear why the CRC screening rates remain low. Among the most documented barriers to screening for CRC are:

- The cost and lack of access to health care. Populations at highest risk are new immigrants, individuals born outside the United States and those with limited English proficiency. They are also the least aware of the need for CRC screening (Beydoun, 2008). Patients residing mostly in rural areas are less likely to have access to CRC care services close to their homes (Bass et al., 2011; Baldwin et al., 2008).
- The lack of communication about the importance of screening with healthcare providers. Physician's recommendation increases the chances of screening among both insured and uninsured populations (Doubeni et al., 2010).
- Differences in individual vs. physician testing preferences (Gupta et al., 2013; Ling et al., 2008). Patients are more likely to undergo a specific screening test if the physician recommends the same test the patient prefers.
- People with lower income levels and lower educational attainment face multiple barriers to screening, such as the lack of insurance, transportation to the hospital or transportation to colonoscopy sites (O'Malley et al., 2004; Beydoun, 2008; CDC, 2008).
- In addition to having the highest incidence and mortality CRC rates, African-Americans face further barriers to CRC screening due perceived bias in the

delivery of healthcare care and overall suspicion of the medical system (Jernigan et al., 2001).

- Personal barriers such as fear of a cancer diagnosis, fatalism, low self-efficacy and embarrassment (O'Malley et al., 2004).

Even though many studies have shown that colorectal cancer is preventable with regular screening (Mandel et al., 1993; Winawer et al., 1993), the rates of the United States population adhering to CRC screening recommendations are still low. This can be attributed to a wide range of factors.

While several factors act as barriers to CRC screening, some factors serve as screening promoters. Among these are the following:

- Having medical insurance has been shown to correlate strongly with CRC screening adherence (Lewis & Jensen, 1996; Breen et al., 2001).
- Media-famous personalities. While the role of celebrity spokespeople in influencing public opinion has been extensively documented (Kessler & Pines, 1990; Cram et al., 2003; Brown & Potosky, 1990), the impact of public and online campaigns on preventive care measures have been less explored. In the context of CRC, there is evidence that a celebrity spokesperson - may significantly influence the public participation in CRC preventive efforts. Two different datasets show an increase in colonoscopy use that was sustained up to 9 months post a televised series campaign by Katie Couric promoting CRC screening (Cram et al., 2003). Similarly, President Reagan's colon cancer episode of July 1985 led to an increase

in colon cancer inquiries to a national hotline, as well as to an increase in the use of early detection tests (Brown & Potosky, 1990).

- Media - "The Super Colon" is an inflatable installation with a height of 8 feet and a length of 20 feet that replicates the human colon. The aim of it is to educate people about the CRC risks and symptoms as well as the importance of early detection and treatment. Studies showed the walk-through model to be effective in educating the population about CRC and motivate them for screening (Finn et al., 2013; Fickle et al., 2015).
- Word of mouth from people who underwent CRC screening is an important facilitator, in particular among the African American population (Bass et al., 2011; O'Malley et al., 2004).

Interventions to increase CRC screening rates should consider the factors that promote, as well as the ones acting as barriers to maximum efficiency. While screening for CRC detection is desirable, it does not always happen. When screening fails and CRC continues to develop and progress – several treatment options are available. The next section will discuss the existing treatment options for CRC by stage at diagnosis. The sooner CRC is detected, the higher the likelihood of a successful outcome.

3.5 CRC Treatment and Survival

CRC can largely be cured if caught early (Winawer et al., 1997). Treatment options for CRC depend on the stage at which the disease is diagnosed. Patients have the possibility to choose from the existing CRC treatments as well as to participate in clinical trials. Six standard treatments are available for patients with diagnosed CRC (NCI, 2015).

Surgery means removing cancer in an operation. It is the most frequently used treatment for all stages of CRC. After removing cancer during surgery, it is common for doctors to offer patients chemotherapy or radiation treatment to destroy any cancer cells that may be left post surgery. The treatment that is provided in addition to the primary treatment to lower the risk of cancer to come back is called adjuvant therapy and may include chemotherapy, radiation therapy, hormone therapy, targeted therapy, or biological therapy. Common side effects of CRC surgery include fatigue, constipation or diarrhea, and sexual dysfunction.

Radiofrequency ablation involves destroying cancer cells using a special probe with electrodes. The probe is introduced straight through the skin with local anesthesia. The cryosurgery (cryotherapy) treatment uses an instrument to freeze and kill cancerous cells.

Chemotherapy involves administering drugs to prevent cancer cells from growing by killing the cells or stopping them from dividing. There are several ways of administering chemotherapy depending on the type and stage of the tumor. Side effects include hair loss, mouth sores, swelling and rashes, nausea, diarrhea, and fatigue.

Radiation therapy involves applying high energy x-rays or other radiation methods for destroying cancer cells or preventing them from growing. Common side effects include skin, bladder or rectal irritation, nausea, diarrhea, fatigue or sexual problems.

Targeted therapy is a relatively new method of treatment that uses drugs for attacking specific cancer cells without harming the normal cells. It usually triggers fewer side effects compared to chemotherapy. Epidermal growth factor receptor inhibitors trigger acne-like rash, dry skin and swelling/pain in the fingernails or toenails. Vascular

endothelial growth factor inhibitors may trigger kidney damage, bleeding, high blood pressure and artery clots.

Sometimes, the cancer can return after treatment; a process called recurrence. They usually are harder to treat. For stage 0 (at the polyp stage) the standard treatment options include local excision (or polypectomy) commonly performed during the colonoscopy procedure. When the tumor is too large, surgery needs to be performed. For stage I and stage II, the treatment usually involves surgery - resection and anastomosis (a method to combine healthy segments of the colon after the unhealthy portion is surgically eliminated). Stage III calls for a more complex treatment approach, where surgery is also complemented by chemotherapy. Additionally, clinical trials are often available for new chemotherapy routines following surgery. Stage IV and recurrent CRC presents the most complicated treatment challenges. Chemotherapy is often administered before the surgery to shrink the tumor so that surgery can take place. Surgery at this stage is not referred to colon resection only, but may refer to removal of other parts of other organs as well (such as liver, lungs, ovaries or other organs to where cancer metastasized). Radiofrequency ablation and/or radiation therapy may also be administered at this stage depending on the extent of the cancer spread.

CRC prevention and treatment are the first lines of defense against CRC. They both influence survival, as well as mortality rates. Recent advances in cancer care and early detection have led to an improved quality of life and increased survival for many CRC patients. Survival is a term commonly used by doctors to address a patient's prognosis. CRC survival rates depend on many factors. Among the most important ones

are the stage at diagnosis and whether cancer has metastasized and the quality of the surgery. Survival statistics should be interpreted with discretion. Because these statistics are measured for people diagnosed with CRC all over the United States, within long time intervals, and they may not include the progress made in the diagnosis or treatment of this particular cancer type.

The five-year survival rate refers to the portion of people who live at least five years after a cancer diagnosis and is one of the most commonly cited cancer statistics. These rates usually refer to previous outcomes of large numbers of people who previously had cancer. The survival rates of CRC by stage, shown in Figure 9, are based on the National Cancer Institute's SEER database, observing over 28,000 people diagnosed with CRC between 1998 and 2000 (Edge & Compton, 2010).

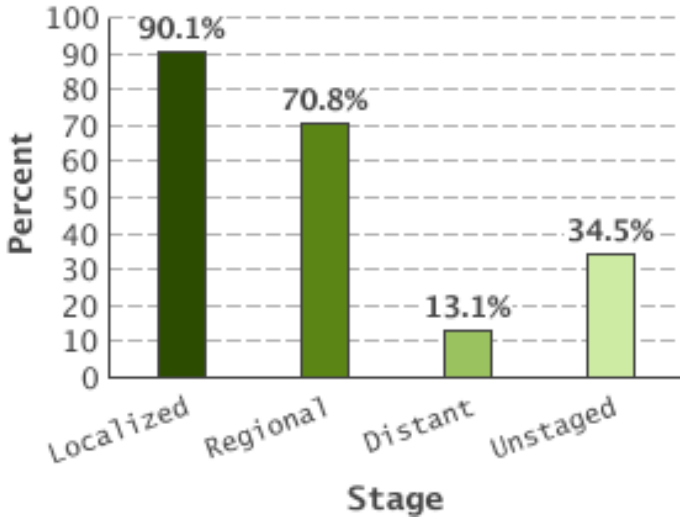


Figure 9: Relative survival for all races, both sexes by SEER summary stage 2000 (Source: SEER, 2016)

The probabilities of cancer survival are hard to predict. Survival and prognosis depend on many factors. Some of these factors (such as treatment selection) are modifiable while others – such as age, sex, and the way cancer responds to treatment, genetic alterations, and personal medical history – are not. The sooner CRC is diagnosed, the better the clinical outcome and the better the patient quality of life.

3.6 CRC Recurrence and Quality of Life

After having CRC successfully treated, the individual are in remission, denoting an absence of detectable cancer cells in the body. While in remission, individuals may experience a CRC recurrence (a return of cancer). Cancer may recur where it began (colon or rectum) or elsewhere in the body. Cancer behavior is largely unpredictable, and there can be no guarantee that cancer will stay in remission forever. Sometimes a few cancer cells that are too small to be detected on tests remain in the body after the treatment. These cells may cause no harm for years. Unexpectedly, something may change in the immune system and "wake-up" those cells. Being active, they grow, divide and become big enough to be detected as recurrences.

Recurrent cancer can occur locally (in the same place where it started), regionally (cancer came back in the lymph nodes near where it started), and distantly (it returned to another part of the body such as lung, liver, bone marrow or brain). The most common sites of CRC recurrences are the liver and the lungs. The risk of recurrence depends on cancer stage as well as on the treatment regimen the patient selected. Generally, the lesser the period when the cancer was considered gone and the time it came back – the more severe the recurrence will be.

Among the determinants of QoL in CRC patients are physical problems such sexual dysfunction due to treatment for men (Phipps et al., 2008) and low income (that correlates with worse physical, social and emotional wellbeing) (Ramsey et al., 2002; Lundy et al., 2009). The presence of a wide social network is positively associated with QoL (Sapp et al., 2003). Moreover, people living alone reported lower perceived wellbeing than those who live in a family (Forsberg et al., 1996). Other factors such as physical activity increase the QoL in cancer patients mainly due to decreasing the levels of fatigue and distress (Buffart et al., 2012). Smoking and alcohol intake are associated with a lower QoL (Blanchard et al., 2004). Several determinants of QoL in CRC patients are modifiable, and interventions should identify and focus on the interactions of factors that are more likely to enhance the QoL of CRC patients (Buffart et al., 2012).

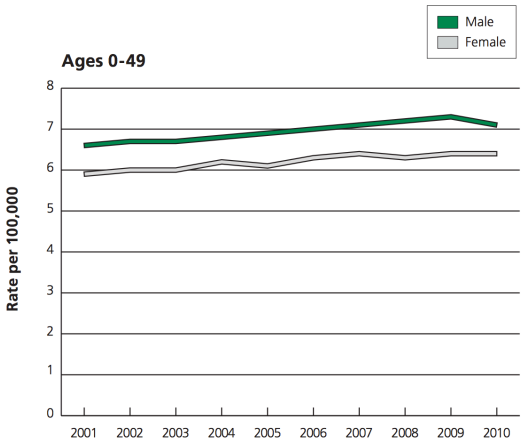
3.7 CRC Disparities

While progress has been made in CRC diagnosis, treatment, screening and even prevention over the past decades, disparities persist in most of these measures of cancer dynamics. Certain populations have higher cancer incidence and mortality rates, benefit less from screening test availability and have higher rates of late-stage cancer diagnoses.

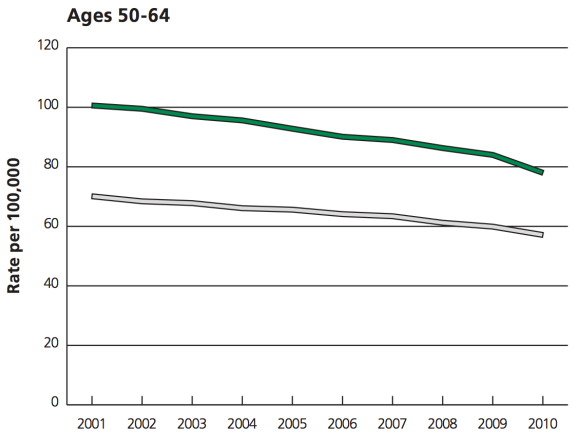
Low socioeconomic status is linked to lack of access to medical care and increased prevalence of particular CRC risk factors such as smoking or obesity. Social inequalities in the United States still influence the interactions between doctors and patients (Willems et al., 2005). Cultural factors such as language, values and traditions can influence health behaviors and affect individual's beliefs about cancer. Poor and medically underserved populations have higher chances of developing cancer. They also benefit less of early diagnosis, when cancer is most treatable (Freeman, 2004).

Individuals from low socioeconomic strata are less likely to benefit from prevention interventions and proper cancer management resulting in less favorable cancer outcomes compared to their counterparts (Ward et al., 2004).

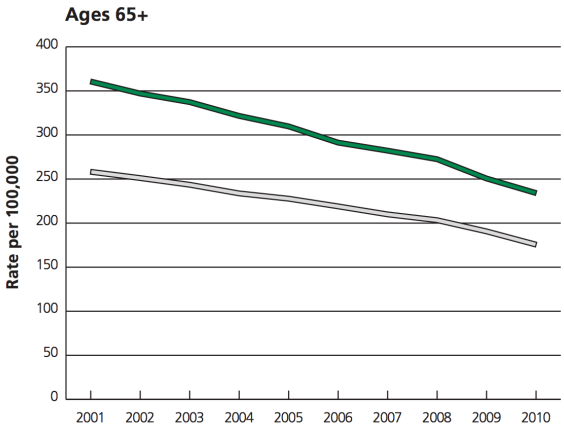
If detected early, the survival rate for CRC is around 90% (SEER, 2016). However, only 37% of CRC cases are discovered that early. Between 2006-2010, CRC incidence rates in African Americans were about 25% higher than those in Whites (Siegel et al., 2014). CRC mortality rates were 20% higher in African American population compared to Whites. Differences can be noticed among the age at which people contract CRC. Incidence rates are declining in adults of age 50 and over and are increasing in adults younger than 50, as shown in Figure 10.



CRC Incidence Trends by Age and Sex, 2001-2010 Ages 0-49 Rates are age adjusted to the 2000 United States standard population.



CRC Incidence Trends by Age and Sex, 2001-2010 Ages 50-64. Rates are age adjusted to the 2000 United States standard population.



CRC Incidence Trends by Age and Sex, 2001-2010 Ages 65+. Rates are age adjusted to the 2000 United States standard population.

Figure 10: CRC incidence trends by age and sex. Rates are age adjusted to the 2000 United States standard population (Source: American Cancer Society, 2014)

CRC incidence rates also vary geographically – mainly due to the regional variation in risk factors and access to screening and treatment. Other important factors that may have contributed to these disparities are socioeconomic factors, policies, and proximity to medical services. CRC incidence is about 40% higher in men than in women

(Siegel et al., 2014). Only 40% of the CRC patients are diagnosed early approaching a 5-year survival rate of 90%. Survival declines to 70% for patients with local disease and to 13% for those with distant CRC. The distribution of CRC stages and survival also varies by race and ethnicity, as shown in Figure 11.

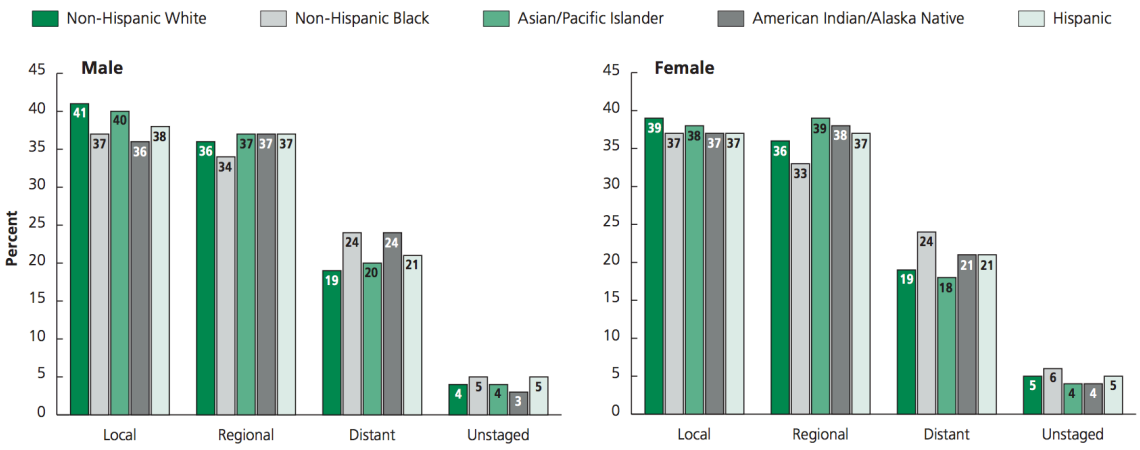


Figure 11: CRC stage distribution (%) by race/ethnicity, 2003-2009 (Source: NCI, 2013)

Being African American or having low educational attainment combined with the location of residence in the United States largely determines the likelihood to die due to CRC. African American men and African American women have higher CRC incidence and mortality rates compared to Whites and have more advanced stage at diagnosis. The most significant contributors to this disparity were identified as socioeconomic status and health insurance (Berry et al., 2009). A better understanding of who benefits from screening is needed, and whether high-risk African Americans get screened for CRC as recommended by current guidelines.

The combination of poverty and cancer is lethal (Freeman, 2004). Poor people are more likely to die of cancer due to lack of quality health care access and/or health insurance. Lacking healthcare insurance and/or the ability to pay for care, individuals may delay seeking medical attention for their symptoms. Poverty and lower socioeconomic status influence access to early detection tests, access to timely and high-quality treatment and is associated with increased the prevalence of other illnesses (comorbidities) (Edwards et al., 2014; Le et al., 2008). Moreover, these population groups are underrepresented in clinical trials.

CRC disparities pose several research challenges. Interventions that address the drivers of disparities for different population groups are needed. To reduce or eliminate CRC disparities, a multilevel approach that addresses the multi-factorial nature of CRC disparities and determinants across time is needed.

3.8 Delay in Diagnosis and Treatment

Delay in the diagnosis and treatment of illness has been of scientific concern throughout the last century. The term "delay" is used to describe the waiting time (usually measured in days or weeks) in cancer investigation and treatment procedures. Unlike the case for other cancers types, the benefit of early detection of colorectal cancer is astonishing. The early diagnosis of CRC leads to timely treatment, better patient outcomes, and better survival rates.

In the context of CRC, the impact of delay on survival has been studied since the 1960s (Ramos et al., 2007). The delay in CRC treatment is amongst the most common causes for malpractice claims (Singh et al., 2012). Given that greater delays lead to a more advanced stage at diagnosis and poorer survival, delay in cancer care is an

international priority (Richards, 2009). Several factors (such as age, gender, cultural beliefs, patient-physician communication, anxiety, and fear), directly or indirectly, influence timely seeking of medical care. These factors also serve as important determinants of outcomes for cancer patients.

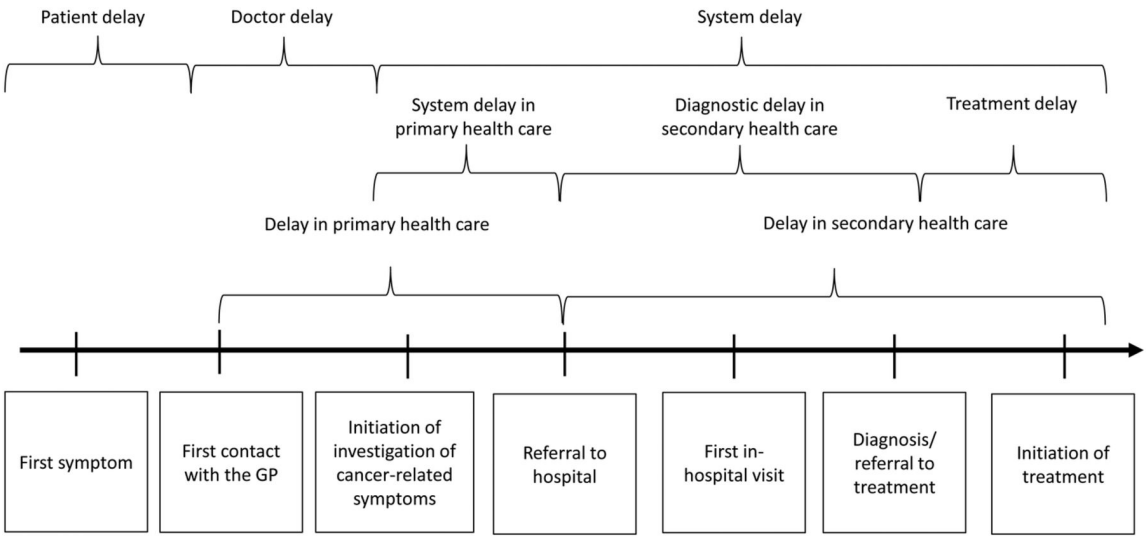


Figure 12: Categorization of delay (Source: Hansen et al., 2008)

The most widely cited theoretical framework for modeling delay, shown in Figure 12, was initially proposed by Safer et al. (1979), and has been successively continued and further developed by Andersen (1995). The resulting Andersen Model of Total Patient Delay has widely been used to study conditions ranging from cancer to myocardial infarctions (O'Carroll et al., 2001). A systematic review by Walter et al. (2012) investigated ten papers reporting the application of Andersen's model of patient delay assessing the delay in a cancer diagnosis. The authors found clearly identifiable stages

among symptom detection, presentation to a primary care physician, diagnosis and starting treatment.

Safer et al.'s (1979) model measures the total delay using a three-stage model, from the moment of noticing symptoms to the point of seeking treatment, as shown in Figure 13. A cohort study of CRC cases diagnosed between 1/1/2000 and 3/1/2007 found several missed possibilities for early diagnosis (Wahls & Peleg, 2009). About half of the individual cases diagnosed with cancer late, or lacking any CRC screening, were due to patient-related factors.

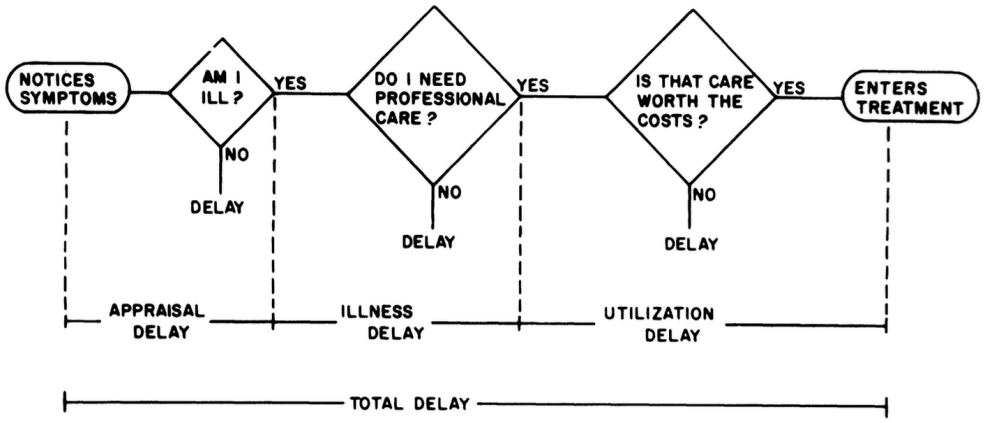


Figure 13: Decision to seek delay care in three stages of an illness episode (Source: Safer et al., 1979)

The framework proposed by Safer et al. (1979) measures the total time from symptoms onset to seeking treatment, initially employing a three-stage model. Andersen et al. (1995) further extended the model to include the time between the decision to seek medical care and to act on that decision. The more detailed model, shown in Figure 14, incorporates not only the delay components but also decision-making mechanisms. Most

studies on CRC delay distinguish between patient delay, physician delay and system (or hospital) delay.

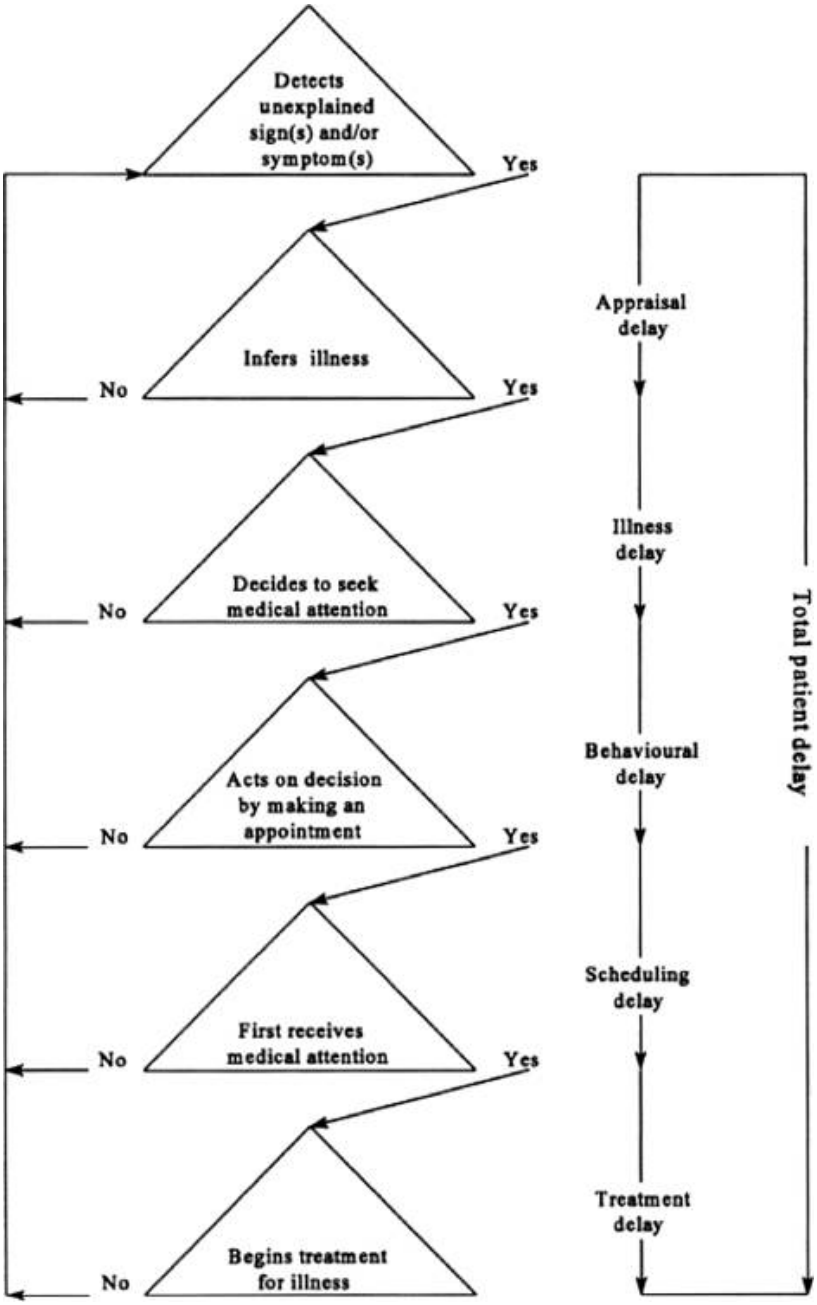


Figure 14: Andersen model of total patient delay (Source: Andersen et al., 1995)

Patient delay

The period from symptoms onset to the first consultation with a primary health care physician is referred to as patient delay (Weller et al., 2012; Hansen et al., 1997; Olesen et al., 2009). Knowledge about CRC has been shown to be a strong predictor of patient delay concerning CRC screening (Gimeno-Garcia et al., 2009; Cockburn et al., 2003). Before entering the health care system, patients initially notice and interpret their symptoms as severe or requiring medical attention. Underestimation of symptom seriousness is also a commonly reported cause of the patient delay (Courtney et al., 2012). When symptoms are not perceived as serious the patient is less likely to seek medical care. In the case of CRC, it is common for people to attribute their symptoms to other ailments, diet, stress, hemorrhoids or aging processes (Siminoff et al., 2014). The period between symptoms onset and recognizing these symptoms as serious is commonly referred to as appraisal delay.

Financial concerns were shown to have a direct effect on the use of healthcare services. Several studies focused on patients delaying to seek medical care due to perceived costs and/or lack of health insurance (Becker, 2004; Ward et al., 2008; Sabatino et al., 2006; Weaver et al., 2010). Also, financial barriers were found to prevent people without health insurance from seeking medical attention unless they were in severe pain (Becker, 2004).

Doctor (Provider) & System delay

Doctor delay refers to the waiting time between the first visit to the primary care physician and the initiation of cancer-related diagnostic tests or specialist (usually oncologist) referral. Among the factors commonly attributed to provider delay are: (1)

misdiagnosis – attributing symptoms to conditions other than CRC, (2) insufficient patient examination for malignancy, and (3) patient characteristics.

System delay starts from the initiation of cancer-related investigations until the beginning of treatment, or patient's decision to refuse treatment. It is desirable that the delay in CRC diagnosis as well as treatment delay to be as small as possible. The Andersen Model of Total Patient Delay has been further refined by Walter et al. (2012), as shown in Figure 15.

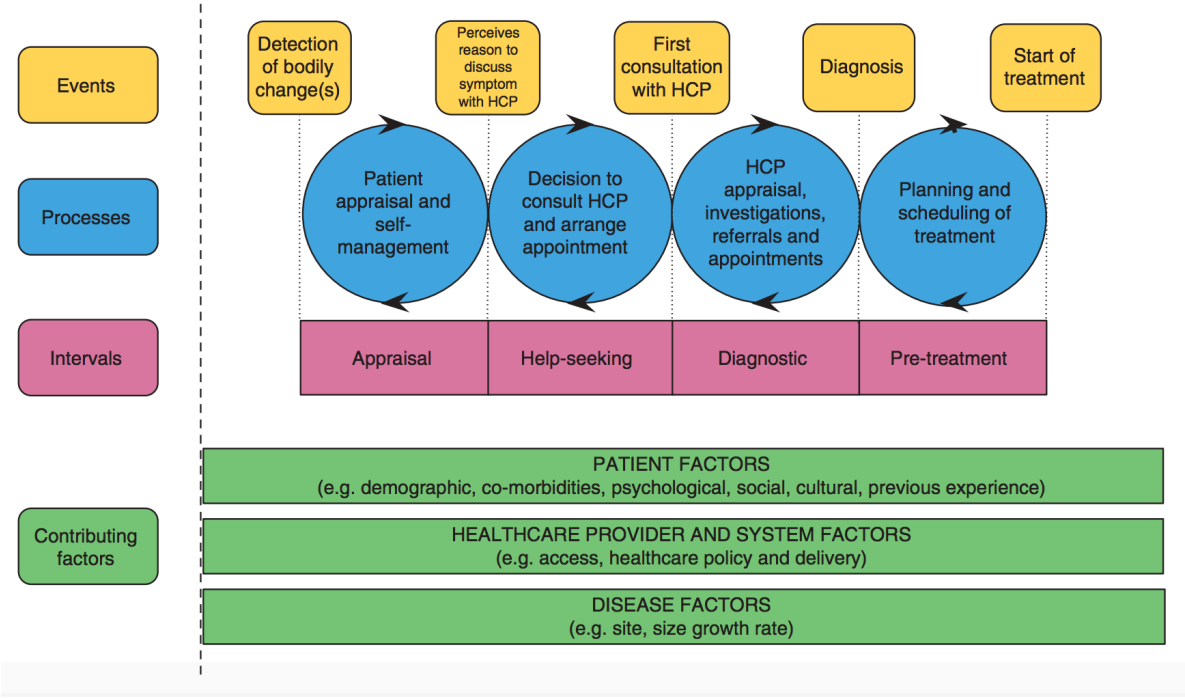


Figure 15: Pathways to treatment model (Source: Walter et al., 2012)

The authors combine illness and appraisal delay into "appraisal interval", behavioral and scheduling delay into "help seeking interval", and the time frame between

the first appointments with a physician up to a formal CRC diagnosis as "diagnosis delay", as illustrated in Figure 15. The time between the formal diagnosis and treatment initiation is combined into "pre-treatment interval". Within the revised model, the events marking the start and ending of each interval are clearly depicted.

3.9 Summary

This chapter provided an overview of colorectal cancer as it relates to this dissertation. It presented an overview and discussed CRC staging and the risk factors contributing to CRC incidence and mortality. The screening and testing methods currently available to detect CRC, and their sensitivity and specificity were discussed. CRC diagnosis the treatment options were considered. Survival, as well as recurrence and quality of life following a CRC diagnosis, have been provided, and the disparities across the CRC continuum have been mentioned, in regards to screening, incidence, treatment, and mortality. They were addressed in section 3.7. The chapter concluded with an important aspect of cancer care – delay in diagnosis and treatment. Delay in CRC care can occur at patient, physician and hospital level. This is of critical importance because of the association between delay and treatment/survival. The next chapter discusses the methodology employed in this work.

CHAPTER 4: METHODOLOGY

Contents

4.1 Introduction 73

4.2 Computational Social Science Modeling Approaches 74

4.2.1 Cellular Automata (CA) 74

4.2.2 System Dynamics (SD) 75

4.2.3 Microsimulation Models (MSM) 75

4.2.4 Agent Based Models (ABM) 76

4.2.5 Geographic Information Systems (GIS) 77

4.2.5 Social Network Analysis (SNA) 83

4.3 Data-driven ABM – a "bottom-up" Approach to SDC 83

4.4 Calibration 87

4.4.1 Input vs. Output Calibration 88

4.5 Verification and Validation of ABMs 89

4.6 "Understanding our Creations." 93

4.6.1 UML vs. Pseudocode 93

4.6.2 The ODD Protocol 95

4.7 Summary 96

> "The more we study the major problems of our time, the more we come to realize that they cannot be understood in isolation. They are systemic problems, which means that they are interconnected and interdependent."
>
> (Fritjof Capra, 2014)

4.1 Introduction

Computational Social Science (CSS) methodologies have evolved to better address the increasing complexity of human social systems, as well as to incorporate the growing availability of large-scale social data about the various aspects human activity. Traditional epidemiological methods employed in addressing the social determinants of cancer have strong limitations. To make further progress in this discipline, new "bottom-up" methodological approaches that account for the dynamic, multi-level complexity inherent in the SDC are needed.

CSS is the interdisciplinary science of complex social systems and their investigation through computational modeling and related techniques. Spatial, social, economic, biological, cultural and other processes that affect the health of an individual are complex, not nearly decomposable and can not be studied in isolation (Epstein & Axtell, 1996). CSS has been proposed as a "the third way of doing science" along with induction and deduction (Alexrod, 1997). This section will present the existing CSS methodologies and the motivation behind the employing an agent based computational approach to address the SDC. It will introduce the logic of simulation as a method, and discuss the process of building a model, as well as present the ODD framework for describing an ABM. Chapter 4 concludes with providing an introduction for the study case, and the study area in particular, for the model.

4.2 Computational Social Science Modeling Approaches

Several CSS approaches are increasingly being used in many disciplines for gaining a deeper understanding of complex social systems. Among the most common ones are: ABM, Microsimulation models (MSM), System Dynamics (SD), Geographic Information Systems (GIS), Social Network Analysis (SNA). CSS methodologies are particularly suitable for addressing the complexity inherent in the SDC, and they are constantly evolving to include state of the art approaches from other disciplines. When addressing complex systems, a combination of these tools is often employed for a more comprehensive analysis.

4.2.1 Cellular Automata (CA)

Cellular automata (CA) were created in the 1940s by Stanislaw Ulam and John von Neumann. A cellular automaton is a dynamic system composed of a grid of cells. The cells can be in one of a number of states (that is finite), that may change based on certain rules. Even though CA are composed of identical, simple components, they are capable of producing complex system behavior (Wolfram, 1984). The cooperative effects of the CA components may achieve this systemic complexity. CA approaches have been successfully used to model different biological phenomena (Ermentrout & Edelstein-Keshet, 1993). In cancer research, they have been used for modeling cancerous growth at cellular level (Qi et al., 1993), improving cancer therapy (Ribba et al., 2004) or modeling prognosis and treatment (Quaranta et al., 2005). Even though CA have been applied to a broad range of areas at biological and cellular level, they are limited in their ability to account for the heterogeneity or the rich behavioral adaptations characteristic to complex social systems.

4.2.2 System Dynamics (SD)

The first expansion of computer simulation in the social sciences consisted of simulations mainly based on SD (Gilbert & Troitzsch, 2005). The core idea behind SD is the attempt to simulate the target system as a combination of stacks and flows (equivalents for state variables and rates of change) and delays. The system itself is the one and only simulated unit. However, modeling in the social sciences involves analyzing subpopulations, individuals, groups, firms, and interactions between the above – hence the necessity for more refined approaches. SD need to extend their capabilities to incorporate the many levels involved in an ecological approach to health, accounting for disease outcomes, risk factors and other health resources (Homer & Hirsch, 2006). SD models have been extensively used to model patient flows in emergency care departments (Lane et al., 2000) and have also been applied to model substance abuse (Levin et al., 1975), as well as screening interventions for cervical cancer (Royston et al., 1999) among others.

4.2.3 Microsimulation Models (MSM)

While SD emerged in social sciences in the early 1960s, MSM appeared in the 1970s to address the shortcomings of the SD approach. In MSM, each modeling unit (be it individuals, firms or households) is treated individually (Orcutt, 1957). MSM enables us to examine the impact of policy changes on the individual decision-making units (Crooks & Heppentall, 2012). Nonetheless, a weakness of MSM is that it does not consider the interaction between individuals; their behaviors and intentions are not addressed. Subsequently, this issue is tackled later with the surfacing of ABM. A review of several MSMs applied to healthcare is provided by Birkin & Wu (2012). In the context of cancer care, MSM have been applied to a wide range of areas, such as modeling the cost-

effectiveness of mammography screening (Szeto & Devlin, 1996) or modeling the effects of rising chemotherapy costs (Lansdorp-Vogelaar et al., 2009).

4.2.4 Agent Based Models (ABM)

Although both SD and MSM are considered top-down modeling approaches, ABM models agents from a "the bottom-up" where fundamental social structures emerge from the interaction of individual actors. ABM starts with assumptions about the agents and their interactions and generates historical trajectories that uncover the dynamic consequences of initial assumptions. ABMs usually have at least three components: the agents (which can be individuals, firms, families), the rules that govern the interactions among agents and the environment in which the agents are situated. Agents can have internal states (represented by discrete or continuous variables). They can interact with other agents as well as with their environment. The system evolves over time, which can be represented either continuous or discrete. Higher-level patterns may emerge as a result of the agent interactions. There are several advantages of the ABM methodology over other modeling techniques. Firstly, ABM allows for emergent phenomena to develop from local, simple interactions. Also, among the main benefits of ABMs over approaches such as cellular automata is the possibility to incorporate heterogeneity of agents, in addition to the heterogeneity of their decisions, and the possibility of movement in space while interpreting the social content (such as cognition, for instance).

During the last decades, ABM has become a reliable technique for building models to study macro level outcomes in several disciplines. It has been successfully utilized to study several other chronic conditions such as obesity (Hammond, 2009) and smoking (Kasaie et al., 2013) among others. Auchincloss et al. (2011) proposed an ABM

that highlights the challenges of producing positive behavior change when individuals and resources manifest residential segregation and behaviors are triggered/restricted by various factors.

4.2.5 Geographic Information Systems (GIS)

Among the main advantages of GIS is the capability to analyze spatial information. It offers the possibility to capture real-world phenomena in multiple layers (three-dimensional) and resolutions (spatial scales) over time (Gimblett, 2002). Spatial models have the advantage to illustrate the distribution of individuals over space adequately, and visually represent how they relate to each other – therefore offering a rich representation of the environment. Three approaches for integrating information about space and time within a GIS can be distinguished: (1) location-based, (2) time-based, and (3) entity-based. A space-time representation of data within GIS can therefore be viewed as a series of "snapshots". Linking the two modeling methodologies allows us to create models that are directly connected to space, enabling us to compare aggregate outputs with real world data. The disadvantages of GIS are their limited dynamic modeling capabilities, and the inability to appropriately handle time (Crooks & Castle, 2012).

The United States atlases of cancer mortality highlight geographical variation in cancer incidence rates and risk factors (Leck, 1976). These patterns led to identifying spatial areas with higher carcinogenic exposures that influence cancer risk (Fraumeni, 1988). Even though GIS are still in the early stages of application in cancer research (Seidman, 2006), there are several examples of successfully applying GIS systems to different aspects of cancer control. GIS systems have shown to be useful in several areas, such as evaluating cancer clusters (Kulldorff et al., 1998), assessing breast cancer risk

and past exposure to pesticides (Brody et al., 2004), fogger trucks and the development of breast cancer, (White et al., 2013) or analyzing geographic patterns of cancer incidence (Rushton et al., 2004).

In situations where the complex system under study involves spatial components, it is useful to use GIS in combination with ABM. This combination can potentially be more advantageous and produce better results than using each of them in separation. Using ABM and GIS together provides the ability to model both patterns and process simultaneously (Rand, 2012).

4.2.5.1 Linking GIS and ABM

Linking ABM and GIS allows agents to be associated with actual geographic locations, and provides the opportunity to model emergent phenomena while preserving the individual interaction features available through ABM in a GIS over space and time. GIS and ABM are used together in situations where space, bounded rationality, and individual variability matters, when we are dealing with non-equilibrium systems and spatial distribution is among the objects of interest.

Spatial data is important for ABMs because it allows (via cells, networks, or a continuous space) for documenting macro-phenomena, informing micro-level process modeling, deriving maps or different demographic variables of agents and their environments. In ABMs, space has many purposes, such as: containing the agents, and defining spatial relationships between them, as well as between agents and the environment. It is also used to capture various nuances of spatially distributed patterns (Brown et al., 2005).

Both ABM and GIS are methodologies that can help advance our understanding of complex social systems. The purpose may vary: from gaining insight into the structure or behavior of a real world system, to forecasting the behavior of systems that are difficult to explore in the reality.

4.2.5.2 Coupling vs. Integration/Embedding

Coupling refers to the linkage of two stand-alone systems by data transfer (Crooks & Castle, 2012). Integration, on the other hand, implies embedding the necessary functionality of either ABM or GIS within the dominant system. Depending on what integrates what, the final system is referred to as either GIS-centric or modeling-centric. Coupling itself is accomplished using one of three approaches: loose, moderate and tight. Loose coupling refers to the asynchronous operation of functions within each system, with data being exchanged between systems in the form of files. According to this approach, the inputs are prepared within the GIS. Hereafter, they are sent to the simulation system that performs the program execution and sends the data back to the GIS. GIS would perform further explorations and visualization tasks. It is essential that the format of data that is transmitted by GIS and the format of the data the ABM can read to be the same. Moderate coupling is conceptually situated between the loosed and tight coupling. It includes database access links and procedure calls between the GIS and ABM, and it increases the execution speed. Tight (close) coupling involves simultaneous operations of systems, allowing direct communication between systems during the program execution.

An alternative approach would attempt to connect the two systems. This middleware-based method (which essentially represents tight coupling) allows for

appropriate choice of the functionality available in the GIS or ABM for a given mission. ABM could be used for simulation purposes, while GIS for visualization, for instance. GIS-centric or modeling-centric should be utilized in the case of studying geospatial agents consisting of a vast number of complex interacting agents. According to Brown et al. (2005), the four approaches considered for ABM and GIS coupling are:

(1) Identity relations – referring to the relation between agents and the spatial database,

(2) Causal relations – addressing the changes in spatial features that may happen because of agent interaction,

(3) Temporal relations – where information updating could be challenging, especially when it is synchronous, and

(4) Topological relations.

Coupling can be useful in circumstances where ABM and GIS systems already exist in isolation, or when the cost of building the functionally of one system into another would be too high (Maguire, 2005). Coupling has often been the preferred approach in linking ABM with GIS. Nonetheless, there is a disadvantage when compared to integration: coupling often results in rather specialized and isolated solutions, which prevents a standardization linkage between ABM and GIS to emerge. An advantage of integration, compared to coupling, is that an integrated system is user-friendlier. Modeling-centric approach became popular with the advance of scripting capabilities, and the fact that it is relatively straightforward to be programmed (Gilbert & Bankes, 2002). Moreover, the functions available from GIS libraries reduce model development

time and improve the model transparency. Integration does not come without disadvantages, posing a risk for the designed systems to be closed, monolithic and costly (Fedra, 1996).

A few of the challenges faced by GIS and ABM integration are model communication and validation, linking quantitative and qualitative methodologies and real-time simulation. Crooks et al. (2008) compiled a comprehensive list of the key challenges in combining ABM with GIS:

- The goal of the model,
- Rooting the model in theory,
- The extent to which it is possible to replicate it,
- The scope of verification, validation and calibration
- Representing model dynamics and the interaction between components and agents,
- The degree to which the model is operational, and
- Model communication.

As a result of integration, GIS users get the possibility to model emergent phenomena through individual interactions, and ABM modelers benefit from having agents tied to specific geographical locations. Loose coupling is widely used in linking ABM with GIS, but it is suggested that much tighter coupling is needed to simulate real world spatial/temporal dynamics (Brown et al., 2005). Each of the models discussed can be placed within a power vs. difficulty continuum of model development as shown in

Figure 16. Linking GIS and ABM is still non-trivial, and many challenges that need to be addressed and considerations to be taken into account (Gilbert, 2007).

		Agent	
		Designed	Analysed
Environment	Designed	Model description – Abstract Purpose/intent – Discovery of new relationships – Existence proof Verification and validation strategy – Theoretical comparison – Replication Appropriate development tools – Easy to implement simulation/ modelling system Example model – Filatova et al. (2009)	Model description – Experimental Purpose/intent – Role-playing games among stakeholders – Laboratory experiments Verification and validation strategy – Repetitions – Adequacy of design Appropriate development tools – Flexible simulation/modelling systems with well developed user interfaces Example model – Mooij et al. (2002)
	Analysed	Model description – Historical Purpose/intent – Explanation Verification and validation strategy – Qualitative: goodness of fit Appropriate development tools – Advanced simulation/modelling systems linked with GIS Example model – Mathevet et al. (2003)	Model description – Empirical Purpose/intent – Explanation – Projection – Scenario analysis Verification and validation strategy – Quantitative: goodness of fit Appropriate development tools – Low-level programming languages Example model – Jackson et al. (2008)

Figure 16: Power vs. difficulty continuum of model development (Source: Crooks & Castle, 2012)

4.2.5 Social Network Analysis (SNA)

Social relationships are a fundamental element of human life. SNA addresses social relationships through the lens of network theory, where the network "nodes" are the actors (that may represent people, families, organizations) and the "ties" represent relationships between actors. For a better understanding of network data, social networks are often visually represented as diagrams of points and lines, where points denote the actors and lines represent the ties between actors. Every health topic can be viewed through the network perspective (Valente, 2010). This methodology has been successfully applied to areas such as HIV transmission through sexual contact networks, substance abuse, smoking, and obesity. The process of applying social networks to behavior change has been referred to as "network interventions" (Valente, 2012). Distinguishing between the several types of network interventions and individual social context is crucial for harnessing the power of social interaction to achieve positive behavior change.

Social networks are determined by individual characteristics such as age, sex, geographic location, education, income and occupation among others. The goal is to gain insight into how social networks determine the types of relations people have, as well as disentangling their individual and group behaviors.

4.3 Data-driven ABM – a "bottom-up" Approach to SDC

The slogan "Keep it Simple, Stupid" (or KISS) is a widely adopted principle in ABM (Axelrod, 1997). Similar to Occam's razor, it states that the simpler the model – the better. The simplicity should be specifically expressed in the assumptions of the model, rather than in the model results. For the more complex models, Edmonds & Moss (2005)

recommended the KIDS approach (standing for "Keep it Descriptive Stupid"). The KIDS approach, as opposed to the KISS one, addresses the increasingly descriptive nature of ABMs (which can get quite complex). It is an appropriate approach for models based on empirical data, that aim at maintaining a high fidelity to the system being modeled, and focus less on abstraction and more on the descriptive aspect of model building (Edmonds & Moss, 2005). Simple (toy) models, as opposed to the high fidelity ones, are models that represent stylized and highly abstracted representations of real world systems. An example of such model is Schelling's model of segregation (Schelling, 1969). In its original version, Schelling's model has shown that simple, individual racial preferences, up to some threshold, can lead to higher levels of total segregation. While Schelling's model has explanatory power, most toy models act as proof of concept models. While useful in some cases, they are mere abstract caricatures of the real world systems. For addressing complex problems involving heterogeneous agents and multilevel interactions, data-driven models that account for this complexity, maintaining higher fidelity to the real world phenomenon are needed.

The capacity to handle and investigate large amounts of data revolutionized numerous areas. However, the emergence of a data-driven "computational social science" has been developing much slower (Lazer et al., 2009). The increasing amount of data about social characteristics calls for new approaches able to integrate data from multiple sources into a dynamic model. To address the reductionist view predominant in the study of SDC, a data-driven bottom-up approach aiming to shed light into how social systems may influence the dynamics of cancer is implemented.

Abstraction is a vital part of the model building. “No substantial part of the universe is so simple that it can be grasped and controlled without abstraction.” (Rosenblueth & Wiener, 1945). The following step aims to distinguish which parts of the real world system are important for modeling the phenomenon we wish to understand. That phenomenon is then replaced with a comparable, but simpler structure. How much detail to include and how much to set aside largely depends on the purpose of the model. Alternatively, as Einstein put it, “Everything should be as simple as possible, but not simpler.”

There are many ways in which data can be incorporated into the process of building ABMs and informing the conceptual dynamic models. One of the benefits of ABM is that when there is not enough data; the data can be generated during multiple simulation runs. The logic of the simulation as a method is presented in Figure 17:

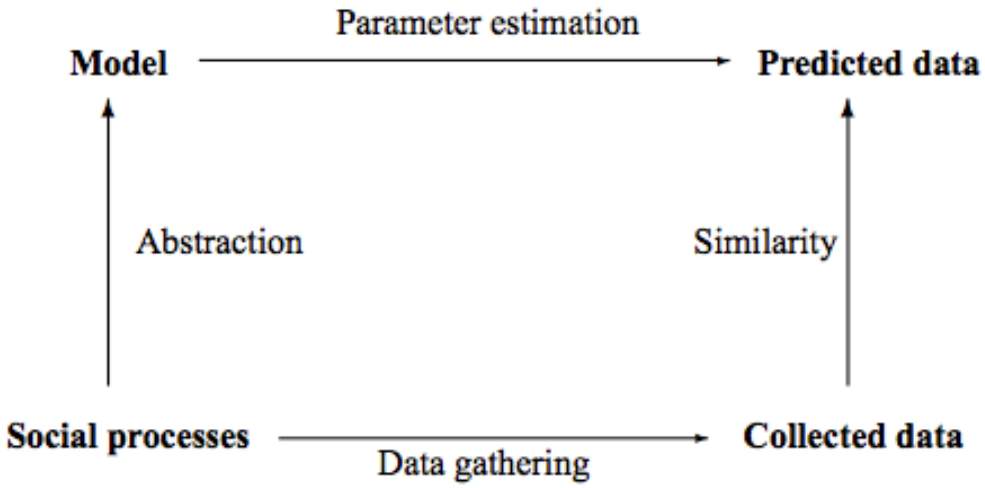


Figure 17: The logic of simulation as a method (Source: Gilbert & Troitzsch, 2005)

The target is the real world phenomenon we are interested in modeling. After applying abstraction to the target, we obtain the model – which is a simplified view of the phenomenon. As a result of simulating the model, we obtain simulated data that can be compared, either qualitatively or quantitatively to data collected from the real world target with the purpose of validating the model.

Following the KISS principle makes adopting a data-driven approach to ABM challenging. By employing such a method, we add additional levels of complexity to the model – conflicting with the KISS principle (Edmonds & Moss, 2005). Nevertheless, if the model is to be used for informing policy and decision-making, validation against data is recommended, where the models can be directly compared with real world data. A middle line should be sought between building a complex data-driven model and maintaining simplicity to a reasonable degree. "The goal of science is to make what is wonderful and complex, understandable and simple - but not less wonderful." (Simon, 1996).

By using data, we build models that are closer to the target system. The model becomes accurate, embracing the complexity of the real world phenomena. Data may be used at multiple stages of model development, from model initialization (as opposed to opting for initializing the model from a uniform random distribution, that is rather common in the ABM field) to validation – where we compare the model results with existing data. One of the benefits of ABM is that it can integrate data from multiple sources and disciplines, from GIS maps and sample surveys to social media data and

meta-analyses. Calibrating the model to real world data increases the model credibility as well as improves its validity.

Survey data have been increasingly used for both informing ABM as well as for validation purposes (Hassan et al., 2008). The increasingly preferred method for building models is employing a hybrid approach involving several existing CSS methodologies, as they complement each other to accomplish the goal of creating useful models. Robinson et al. (2007) mention the following empirical approaches for specifying micro level processes within ABMs: representation surveys, participant observation, field and lab experiments, companion modeling, GIS and remotely sensed spatial data. These approaches can be used for both documenting the macro level phenomena as well as for informing the micro processes within the model. Given the dynamical nature of ABMs, survey data repeatedly collected at different points in time can be particularly valuable for both informing ABMs as well as for validating them.

4.4 Calibration

Calibration is considered to be one of the greatest challenges in building ABMs (Crooks et al., 2008). Model calibration should be realized before the model output is validated (Ngo & See, 2012). It "refers to the process of creating a model such that it is consistent with the data used to create the model" (Verburg et al., 2006). This modeling phase involves fine-tuning the model parameters to identify that unique set of parameters that aligns the model to data. An example of calibration in its simplest form is obtaining observations by tuning model parameters and observing how it compares to the real world observations. Calibration is, therefore, an iterative process, comparing the model to actual system behavior. If we have empirical data, we can use calibration to find the set

of parameters that gives us the best fit to the observed data. Among the objectives of calibration is to improve the agreement with experimental data.

One of the challenges faced at the calibration stage is the difficulty in obtaining the necessary data for proper calibration (Manson et al., 2012). Also, properties like path-dependency, and other complex system characteristics pose further challenges. The model is considered valid if it can correctly answer the question it aims to answer (Casti, 1997).

4.4.1 Input vs. Output Calibration

When calibrating the model against real data we distinguish between two types of calibration: input and output. The difference between these two calibration types lies in the following: the input calibration refers to the assessment of the realism of models' initial conditions and fundamental propositions while output validation involves comparing the outputs produced by the model with real-world data. Input and output calibration can, and should be used in combination.

Because of the non-linear properties of ABMs and due to their high dimensionality, fitting the model output to real data for calibration/validation purposes is often a challenging task. Consequently, computer simulations utilize 'ad-hoc' approaches to parameter estimation, validating the model via a qualitative comparison of model predictions with real data (Dancik et al., 2010). Alfarano et al. (2005) proposed a direct estimation method of the underlying parameters of a relatively simplistic ABM using a parametric approach. Only a few (relatively simple) examples exist on estimating ABMs. They provide ABM estimations that are simple enough to derive a closed form solution for the distribution of relevant statistics. Winker et al. (2007) address the issue on

moment selection, suggesting a set of statistical data on exchange rate returns to estimate models of the exchange rate.

Identification problems come onto the scene when we deal with models that are consistent with the data that is used for empirical validation. While circulating in the econometrics literature as 'identification', in the philosophy of science this phenomenon is called 'under-determination' (Fagiolo et al., 2007). Identification is tightly connected to the Duhem-Quine thesis (Earman, 1983), which asserts that any hypothesis is inevitably tied to some auxiliary assumptions, and it is therefore not possible to test and falsify a single hypothesis in isolation. A suggested path to circumvent this challenge is adding equations (based on economic theory) to the models to the point when the rank of the coefficient matrix is similar to the amount of independent variables (Moss, 2008). According to Hansen and Heckman (1996), microdata could offer one potential avenue for resolving the identification problem, though no clear formal way would demonstrate how access to microdata would solve the identification issue.

4.5 Verification and Validation of ABMs

Verification - means building the model "right" (Balci, 1998). It refers to the accuracy with which the model logic is implemented. At this stage computer code is tested for its formal logic. Verification errors could be the cause of emergent model artifacts that are not the result of the models' true behavior. When verifying a model, we usually look for logical errors or "programming bugs" that might occur during the implementation of the model design into computer code. Examples of unpleasant surprises that can happen are floating-point errors like for instance adding 0.05 twentyfold can produce a total that differs from 1 (Polhill et al., 2004). Proportionality and graceful degradation are also

considered subsections of 'verification', and so do measurement error, sampling regimes, observer bias, and data availability. One way to address verification issues is with sensitivity analysis, where the model parameters are varied over multiple runs of the model runs so that we can examine the changes the performance of simulations (Manson, 2003). It is, therefore, important to acknowledge that artifacts could emerge as a potential result of including space heterogeneity, or applying various randomization techniques. The goal of verification is ensuring that the model works the way it was intended to work and that it was programmed correctly without any critical errors/bugs.

Validation – refers to building the "right" model (Balci, 1998). That is the extent of which the model adequately represents the target system. "The extent to which" indicates that it would be irrelevant to classify a model as valid or invalid. Instead of treating it as a binary event, one should rather consider the existence of certain degrees of validity (Law & Kelton, 2000). Two types of validation can be distinguished: internal and external validation.

Internal validation addresses the issue of whether the model works in a similar way to the real world system being modeled. Achieving internal validity is harder than it may seem (attention is needed for analyzing and understanding if the results are indeed the actual result of the model and not anything else) (Axelrod, 1997). It is necessary to stress the value of replication in general and for validation purposes in particular. Replication has been referred to as "one of the hallmarks of cumulative science" (Axtell et al., 1996). It helps us test the robustness of the inferences based on models. Alignment or docking refers to comparing two models and check if they can produce the same

results (Axtell et al., 1996). Docking can also be considered as a part of the internal model validation step.

External validation refers to comparing the behavior of the target system with the behavior of the model. Depending on the purpose of the model, external validation can be qualitative or quantitative. Qualitative validation refers to comparing the replication of a real-world pattern with its modeled equivalent. As an illustration, we could take the simple version of Schelling's segregation model, where agent interaction at one level, gives birth to segregation patterns at a higher level. Output validation can be listed as an external validation mechanism, as it involves graphically and statistically matching the model's predictions against a set of real data (Ngo & See, 2012). Fitting the model to actual data poses a significant challenge for validation, as in social simulation we are dealing with agents that interact and adapt. However, errors in model input data can cause errors and uncertainties in model output. Therefore, it is essential to think about means of minimizing them (Evans, 2012).

Validation is an iterative process requiring a step-by-step iterative approach (Ngo & See, 2012). As shown in Figure 18, the process of validating a model may start with face validation. After face validation has been achieved and the model looks right prima facie, sensitivity analysis is performed, followed by model calibration. Following these stages, the results of the model are compared to the real world data. If the predicted results match reality, according to Ngo & See (2012), the model is considered to be fully validated. Among the most important model validation stages is the final stage, when the model results are compared with the real data, as this shows the extent to which the

processes have been correctly implemented and the level of fidelity to which the real world system have been represented. This helps to ensure that the model was built correctly and that the behaviors it manifests are not the results of artifacts or errors committed during the many phases of building a model (Troitzsch, 2004). One of the main goals of validation is to measure the performance of the model, indicating the extent to which the model is to be trusted (Verburg et al., 2006).

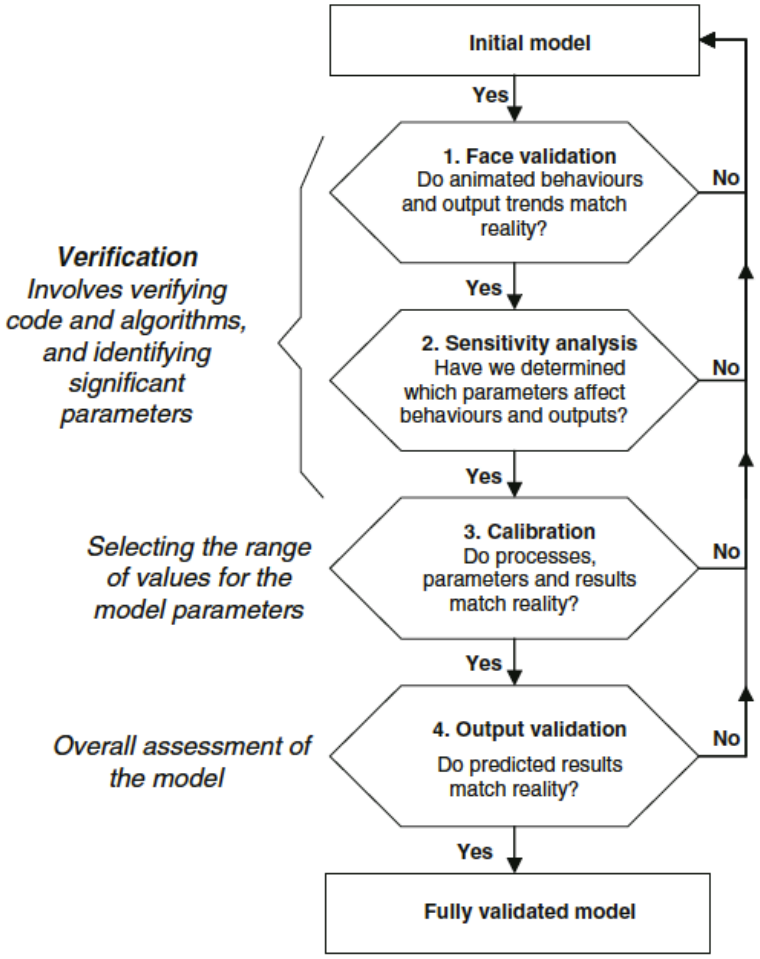


Figure 18: A general validation process of ABMs (Source: Ngo and See, 2012)

4.6 "Understanding our Creations."

ABM is a tool that can help us gain insight into the overwhelming complexity of the different parts of our world. However, understanding models – their purpose and results - presents another challenge in itself. As ABM becomes more widely used as a methodology, and as models get more complex, it is becoming increasingly challenging to make sense of the plethora of emerging models that are built on different platforms, based on various conceptual frameworks and present simulation results in dissimilar ways. Numerous efforts to improve the process of describing agent systems have been made. Diagrams, flowcharts, UML, and pseudo-code are commonly used for the purpose of describing models. Object-oriented design and programming promote reasoning in a way that models the way we think and interact in the real world, which is a strong point for modeling agent-systems. While there has been a clear agreement that the object-oriented paradigm is a natural way to model agent systems, UML has largely been absent from ABM publications (Bersini, 2012).

4.6.1 UML vs. Pseudocode

UML allows creating diagrams that encompass both state variables of model entities and their processes. It provides a higher level of abstraction than Object Oriented Programming (OOP), and its main goal is to provide good graphical visualization for object-oriented development. The UML diagrams that are most suitable for describing the ABM development are class, sequence, state, and activity diagrams (Bersini, 2012). We can use diagrams to show the hierarchy of objects in a software program; we can show how one thing leads into the next with sequence diagrams or use activity diagrams, which

basically have the same purpose as flowcharts. UML has the potential to become the 'lingua franca of ABM' (Miller & Page, 2009).

Compared to diagrams – pseudocode is linear, rather than graphical. It allows us to combine computer code with general wording and represents a good way to present algorithms to readers without familiarity with computer code. It can be used when we are dealing with models that are straightforward and small. Among the reasons why software developers may prefer pseudocode to UML would be the unwillingness to expose personal coding style. Also, software programs may not be easily portable from one machine or operating system to another, and sharing the pseudocode only (instead of language-specific code) would be more appropriate (Gilbert & Terna, 2000).

While pseudocode may be a good option when working at lower levels of abstraction (Booch, 2006), UML takes over when we have to deal with higher levels of abstraction. In a study by Scanlan, (1989) flowcharts outperform pseudocode as they produced fewer errors in perception and understanding, giving more confidence in accepting the algorithm, reducing the time spent on answering questions about an algorithm, and lessen the number of times needed to look at an algorithm.

Another good practice in developing agent systems is pattern-oriented modeling (POM). This approach offers a unifying framework for understanding the inner workings of agent-based complex systems, making bottom-up modeling more rigorous and comprehensive (Grimm et al., 2005). POM helps identify patterns of underlying structure, processes and functions, and incorporates them into building the model. Two issues of concern may emerge at this point: first, if the problem drives the model, it will

be too simple; and second - if all available data motivate the model design, the model will end up being too complex. Therefore, POM advices observing patterns, at different spatial and temporal scales, and having them refine the research questions.

4.6.2 The ODD Protocol

Ontologies and protocols assist in understanding models and possibly accelerate their development. The ODD protocol (that stands for: overview, design concepts and details) has been proposed as an alternative way to describe agent-based social simulations (Grimm et al., 2010). It involves seven sub-elements, and each of them must be documented good enough for the model's purpose and design to be clear and replicable, as shown in Table 6:

Table 6: The ODD protocol overview

Overview	Purpose
	State Variables and Scales
	Process Overview and Scheduling
Design Concepts	Design Concepts
Details	Initialization
	Input
	Sub-models

Originating in ecology, the ODD is one of the most commonly used protocols for describing ABMs in academic literature. It enforces a logical ordering when presenting

agent system, improving readability not only for the purpose of understanding but regarding replication purposes as well. According to Grimm et al. (2010), among the complaints about ODD are the fact that ODD can be redundant and that it is overdone for simple ABMs. More recently, Müller et al. (2013) extended the ODD protocol to account for human decision-making. The updated protocol, named ODD+D, provides a standardized practice aimed at clarifying the empirical foundation of the choice of the decision model (Müller et al., 2013). This protocol seeks to improve the current way of describing ABMs so that models can be communicated more efficiently. Among the emergent benefits of using this protocol are rigorous model formulation, facilitating model comparison and promoting holistic approaches to modeling (Grimm et al., 2010). With the increase in popularity of the ABM methodology, it is important to have a good way of communicating models. Following these guidelines and practices helps ensure model extensibility, maintenance, and replicability.

4.7 Summary

This chapter revised some of the existing computational social science methodologies (CA, SD, MSMs, ABMs, GIS, SNA) with a detailed focus on ABM. The aspect of linking GIS with ABM is discussed, together with the distinction between coupling and integration of the two methodologies. The benefits of data-driven modeling, as opposed to caricature, or toy models, are examined together with the general process of building a model. Section 4.5 focuses on Calibration while the following section discusses model verification and validation. The difference between internal and external validation is presented. Once the model has been designed, implemented, calibrated, tested, verified and validated, I proceed to presenting the next section – that is communicating the model.

Section 4.7 touches on communicating the model to the scientific community and existing protocols that facilitate this phase. While UML and pseudocode can be used to enhance model presentation, the most commonly used protocol in describing ABMs - the ODD protocol – is briefly introduced as it will be utilized in the next section for the purpose of describing the ABM introduced in this work.

This chapter builds on the previous chapters in laying the theoretical foundations for building the model. The following chapter moves on to present the study area and the social, economic, and health characteristics of the population for the study case.

CHAPTER 5: STUDY AREA

Contents

5.1 Introduction 99

5.2 The Geography of the District 100

5.3 Socio-Demographic Characteristics of DC Residents 103

5.3.1 Age distribution 104

5.3.2 Race/Ethnicity 107

5.3.3 Marital Status 110

5.3.4 Education 111

5.3.5 Income and Poverty 114

5.3.6 Employment 118

5.3.6 Occupations 120

5.4 Health Characteristics of DC Residents 121

5.4.1 Life Expectancy 123

5.4.2 Overweight and Obesity 124

5.4.3 Physical Activity 125

5.4.4 Smoking/Tobacco Use 127

5.4.5 Alcohol Use 128

5.4.6 Health Insurance/ Access to Heath Care 130

5.5 Health Disparities 133

5.6 Colorectal Cancer in DC 135
5.6.1 CRC Status in DC 135
5.6.2 CRC Incidence in DC 138
5.6.3 CRC Mortality and Survival in DC 141
5.6.4 CRC Screening Status in DC 147
5.6.5 CRC Disparities Summary 151
5.7 Summary 154

5.1 Introduction

While the previous chapters covered the theoretical background for this work, in this chapter, I will elaborate on the study area of the model. The chapter starts with introducing the geography of the study area - Washington D. C., and the characteristics of district residents relevant to the model. Washington D. C. was chosen as the study area due to the fact that from 2000 to 2004 it ranked among the highest in the United States for CRC deaths (ACS, 2008). Initially, the demographic and socioeconomic characteristics of district residents are introduced, followed by health characteristics and the risk factors distribution among the district residents. Health disparities predominant in the district are briefly reviewed. The last section of the chapter reviews the status of CRC in the district and presents the data on CRC incidence, mortality, survival, as well as reviews the screening rates and other CRC-related statistics relevant to the model. The disparities predominant in CRC care are also briefly reviewed. The demographic, socioeconomic and health data presented in this chapter are all based on the year 2000, as

this serves as the initial year of the simulation. The next chapter will introduce the details of model structure and development, applying the ODD protocol.

5.2 The Geography of the District

Washington, D. C. is the urban center of the Washington Metropolitan Statistical Area (MSA), as shown in Figure 19, and the capital of the United States. Washington, D. C. is bordered by Arlington and Alexandria (Virginia) to the south and west, Montgomery County (Maryland) to the northwest, Prince George's County (Maryland) to the east and the Potomac River. It has a total area of 68.3 square miles (177.0 km²).

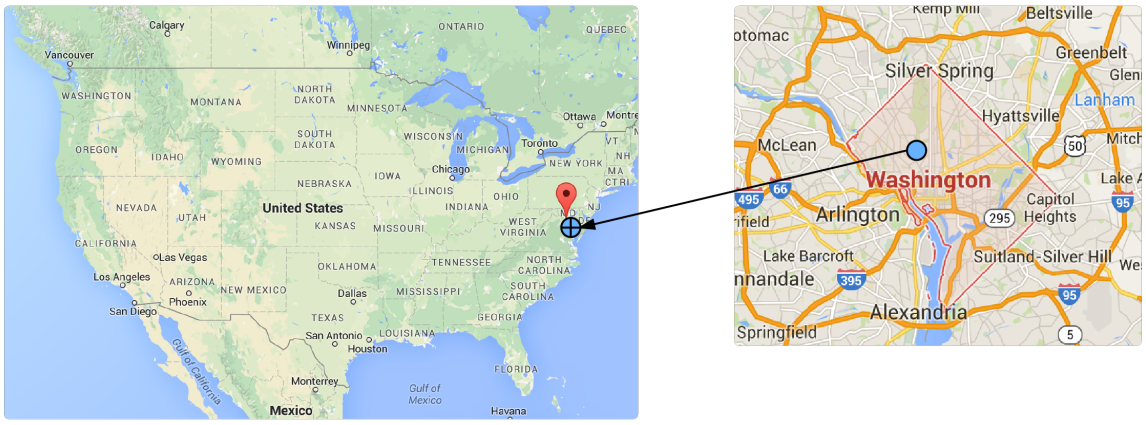


Figure 19: The study area (Source: Google Maps, 2016)

According to the U. S. Census Bureau (2000), in 2000 Washington, D. C. consisted of eight wards and 188 tracts, as shown in Figure 20. The district neighborhoods are grouped in 39 “neighborhood clusters”, each cluster consisting of three to five neighborhoods. District government planning, budgeting and service

delivery is focused on these clusters. A census tract is a small and relatively permanent statistical subdivision of a county, meant for long-term statistical comparisons among censuses.

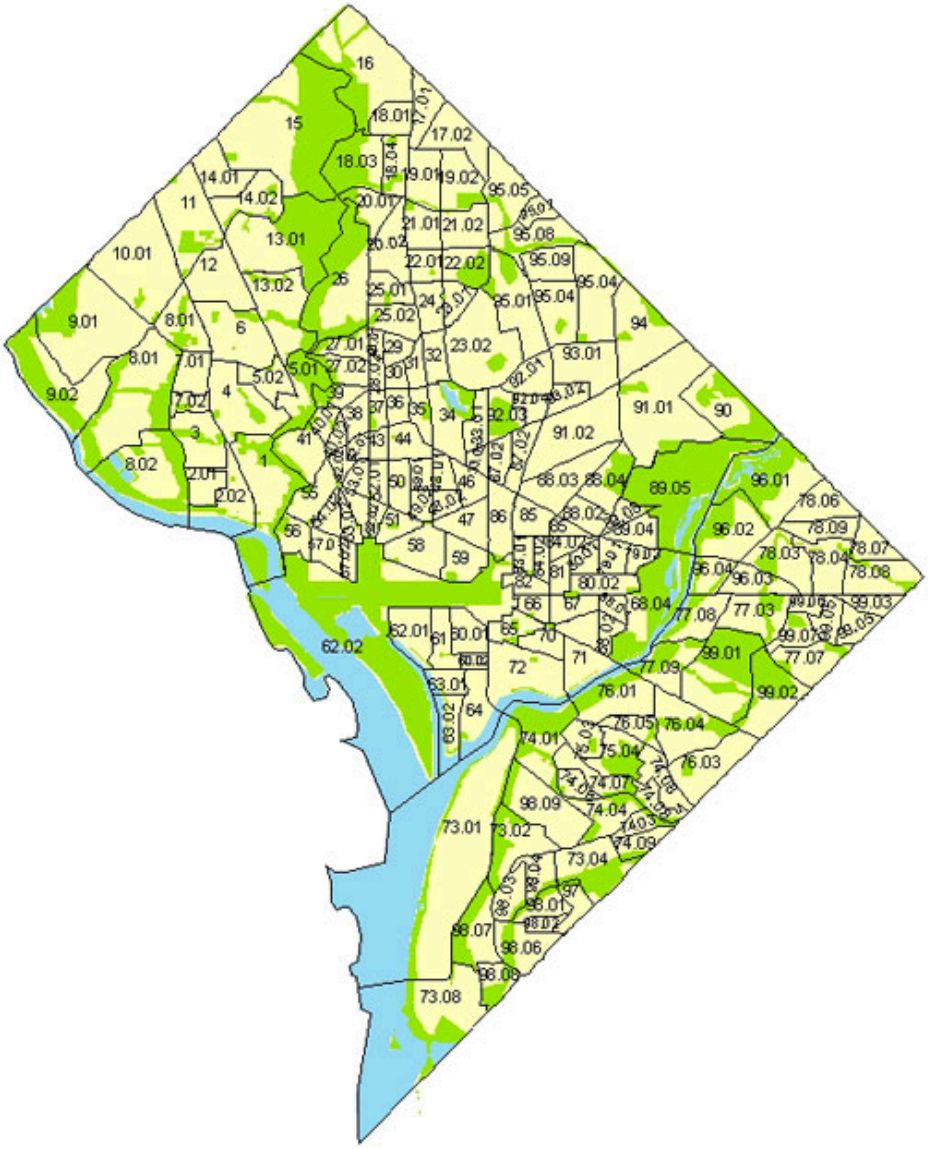


Figure 20: Washington, DC 2000 census tracts (Source: NeighborhoodInfo DC, 2016)

A census ward is another type of census population enumeration districts. The U. S. Census Bureau planned to report demographic data by wards following the 1980 census (U. S. Census Bureau, 2000). However, wards appeared to have certain statistical downsides, given the considerable variation in their size and geographic configuration.

Based on the 2000 U. S. Census, on April 1, 2000, there were 572,059 ethnically and culturally diverse individuals in Washington, D. C. In 2000, nation's capital experienced a fast suburban growth, experiencing a moderate population decline in the 1990s (Singer, 2004). There were 274,845 housing units and 248,338 households in DC in 2000.

An overview of Washington, D. C. profile showing race breakdown, population distribution per tract, and the population distribution by age and sex according to the 2000 U. S. Census Bureau is presented in Figure 21. Population density also varies by tract, with residents being more concentrated in the center of the district in 2000, as shown in Figure 21.

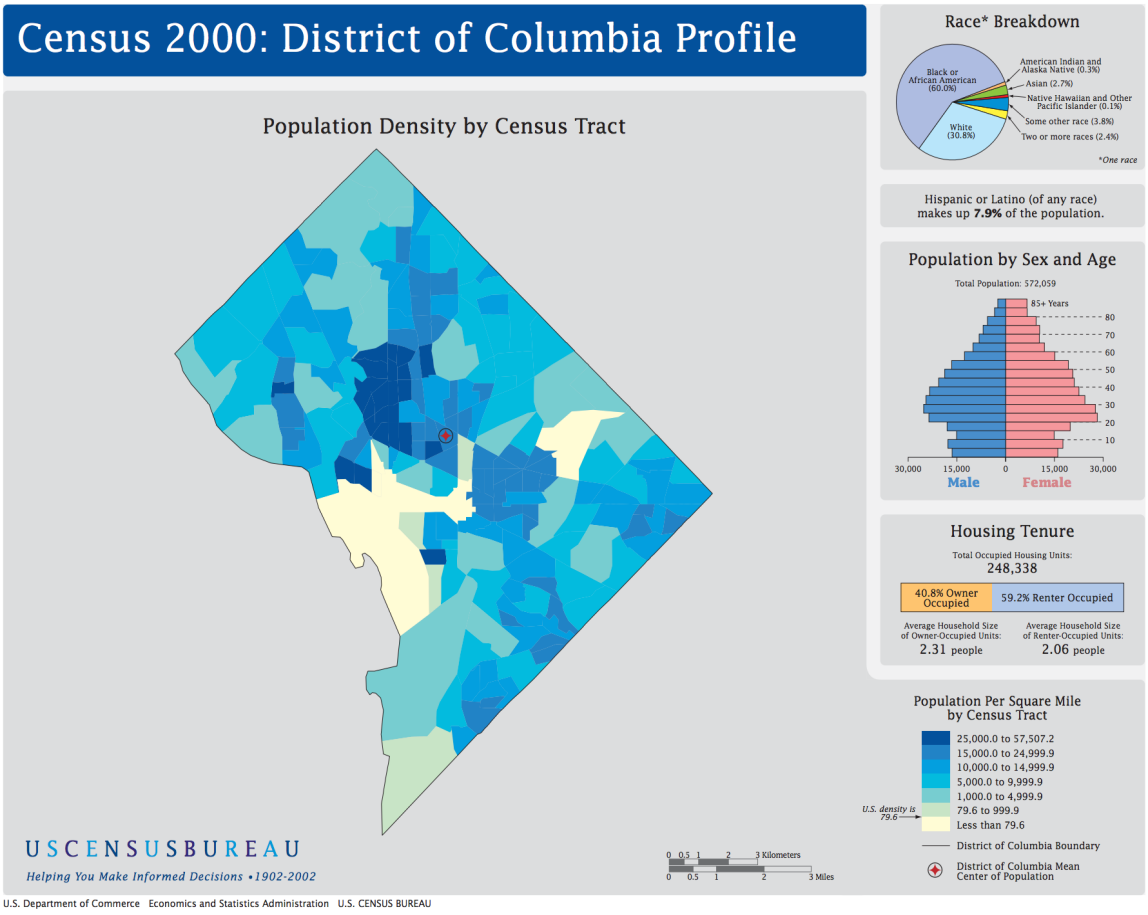


Figure 21: The 2000 Washington, DC Census profile (Source: USCB, 2000)

5.3 Socio-Demographic Characteristics of DC Residents

The socio-demographic profile of the DC residents is compiled from multiple sources. The main data sources used in the model are: the 2000 Decennial Census (U. S. Census Bureau, 2000) data for DC – which is the broadest source of information on the United States population, the National Health Interview Survey (NCHS, 2000), the NCI Surveillance, Epidemiology, and End Results (SEER, 2014) and the 2000 BRFSS data for Washington, DC (CDC, 2000). To complement the main data sources, multiple

secondary sources are used (such as CDC reports, U. S. Census data, qualitative research data, and also other surveys and studies that focus on the social determinants of cancer).

5.3.1 Age distribution

Differences in age distribution may account for variations in prevalence, incidence and mortality rates of certain health disorders. Younger population groups are usually more susceptible to substance abuse, STDs, or crime, while the older population tends to suffer from chronic conditions such as diabetes, cancer or heart disease. In 2000, the percent of young people under 18 was 20.1%, and the proportion of individuals between 18-64 was 67.7 (compared to the United States average of 61.9% for the same population group). Population distribution by age groups in DC in 2000 is presented in Figure 22.

The Census Bureau also provides demographic data at several geographic levels, and the tract level data was used for creating and distributing the agent population across the geographic environment of the model. The gender division is fairly even, with females constituting 52.9% and males – 47.1% of the total DC population in 2000, as shown in Figure 23.

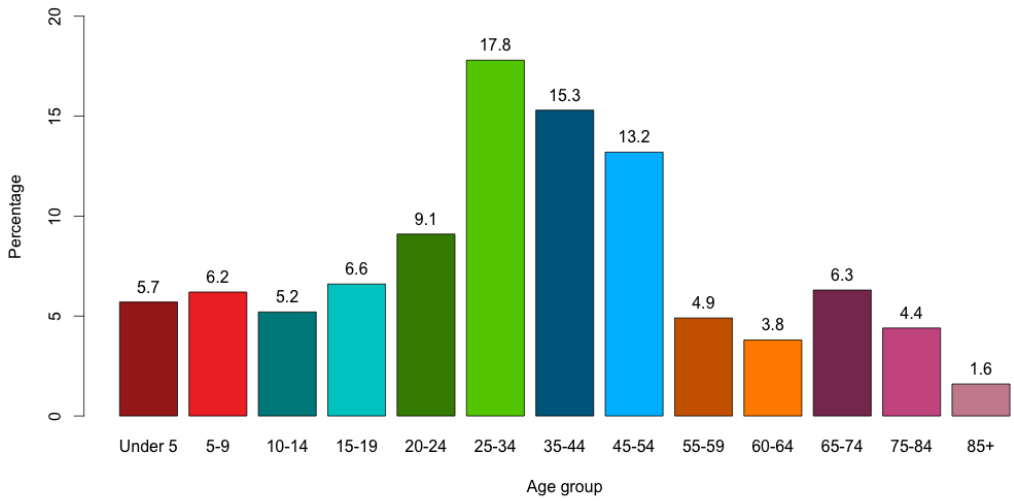


Figure 22: Age groups distribution for 2000 in DC (Source: USCB, 2000)

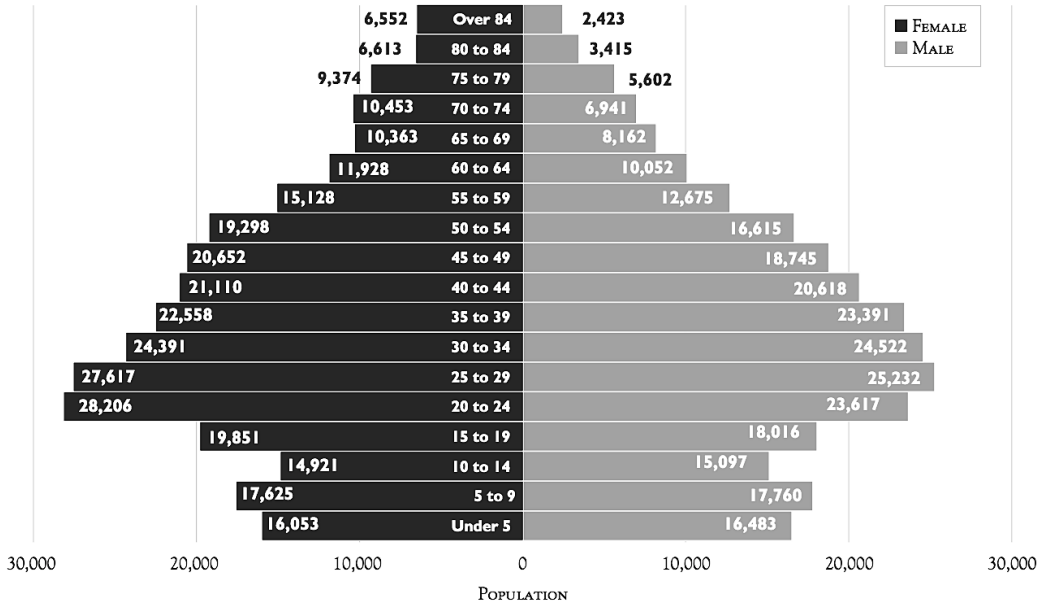


Figure 23: Population by 5-year age groups in 2000: Washington, DC (Source: Singer, 2004)

Washington, DC follows the overall US pattern of women tending to live longer than men. A broad decline of the middle class has been observed in the '90s (Berube et al., 2003) when the district witnessed a population decrease and continued struggles for low-income households. While the older population grew in most states between 1999 and 2000, Washington, DC's population of age 65 and over experienced a decline (Hetzel & Smith, 2001). This may be explained by the earlier migration of the population to retire to the district's suburbs. The spatial pattern of the Washington, D. C. seniors exhibits an interesting pattern: with most population located in the Far Northeast DC, southern Montgomery County and partially Fairfax County, as shown in Figure 24.

Figure 24: Share of population 65 and older, 2000: Washington, DC (Source: Berube et al., 2003)

The risk of developing cancer increases with age (Armitage & Doll, 1954). Aging is considered to be among the most important risk factors for cancer (Jones & Baylin, 2002). This may be associated with the physical changes at the body level that occur as people age, with less active defense and repair mechanisms within the body and a longer overall exposure to risk factors that increase cancer risk. Advancing age also increases the risk for other health conditions besides cancer (Cairns, 1978), and understanding the age profile and age changing patterns of a population is essential for gaining insight into the lifestyle and behavior choices as well as consumption tendencies that put some individuals at higher risk of cancer compared to others.

5.3.2 Race/Ethnicity

Understanding the distribution of race and ethnicity in a given population is of particular importance for its health assessment and for designing effective policy interventions. The population of Washington, D. C. in 2000 was predominantly African American. In 2000, the district had the 4th highest proportion of African Americans among the 23 living cities in the United States (Berube et al., 2003). In 2000, the U. S. Census Bureau allowed people the option to select more than one race for the first time (Jones et al., 2001).

About 60% of the DC population in 2000 was African American, and only 27.8% of the district's population was constituted of Whites, as shown in Figure 25. The proportion of Hispanic and Asian residents was increasing compared to previous years, making the district a "majority minority" (with more than half of the population being non-White or Hispanic).

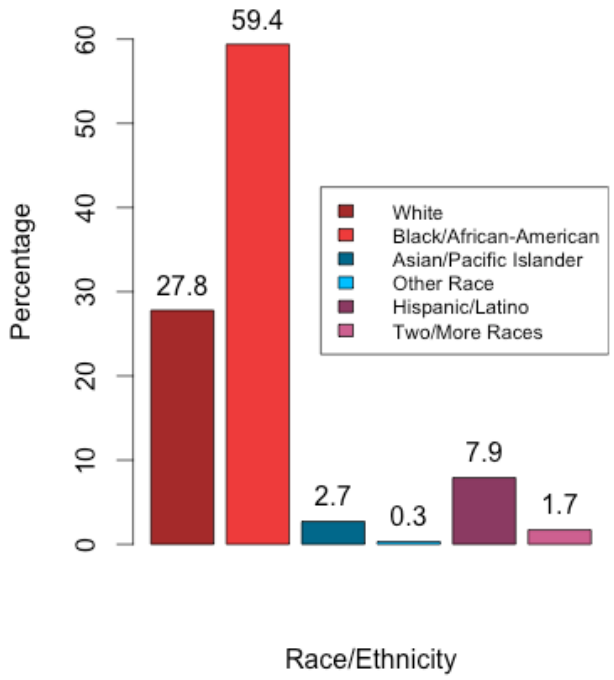


Figure 25: Population proportion by race/ethnicity in Washington, DC, 2000 (Source: USCB, 2000)

The spatial distribution of the DC's ethnic and racial profile is revealing interesting patterns, as shown in Figure 26. For instance, African Americans live primarily on the east of 16th street, while Hispanics are concentrated in Silver Spring, Adams Morgan and Arlington/Fairfax, VA. Racial and ethnic minorities not only suffer disproportionately from certain diseases but also have reduced health care access and overall reduced access to resources.

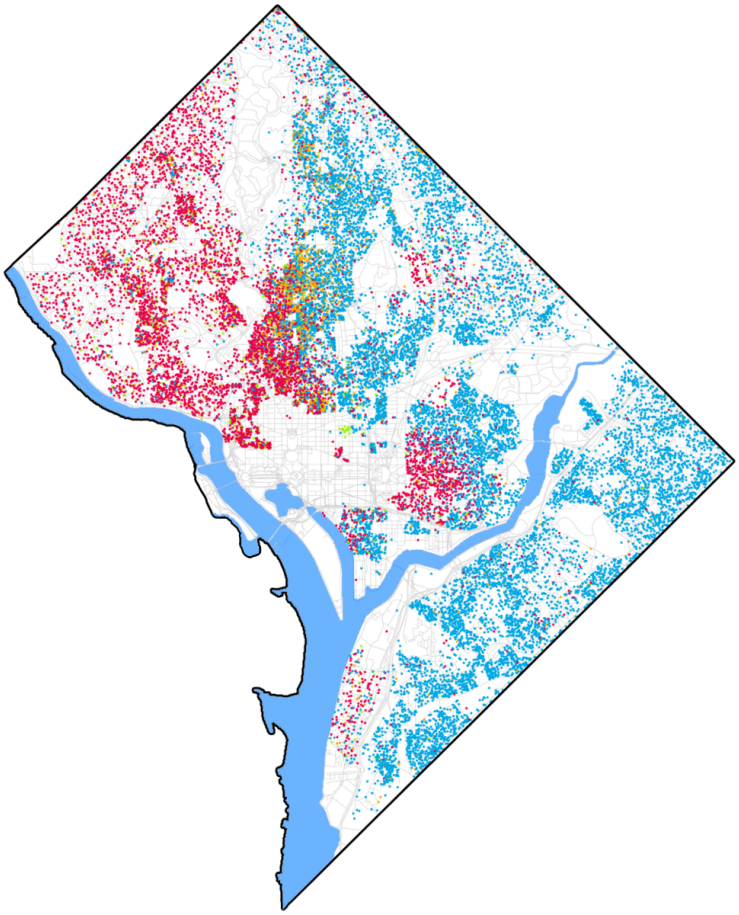


Figure 26: Race and ethnicity Washington, DC profile, 2000 (Source: Eric Fisher, Data Source: Census 2000). Red represents the White population, Blue – the African Americans; Green – Asian; Orange represents Hispanics and Gray refers to the population ethnicity marked as Other in the census. Each dot marks 25 people; created by Eric Fischer

There is high variation regarding cancer incidence and mortality by race. African Americans had the highest cancer incidence rates in the nation (Landis et al., 1999). They were also twice more likely than Whites to report poor health (Lillie-Blanton et al., 2003). Cancer mortality rates among African Americans are about 34% higher compared to Whites (Landis et al., 1999). The root causes of health disparities between the different races are largely unknown, but late stage at diagnosis may be a significant contributor to the survival disparities between Whites and African Americans (Mayberry et al., 1995).

Understanding the racial and ethnic profile of evolving populations is essential for addressing the increasingly diverse needs of the people and to deliver services adjusted for the diverse population groups accordingly.

5.3.3 Marital Status

According to the 2000 U. S. Census Bureau, "nonfamilies" were the dominant household types in the United States, in cities as well as in suburbs. A "nonfamily" consists of people living alone or with non-relatives. The average household size in Washington, D. C. was 2.16 in 2000, slightly decreasing comparing to the 2.26 in 1990. Among the living cities, Washington D. C. had the highest percentage of individuals living alone and the lowest proportion of married couples (Singer, 2004). The rate of marriage in the district was 5.4, and the fraction of divorces was 9.7%. About half of the population 18 and over were never married, 29.9% were married, and nearly 10% were divorced in the district in 2000, as illustrated in Figure 27.

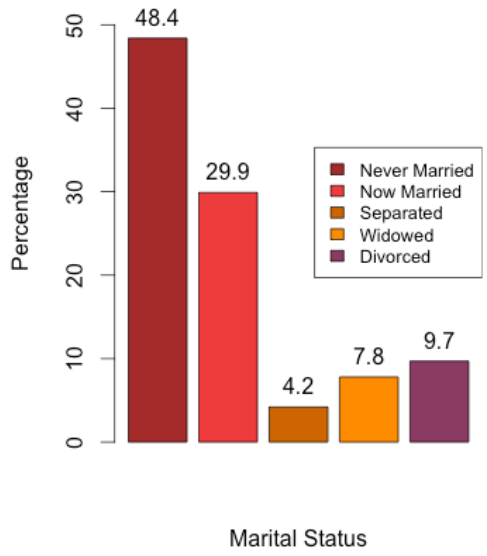


Figure 27: Population proportion by marital status in Washington, DC, 2000 (Source: USCB, 2000)

The “Now Married” category includes all married families, both with the spouse present and those with spouse absent. The “Separated” category refers to the people living without a spouse. Marital status (being married) has been associated with lower incidence of cancer at all sites (Ernster et al., 1979). Married individuals were less likely to present with metastatic cancer (Aizer et al., 2013), and had lower cancer death rates than unmarried patients, even after adjusting for demographics, cancer stage, and treatment received. With the mounting evidence of the role of marital status on health (Verbrugge, 1979), knowledge about the distribution of marital status across the population of interest is essential. Marital status has implications for both health and mortality, with unmarried population reporting poorer health and higher mortality risks compared to married individuals (Robards et al., 2012).

5.3.4 Education

Education has been consistently linked to healthier lives (Ross & Wu, 1995). Several pathways can influence this association. Individuals with lower education levels, for instance, tend to have poorly paid jobs or experience difficulty in maintaining continued unemployed for longer periods. Lower education levels have been associated with increased likelihood to reside in neighborhoods that pose various health risks or work in hazardous conditions (Beydoun, 2008). However, education does not act on individual health and wellbeing in isolation (Feinstein et al., 2006). Other factors, such as income or environmental conditions among others – interact with education to influence health. Education may influence income or occupational choice. Increased levels of education may also lead to different thinking and decision-making patterns regarding individual health (Cutler & Lleras-Muney, 2006).

Educational attainment has been linked to increased mortality rates for all cancers in African American and White populations for both males and females. African American males with 12 or fewer years of education reported more than twice higher prostate cancer death rates compared to African American men who completed more levels of education (Albano et al., 2007).

According to the 2000 Census, Washington, D. C. is one of the most educated areas in the United States. In 2000, most district residents held a bachelor's degree (Singer, 2004). About 20.6% of the DC residents were high school graduates and around 15.4% had some college experience but no official degree, as shown in Figure 28. According to the 2000 U. S. Census Bureau (2000), about 39.1% of the population of age 25 or over held a bachelor's degree or higher in 2000 in DC. The substantial proportion corresponds to the relatively large number of residents employed in the government. However, the district was behind the national average in regard to the percentage of adults with a high school diploma (Berube et al., 2003). While Washington, D. C. is considered one of the most educated areas in the United States, inequalities persist between the different races and ethnicities in the district.

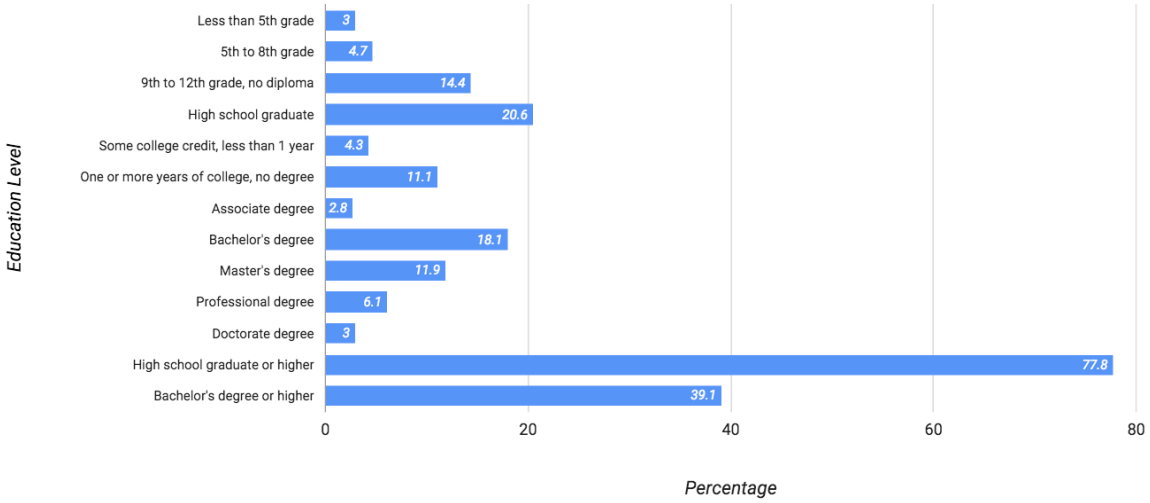


Figure 28: Proportion of population of age 25 and over by education level in DC, 2000 (Source: USCB, 2000)

Addressing the education gap requires new approaches that would account for the complex interplay among the several factors that interact in non-obvious ways to influence the relationship between education and health. Also, policies should be put in place to address the impact of education on overall population health. Focusing on education is particularly important if the goal is to decrease the health disparities predominant in the district, and positively influence the health the DC residents.

Over 80% of the White adults in DC held a bachelor's degree in 2000, compared to only 17% of the African-Americans, as shown Figure 29. Education and race are part of a larger and more complex network of socioeconomic conditions that influence health. Health interventions need to address the differences in education among the diverse racial and ethnic groups in the district.

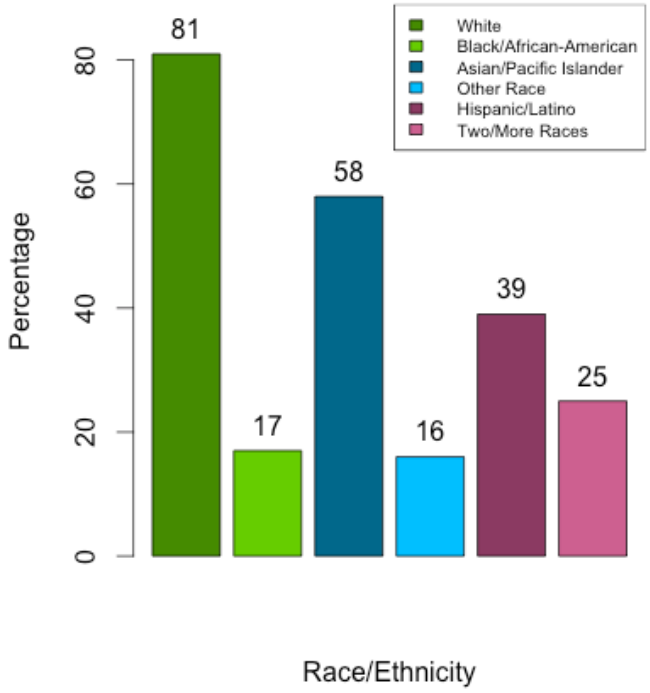


Figure 29: Proportion of population of age 25 and over with at least a bachelor's degree by race/ethnicity, 2000 (Source: USCB, 2000)

5.3.5 Income and Poverty

Income levels represent the ability of individuals to provide for themselves and their families. The income per capita in DC was $28,659, compared to the United States income per capita average of $21,587 in 2000 (for the year 1999). Per capita income for 1999 is presented in Figure 30. While DC is one of the areas with the highest income per capita, it is striking that the district also had the highest poverty rates in the United States – about 20% of the DC residents were living in poverty in 2000, that would account for 109,500 people.

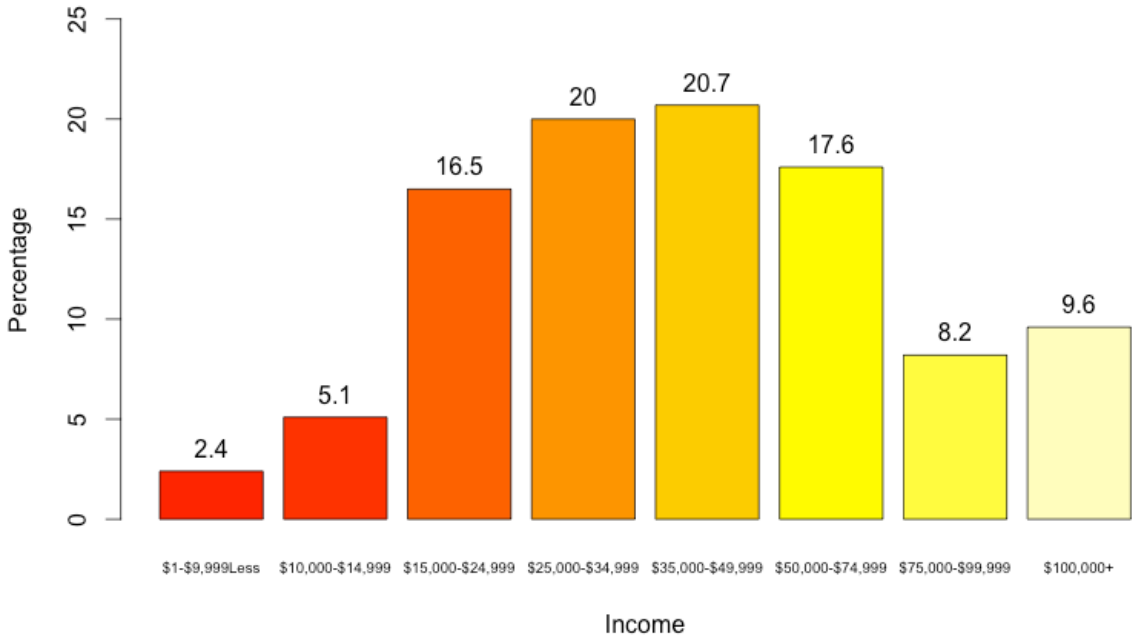


Figure 30: Per-capita income for 1999 (Source: USCB, 2000)

Significant disparities can be observed regarding income as well. The annual average earnings for a White household was $67,000, while the average annual income for a Hispanic household was about $36,000. The disparities become more prominent when looking at the African American households, which annually earned about $30,000 on average in 2000, as shown in Figure 31. A large proportion of the district's minority households struggled financially.

The influence of income on health has widely been documented (Marmot, 2002). Its role on health becomes increasingly important when income is linked to physical deprivation and restricting the opportunities for social participation. Usually, income does not influence health in isolation. In combination with race or ethnicity, the effect of income on health may be especially damaging.

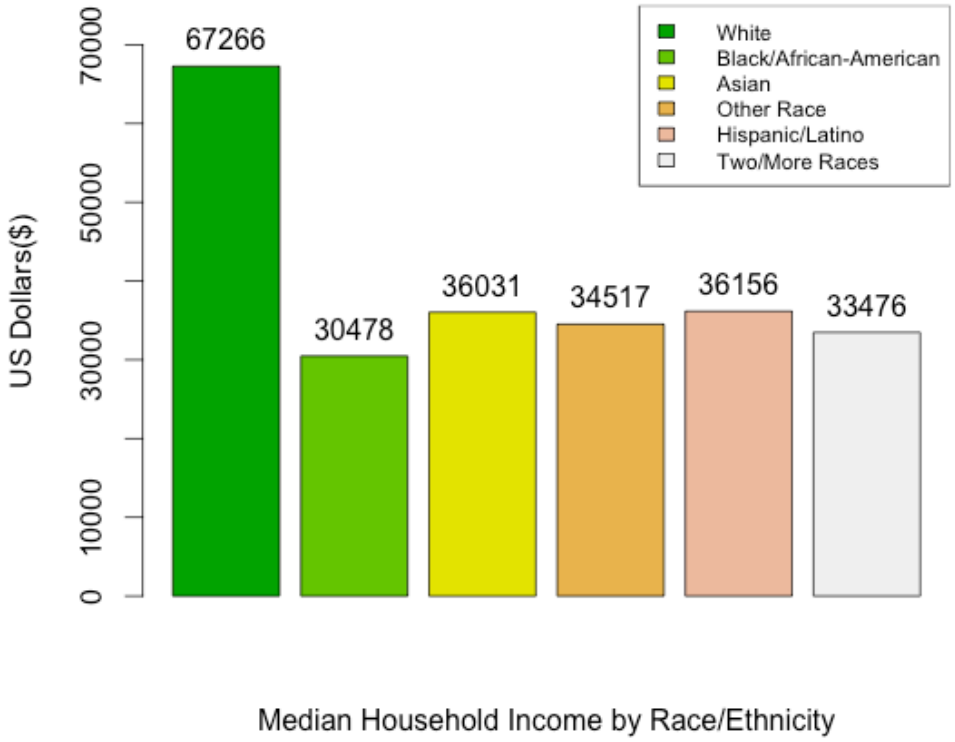


Figure 31: Median household income by race/ethnicity in DC, 2000 (Source: USCB, 2000)

Poverty is unequally distributed geographically, as shown in Figure 32. Most neighborhoods with high poverty levels in the DC area are highly concentrated towards the Southeast DC, which corresponds to the geographic area where African Americans predominantly reside. It is common for education levels to modulate individual income (Cutler & Lleras-Muney, 2006). Higher educational levels correlate with higher individual income levels that in turn affect health. Poverty also varies by census wards, with Ward 8 having the highest poverty rates and Ward 3 having the lowest poverty rates in DC in 2000 (Chandra et al., 2013).

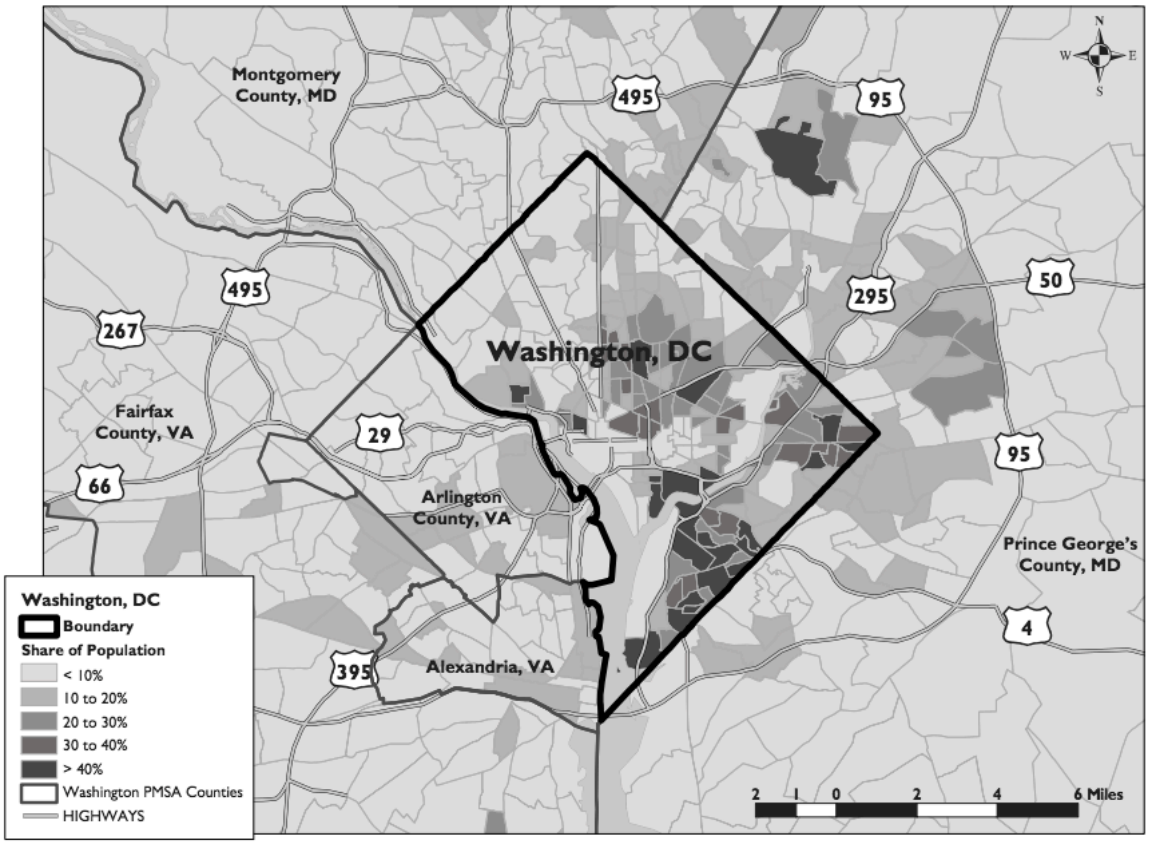


Figure 32: Proportion of individuals living in poverty in DC, 2000 (Source: Berube et al., 2003)

Disparities have been observed regarding income per capita and health status among world nations (United Nations, 1988). Cancer incidence rates vary depending on the income and socioeconomic status across nations as well (Parkin & Muir, 1992). There are several pathways by which income may influence health. Lower income levels may affect cancer incidence and mortality rates by several barriers such as the lack of obtaining and maintaining health insurance coverage, difficulties in transportation to hospitals or transportation to various sites for cancer-related screening or treatment procedures (O'Malley et al., 2004; Beydoun, 2008).

5.3.6 Employment

Being employed is a good indicator of the residents' ability to obtain health insurance and have access to quality health care. The vast majority of the individuals obtain health insurance when they get employed and lose it when they become unemployed (Chandra et al., 2013). Unemployment has been linked to increased stress levels and various adverse health conditions. It has also been linked to poor nutrition, poor living conditions and other factors that affect the health of the individuals.

The unemployment rate in Washington, D. C. was about 11% in 2000. While the proportion of district residents active in the labor force was similar to the average across the nation in 2000, the unemployed rate was lower than the national average (Berube et al., 2003). About 63.6% of the DC residents of age 16 or older were in the labor force in 2000, as shown in Table 7.

Table 7: Employment status of residents of age 16 and over, DC 2000 (Source: USCB, 2000)

Subject	Number	Percent
EMPLOYMENT STATUS		
Population 16 years and over	**469,041**	**100.0**
In labor force	298,225	63.6
Civilian labor force	294,952	62.9
Employed	263,108	56.1
Unemployed	31,844	6.8
Percent of civilian labor force	10.8	(X)
Armed Forces	3,273	0.7
Not in labor force	170,816	36.4

Furthermore, there were differences regarding unemployment between wards, as shown in Figure 33. Ward 8 accounted for the highest unemployment rates in DC, and

Ward 4 for lowest unemployment rates in 2000 – with only 6.6% of the ward population reporting as being employed.

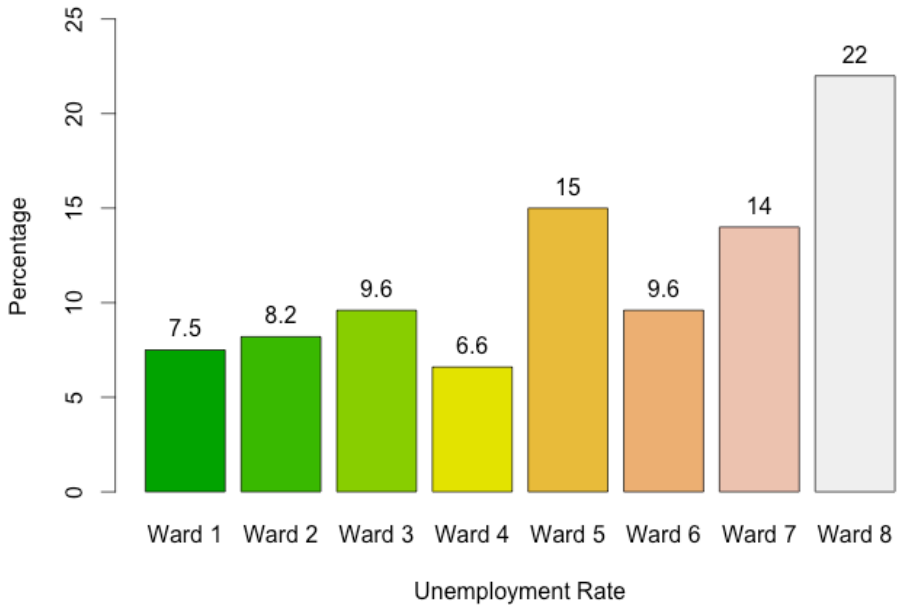


Figure 33: Unemployment rates by wards in DC, 2000 (Source: USCB, 2000)

The association between unemployment and adverse health outcomes has been well established. While unemployment at the population level may lead to recession and/or economic decline (Brenner & Mooney, 1983), at the individual level it can result in increased stress causing a series of health conditions, and limiting access to health care, with most people obtain their health insurance through an employer. As most DC residents obtained their health insurance coverage through their employers (see Section 5.4.6), maintaining continuous employment is of particular importance. Employment also plays a role in a cancer diagnosis. About one-fifth of the patients previously employed experienced difficulties in finding job opportunities because of their cancer diagnosis

(Hewitt et al., 1999). Moreover, those who have already had a job experienced a phenomenon called "job lock" – being unable to change jobs due to fear of losing health insurance coverage (Hewitt et al., 1999).

5.3.6 Occupations

The influence of occupation and workspace on health and cancer, in particular, can be traced back to Percival Pott, who noticed an increased incidence of scrotal cancer in chimney sweeps (Hill, 1965). Workspace characteristics (such as job strain, night shifts, and even job insecurity) also influence health via stress and immune modulation pathways (Ganster & Schaubroeck, 1991).

Management, professional, and occupations accounted for 51.1% of the workers, as most of the employees are federal workers as well as other occupations supporting those fields. Educational, health, and social services accounted for 18% and the public administration jobs for 15%. In 2000, Washington, DC had the lowest percentage of agriculture, production and transportation workers.

In 2000, professional, scientific, management, administrative accounted for 18.8% of the employed civilian population 16 years and over in the district, as shown in Figure 34.

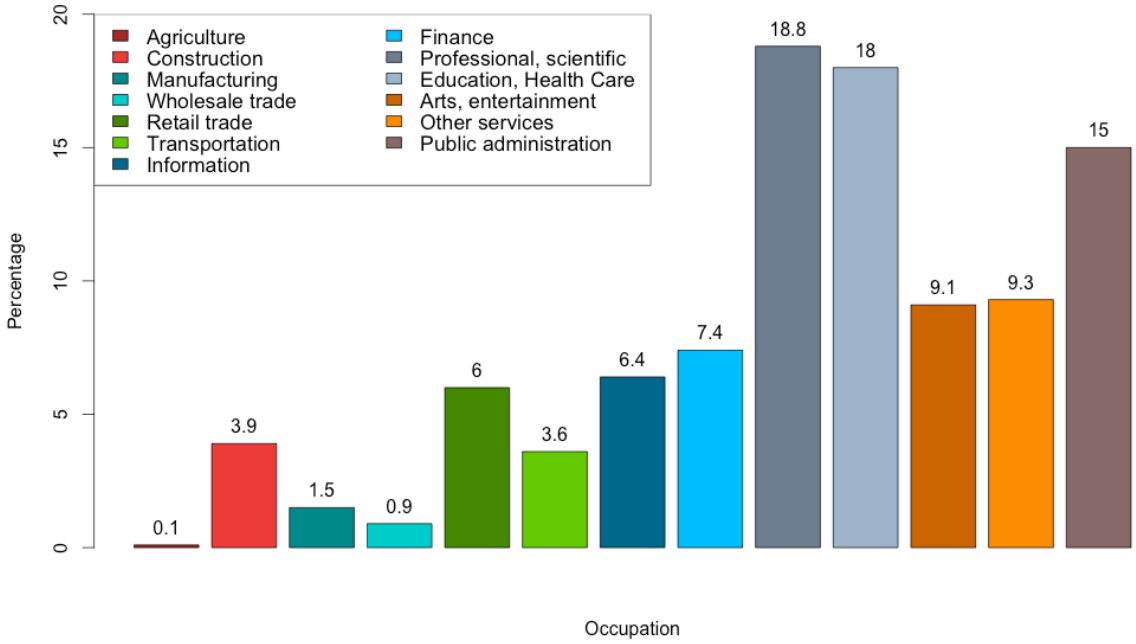


Figure 34: Proportion of employed civilian population 16 years and over by occupation in DC, 2000 (Source: USCB, 2000)

Occupation related physical activity was shown to decrease cancer risk, by increasing the influence of protective factors that modulate cancer risk (Samad et al., 2005). By contrast, sedentary occupations are correlated with increased risk of cancer (Gerhardsson et al., 1986). According to Garabrant et al. (1984), colon cancer risk in men with sedentary jobs was shown to be over 1.6 times greater compared to their counterparts. Subsequently, the risk increased as physical activity levels decreased.

5.4 Health Characteristics of DC Residents

The demographic and social characteristics of a population largely determine the overall health of its residents. Among the pathways by which socio-demographic characteristics may affect health are: shaping individual decision-making and health behaviors, influencing access to resources and contextualizing risk factors. Socio-demographic

characteristics interact with other determinants to affect not only individual health but also the health communities and of the population as a whole. This section builds on the characteristics of DC residents provided earlier in this chapter to provide an overview of the health-related characteristics of the district population in 2000.

Any death that occurs before the national life expectancy (that is 76.9 years) is considered premature. Heart disease and cancer were the leading causes of premature death among the DC residents in 2000, as shown in Figure 35. About half of the deaths can be attributed to chronic conditions, such as heart disease, diabetes, stroke, and cancer. Heart disease was the leading cause of death both in DC and nationally. About one in five district residents died of cancer in 2000.

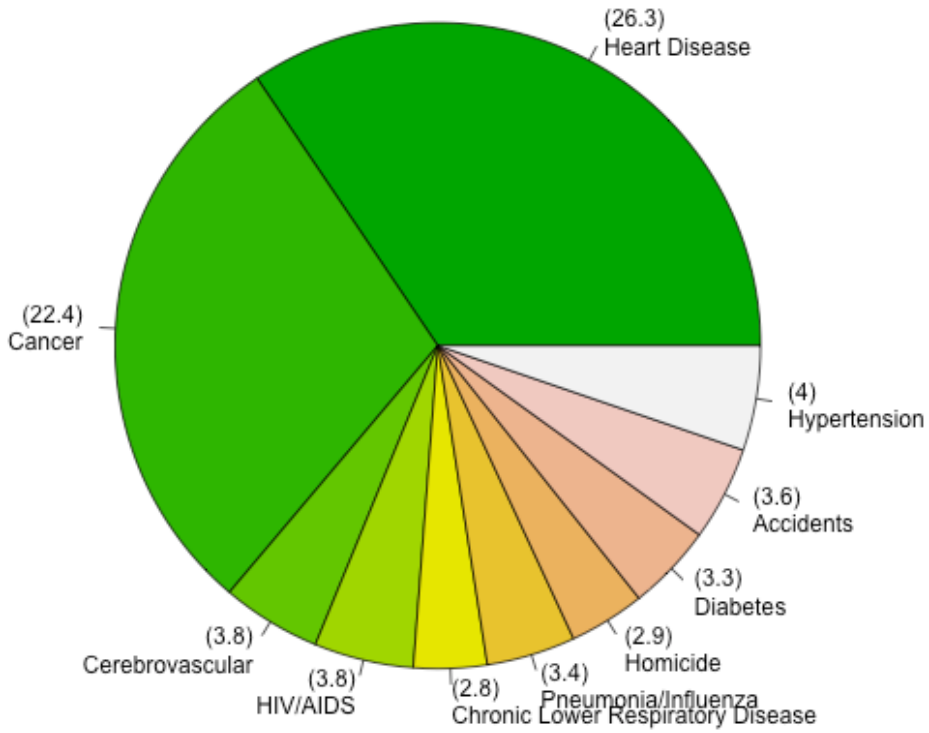


Figure 35: Top ten leading causes of death in DC, 2000 (Source: DC DOH, 2002)

5.4.1 Life Expectancy

Life expectancy is a measure of longevity, consisting of the average number of years a person has left to live, given that he/she has survived to that age, based on the age-specific death rates. DC residents are on average less healthy than the U.S. population as a whole (Lillie-Blanton, 2003). In 2000, the average life expectancy in the district was 72.6 years, compared to a United States average life expectancy of 76.9 years, making it the places with the lowest life expectancy in the nation (Wei et al., 2012). There were evident disparities in the life expectancy at birth not just among men and women, but also among the various races/ethnicities in DC, as shown in Figure 36 and Figure 37. On average, females tend to live longer than males and Whites tend to live longer than African Americans both in the United States and the district in particular.

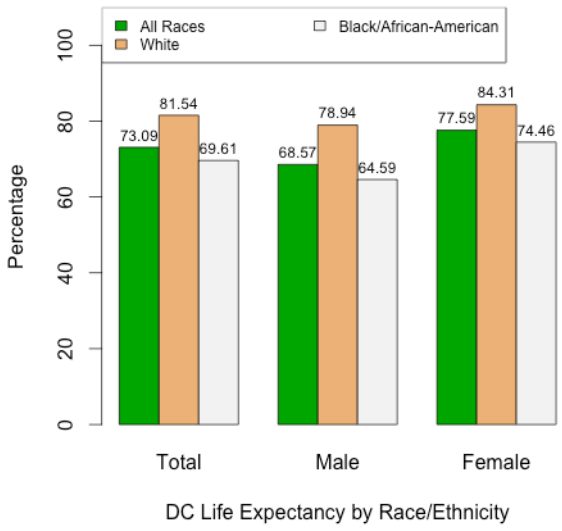


Figure 36: Life expectancy at birth, by race and sex in DC between 1999–2001

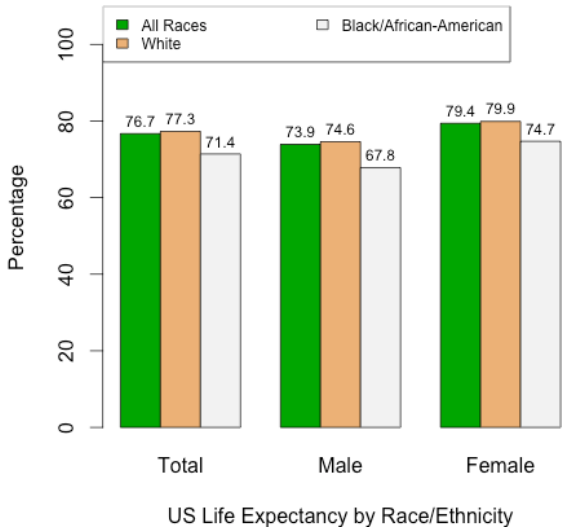


Figure 37: Life expectancy at birth, by race and sex in U.S. between 1999-2000

Life expectancy at birth is an important indicator of the fundamental health of a community. As mortality rates decrease, the increase in life expectancy over time may indicate improvements in the health status of the society as a whole.

5.4.2 Overweight and Obesity

Overweight and obesity are measures attributed to the weight of an adult that is greater than what is generally accepted as "normal" or healthy for a given individual height. They have been shown to increase the risk for a broad range of health conditions. Overweight and obesity are determined using weight and height of an adult, calculating the "body mass index" (BMI), which correlates to the amount of body fat. A person with a BMI is within the range of 25 and 29.9 is considered to be overweight, and a person with a BMI of 30 or higher is considered obese.

In the past decades, United States experienced an obesity epidemic (Mokdad et al., 1999). According to CDC BRFSS (2000) data for 2000, the overall prevalence of obesity in DC (21.5%) was higher compared to the nationwide prevalence (20.1%). The prevalence of overweight population in DC was 31.7% in 2000. Overweight levels were higher among males compared to females. Both overweight and obesity were more prevalent in African Americans, as shown in Figure 38 and Figure 39.

According to the BRFSS data, 31.7% of the DC residents were overweight, 21.5% of the residents were obese, and 46.8% reported as being neither overweight nor obese. Overweight and obesity are influenced by the ability of individuals to maintain a healthy diet. BRFSS collects data on the adults who consumed fruits and vegetables five or more times a day. In 2000, about 68.1% of the DC residents reported to consume less than the recommended five servings of fruits and vegetables per day.

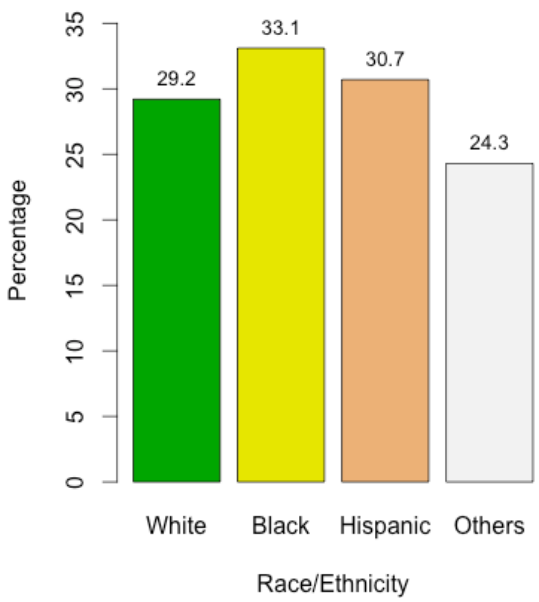


Figure 38: Overweight prevalence by race, DC 2000

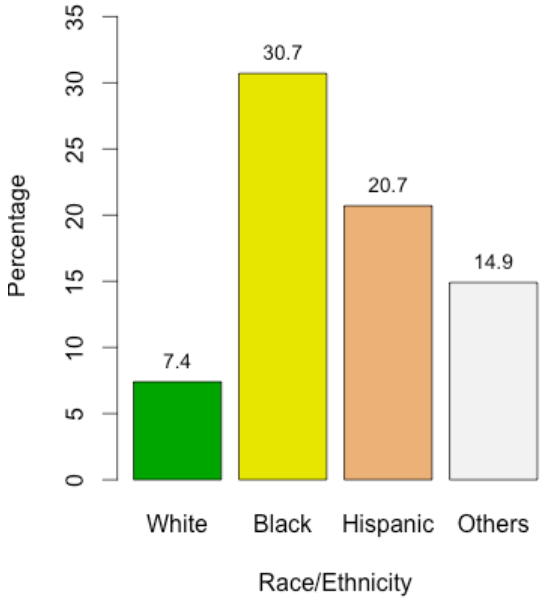


Figure 39: Obesity prevalence by race, DC 2000

5.4.3 Physical Activity

Exercise has been shown to considerably improve the quality of life if practiced regularly (Rejeski et al., 1996). Physically active people have a 25% lower risk of developing CRC comparing to the less physically active individuals (Boyle et al., 2012). Moreover, in the context of colon cancer, physical activity was shown to improve survival and reduce the risk of recurrence (Meyerhardt et al., 2006). Among the mechanisms linking physical activity to cancer are: decreasing sex and metabolic hormones, reducing the levels of inflammation, and improving the overall immune function (McTiernan, 2008). There is an increase in the body of knowledge associating increased levels of physical activity with improved health and decreased the risk of several types of cancer as well as other diseases (McTiernan et al., 1998).

Current recommendations indicate that adults should engage in at least 30 minutes of daily exercise (US DHHS, 1996). Persistent physical activity has been shown to be beneficial to health, reducing mortality and the risk of several health conditions such as cardiovascular disease, colorectal cancer, diabetes, osteoporosis, obesity and depression (McTiernan et al., 1998). In 2000, the levels of physical activity in the district were low. According to the BRFSS data, 20.8% of the DC residents reported "no physical activity" compared to a United States average of 26.9%. Disparities regarding the prevalence of physical activity in DC were observed. White people were more likely to be physically active compared to other ethnicities, as shown in Figure 40. Additionally, there were gender disparities in the district regarding physical activity, with females reporting higher levels of physical activity compared to men.

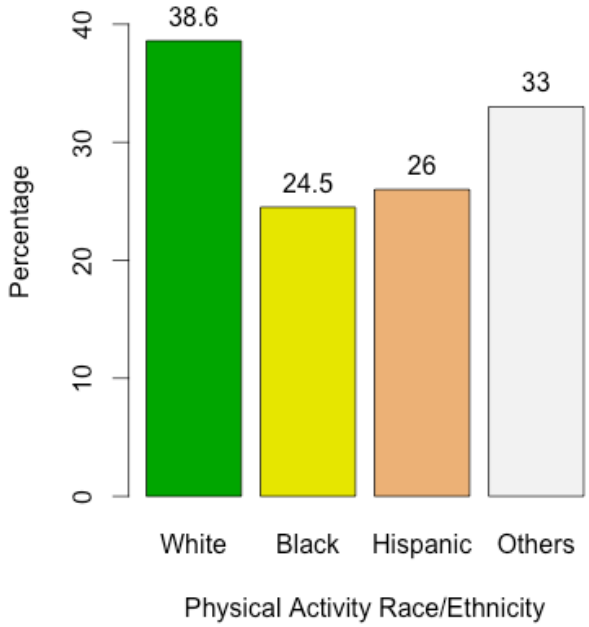


Figure 40: Physical activity prevalence by race in DC, 2000

Physical inactivity, combined with excess weight, significantly increases the risk of both morbidity and mortality. Policies are needed to address this interaction and to provide opportunities for individuals to be more physically active and to follow healthy diets. This is important not just for weight control, but also for the overall health and well being of the residents.

5.4.4 Smoking/Tobacco Use

Smoking is one of the most preventable causes of death in the United States (Danaei et al., 2009). Apart from being the leading cause of lung cancer, smoking increases the risk for other diseases such as other types of cancer, heart disease, and respiratory illnesses. Particularly, women who smoke tobacco are at increased risk for infertility, spontaneous abortions and various conception problems (Cnattingius, 2004). In nation's capital, the prevalence of smoking was lower compared to the national average in 2000 (Human Services, 2000). However, the proportion of residents who smoked tobacco varied by race and ethnicity, with a higher prevalence of daily smokers among African Americans, as shown in Figure 41.

According to the BRFSS, only a proportion of 20.4% of the DC residents reported to be current smokers in 2000. Smoking was less prevalent in adults 65 years of age and over. This population group, however – was shown to have smoked previously. In 2000, males were more likely than females to be current smokers, and the age group with most smokers was between 45 and 55 years. Smoking is a major public health problem that also affects individuals exposed to second-hand smoke.

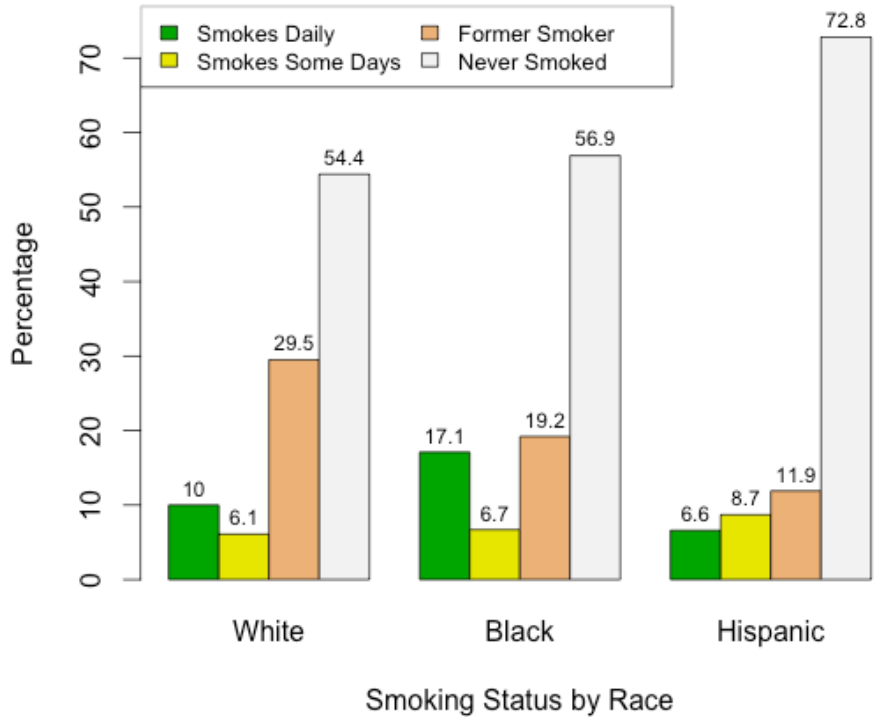


Figure 41: Smoking frequency by race/ethnicity for adults 18 and over in DC, 2000

5.4.5 Alcohol Use

The damaging effect of alcohol use on human health is a global problem, which results in millions of preventable deaths and thousands of years of life lost, contributing to the leading causes of about 200 health conditions (WHO, 2014). Alcohol abuse not only contributes to many diseases but also influences social life, negatively impacting individuals, families and communities. When used in moderation, alcohol has little to no adverse effects on health. Heavy and binge drinking, however, contribute to serious health problems both in the short and in the long run. Heavy alcohol drinking contributes to several health conditions such as liver problems, violence, colorectal cancer, oral and

throat cancer, pneumonia as well as anxiety, depression, violence and alcohol dependence (Baan et al., 2007).

While smoking in DC was below national average, binge and heavy drinking were more common in DC compared to the United States average. The distribution of alcohol use across age groups is presented in Figure 42. While alcohol consumption is more predominant in adults 18 and over, adolescents have also shown to consume alcohol in the past month, year and during their lifetime as well. Data on alcohol/tobacco use in BRFSS is not available for the year 2000. Therefore, the 2001 data will be reported.

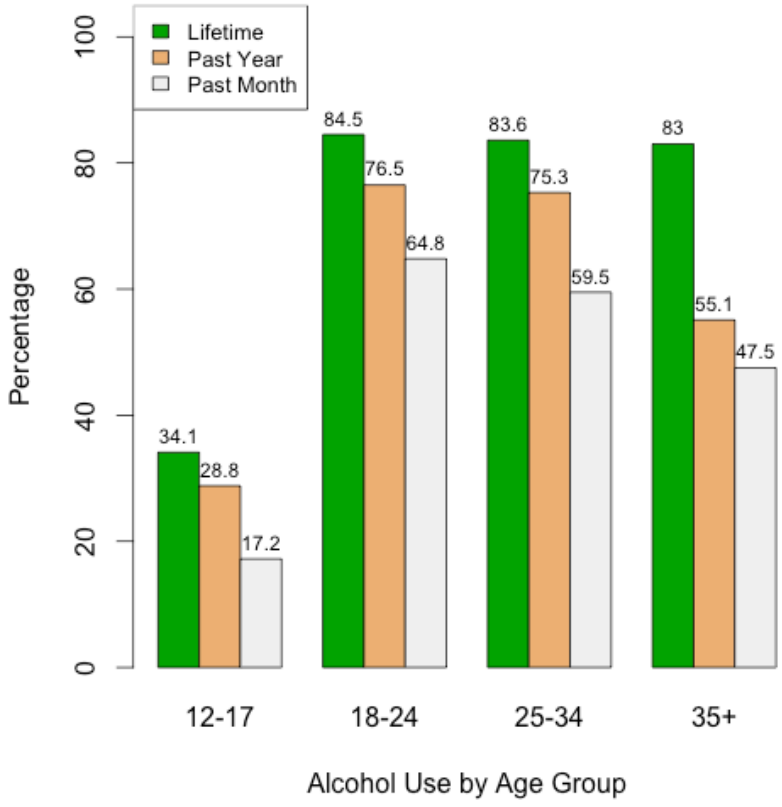


Figure 42: Alcohol use by age group, Washington DC, 2000 (Source: DC Department of Health, 2001)

As it is the case with smoking, physical activity, and overweight, racial disparities are also observed regarding alcohol consumption. According to the 2000 CDC BRFSS (2000), the proportion of White adults who consumed at least one alcoholic beverage in the past 30 days was 83.2%, compared to the percentage of African American district residents who was just 45.9% in 2000.

Females were more likely to be heavy drinkers than males. Heavy drinking is defined differently for men and women: males being considered heavy drinkers if they had more than 15 or more drinks per week while women are considered as "heavy drinkers" if they consumed more than eight alcoholic beverages a week. Adults having five or more drinks on one occasion are considered binge drinkers. According to the CDC BRFSS (2000) data, White residents were twice more likely to be binge drinkers comparing to African Americans, with a consequent proportion of 22% and 8.4% correspondingly. The percentage of Hispanic residents who were binge drinkers in 2001 was also relatively high – 18.1%. Given the toll the alcohol abuse imposes on individuals, families and communities it is important to tailor policies that address the short-term as well as the long-term effects of excessive alcohol consumption.

5.4.6 Health Insurance/ Access to Heath Care

Having access to health care services is considered vital for good health at both individual and community level (Lewin-Epstein, 1991). While health insurance rates in DC were higher compared to the national average, the disparities regarding health insurance coverage by race and terms of access to health services were prominent. Washington, DC has high employer-sponsored health insurance coverage, as most of the DC residents are educated professionals, as shown in Figure 43.

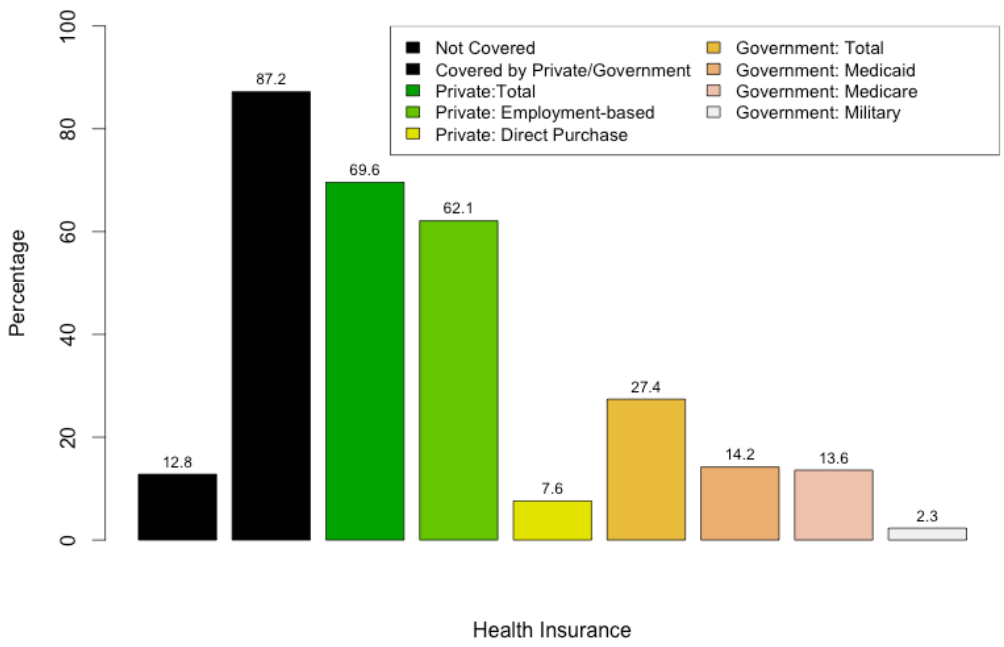


Figure 43: Health insurance coverage status and type of coverage, DC 2000

The proportion of the DC population covered either by private or by government insurance was 87.2% in 2000. The total percentage of DC residents holding private health insurance was 69.6%, and the proportion of individuals holding government health insurance in 2000 constituted 27.4%.

While Washington, D. C. was one the districts with the highest levels of health care coverage in the United States, not all residents benefited from this advantage equally. The proportion of African Americans in the district with any health insurance coverage was 85.8%, compared to the 96.5% for Whites and just 70% for Hispanics, as shown in Figure 44.

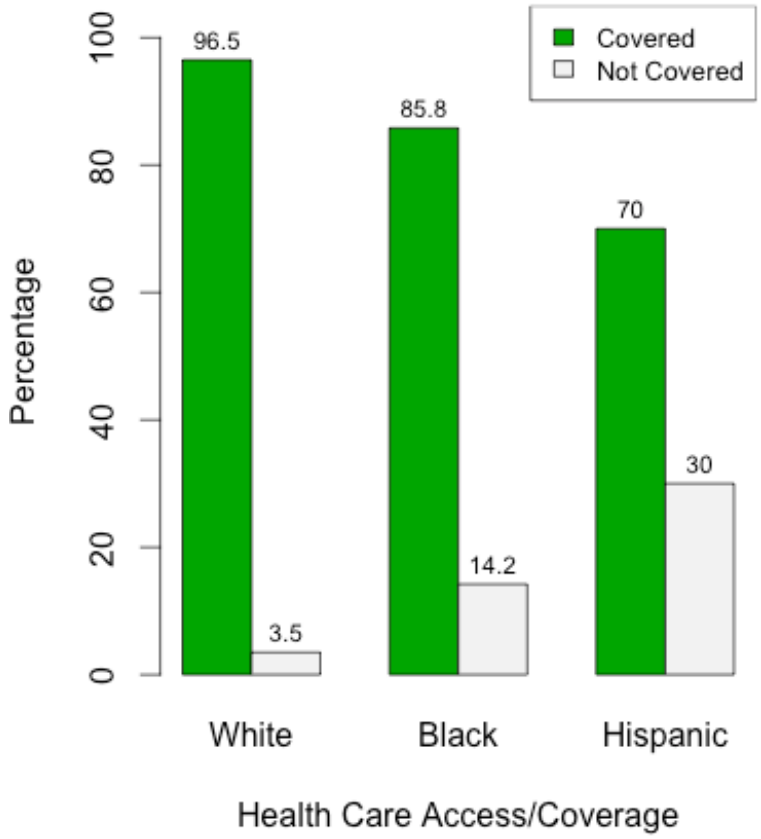


Figure 44: Proportion of individuals with any type of health care access/coverage, DC 2000 (Source: CDC BRFSS, 2000)

As the United States 2000 Census revealed high residential segregation for African Americans in DC, health problems, and reduced health insurance access is more evident in the inner city. The lack of health insurance decreases the likelihood for individuals to have a regular source of health care, either because it is unaffordable or not available in their immediate neighborhood (Diaz, 2014).

There are a few problems with health care accessibility showcasing the inequality predominant in DC regarding heath care access. DC manifests disparities concerning the spatial accessibility of primary care services. The more prosperous areas of the district have more than three times the number of Primary Care Physicians (PCPs) compared to

the neighborhoods living in poverty (Ormond, 1999). The shortage of PCPs in poorer neighborhoods may be attributed to multiple disincentives for them to work in economically disadvantaged neighborhoods with more residents relying on Medicaid, with higher poverty and violence rates and higher frequency of missed appointment and non-compliance (Andrulis & Carrier, 1999). Limited health care access to primary health care services for African Americans living in areas with concentrated poverty may negatively influence timely access to medical assistance when it is needed, delays in diagnosis and treatment and overall poorer health. Access to care and health insurance availability is also linked to financial factors, such as individual and household income, as discussed earlier in this chapter.

5.5 Health Disparities

Health disparities are defined as the inequalities that emerge when certain population groups do not benefit from similar health advantages as members of other population groups. They may refer to inequalities in the distribution of the determinants of health among different population groups, and are most commonly observed among racial and ethnic minorities, females and the elderly. Accounting for the disparities within a community is essential for promoting health behaviors and implementing prevention policies to reduce the overall burden of disease.

The health disparities showed up in this chapter for most health characteristics of the district residents. While disparities within the health characteristics of the district population are consistent with the ones at the national level, the racial and ethnic disparities in DC expose a gap different from the rest of the nation (Lillie-Blanton, 2003). Even though Washington DC residents are more likely to consume the recommended five

servings of fruits and vegetables per day comparing to the national average, according to BRFSS data, African Americans and Hispanics are less likely to follow these guidelines comparing to Whites (CDC, 2000).

Across the United States, African-Americans are two times more likely than Whites to report fair or poor health (Lillie-Blanton et al., 2003). In DC, they were four times more likely to report fair or poor health compared to Whites. Also, African-Americans were several times more likely than Whites to report a disability or a chronic health condition that is limiting their activities (Whitson et al., 2011). The health disparities among the different races/ethnicities regarding cancer incidence and mortality, in particular, may be partly caused by the disparities in the social and demographic characteristics of population groups (Freeman, 2004; Krieger, 2005).

Low income is a risk factor for a broad range of chronic conditions. DC residents with lower income reported poorer health than higher income residents. Gender disparities are reflected in life expectancy between men and women, with women living longer than males, which is consistent with the overall United States trends. Moreover, disparities can be observed among men and females seeking health for the same health condition, with women being more frequent health-care inquirers (Nelson, 2002).

African Americans benefit less from health care access both regarding geographic proximity, affordability, and accessibility. They also have higher obesity rates, are less likely to exercise or consume the daily recommended servings of fruits and vegetables. The causes of health disparities greatly vary, and may include cultural differences, genetic predispositions, health-related behaviors or the environment. From 2003 to 2006,

the joined expenses on health disparities and early mortality in the United States constituted $1.24 trillion (Murray et al., 2006). Health disparities continue to affect the population health and the health care system as a whole.

5.6 Colorectal Cancer in DC

This section will present the CRC status in DC, presenting the incidence rates of this cancer site by race and gender, as well as the overall CRC mortality trends. The aggregated data on the burden of colorectal cancer in DC reported in this section, as well as the incidence and mortality data was obtained from the DC Department of Health (2000). In this chapter, I also discuss the coverage of screening for CRC in the district and the disparities across the CRC continuum. The data on CRC screening rates in the DC (particularly on FOBT, colonoscopy and sigmoidoscopy) was obtained from the CDC BRFSS (2002) data.

5.6.1 CRC Status in DC

Colorectal cancer is the second leading cause of cancer deaths in the United States (Jemal et al., 2004). In 2000, cancer was one of the leading causes of death in DC, and in the nation. Between 2000 and 2004 DC had among the highest rates of cancer deaths in the United States, ranking third highest in CRC deaths in the United States (ACS, 2008), as shown in Figure 45. The incidence of CRC decreased for men and women of all races in DC (et al., 2015) between 2000 and 2009, with a notable exception – a sharp spike in CRC incidence in 2002.

However, the prevalence of CRC in DC remains high. The cancer-related services available to DC residents are either not accessible or not affordable for the individuals without health insurance (or underinsured) or below the poverty line. To get an overall

picture of CRC status in DC, it would be helpful to address the multiple aspects of the continuum of cancer prevention, treatment, and outcomes as well as the disparities among the factors that contribute to the different stages of the continuum.

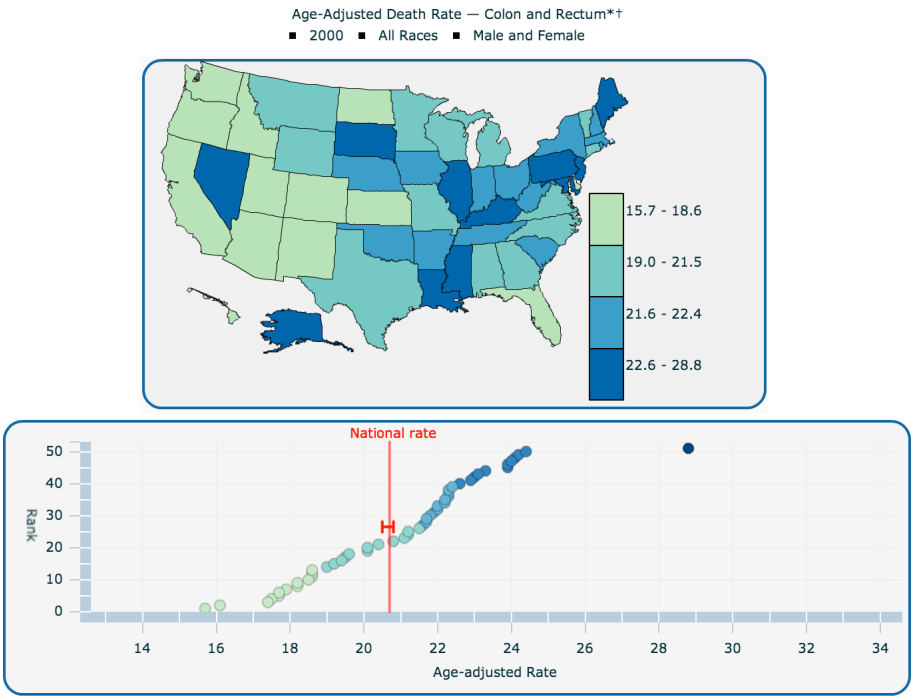


Figure 45: National Vital Statistics System Data on age-adjusted CRC death rates (Source: U.S. Cancer Statistics Working Group, 2015) *Rates are per 100,000 persons and are age-adjusted to the 2000 U.S. standard population

The median age for CRC diagnosis is 71 years for both sexes, with 70 years for males and 73 for females (Ries et al., 2007). There are disparities in terms of race/ethnicity for median age at diagnosis, with Whites having an average age at diagnosis of 72 years, and African Americans 67 years.

Cancer risk increases with advancing age, and as mentioned in Section 5.3.1 age is among the most significant risk factor for most cancer types. CRC is more predominant

among the elderly population. As shown in Figure 46, the risk of developing CRC increases as the population ages.

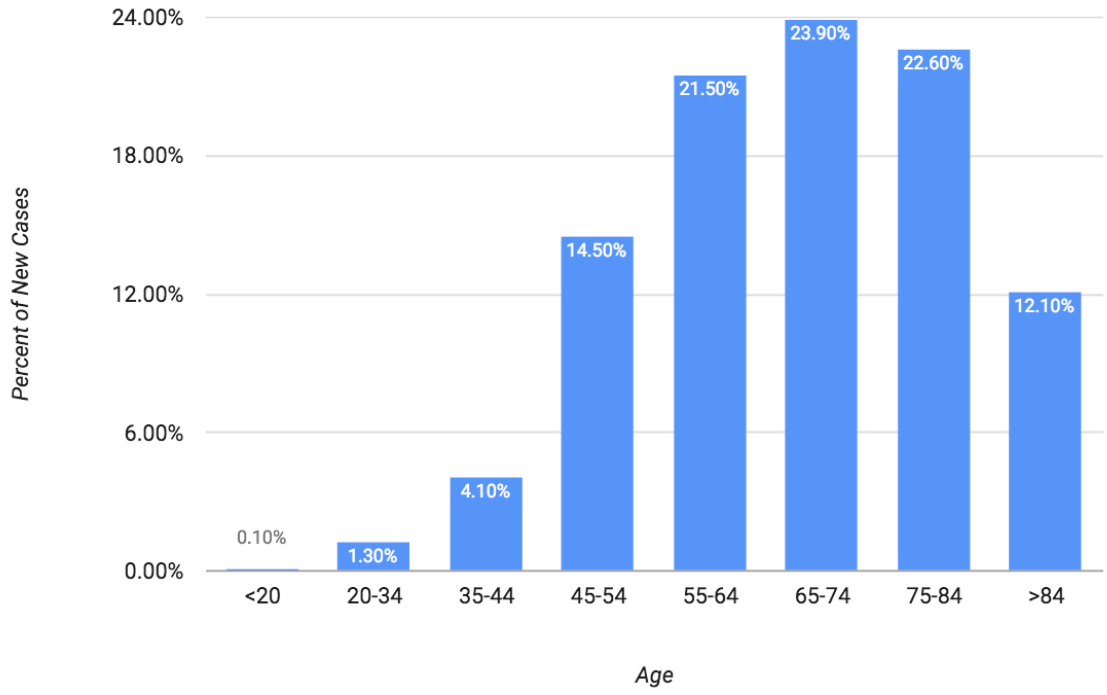


Figure 46: CRC incidence by age groups for all races, both sexes (Source: SEER, 2016)

The age-adjusted overall CRC incidence and mortality rates between 2000 and 2004 are presented in Figure 47. The spike observed in 2002 has been attributed to federal changes in insurance coverage (Chatterjee et al., 2015).

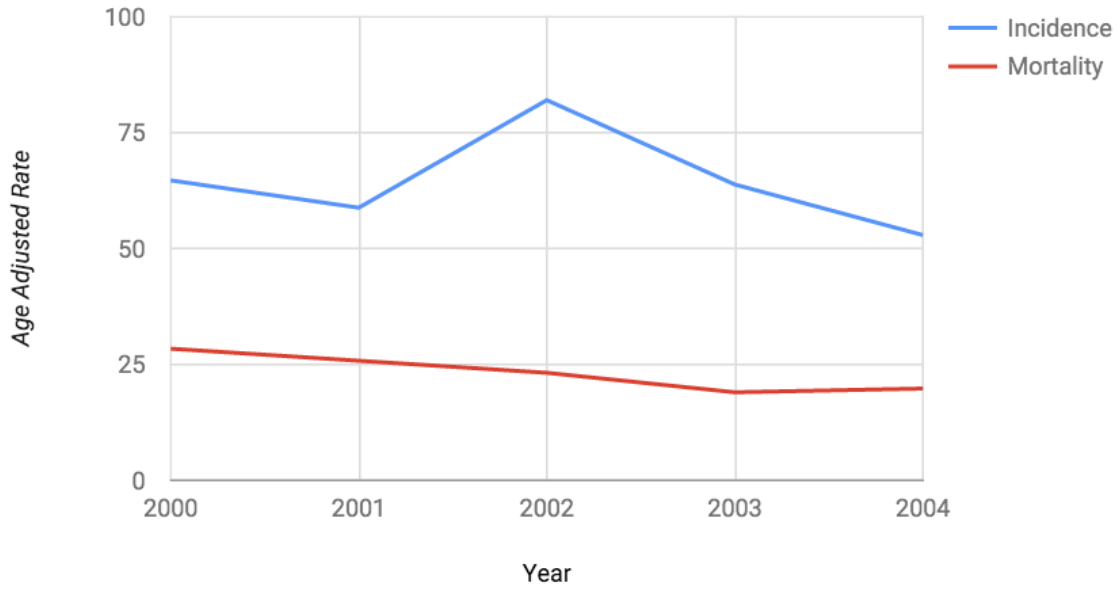


Figure 47: Age-adjusted overall CRC incidence and mortality rates in Washington, DC between 2000-2004 (Source: DC Department of Health, 2000b)

In particular, the spike in incidence observed in 2002 is associated with DC mandating healthcare plans to cover CRC screening procedures. A positive association has been observed between state mandates and the use of recent CRC screening tests for individuals covered by health insurance. Washington, D. C. implemented the mandate on CRC screening coverage on April 13, 2002 (Cokkinides et al., 2011). In the next section, I discuss the CRC incidence as well as the disparities among males and females and between the different races/ethnicities in the district.

5.6.2 CRC Incidence in DC

CRC incidence refers to the number of new CRC cases occurring in a population during a period of time (usually a year). Overall, a declining rate in CRC incidence has been

observed in the district. The decrease has been noticed for both males and females in Washington, D. C., as shown in Figure 48.

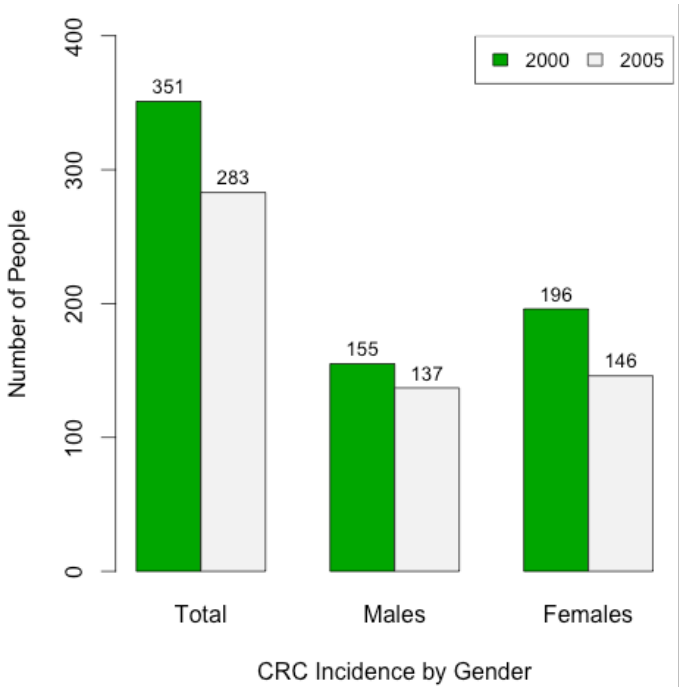


Figure 48: CRC incidence by gender in DC for the years 2000 and 2005 (Source: NCHS, 2015)

Examining the CRC incidence rates by race/ethnicity, disparities have been observed in CRC incidence rates between the White and African American DC residents, as shown in Figure 49. The CRC incidence rates among Whites in DC in 2000 were almost four times higher compared to African Americans. Even though in 2000 the CRC incidence slightly decreased, the racial gap in terms of CRC incidence has not changed. CRC incidence in 2005 among Whites was more than twice higher compared to the incidence among African Americans in DC. The total number of 267 African Americans

contracted cancer in 2000 compared to 211 the year 2000, slightly decreasing. The disparities among the CRC new cases across the different races/ethnicities persist, with African American males having the highest CRC incidence rates in the United States.

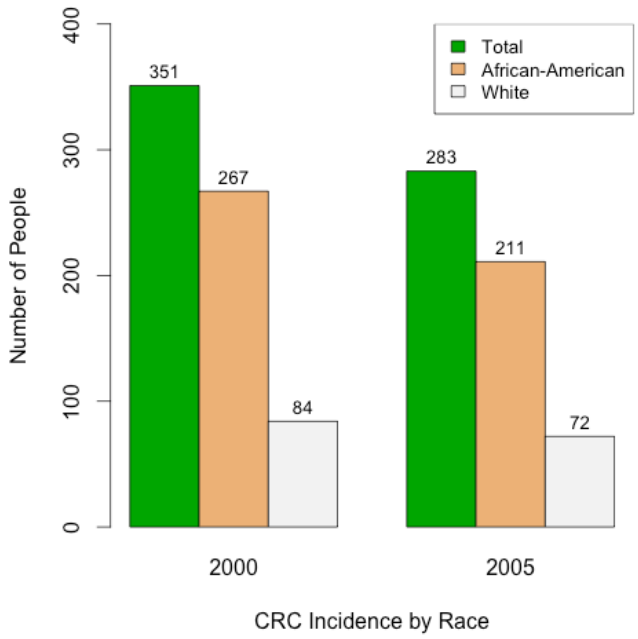


Figure 49: CRC incidence by race in DC for the years 2000 and 2005 (Source: NCHS, 2015)

Furthermore, cancer incidence rates are increasing in the population below the recommended screening age (Seigel et al., 2009). This can be attributed to the increasing trends in risk factor prevalence, such as obesity, smoking, red meat and alcohol consumption, as well as the lack of physical activity. These behavioral risk factors are modifiable, and interventions are needed to address these issues via lifestyle modifications so that primary CRC prevention can be enforced (Chatterjee et al., 2015).

A sharp decrease in CRC incidence rates has been observed in 2002 for the White population in the district. The African American population of the district has not experienced a similar decrease in CRC rates compared to Whites, further deepening the disparity gap between these two population groups.

The opposing trend in CRC incidence for the African American and White population groups of DC has been explained by the late CRC stage diagnosis, as well as by the complex interaction between several factors, such as environmental, social or economic (Alexander et al., 2007). New approaches to address the interplay between these elements are needed to shed light into the disparities that have been predominant in CRC incidence and mortality in DC.

5.6.3 CRC Mortality and Survival in DC

In 2000, DC had the highest CRC five-year aggregate mortality rates in the nation among males and females, with 36.9% and 23.3% respectively (Stewart et al., 2004). Of the cancers affecting both genders, colorectal cancer mortality rates were the highest in the nation (breast and prostate cancers being gender specific). The age-adjusted death rate due to CRC in DC in 2000 was 28.8 (per 100,000), compared to a national average of 20.7 (per 100,000), as shown in Figure 50.

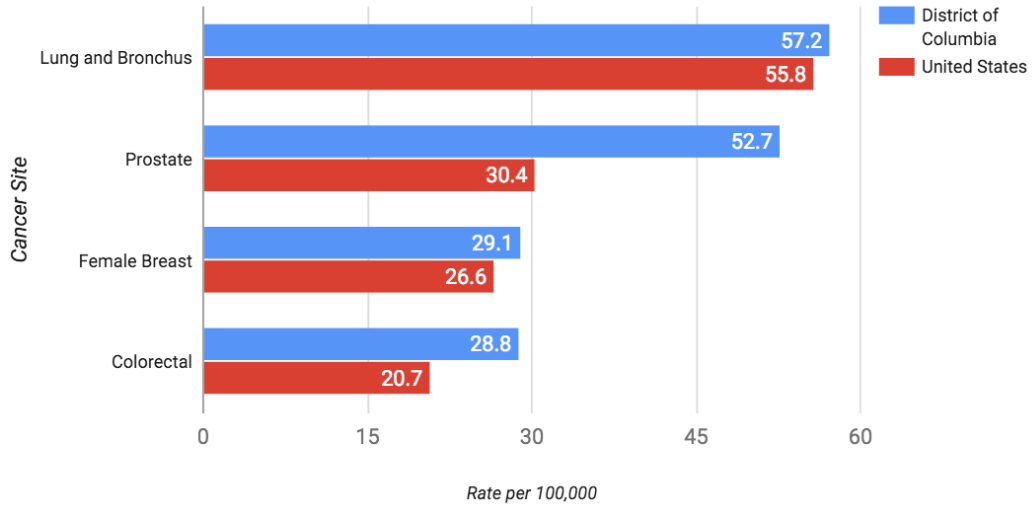


Figure 50: Age-adjusted cancer death rates for the 4 primary sites with the highest rates within DC, for men and women in 2000, DC (Source: USCS, 2015)

The overall mortality rates in CRC are presented in Figure 51.

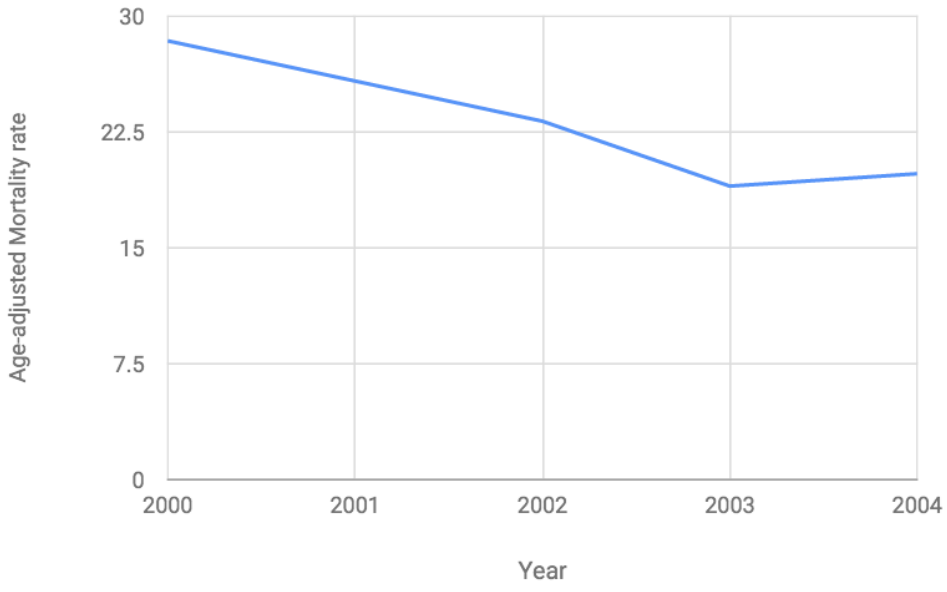


Figure 51: Overall CRC mortality rates between 2000 and 2004, DC (Source: DC Department of Health, 2000b)

Even though there were fewer overall CRC mortality cases in 2004 compared to 2000 in DC, disparities between Whites and African Americans persisted, as shown in Figure 52. The CRC mortality rates were lower among the White population of DC in 2000 compared to African American residents in the same year. Similarly, in 2004, the number of deaths among the White population was significantly lower compared to those among African American residents.

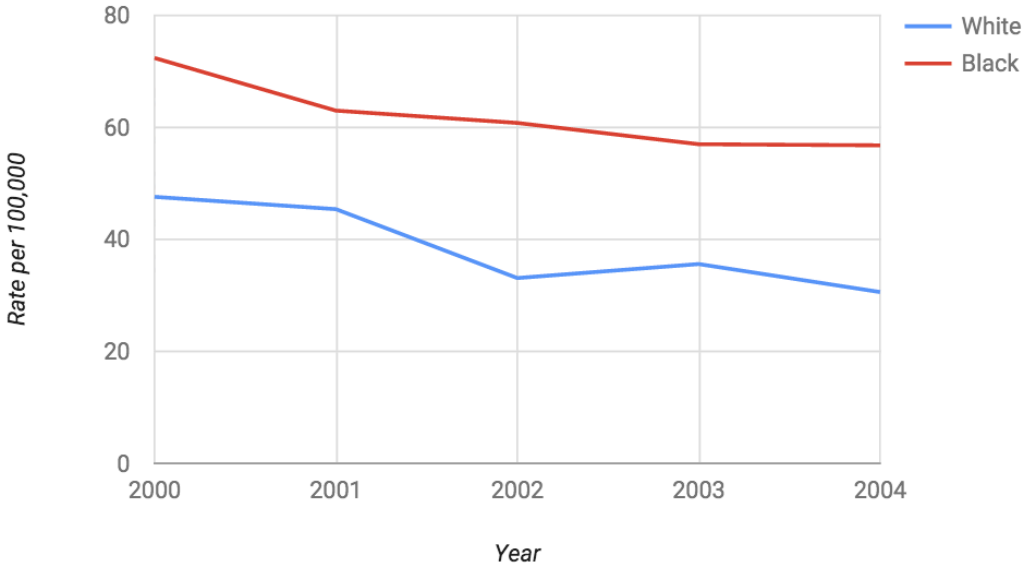


Figure 52: Overall CRC mortality rates for White and African American population groups in Washington, DC between 2000-2004 (Source: DC DOH, 2000b)

African American residents are more likely than Whites to die from CRC in the United States and in DC (Chatterjee et al., 2015). Furthermore, African American males are more likely than African American females to die of CRC. While CRC mortality was

decreasing among both White and African American populations, the gap among the races/ethnicities persisted throughout the years.

Disparities have been observed regarding CRC mortality trends for both males and females between 2000 and 2004, with men having higher death rates due to CRC compared to women, as shown in Figure 53. From 2001 to 2003 there was a slight decrease in CRC mortality rates for both males and females. A sharp decline in CRC mortality has been observed in 2004 for males, and a sharp increase has been noticed in the female CRC death rates.

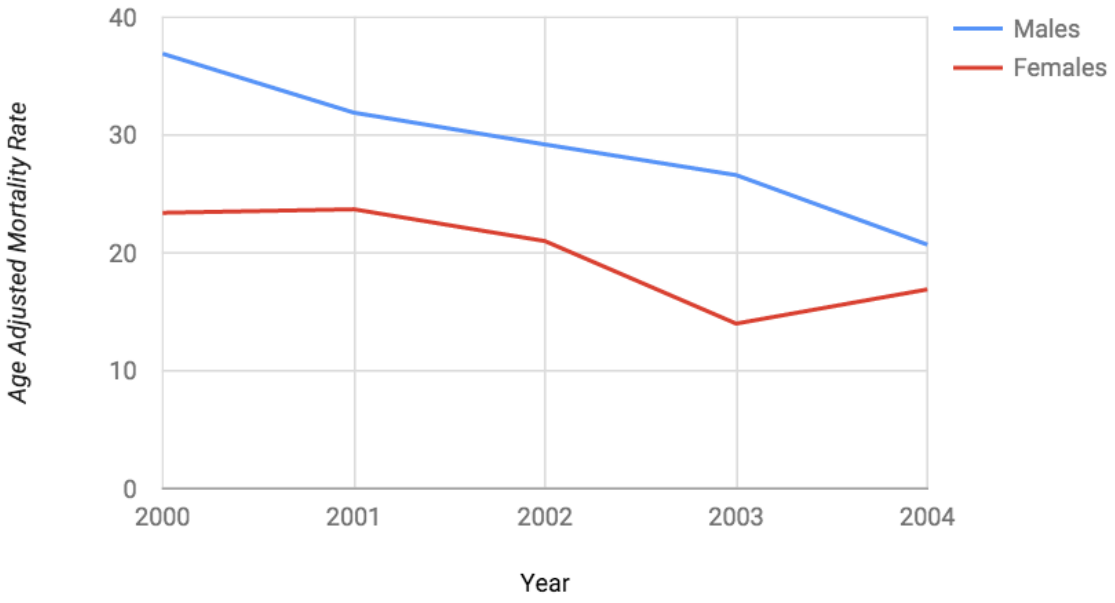


Figure 53: Overall CRC incidence rates for males and females population groups in Washington, DC between 2000-2004 (Source: DC DOH, 2000b)

DC CRC mortality rates for males and females are in general agreement with the overall CRC mortality trends across the nation, as shown in Figure 54.

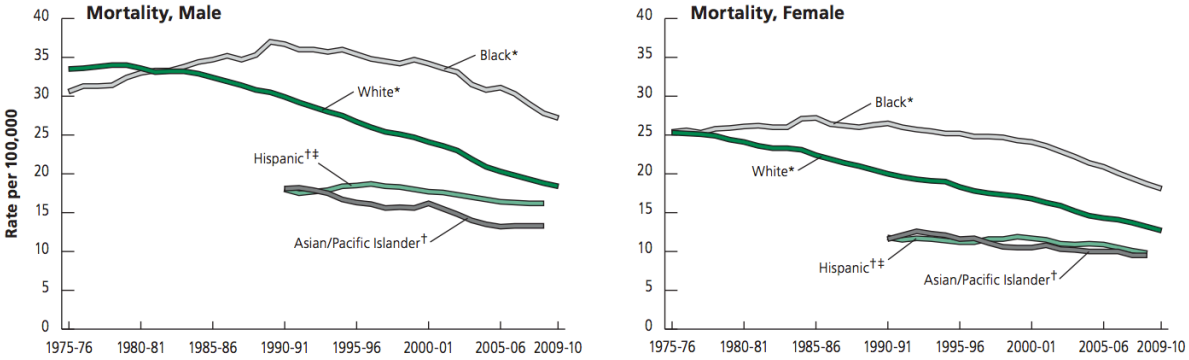


Figure 54: Trends in CRC incidence rates by race/ethnicity and sex, 1975-2010 (Source: SEER, 2013)

Observed survival is the estimate of how many individuals are alive at a point in time subsequent a cancer diagnosis. CRC survival depends on a broad range of factors, with the most significant ones being the stage at diagnosis and the tumor types. Stage at diagnosis reflects the status and the extent to which cancer has spread (or not) inside the body. If CRC is found at the localized (or Stage 1) stage, the likelihood of survival five years post diagnosis is the highest, reaching 90.1% (SEER, 2016). The probability of being alive five years after diagnosis decreases as the stage at diagnosis advances. Therefore, screening for CRC is of utmost importance, as the chances of survival are higher if cancer is found early, when it is most treatable. The prognosis is worse as the stage at diagnosis advances, and the treatment options are limited and less effective when cancer has largely developed and spread to other organs. Disparities have been observed

in CRC survival as well. Asians/Pacific Islanders are the most likely to be alive five years after a CRC diagnosis, as shown in Figure 55.

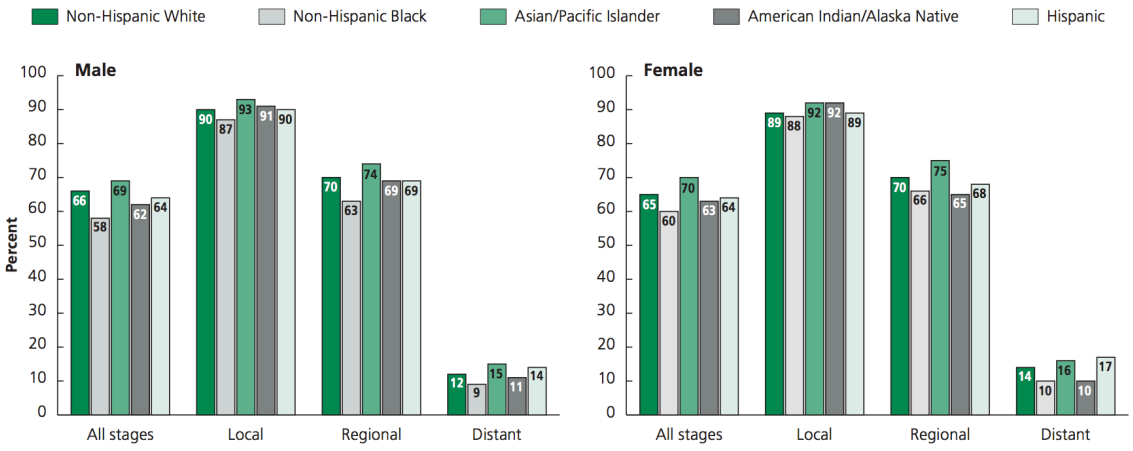


Figure 55: Five-year CRC-specific Survival (%) by race/ethnicity, 2003-2009 (Source: SEER, 2013)
* probability of not dying from colorectal cancer within 5 years of diagnosis

Out of all races/ethnicities, Whites are the most likely to be diagnosed with CRC early, when treatment is most effective (American Cancer Society, 2012). The African-American proportion of the population with a CRC diagnosis is less likely to receive appropriate CRC treatment (Ayanian et al., 2003). However, it is worth noting that the African Americans who do get chemotherapy as part of their CRC treatment, they are as likely to benefit from it as much as the other racial/ethnic groups in terms of survival. They are also more likely to experience less therapy side effects.

Owning private health insurance as opposed being uninsured or on Medicaid may also influence survival. At Stage 2 CRC, people with private health insurance fare better

than those with Stage 1 CRC that are uninsured or are covered by Medicaid, as shown in Figure 56.

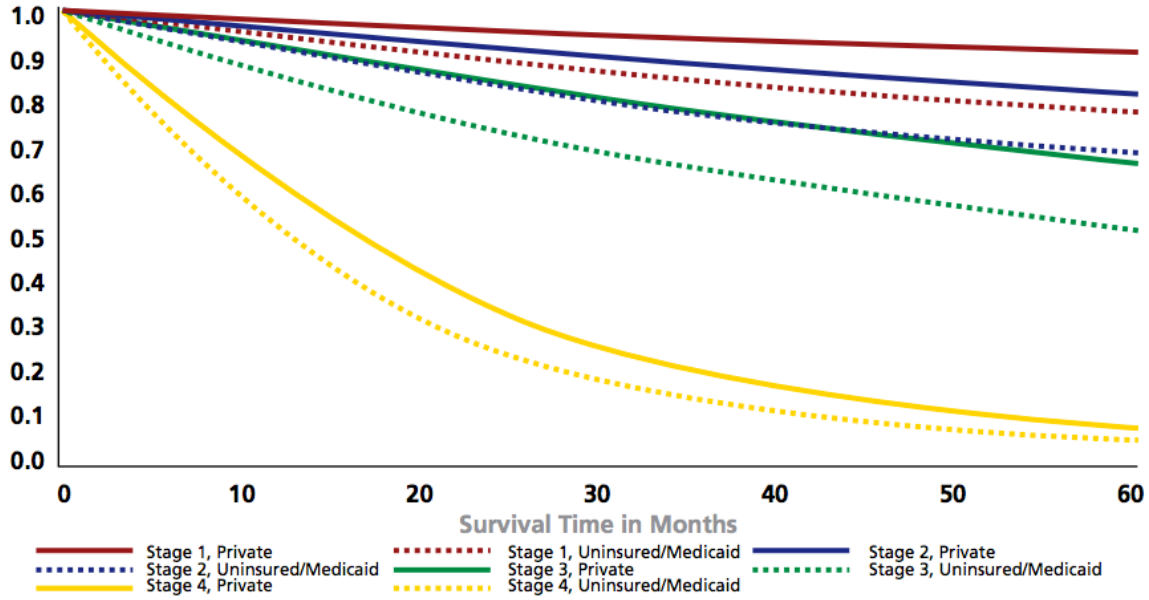


Figure 56: CRC covariate adjusted survival by stage (Source: ACS, 2008)

The disparities in CRC survival are fundamentally driven by socioeconomic inequities (American Cancer Society, 2012). Most of these disparities tend to fade as CRC treatment and care among the different races/ethnicities is the same for individuals with similar CRC stages (Bach et al., 2002).

5.6.4 CRC Screening Status in DC

The importance of CRC screening has thoroughly been discussed in Chapter 3. The goal of screening is to detect cancer before the manifestation of symptoms, in order to prevent it from progressing. CRC screening can considerably reduce the mortality rates due to CRC (Mandel et al., 1999). In 2002, the BFRSS collected data on CRC screening by

state. The prevalence of Blood Stool Test use for CRC Screening in DC during the period between 2000-2002 among adults of age 50 or over increased with increasing age, as shown in Figure 57.

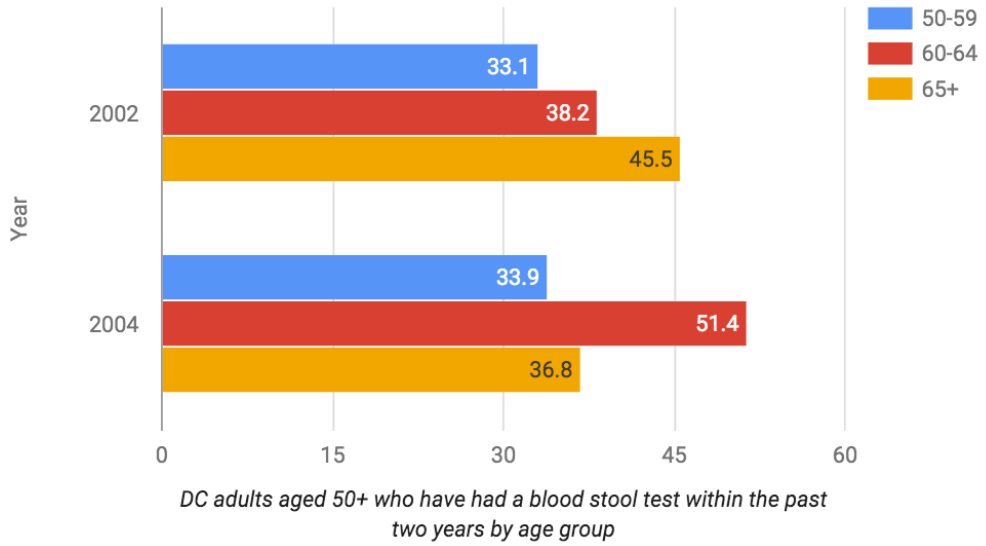


Figure 57: Adults aged 50+ who have had a blood stool test within the past two years in DC by age group, 2002 and 2004 (Source: CRC BRFSS, 2002)

The prevalence data on CRC screening using Blood Stool Tests was collected from adults of age 50 or over. BRFSS reported that about 39.2% of the DC adults 50 years of age and over had an FOBT test within the past two years. The proportion of DC adults between 50-59 years who had an FOBT test in the previous three years was 33.1%, among adults 60-64 years of age – the prevalence of FOBT use was 38.2%, and for adults 65+ years – it reached 45.5%. Males in DC were slightly more likely (40.1%) to have had an FOBT in the past two years compared to women (38.6%), and Whites (47.9%) were

more likely to have used the test compared to African Americans (34%). The likelihood of having had the test in the past year also increases with increased levels of education and improved income.

According to BRFSS, the proportion of DC adults who have ever had a sigmoidoscopy or colonoscopy (procedures jointly referred to as endoscopy) was 61% in DC in 2002. Similar to the distribution of adults who have had an FOBT test in the past two years, the population proportion of 50 years of age and older in 2000 who have ever had a sigmoidoscopy or colonoscopy in DC increased with increasing age, as shown in Figure 58. The percentage of individuals within the 50-59 years age range who have ever had a sigmoidoscopy or colonoscopy was 49.8% in 2002. Among the 60-64 years age range – 65.1%, and among the adults of age 65 or over the proportion reached 71.4%.

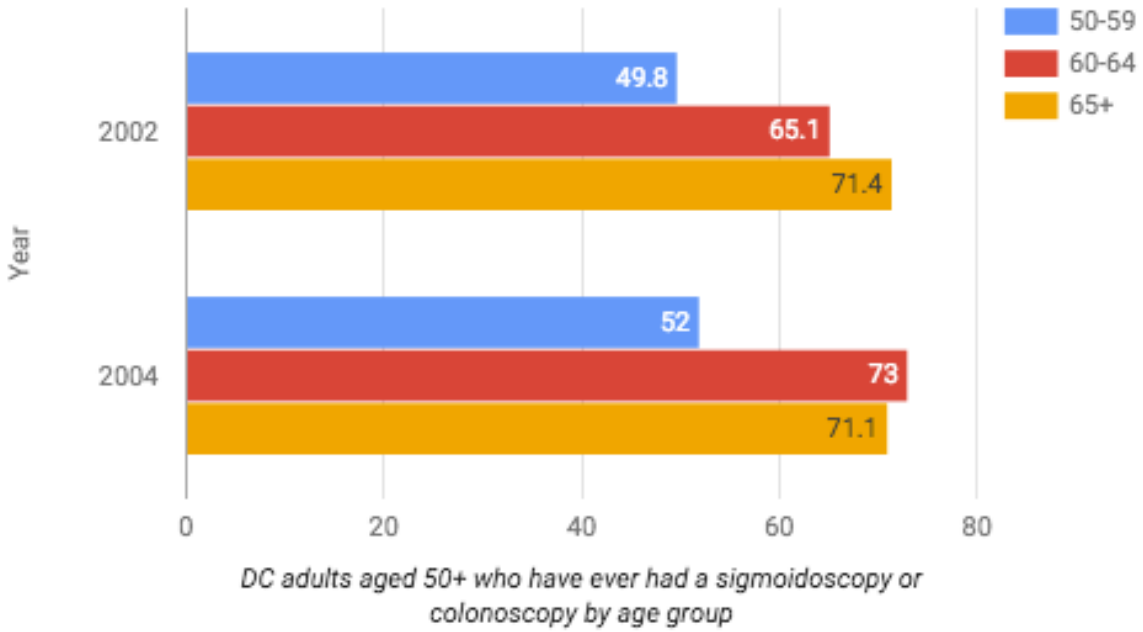


Figure 58: Adults aged 50+ who have ever had a sigmoidoscopy or colonoscopy in DC by age group, in 2002 and 2004 (Source: CDC BRFSS, 2002)

Women, compared to men, were more inclined to have ever undergone an endoscopy. About 63.2% of the women 50 years of age and over in DC reported having ever had a colonoscopy, compared to 59.2% of men within the same age group in 2002 as shown in Figure 59.

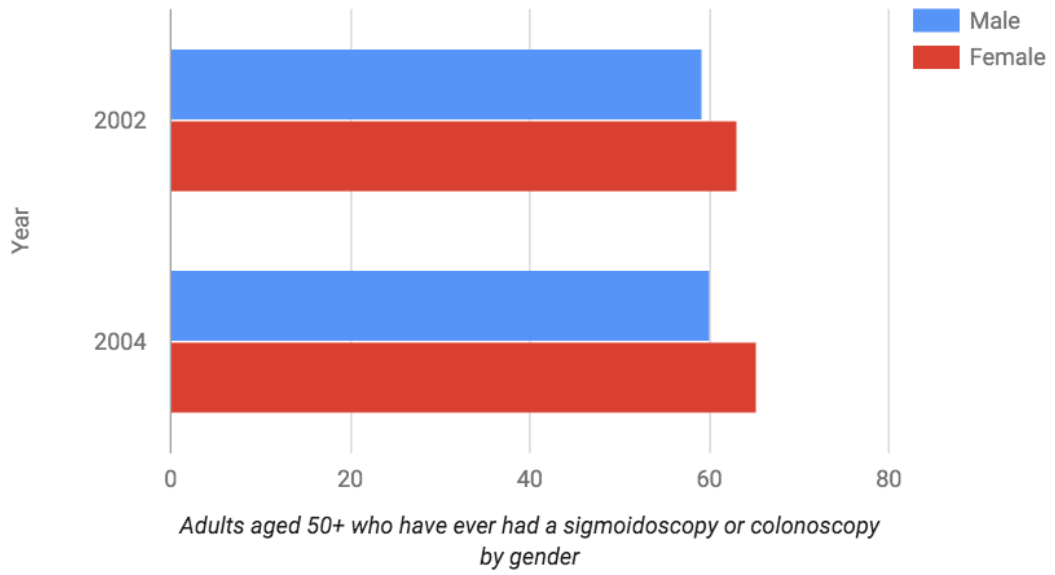


Figure 59: Adults aged 50+ who have ever had a sigmoidoscopy or colonoscopy in DC by gender, in 2002 and 2004 (Source: CDC BRFSS, 2002)

The endoscopy rates in 2004 increased for males compared to 2002. About 65.3% of women reported having ever had a sigmoidoscopy or colonoscopy, compared to about 60.2% among men.

Consistent with the disparities previously observed across the CRC care continuum, Whites (70.5%) were more likely than African-Americans (57%) to have had undergone one of these procedures in the past. The disparities persisted regarding income

as well. Out of the adults who have ever undergone a sigmoidoscopy or colonoscopy, 68.4% had an annual income of $50,000/year or more.

5.6.5 CRC Disparities Summary

In DC, disparities persist in most aspects of cancer care, from access to screening and early detection, to incidence, treatment, and mortality. Access to CRC screening and treatment is highly influenced by where in the district the residents reside. Cancer-related facilities are located mostly in the northwest part of Washington, D. C., making it difficult for residents that depend on public transportation to reach a hospital located the opposite part of the district (Chatterjee et al., 2015). This unequal distribution of facilities providing cancer care in the Washington, D. C. may impact the city's high incidence and mortality trends.

The disparities among health characteristics of DC residents correlate with the ones in CRC incidence, mortality, survival, screening and treatment among the different races and ethnicities, population groups with varying income, education, and access to care, as well as among males and females and within different age groups. African Americans are less likely to have their tumors discovered early, at a localized stage when treatment is most effective, compared to Whites, as shown in Table 8. African Americans were also more likely to be diagnosed late, after their cancer had metastasized compared to individuals of other ethnicities.

	Local	Regional	Distant	Unknown
Male, All Races[1]	**37.9**	**26.3**	**18.2**	**17.7**
White Male	43.7	25.8	14.7	15.8
Black Male	35.9	26.7	19.3	18.1
Female, All Races[1]	**39.9**	**25.6**	**15.6**	**18.9**
White Female	42.9	27.2	12.0	17.8
Black Female	38.6	26.0	16.8	18.6

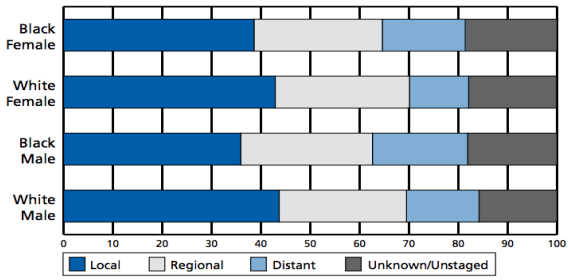


Table 8: DC CRC stage at diagnosis by sex and race, 2000-2004 (Source: Altekruse et al., 2010)

Disparities can also be observed among the prevalence of CRC risk factors in DC. African American and Hispanic populations of the district are at higher risk of health problems due to being overweight and failing to meet the moderate exercise guidelines compared to Whites (ACS, 2008). While the proportion of population receiving timely CRC screening increased over time, the disparities among the racial/ethnic groups increased, as shown in Figure 60. White individuals were more likely to be up to date with CRC screening than any other race or ethnicity.

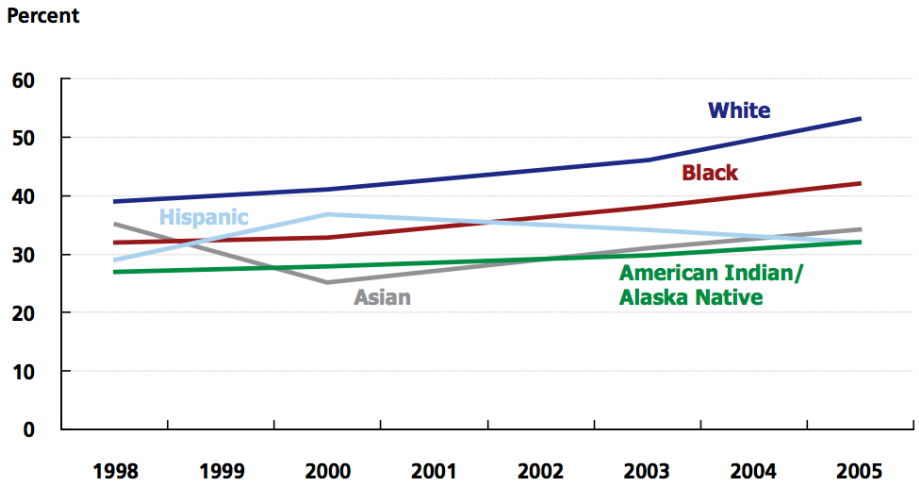


Figure 60: Proportion of people getting CRC screenings by race (Source: NCHS, 2015)

The difference in incidence and mortality between the White and the African American groups of the DC population is outstanding. While in the 1980s the death rates due to CRC were comparable between the Whites and African Americans, the difference in mortality rates among the two population groups increased considerably over time, as shown in Figure 61. White people are less likely to die from CRC compared to African American individuals, and these disparities can be traced to the variances in health behaviors, risk factor distributions, the access to care, as well as to differences in CRC screening practices, CRC treatment, and overall CRC care.

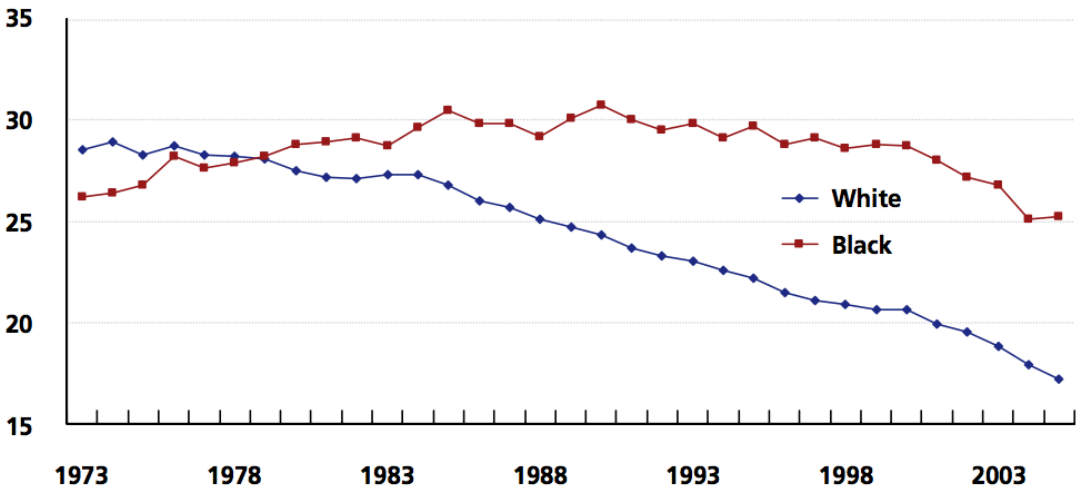


Figure 61: CRC mortality rates, per 100,000 people (Source: NCHS, 2015)

The overall cancer mortality rates in DC differ between men and women, with men having higher cancer mortality rates compared to women both in Washington, D. C. and nationally, as shown in Figure 62.

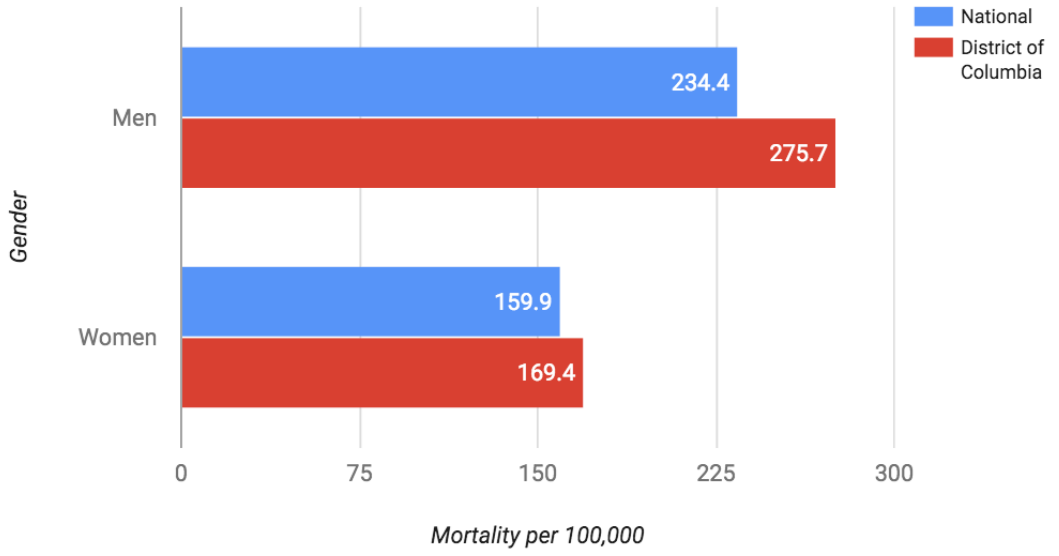


Figure 62: Cancer mortality rates in DC from 2001-2004, per 100,000 (Source: DC DOH, 2000b)

The underlying causes of CRC disparities are interrelated and complex, being linked to social, behavior and economic factors among others. Eliminating the disparities persisting in all the aspects of cancer care continuum is essential for equally improving the health outcomes for all DC residents. Research in CRC disparities is critical to progress in controlling this disease.

5.7 Summary

This chapter focused on describing the study area, starting with presenting the geography of the district, and its territorial division. Further, the socio-demographic characteristics of DC residents were introduced: age distribution, race, and ethnicity, marital status, education, income and poverty, employment and occupations. These characteristics are based on the U. S. Census state data for the year 2000. Next, the health characteristics of

nation's capital were presented. In particular, life expectancy, overweight and obesity status, physical activity, smoking status, alcohol use and health insurance/access to care are discussed. The section on health characteristics is concluded with a discussion of health disparities predominant in the district in 2000; as this affects susceptibilities and trigger further disparities at higher levels, contributing to the overall disease patterns. Section 5.6 discussed the status of CRC in the district. Incidence and mortality rates due to CRC between 2000 and 2004 were presented, as well as the distribution of stages at diagnosis between the White and the African American population groups of the district.

Differences between men and women regarding both incidence and mortality were discussed, and the prevalence of CRC screening test use in the district in 2000 was presented. Statistics about the FOBT test use and colonoscopy use were then discussed, and the difference between test use among the Whites and African Americans, as well as among men and women in the district are considered. The chapter concludes with a summary of CRC disparities regarding incidence, prevalence, screening, and mortality. In the next chapter, the agent-based model exploring the social determinants of colorectal cancer in DC is presented, and thoroughly described using the ODD protocol.

CHAPTER 6: CANCERSCAPE

Contents

6.1 Introduction 157

6.2 Data 159

6.2.1 Census Data 160

6.2.2 GIS Data 161

6.2.3 Joining Data to Shapefiles 162

6.3 State Machines for Specifying Behavior 163

6.4 Follow the Average (FTA) Mechanism 167

6.5 Model Overview 169

6.5.1 Purpose 174

6.5.3 Process Overview and Scheduling 180

6.6 Design Concepts 184

6.6.1 Basic Principles 184

6.6.2 Interaction 185

6.6.3 Emergence 185

6.6.4 Adaptation 185

6.6.5 Stochasticity 186

6.6.6 Collectives 186

6.6.7 Observation 186
6.7 Details 187
6.7.1 Initialization 188
6.7.2 Input 189
6.7.3 Submodels 190
6.8 Summary 220

> "The goal of science is to make the wonderful and complex understandable and simple – but not less wonderful"
>
> (Herbert Simon, 1996)

6.1 Introduction

Building on the review of SDC and CRC provided in Chapters 2 and 3, and the methodological considerations provided in Chapters 4 and 5 - Chapter 6 outlines an ABM that simulates the social determinants of CRC in Washington, DC. The model was designed to incorporate current research on the social determinants of cancer using a methodology that is mostly unexplored in the field. This approach, however - is particularly suited to address the non-linear dynamics, the complexity and the multilevel interactions characteristic to the SDC.

Regarding general implementation details - the model was build using Java Development Kit (JDK) 7 supported by MASON (Luke et al., 2005) and GeoMason (Sullivan et al., 2010) within the IntelliJ IDEA 15 development environment, as shown in Figure 63. GitHub was used for the purpose of source code management (on private repositories), back-up, and version control.

Figure 63: The tools used for building Cancerscape: Java 7, IntelliJ IDEA 15, MASON, JavaScript, R and Github

MASON is a free, open source discrete-event Java-based multiagent simulation system for building Java-based simulations. It provides both a model library as well as 2D and 3D support. Even though MASON has several advantages over other available agent-based simulation platforms: such as separating models from their visualization and providing the possibility for models to be checkpointed and dynamically migrated across platforms – it does not support geospatial functionality independently (Sullivan et al., 2010). Thus, GeoMason (which is an optional MASON extension that adds support for vector and raster geo-referenced data) was used for the purpose of integrating and

manipulating geospatial data in the model. GeoMason[1] makes use of the Java Topology Suite (JTS) for fundamental geometry support.

GitHub is a web-based Git[2] repository service. GitHub Education offers students with a free of charge micro account including 5 private repositories that were used for storing Cancerscape source code and data. GitHub provides a user-friendly code management environment that was used throughout the Cancerscape development. Other languages and programs used for the purpose of this dissertation involve QGIS and ArcGIS (for geographic data pre-processing), R (for statistical data analysis), Visual Paradigm and UML (for diagramming) and JavaScript (for creating a dashboard to interactively display the results of the model). The simple diagrams were generated using Google Sheets functionalities.

6.2 Data

Getting useful data to inform a computational model represents a challenge for agent-based modelers. Cancerscape integrates diverse data types from multiple sources. In order to capture the multilevel complexity of the social determinants of health – data from several surveys, such as United States Census Bureau data, additional secondary sources and academic literature (qualitative research data, CDC reports, surveys and studies that focus on the social determinants of cancer) were used. Several types of data inform the model in regards to agent population generation and characteristics assignment. The types of data required by Cancerscape are summarized in Table 9.

[1] GeoMason is released under the Academic Free License, version 3.0.

[2] Git is a distributed revision control system focused on speed, data integrity, and non-linear workflows, initially introduced developed by Linus Torvalds for the Linux kernel.

Table 9: Data types required by Cancerscape

DATA TYPES	COMMENTS
DEMOGRAPHIC	- Population of individuals grouped by age, sex, employment by tract according to 2000 Census - Household size data, types and distribution - Houses data
SOCIAL	- BRFFS data on general risk factors and behaviors - Survey data on social characteristics
GEOGRAPHIC	- DC GIS population by tract shapefile for the year 2000 - Hospitals shapefile
ROAD NETWORKS	- TIGER road network data. Obtained from the 2000 United States Census Bureau, this data contains the entire DC road network
COLORECTAL CANCER	- Colorectal cancer symptoms, stages, and polyp progression rates - Initial distribution of CRC cases (SEER survey) - CRC risk factor distribution (BRFSS survey) - Other factors relevant to the CRC development and progression

6.2.1 Census Data

Individual-level demographic data for agent characteristics are taken from the 2000 Decennial United States Census Bureau (USCB, 2000). The USCB provides individual-level data based on several geographic levels. The data was extracted with the appropriate tract level identifiers for later processing and GIS file joining. The census demographic data required preprocessing, as the identifiers do not automatically come formatted as "Text" – the format required by the GIS processing system. From the 2000 Decennial Census, the data on households, houses, individuals and other demographic characteristics were obtained for each individual census tract.

6.2.2 GIS Data

The GIS layers integrated in Cancerscape were obtained from two main sources: the U. S. Census Bureau TIGER (Topologically Integrated Geographic Encoding and Referencing) data – that provides the spatial context for decennial United States Census Bureau data collection, and the DC GIS Data Inventory provided by the official web portal of the DC government information and services. TIGER data provides information regarding the boundaries of DC, counties, census tracts, block groups, and blocks. An overview QGIS with of some of the GIS data layers included in Cancerscape is provided in Figure 64.

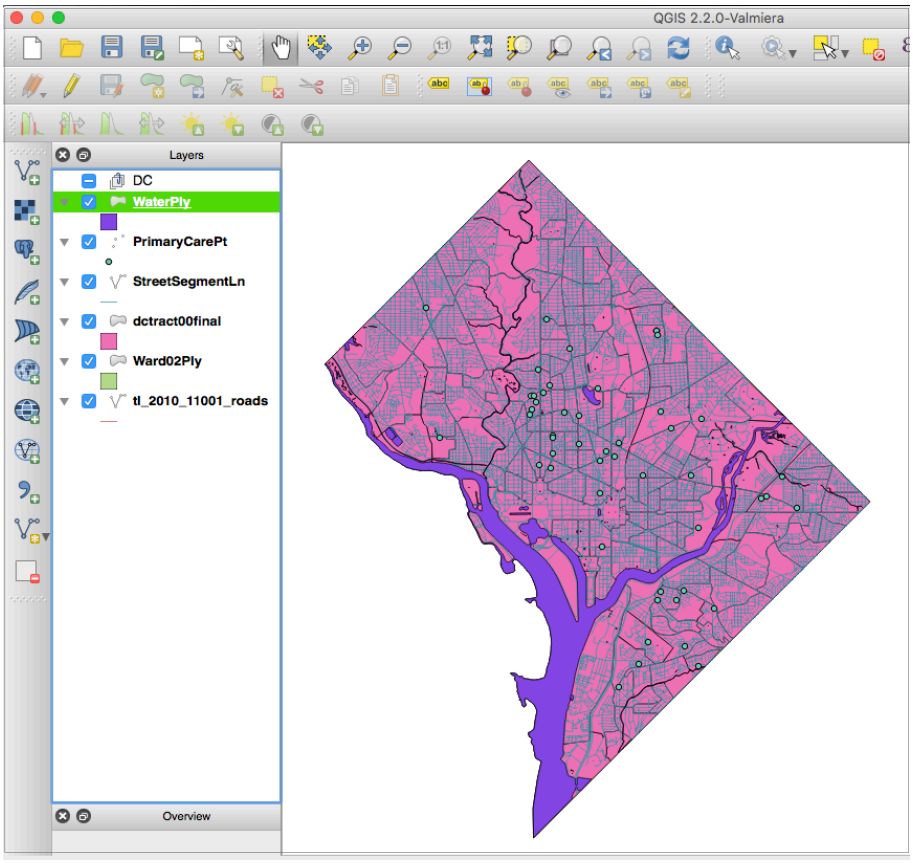
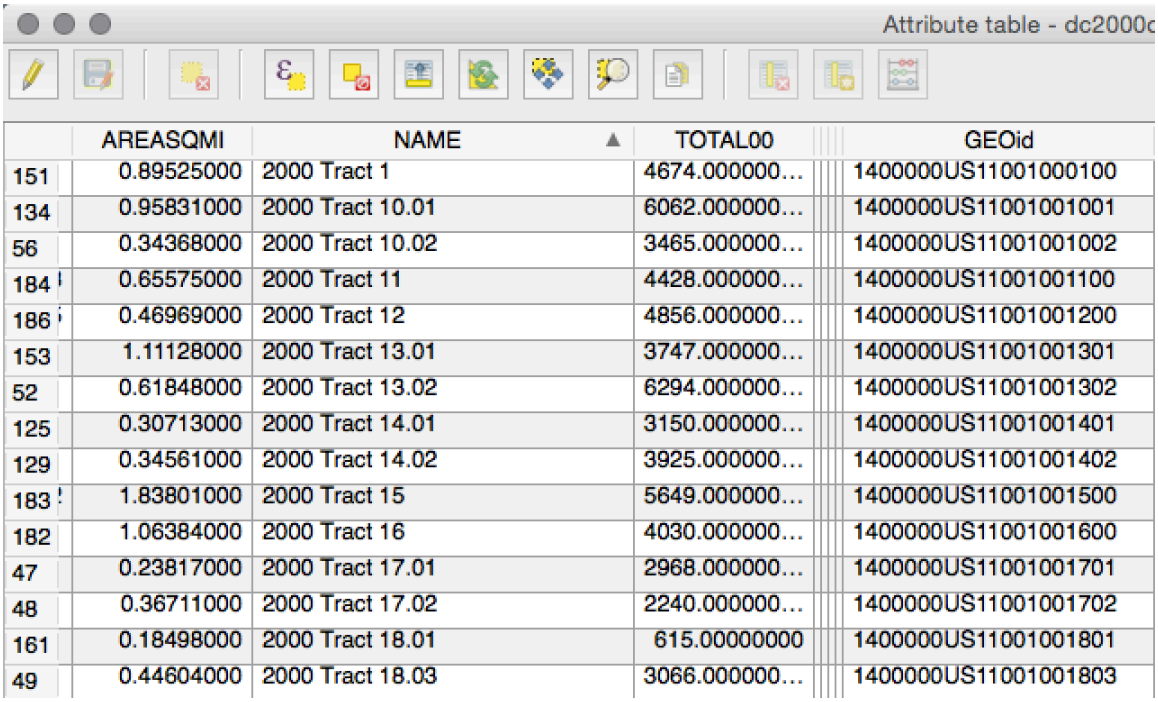


Figure 64: QGIS interface showing prepared shapefiles to be included in the model

6.2.3 Joining Data to Shapefiles

Shapefiles were merged to the United States Census Bureau demographic data using ArcMap 10.1. ArcMap is the main component of the ArcGIS suite. It is commonly used to perform a broad range of GIS tasks. The census CSV (comma separated values) files with demographic characteristics also contain qualified geographic identifiers (Geo.Id and Geo.Id2) that act as shared fields, as shown in Figure 65.



	AREASQMI	NAME	TOTAL00	GEOid
151	0.89525000	2000 Tract 1	4674.000000...	1400000US11001000100
134	0.95831000	2000 Tract 10.01	6062.000000...	1400000US11001001001
56	0.34368000	2000 Tract 10.02	3465.000000...	1400000US11001001002
184	0.65575000	2000 Tract 11	4428.000000...	1400000US11001001100
186	0.46969000	2000 Tract 12	4856.000000...	1400000US11001001200
153	1.11128000	2000 Tract 13.01	3747.000000...	1400000US11001001301
52	0.61848000	2000 Tract 13.02	6294.000000...	1400000US11001001302
125	0.30713000	2000 Tract 14.01	3150.000000...	1400000US11001001401
129	0.34561000	2000 Tract 14.02	3925.000000...	1400000US11001001402
183	1.83801000	2000 Tract 15	5649.000000...	1400000US11001001500
182	1.06384000	2000 Tract 16	4030.000000...	1400000US11001001600
47	0.23817000	2000 Tract 17.01	2968.000000...	1400000US11001001701
48	0.36711000	2000 Tract 17.02	2240.000000...	1400000US11001001702
161	0.18498000	2000 Tract 18.01	615.00000000	1400000US11001001801
49	0.44604000	2000 Tract 18.03	3066.000000...	1400000US11001001803

Figure 65: Sample of the attribute table of the DC 2000 demographic shapefile

The data files were joined with the GIS shapefiles using ArcMap, to create one file that includes the tract-level demographic data to be used by the model for generating the agent population. The GIS software (as well as other Microsoft Windows-only based

application) utilized for the purpose of this dissertation was accessed remotely via George Mason University's Virtual Computing Lab[3].

6.3 State Machines for Specifying Behavior

A finite state machine (FSM) represents a mathematical model of a reactive system. It is composed of a finite set of states and behaviors, with the transitions between the states depending on conditional statements. This mathematical model of computation is used to design algorithms for dynamic systems. FSMs are widely used in designing both software applications and sequential logic circuits.

The system state may be represented by a state variable s:

$$s \in \Sigma$$

where Σ is the finite set of all possible system states.

In the past years, FSMs have been increasingly used for modeling different aspects of ABMs (Borshchev & Filippov, 2004; Carley et al., 2006; Gnilomedov & Nikolenko, 2010; Torrens et al., 2011; Sakellariou 2012). An agent can be in several states during its life cycle. At any time step of the model, the agent needs to be in any one single state. The periodicity in Cancerscape is presented in Table 10.

Table 10: Periodicity in Cancerscape

Every Hour	*Every Day*	*Every Year*
Daily activities	Cancer progression Andersen Model of Total Delay Behavioral decisions	Aging Annual physical exams

[3] https://www.vcl.gmu.edu/index.php?mode=selectauth

Therefore, as part of the initialization stage, agents are attributed one initial state; like for instance all agents are attributed the state *SLEEP* at initialization, as shown in Figure 66, because the model begins at midnight on January 1, 2000. On an hourly basis, (or daily, weekly – depending on the frequency we define at the beginning of the simulation) agents run several FSMs in parallel - that can be interrupted or have their course altered by other external events.

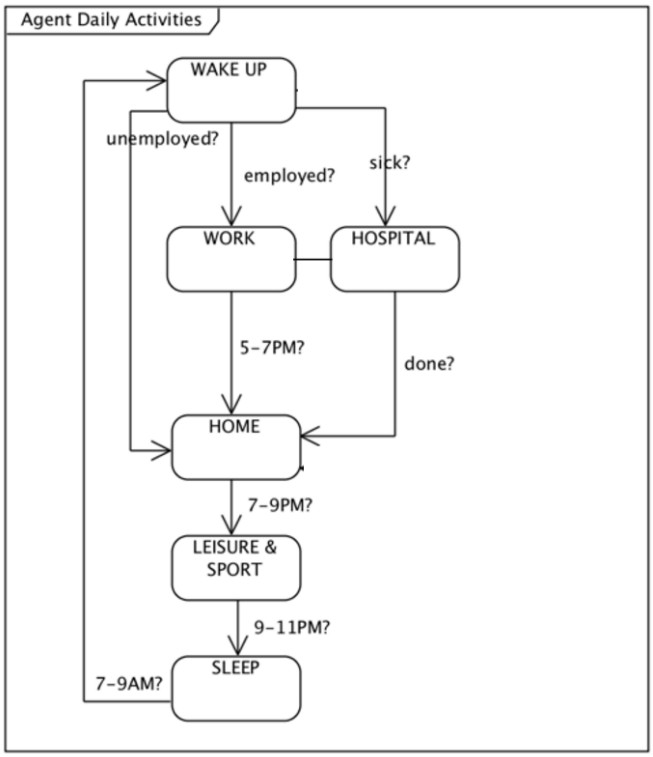


Figure 66: Agent daily activities FSM diagram

For instance, agents run their regular daily activities on and hourly bases, as they select their activities each day depending on the time of the day and individual characteristics. Cancer Surveillance is represented as another FSM that runs in parallel

with the agent daily activities. Cancer Surveillance runs on a daily basis, and the state attributed at initialization for all agents is NOTICE SYMPTOMS. This state implies that the agents are watching their bodies for unexplained changes on a daily basis. This assumption is motivated by the fact that individuals are determined to maintain an explicable physiological condition (Andersen et al., 1995). According to Schachter & Singer (1962), people have the need to evaluate and continuously search for explanations for any unexplained changes in their physiological state.

Modeling agent behaviors using FSMs lets us treat every agent action as a single state, allowing for flexibility in the duration of actions and seamless transitions between states. The rationale for specifying agent daily activities is derived from US Department of Labor's American Time Use Survey (ATUS, 2015). An average day for employed individuals of age between 25-54, based on the ATUS data is presented in Figure 67.

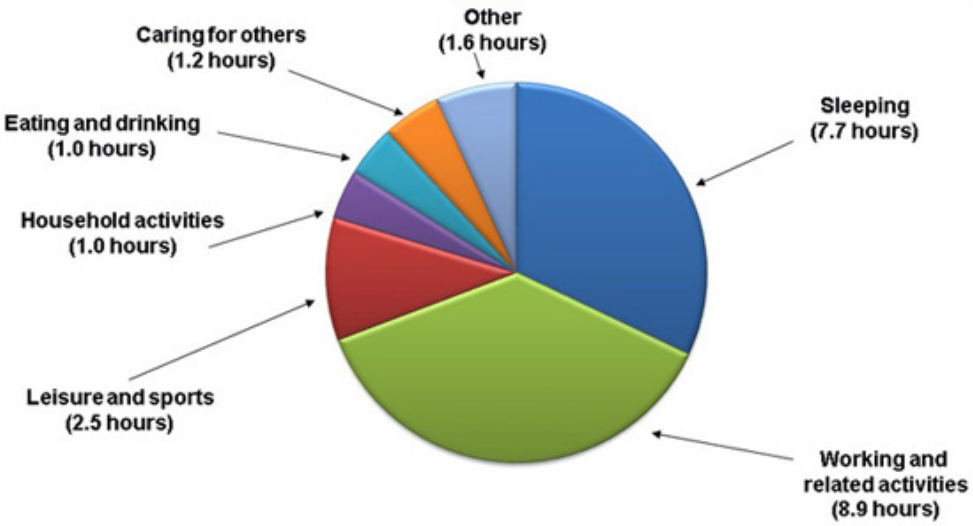


Figure 67: Time used on an average workday for employed persons ages 25 to 54 with children (Source: ATUS, 2015)

The average day presented above refers to all persons within the United States that are of age 15 and over, and combines all days of the week. While this is an overly simplified general snapshot of the time used by employed persons in the US, it offers an appropriate simplification that serves as the starting point for specifying agent activities. The time use slightly differs for persons 55 years and over. Figure 68 shows the hours per day that individuals age 55 and over spent doing selected activities. Because most people within this age range are retired, a large part of their time is spent doing leisure and sports activities, followed by household activities. Agent activities are adjusted for agents to go to the hospital when they feel sick (that is, when they notice symptoms of cancer).

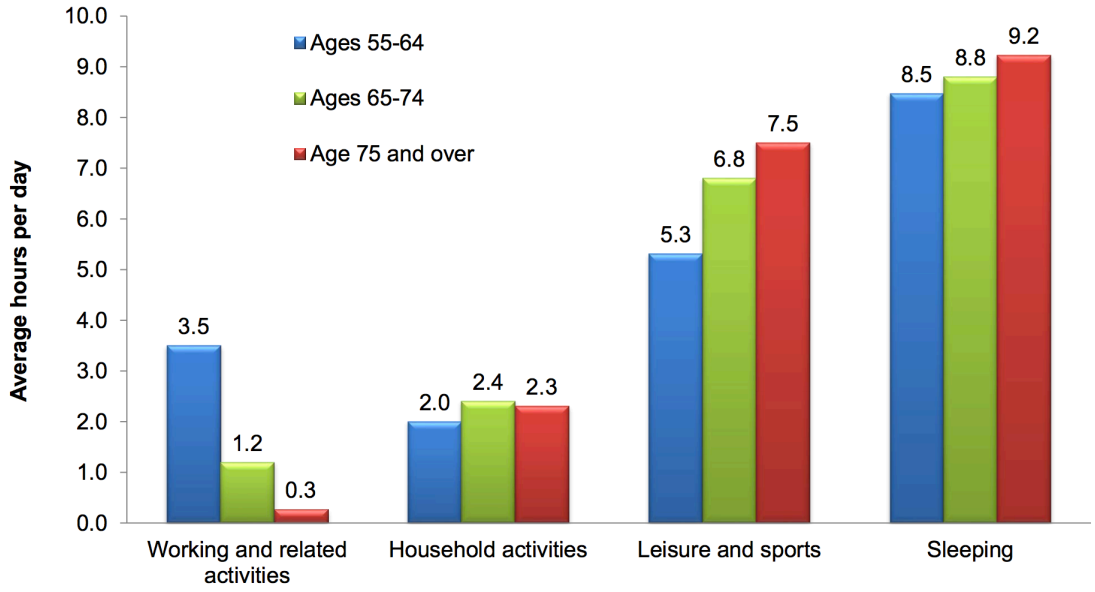


Figure 68: Hours per day that individuals age 55 and over spent doing selected activities (Source: ATUS, 2015)

6.4 Follow the Average (FTA) Mechanism

The same way individuals are interconnected so is their health (Smith & Christakis, 2008). Social norms are self-enforcing behavioral regularities (Axelrod, 1986; Cialdini & Trost, 1998; Lewis, 2008). One interesting aspect about them is that individuals conform to social norms *without thinking about it* (Epstein, 2001). As socially embedded entities, individuals habitually compare themselves to other individuals in their social network. Often, they seek essentially to conform to the established norms of their social group. As discussed in Section 2.4, social networks can affect individual health through a variety of mechanisms. Among the most notable ones are perceived and real social support, social influence (such as norms), social engagement, face-to-face contact (physical exposure to secondhand cigarette smoke) and access to resources (such as access to jobs, money, knowledge) (Berkman & Glass, 2000).

The social network generation for Cancerscape agents was implemented as an ego network for each single random agent. To define how agents communicate with their social environment in the model, the FTA rule was implemented. By following the FTA rule, agents evaluate their own status and adjust it to match the mean predominant in their social network (Hammond & Epstein, 2007). Following a similar example to the one proposed by Hammond & Epstein (2007), the individual's smoking index adjustment rule can then be calculated in the following way:

if my smoking index > average, lose 5 points
if my smoking index < average, gain 5 points
if my smoking index = average, do nothing

The results of how the above rule unfolds is presented in the Table 11:

Table 11: The 'follow the average' (FTA) rule

Agent 1	Agent 2	Agent 3	Agent 4	Agent 5	Mean
100	100	100	100	150	**110**
105	105	105	105	145	**113**
110	110	110	110	140	**116**
115	115	115	115	135	**119**
120	120	120	120	130	**122**
125	125	125	125	125	**125**

One of the FTA rule assumptions is that people have the tendency to be or act like the average of their social network. Also, for simplicity purposes, we assume that the smoking index can be directly adjusted (as opposed to including more complex physiological mechanisms to model this process).

When individuals attempt to 'follow the average', while the initial distribution is skewed the mean actually increases. This mechanism may compound biological, physical, and environmental determinants to produce dynamics of interest to the social determinant of cancer. Moreover, the dynamics of social contagion as it relates to diet or physical activity are part of a broader complex system driven by self-comparison. Normative self-comparison with meaningful others was found as a possibly salient driver of obesity-related health behaviors (Shakya et al., 2015).

While FTA has been fruitfully applied to obesity (Hammond & Epstein, 2007; Wang et al., 2014) it may be useful to other areas and research disciplines where social conformity may be a significant driver of agent behavior, as it has the ability of yielding

unexpected population dynamics. In the remainder of this chapter an ODD framework (Grimm et al., 2010) will be introduced, describing the model overview, conceptualization, design concepts and implementation details.

6.5 Model Overview

Cancerscape models a system where cancer dynamics is influenced by a complex interplay of environmental, human and social systems. The model represents a complex system, reflecting the influence of various other systems on cancer dynamics in a large artificial population. A general overview of the model is shown in Figure 69. It consists of three main phases that are: (1) setup – that includes the data setup, the creation of the spatial environment, as well as population generation and initialization; (2) simulation – this phase that involves starting the simulation and the schedules, where agent daily activities occur, their behaviors and interactions take place; agents engage in cancer-related decision making and dynamics, happening over 5 simulated years, and (3) the 'wrap up' step, where several statistics from the simulation runs are collected. The model starts on January 1, 2000 and ends on the last day of the year 2004.

Cancerscape is a spatially explicit ABM, combining diverse data from multiple sources and of different types. In a similar to the real world way, the boundedly rational agents behave autonomously. Their interactions are further influenced by behavioral and socio-demographic characteristics of their social connections, as well as by their geographic environment.

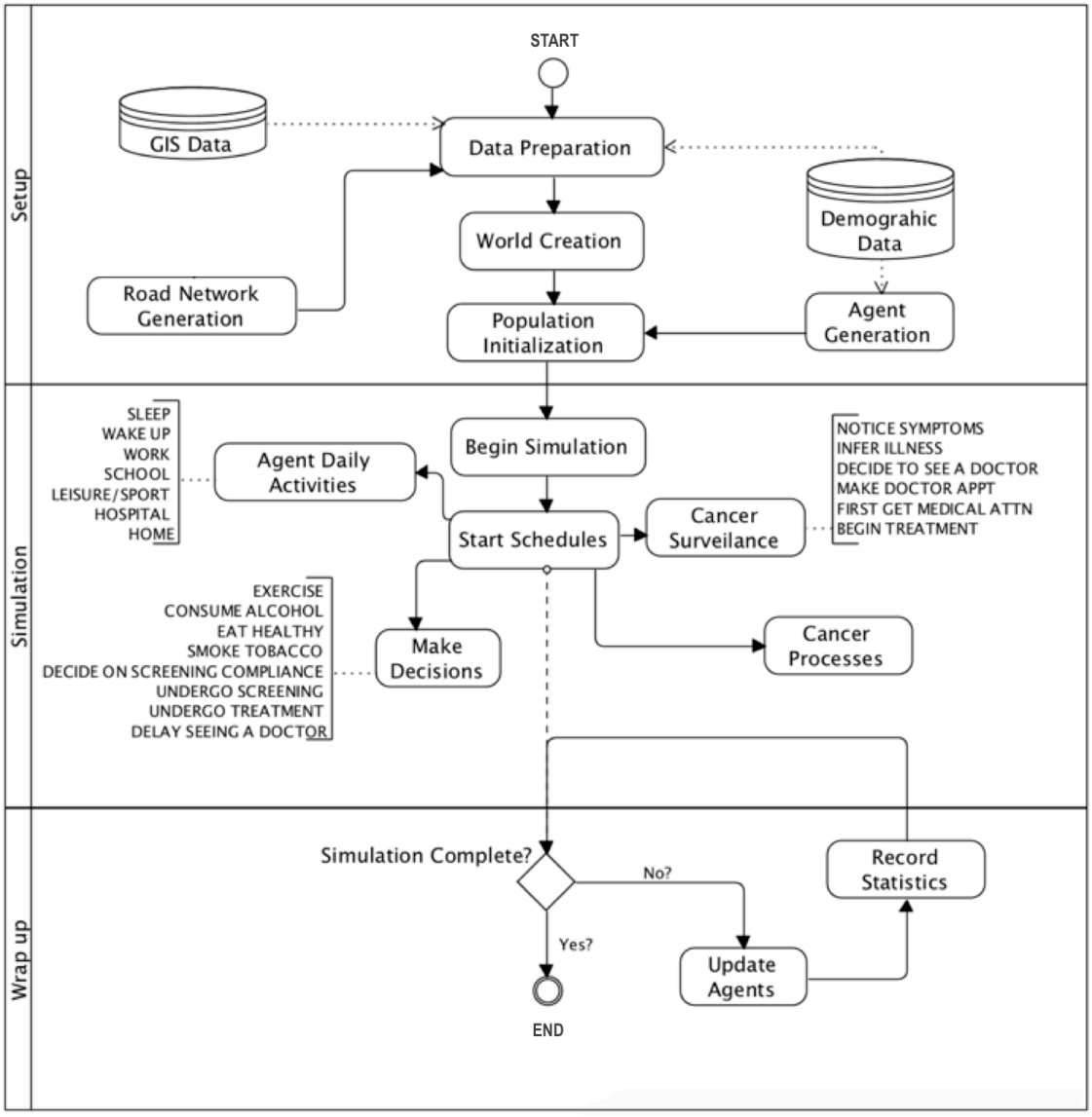


Figure 69: High-level model diagram

A theoretical framework capturing the social determinants of cancer is presented in Figure 70. It consists of four levels of influence: biological, micro, meso and macro. The biological level refers to the individual genetic make-up and the biological process of carcinogenesis. This level received the most attention in cancer research during the past

decades, as discussed in Section 2.2. The focus has been on single individual genes gone wrong, single mutations, single associations. Cancer research largely neglected the importance of emphasizing the individual and society from a systems science perspective. A system, that is composed of, and influenced by, multiple other systems.

The micro, meso, and macro level determinants address these systemic influences. The micro level consists of the individual level determinants such as early life exposures, education, occupation or marital status, as discussed in Sections 5.3.1 to 5.4.6. To account for community effects, the meso level addresses the immediate systems in which the individual is embedded, such as his personal social networks, community, as well as the social and cultural norms.

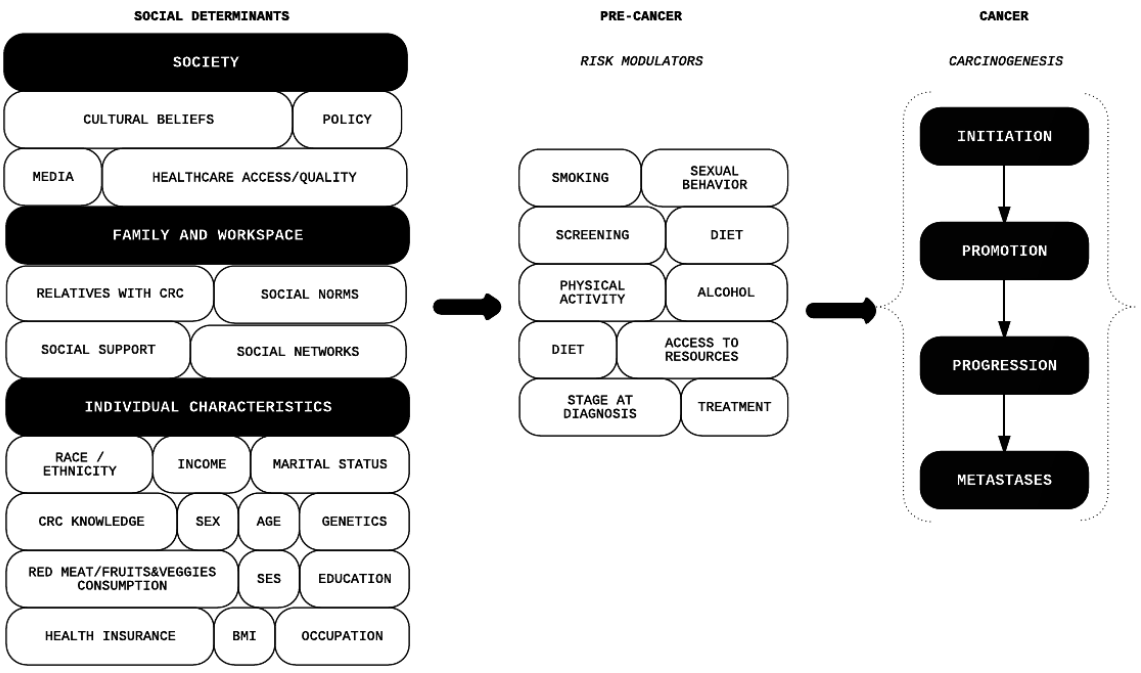


Figure 70: A framework addressing the influence of social factors on cancer

The macro level determinants deal with the broader systems influencing (either directly or indirectly) the individual and the social structures. A high-level overview diagram showing of the modules available in Cancerscape is provided in Appendix D.

The macro level. At macro level, the more general societal characteristics such as cultural beliefs, media or healthcare access are included. Health care access is affected by the society in which the individual is embedded (Wee et al., 2005; Smith et al., 2006). At the initiation stage of carcinogenesis, appropriate screening practices can lead to early detection and CRC prevention. Cultural norms and certain beliefs (such as feelings of shame and stigma) regarding cancer screening may influence the adherence to cancer screening tests. Somewhat important at the macro level is the association between cancer and religion. Religion was shown to affect cancer outcomes in at least two ways: either by protecting against cancer mortality or by affecting the progression of cancer (Stefanek et al., 2005). The mechanisms by which religion relates to cancer patterns are subject to debate. Some religion groups (such as Mormons) are characterized by greater conformity regarding cancer risk factors and behaviors. Arguably, religion and spirituality might help with coping following a cancer diagnosis. Healthcare access and quality (discussed in previous chapters) also contribute at the macro level, influencing health outcomes at individual and community level.

The meso level. At meso level, several family and workplace factors have shown to influence cancer development. Social engagement and support also influence cancer incidence, mortality, and prognosis. Reynolds & Kaplan (1990) examined the relation between social isolation and cancer in a population-based sample of 6,848 adults living in

Alameda County, California, during 17 years of follow-up. The findings indicate that women who were socially isolated were at significantly higher risk of dying of cancer of all sites. The meso level addresses the social structures that directly or indirectly affect and are influenced by the individual, as they facilitate the access to resources, risk-taking behaviors and exposures.

The micro level. Several factors affect cancer development at the micro, or individual level. People from lower socioeconomic status are more likely to contract and die from cancer compared to people from higher SES. The connection between SES and cancer has been widely studied (Krieger et al., 1999). Socioeconomically disadvantaged populations often have more advanced stages of cancer at diagnosis and are less likely to receive aggressive treatment for their disease. They are also more prone to live in poorer neighborhoods, more exposed to pollution and chemicals. Another important modifier of cancer risk is marital status. Being married has been associated with lower incidence of cancer at all sites (Ernster et al., 1979).

Cancer rates vary considerably by education level. Individuals with a lower education level were found to have a higher risk of malignant disease. Low education affects health literacy, which in turn influences health-seeking behaviors and screening practices (Albano et al., 2007). Other factors such poverty, BMI and waist circumference have also been shown to influence individual level cancer risk. The increase in obesity prevalence has become a public health concern. During the past decades, the U. S. population experienced an obesity epidemic (Mokdad et al., 1999). Obesity is associated with increased risk of cancer comorbidities, modulating cancer risk at various stages of

development, influencing its development and progression (Larsson & Wolk, 2007). Equally important is the effect of diet and physical activity, alcohol consumption, and tobacco smoke on cancer risk.

The micro, meso and macro level determinants interact within each level and between levels to modulate cancer risk. It is this interaction that contextualizes and shapes the individual and group behaviors, exposures and access to resources to affect the population measures of cancer incidence, prevalence, morbidity and mortality.

6.5.1 Purpose

The purpose of the model is to address the limitations of existing methodologies predominant in the social determinants of cancer by employing a novel, agent-based modeling approach (a methodology relatively unexplored in this field), combined with social network analysis and GIS, aim to shed light on how social factors influence cancer dynamics in a realistic population. The model uses data from multiple sources to simulate the influence of social factors on the dynamics of cancer in a large population of agents. It incorporates the most recent research on the social determinants of cancer.

6.5.2 Entities, State Variables, and Scales

When the number of state variables is large, using UML Class Diagrams is recommended for a clearer overall representation of model entities (Grimm et al., 2006). Given that this is the case for most Cancerscape entities, UML is used as the method of presentation. For instance, the UML diagram for a section of the *Person* class is presented in Figure 71, highlighting a few state variables of this class in a UML Class Diagram. The agents interact with their environment, with their social networks, as well as with other entities such as the primary care physician (PCP).

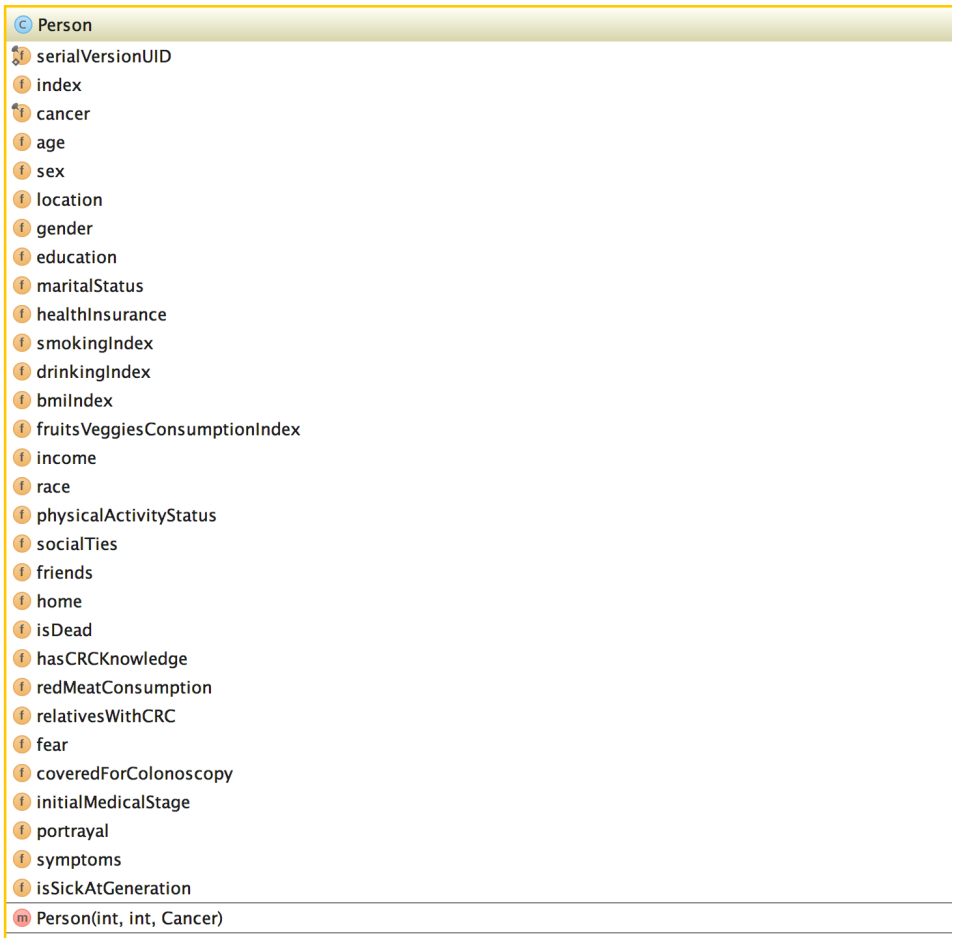


Figure 71: Section of class diagram for person

According to the AAFP (2015), a primary care physician (PCP) is a general physician who offers comprehensive care to the undifferentiated patient. The PCP takes care of the patient from the first interaction, continuing to do so over a longer period. He is the medical practitioner whom a person with an undiagnosed health condition initially usually contacts. In Cancerscape, a PCP has the following roles: (1) he performs annual routine health exams, (2) determines (based on the agent age, race, and risk category) if the agent should undergo CRC screening tests, (3) advises the person on healthy lifestyle

behaviors. Also, if the agent is eligible for CRC screening, the PCP shares knowledge about CRC with the agent, and (4) refers the agent to a specialist if cancer is suspected.

As opposed to PCPs, who generally have a wide scope of practice, Specialists are trained in a specific area of medicine. Depending on the type of care the cancer patient needs, there can be distinguished at least three types of specialists: Medical Oncologists, Radiation Oncologists, and Surgical Oncologists. For the sake of computational parsimony, Cancerscape models the Specialist as a generic entity, who could represents a medical, radiation or surgical oncologist. The actions performed by the Specialist in Cancerscape are: (1) assigning a cancer diagnosis to the patient, (2) resecting the colorectal polyp (via the colonoscopy procedure), should the agent present with one, and (3) applying CRC treatment to the agent depending on the cancer stage and other individual agent characteristics.

All agents have a cancer progression state. Cancerscape agents can be in either one of the following states: “NO_CANCER”– for agents with no sign of cancer, “POLYP_5” – state attributed to agents who have a colorectal polyp of 5mm or more, and “STAGE_1”, “STAGE_2”, “STAGE_3” and “STAGE_4” respectively, depending on the value of the cancer index that will be described in detail later in this chapter. An UML diagram showing the relation between the Person, the cancer progression and the cancer itself is presented in Figure 72.

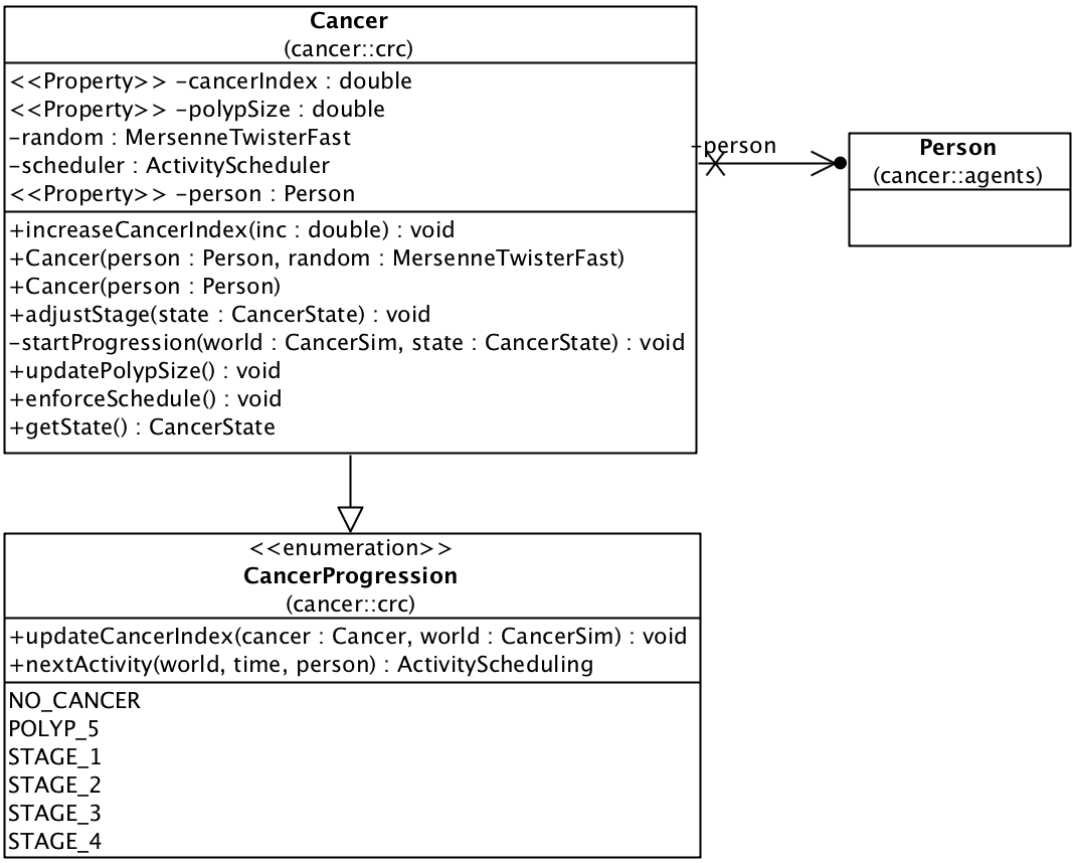


Figure 72: UML diagram showing the relation between the *Person*, *Cancer* and *CancerProgression*

CRC development has long been thought to be a multistep process, and most commonly begins from a polyp in the colon or rectum (Muto et al., 1975). In Cancerscape, cancer progression is correspondingly modeled as a multistep process. A person with no cancer can initially develop a polyp – which if it is not found and timely removed during a colonoscopy procedure, will eventually develop into CRC over time (Winawer et al., 1993). Small polyps (< 1cm) are extremely common and the probability of them turning into CRC is very low (Bond, 2000). As polyps continue to grow, their potential for malignancy increases. Individuals with colorectal polyps greater than 5mm

in size are at high risk of developing CRC over time (Panish, 1979; Atkin et al., 1992). In Cancerscape the polyp to cancer progression pathway was modeled in a similar way, as a multistep process beginning from a small polyp that over time may turn into cancer.

The Cancerscape spatial environment provides the context for agents to undertake their activities is presented in Figure 73.

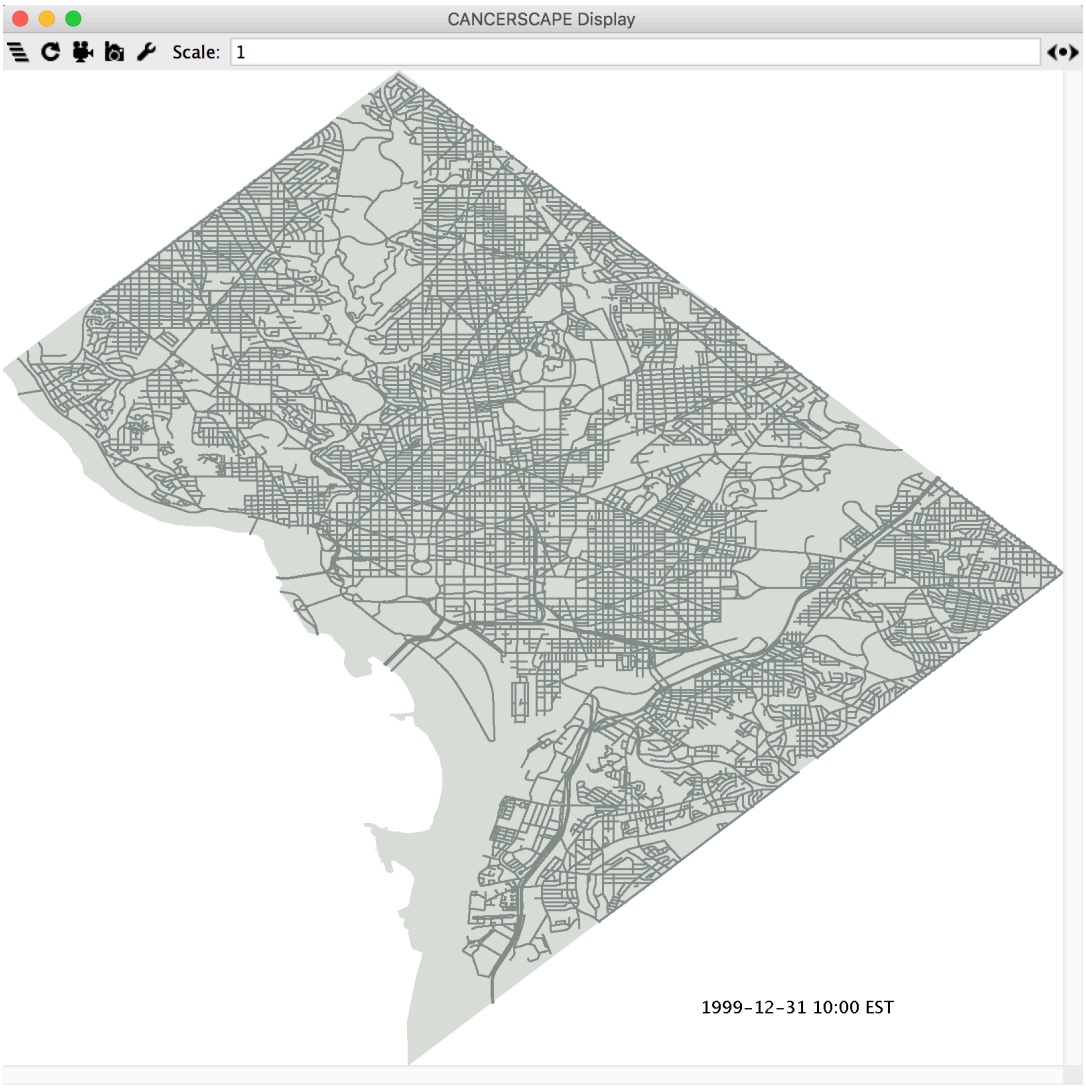


Figure 73: Model screenshot with generated tracts and roads

Figure 73 showcases the DC tracts and the road network only (spatial layers are disabled for visualization purposes). A close-up of the environment with the tracts, road networks, the primary care units (colored in magenta) and workplaces (colored in teal) as they appear within the Mason development framework is presented in Figure 74.

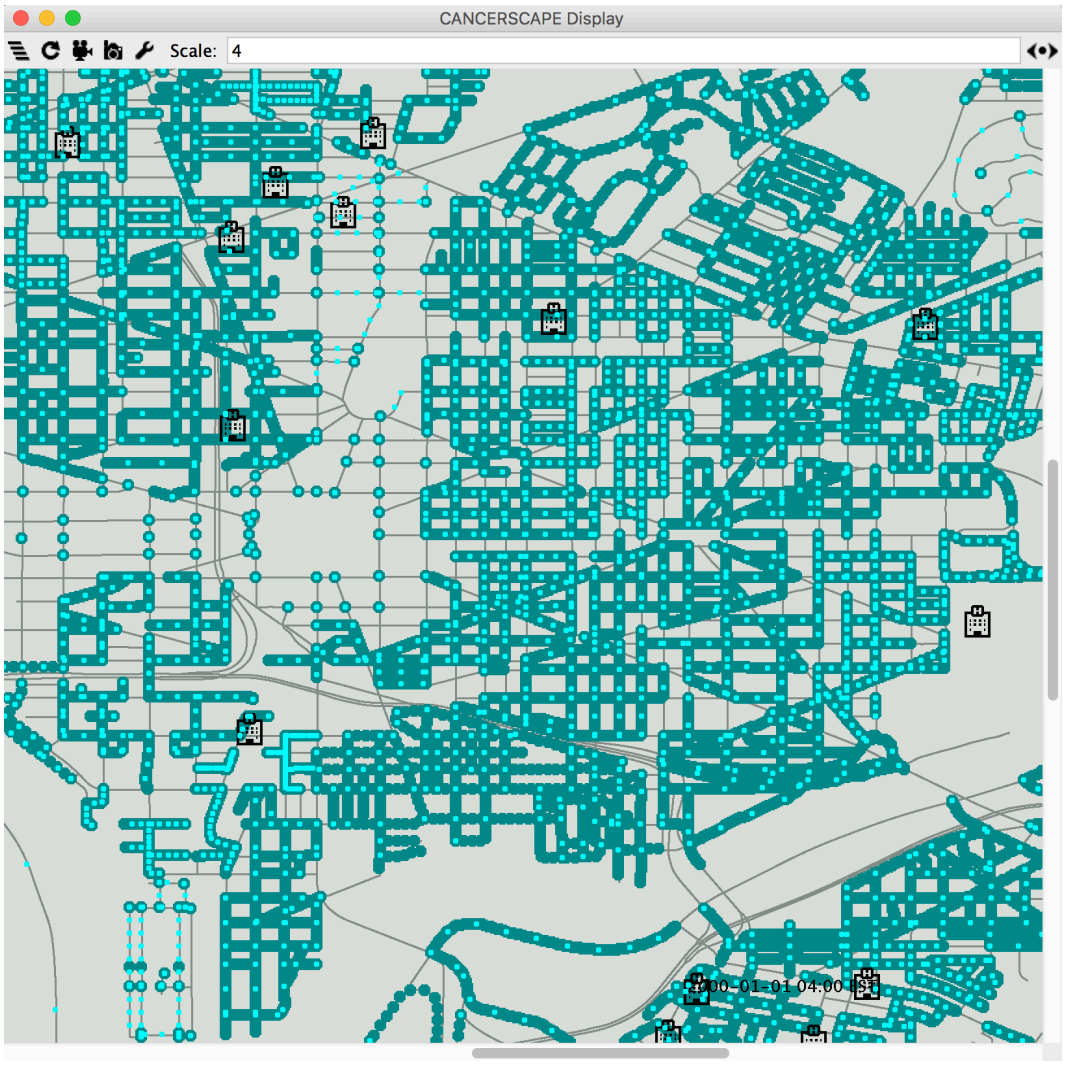


Figure 74: Cancerscape model close-up including primary care units (hospital icon), houses (dark cyan) and workplaces (light cyan)

The model also includes a few "global" parameters. Their values are initially set to values taken from academic literature, but the user can change them for testing purposes and scenario exploration. A comprehensive list of model parameters is presented in Appendix F. The next section discusses the overview of the processes in Cancerscape and the implementation of the scheduling procedures in more detail.

6.5.3 Process Overview and Scheduling

The default time unit in Cancerscape is one hour (60 minutes per one tick), but can be set to any periodicity of interest to the user (from every few hours to weeks or months). All agents take behavioral decisions with conditional periodicity. Additionally, the *Person* agent has several schedules that are running in parallel. They are presented in Figure 75.

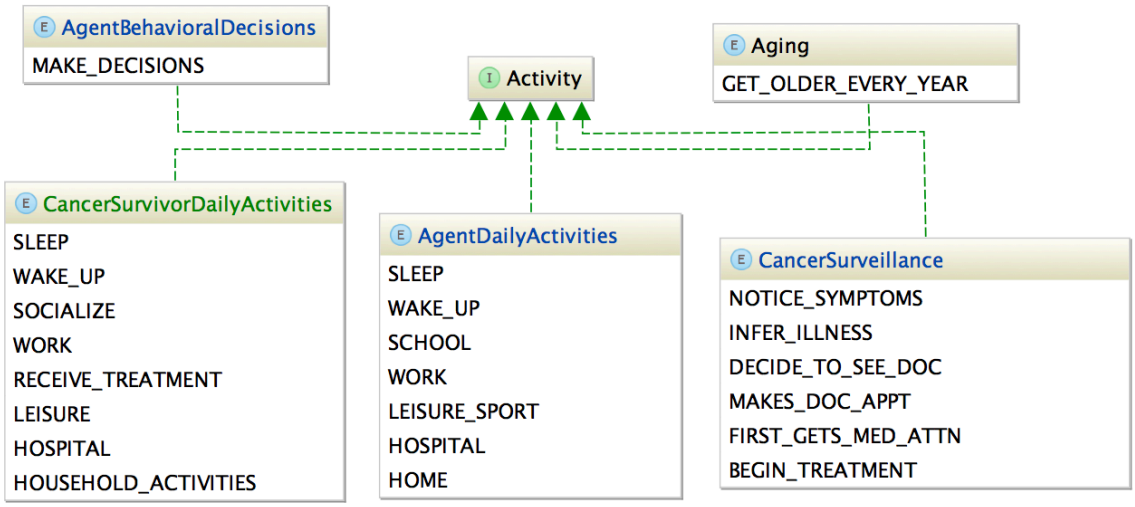


Figure 75: Person schedules in Cancerscape

The behavioral decisions depend on the time of the day and the agent's planned activities. The *Aging* schedule runs for all agents in parallel and has the role of increasing

the agents' age as years go by, to mimic the natural aging process in a grossly simplified manner.

Cancer Surveillance is another schedule run by the person. Agents are initially scheduled to *NOTICE SYMPTOMS* – which is the corresponding state of the "Detect unexpected signs and body changes" in the Andersen Model of Total Delay (Andersen et al., 1995). This is backed up by the Psychophysiological Comparison theory, which states that individuals are motivated (by genetic predisposition and social learning) to maintain an explicable physiological state (Cacioppo et al., 1988; Cacioppo et al., 1989). The scheduling order between the agent, his PCP and the specialist are presented in Figure 76. After the agent develops symptoms and infers illness, he decides to address his symptoms by seeking medical attention. Between the decision to seek medical attention and to act on that decision, there may be further delays, based on the agent health insurance status or the ability of the agent to cover medical costs. Upon acting on the decision to make an appointment with a PCP, the agent first receives medical attention. If the PCP suspects cancer and if the agent is covered by health insurance, the PCP refers the agent to a Specialist. Additionally, the PCP advises the agent on health behaviors and "increases" the agent knowledge level about CRC.

The agent, who has a Specialist referral, goes to the Specialist, who runs diagnostic tests on the agent (mostly screening tests). If the agent has a colorectal cancer polyp and if that shows up during the colonoscopy procedure, the polyp is removed. If the procedure was successful, the agent recovers. If a polyp exists but it is not found during the colonoscopy procedure, it continues to progress to cancer until it triggers

severe symptoms. If the specialist ends up detecting cancer, it recommends a treatment regimen that the agent can accept or reject (depending on his CRC knowledge and on his ability to cover treatment costs).

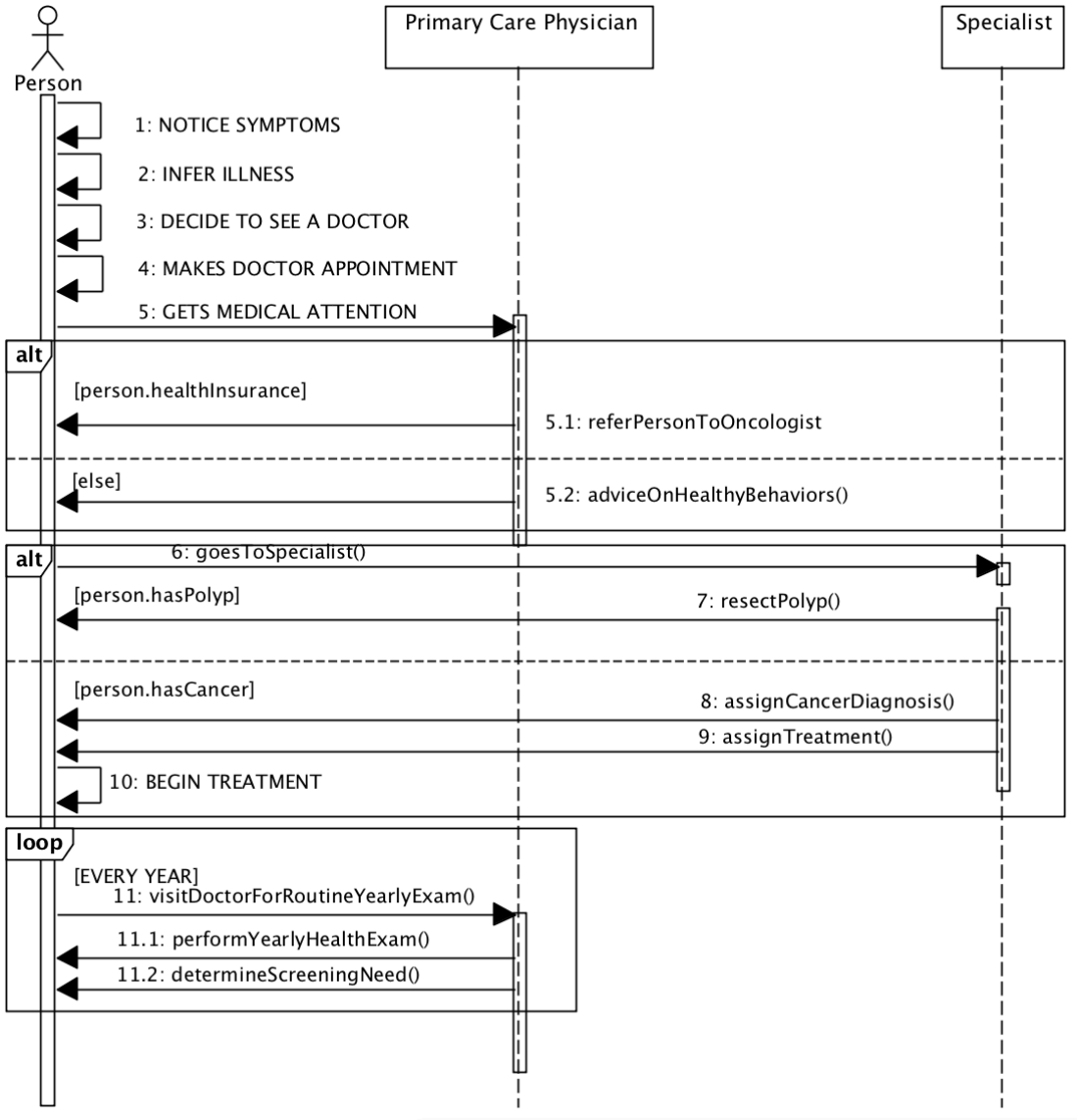


Figure 76: UML sequence diagram presenting the interaction between the Person, Primary Care Physician and the Specialist

If the agent decides to accept the recommended treatment, he then undergoes one of the following treatment regimens: polypectomy (colorectal polyp removal during colonoscopy), surgery, chemotherapy, radiation – or a combination of these options based on individual characteristics and CRC stage at diagnosis. For stages 1 and 2 the most common treatment is polyp removal, sometimes combined with surgery. Surgery is the recommended method of treatment for stages 1 to 3. At Stage 3, surgery is generally combined with chemotherapy. For Stage 4 it is typical for chemotherapy to be administered first in order to shrink the tumor so that surgery can later be performed.

Each treatment method has a probability of success that is considered and implemented in the model. The probability of success following a cancer treatment procedure depends on several factors, such as patient age, comorbidities or delay in treatment (Fevang et al., 2000). For instance, the success rate of the colonoscopy procedure can get as high as 97.2% (Nelson et al., 2002). The success rate for surgery procedures alone highly varies, and can get as high as 95%. However, it largely depends on the current CRC stage of the agent. The more advanced the CRC stage the lower the probability of success for surgery (Sullivan et al., 2012). The probability of success of surgery combined with chemotherapy was found to be around 55% (Vaira et al., 2010). For palliative chemotherapy, which is administered during the later stages of CRC progression with the purpose of improving the quality of life, the success rate is around 10% (O'Brien et al., 2006). Upon surviving treatment, the agent enters the recovery period. However, there is a possibility of recovery or relapse for cancer survivors who underwent treatment.

Another separate schedule that runs in parallel on the agent is the routine yearly health exam. The agents visit the PCP for a routine health exam on an annual basis. During the health exam the PCP advices the agent on health behaviors and determines if the agent is eligible for CRC screening, depending on agent individual characteristics such as age, family history of CRC and risk status of the agent. The schedules have been prudently designed in order to ease the computational load of simulating a little over a half million agents over a GIS space. Appendix E presents the interaction between the several modules in Cancerscape.

6.6 Design Concepts

In this section the design concepts of Cancerscape are explored according to the ODD protocol (Grimm et al., 2010). Cancerscape can be described in terms of seven core complex system modeling concepts: basic principles, interaction, emergence, adaptation, sensing, stochasticity, collectives, and observation that will be individually addressed below.

6.6.1 Basic Principles

Cancerscape draws on theories from scientific literature on the social determinants of cancer. The model is designed to address the following research question: "can a bottom-up, agent-based approach help to get insight into the social determinants of cancer?". The scenarios analyses are designed to shed light into how social determinants of cancer affect the incidence and mortality due to CRC. The purpose of the model is to apply novel, CSS methodologies to better address the complexity inherent in the social determinants of CRC.

6.6.2 Interaction

In order to model complex systems such as the social determinants of cancer, it is the interactions between the model entities that are of particular importance. Cancerscape is massively intricate – with a large number of agents interacting with each other and with the environment. Agents interact with their social networks, and with their primary care providers. This interaction influences agent behaviors and their decision-making processes, which in turn influences their cancer index. The interaction in Cancerscape not only determines agent behaviors but also can alter agent characteristics such as the access to cancer knowledge.

6.6.3 Emergence

Cancer dynamics, including measures such as CRC incidence, screening rates, and mortality among others, emerge based on the lower level interaction among agents and their social networks across the environment. The distribution of screening rates and delays are also the result of decentralized decisions taken by autonomous, boundedly rational agents. All the macro-level patterns and outcomes that result from the interaction of agents at the micro level can be considered emergent phenomena in Cancerscape.

6.6.4 Adaptation

Adaptation is an important aspect in Cancerscape, given that different adjustments are made at different times in the model that influence the agent decision-making and the overall model dynamics. Agents adapt to changes in their environment, and to the characteristics and behaviors of other agents as well. Certain behavioral rules – such as deciding to see a doctor for a yearly routine health exam may trigger changes in the agent himself (such as increasing his probability to adopt healthy behaviors) and in the

environment, as the PCP may suspect CRC also during routine health exams as well, and recommend the agent to get screened for CRC.

6.6.5 Stochasticity

Stochasticity is another important concept in Cancerscape. It can be observed both at cancer and at the agent level. Firstly, the individual agents are activated in random order. There is randomness involved in agent behaviors (such as their decision to comply with CRC screening, accept or reject the treatment recommended by the doctor, among others). However, for consistency between runs – an independent random number generator with a fixed seed number is used for model runs.

6.6.6 Collectives

Collectives in Cancerscape are represented as an intermediate level of organization that influences agent behaviors. Collectives are represented as individual social networks, connecting agents in Cancerscape. Also, agents can be thought as collectives themselves.

6.6.7 Observation

Observation in Cancerscape is accomplished through a graphical display of the model GUI and a series of plots displaying the statistics collected throughout the simulation. Several GUI layers are available for display, as shown in Figure 77. Furthermore, to enhance the observation processes, the results of the model are additionally displayed within a separate interactive dashboard, available for 30 individual separate run, as illustrated in Appendix A. JavaScript is used for creating the dashboard for a user-friendly interaction with model results.

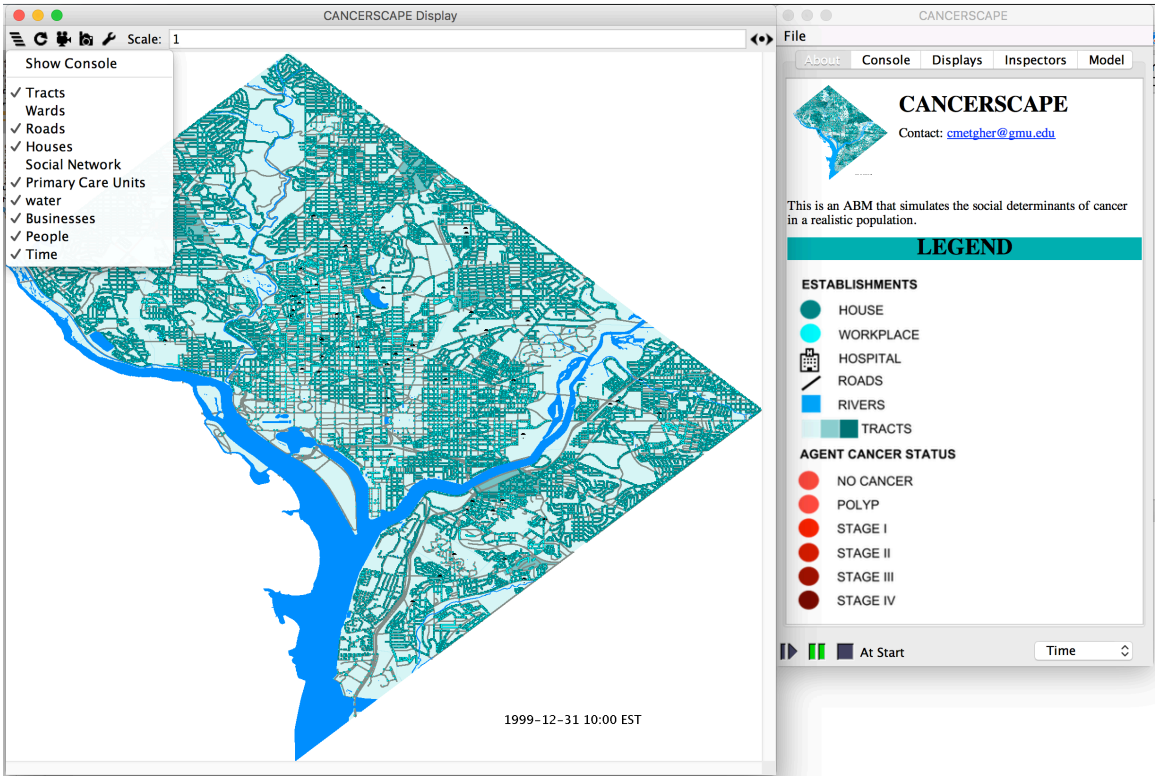


Figure 77: Close up of Cancerscape GUI showing the simulation layers available for display

A few examples of statistical plots provided in Cancerscape include CRC incidence, CRC prevalence, CRC mortality and a series of plots exploring the delay in CRC diagnosis by race and health insurance. The statistics in terms of CRC screening rates and the disparities persisting across the CRC continuum are described in great detail in the next chapter.

6.7 Details

Following the principles of the ODD protocol, in this section the details of the model initialization, input and sub-models will be provided.

6.7.1 Initialization

The model is initialized with the population of Washington, D. C. of the year 2000. According to the United States Census Bureau demographic data, the population of DC in 2000 consisted of 572,059 individuals. The generated population resembles the same characteristic as the demographic characteristic of DC in 2000 by tract (such as the same age, racial and income distribution among others). The simulation starts by reading in the spatial environment files: the GIS layers, the roads data shapefile, and the demographic data. At the beginning of each simulation, the user can set the desired global model parameters. Therefore, at the initialization stage the users vary the input parameters and then observe the impact of these on the model results. The initial number of individuals living with CRC and their distribution by stage, for the year 2000 was obtained from the National Center for Health Statistics (NCHS, 2015) public use data sets for Washington, D. C. The initial distribution of CRC polyps by age and race within the initial population is estimated according to the data provided by Gopalappa et al. (2011), and presented in Table 12, where F is described as a random variable denoting the family history of CRC for a randomly selected individual.

Table 12: Percentage incidence of polyp≤5mm at age [a, b], given R = r and F = f, for polyp pathway of CRC development (Source: Gopalappa et al., 2011)

	All Race		R = Caucasian		R = African American	
Age Group $[a, b]$	$F > 0$	$F = 0$	$F > 0$	$F = 0$	$F > 0$	$F = 0$
$[40, 49]$	0.74	0.12	0.73	0.13	0.90	0.10
$[50, 64]$	4.33	0.70	4.25	0.74	5.28	0.58
$[65, 74]$	4.54	0.73	4.46	0.78	5.53	0.61
$[75,]$	2.13	0.34	2.09	0.36	2.59	0.29

After the GIS layers have been loaded and the road network has been generated, the population of artificial agents is created by tract. The generated agents are contextualized in the environment and initialized. After initialization, the agents scheduling, decision making, and behaviors take place and the model dynamics unfolds. The statistical data collectors are initialized and scheduled accordingly, as described in Sections 6.4 and 6.5.

6.7.2 Input

After initialization, the environmental conditions in Cancerscape remain constant over space and time. To create the spatial environment, Cancerscape takes several shapefiles as input. As mentioned previously, the shapefiles have been pre-processed within the ArcGIS system before being used as input in the model.

Cancerscape is a data-intensive simulation, taking as input a large quantity of information from multiple sources. To recap, the following types of data go into the model:

- ***Demographic:*** The population of agents by age, sex, employment by tract, household size, types and distribution, houses data.
- ***Geographic:*** Tract-level data for Washington, D. C. from the U.S. Census Bureau (2000), the DC GIS shapefiles (hospitals and street segments) for the year 2000 (DC GIS, 2015).
- ***Road Network:*** TIGER road network data. The 2000 United States Census Bureau data provides the entire DC road network including the residential roads.
- ***Colorectal Cancer:*** Colorectal cancer symptoms, staging, polyp progression rates,

and the initial distribution of CRC cases (NCHS, 2015). CRC risk factors and behavioral characteristics distribution (CDC BRFSS, 2000) as well as other relevant factors relevant to the CRC development and progression are included.

6.7.3 Submodels

Cancerscape is supported by several submodels that are interconnected, with each of them contributing to proper model functioning.

6.7.3.1 Agent Daily Activities

The agents' daily activities were chosen to partially match the activities from the American Time Use Survey (ATUS) for the year 2000. When selecting their activities each day, the agents take into consideration their individual demographic attributes (such as age and employment status) and the time of the day. The employed agents consider the time of the day when they have to leave for work. While at work, they have opportunities for interaction with other agents that can, in turn, influence the agent behaviors and decision-making. Agents who are unemployed are left to stay at home, doing household activities, looking for jobs or spending time categorized as 'leisure'. All agents come back home between 5 and 7 PM and some of them, depending their attributes – choose to either engage in recreation activities, physical activity or participate in household activities. Afterwards, the agents pick a random time between 10 PM and midnight and go to sleep. The agents then wake up in the morning between 7 and 9 AM. The agent's daily activities schedule is implemented as a Java Finite State Machine. A graphical representation of the Agent Daily Activities is provided in Section 5.3.

6.7.3.2 Cancer Surveillance

On a daily basis, the agents run the Cancer Surveillance schedule. Cancer Surveillance is based on the Andersen Model of Total delay that was described in more detail in Section 3.8. Several reasons may lead to delaying a cancer diagnosis or treatment, such as:

- *Detecting unexplained signs and symptoms*: that occurs when individuals become aware of changes in their bodies regardless of their perceived meaning attributed to symptoms. For an individual to proceed with inferring illness, the symptoms first need to be perceived as serious (Byles et al., 1992). One factor influencing delay at this stage is knowledge about cancer (Mor et al., 1990). If the agent has knowledge about CRC, he is more likely to link his symptoms to the disease, as opposed to attributing symptoms to other common illnesses.
- *Inferring illness*: After detecting unexplained changes, agents can either seek medical attention or self-medicate (Bernal et al., 2006). In the context of colorectal cancer, individuals with symptoms such as rectal bleeding tend to attribute their symptoms to hemorrhoids, self-medicating and increasing the delay in presenting to a healthcare professional (Talley & Jones, 1998). The time between detecting unexplained changes and inferring illness is referred to as the "appraisal delay". Factors such as pain or bleeding were found to increase the appraisal delay (Safer et al., 1979). For many patients, friends and family play an important role when trying to make sense of symptoms. Individuals who discussed their symptoms with friends were more likely to experience diminished

patient delay (Smith et al., 2008). Individuals proceed to decide to see a medical doctor only after the symptoms fail to naturally resolve.

- *Deciding to see a doctor:* As time goes by and the symptoms fail to resolve, the agent takes the decision to see a PCP. This waiting period in this process is also known as "illness delay". Fear may serve as a stimulating factor in health-seeking behaviors (de Nooijer et al., 2001). Discussing symptoms with others may help the agent to better interpret their symptoms and seek medical help (Smith et al., 2008). An individual may decide to postpone seeking medical care due to a variety of reasons, such as health care affordability, health insurance status or peer pressure (Cacioppo & Petty, 1982).
- *Making a doctor appointment:* The time between taking the decision and proceeding to make that appointment with the doctor is also referred to as "behavioral delay". This type of delay can be influenced by a variety of factors, such as health insurance status, proximity to a healthcare facility, and/or the inability to find time for the doctor visit (de Nooijer et al., 2001).
- *First getting medical attention:* This refers to the time since the patient made an appointment until he got his first consultation with the PCP. This period is also known as the "scheduling delay". On the patient side, scheduling is often delayed at this stage due to fear – specifically due to the fear that symptoms might be linked to cancer. On PCP side, it is commonly delayed due to the lack of providing flexibility in making an appointment at the times that are suitable for

patients (de Nooijer et al., 2001). Scheduling delay has generally been understudied (Walter et al., 2012).

- *Beginning treatment / dying:* This delay stage refers to the period of time from receiving medical attention for the first time regarding the initial complaint, and the actual beginning of treatment. This type of delay focuses mostly on doctors/medical institutions, and less attention is given to it in this work.

The Cancer Surveillance is scheduled to run on a daily basis, but can be interrupted or redirected by external events. In a similar manner, the states of the Cancer Surveillance schedule can trigger other states and processes, as shown in Figure 78.

The Andersen Model of Total Patient Delay is one of the most widely used models for studying delay in establishing a cancer diagnosis (Walter et al., 2012; Ford & Farah, 2013). The factors influencing the transition from one state to the next may differ from one disease to another, as well as from one cancer type to another. The appraisal delay often constitutes the majority of total patient delay (Andersen et al., 1995). To the best of author's knowledge, no research has been yet published involving a computational model based on the Andersen Model of Total Patient Delay. This dissertation may be the first attempt to do so.

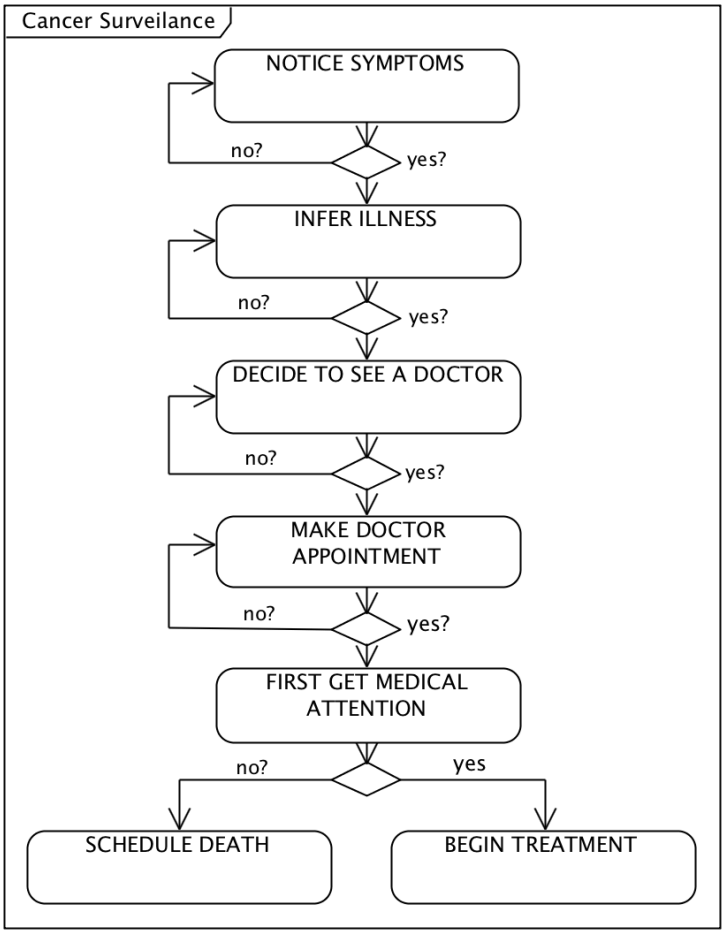


Figure 78: Cancer surveillance in Cancerscape

6.7.3.3 Modeling Colorectal Carcinogenesis

Carcinogenesis is known to be a multistep process. The first model of a multistage carcinogenesis emerged in the 1950s (Armitage & Doll, 1954; Fearon & Vogelstein, 1990; Vogelstein & Kinzler, 1993), and it was focused on CRC, as shown in Figure 80.

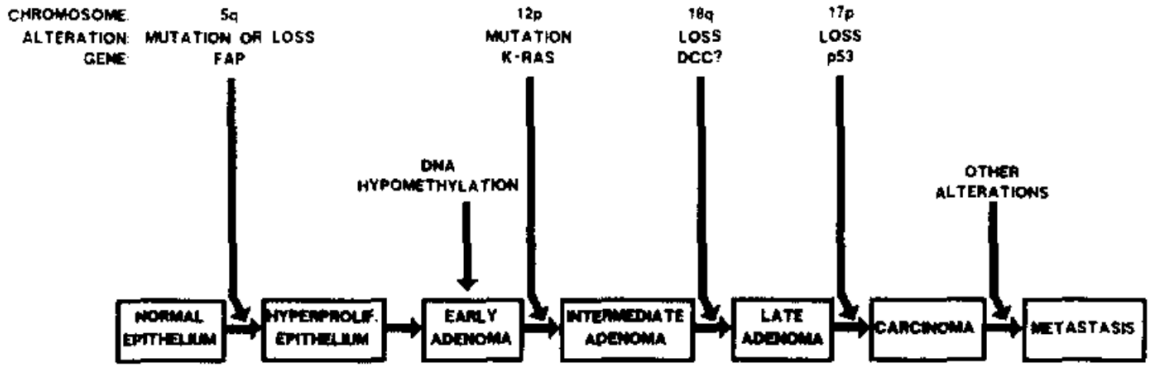


Figure 79: The Vogelstein model for multistep carcinogenesis (Source: Vogelstein & Kinzler, 1993)

In this initial model, each stage is associated with mutations in additional genes, with CRC progressing from adenomatous colon polyps to cancer through a series of steps. The multistep colorectal carcinogenesis is a system in itself; that proceeds from normal colon/rectal mucosa to a small polyp that grows developing bigger in size and ending up advancing into invasive cancer. It is a stepwise process, with each step being driven by well-defined genetic alterations.

Similarly, the colorectal cancer progression in Cancerscape was implemented as a stepwise process that begins with a polyp and develops into clinically detectable CRC within the model, as shown in Figure 80.

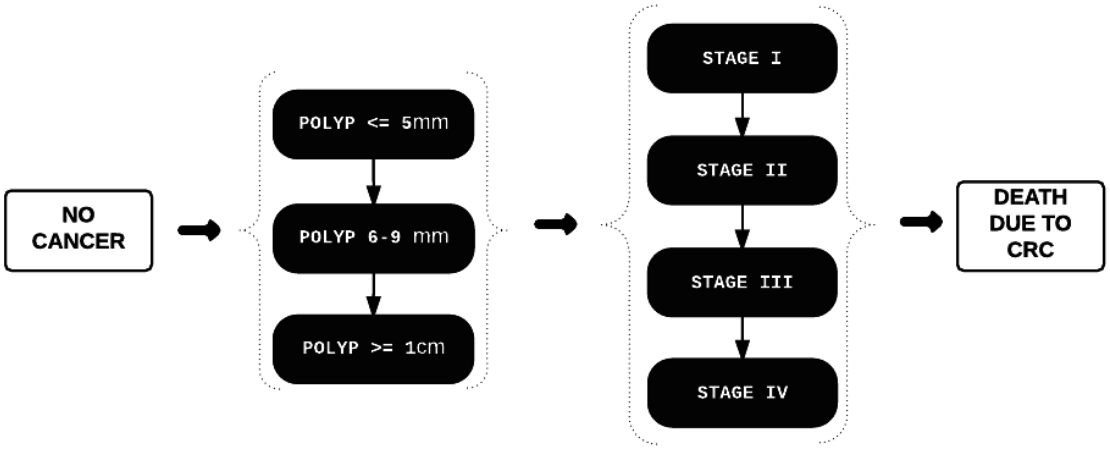


Figure 80: Natural history of CRC

It takes a long period of time for cancer to develop, and the duration of time a person spends in one stage before transitioning to the next stage largely varies depending on individual socio-demographic characteristics (Brenner et al., 2007). Gopalappa et al. (2011) estimated the transition time units (e.g., days, weeks) between stages, as shown in Table 13, with F being a random variable denoting the family history of CRC for a randomly selected individual.

Table 13: Mean times to progress from event *i* to event *j*, given R = r and F = $f\left(\frac{1}{\lambda^{j}_{i|rf}}\right)$ (Source: Gopalappa et al., 2011)

	All Races		R = Caucasian		R = African American	
event i → event j	$F > 0$	$F = 0$	$F > 0$	$F = 0$	$F > 0$	$F = 0$
p_5 → in-situ [a]	23	41.6	21.5	39	29	50
in-situ → local	3.4	3.4	3.5	3.5	3.1	3.1
local → regional	5	5	4.5	4.5	3.5	4
regional → distant	0.95	0.95	0.95	0.95	0.88	0.9

While the initial number of CRC cases for the year 2000 was derived from the SEER data, the initial population with polyps was initialized with the estimated polyp

incidence rates, as explained in Gopalappa et al. (2011). The distribution of CRC stages for both men and women by race/ethnicity is presented in Table 14. Not all polyps may progress into cancer. Social conditions and behavioral risk factors also modulate individual cancer risk (Link & Phelan, 1995).

Table 14: CRC stage distribution (%) by race/ethnicity, 1999-2006 (Source: Altekruse et al., 2007)

	MEN				**WOMEN**			
RACE/ETHNICITY	*Stage I*	*Stage II*	*Stage III*	*Stage IV*	*Stage I*	*Stage II*	*Stage III*	*Stage IV*
Non-Hispanic White	42	35	19	4	40	36	19	5
Non-Hispanic African American	36	34	24	5	36	34	24	6
Asian American / Pacific Islander	40	37	19	4	38	41	18	3
American Indian / Alaska Native	36	37	24	3	40	39	17	3
Hispanic	37	36	22	5	37	36	21	5

The development and progression of a colorectal polyp in an otherwise healthy agent depends on the risk factors and behaviors: an agent who is a heavy drinker will be more likely to develop a polyp compared to a non-drinker – everything else being the same. Once the agent enters the carcinogenesis process, symptoms are attributed according to their cancer stage. The distribution of the most commonly observed symptoms per stage is provided in Table 15.

Table 15: Relation between stage and the most frequently observed symptoms (Source: Gonzalez-Hermoso et al., 2004)

	TNM stage				
Parameter	1	2	3	4	*p*
Rectal bleeding					< 0.001
No	27 (7.7%)	143 (40.6%)	92 (26.1%)	90 (25.6%)	
Yes	72 (23.4%)	123 (39.9%)	70 (22.7%)	43 (14.0%)	
Change in bowel habit					0.56
No	84 (15.7%)	216 (40.4%)	132 (24.7%)	103 (19.3%)	
Yes	15 (12.0%)	50 (40.0%)	30 (24.0%)	30 (24.0%)	
Abdominal pain					0.009
No	93 (16.9%)	223 (40.5%)	128 (23.3%)	106 (19.3%)	
Yes	6 (5.5%)	43 (39.1%)	34 (30.9%)	27 (24.5%)	
Anemic syndrome					0.06
No	98 (16.6%)	241 (39.4%)	150 (24.5%)	123 (20.1%)	
Yes	1 (2.1%)	25 (52.1%)	12 (25.0%)	10 (20.8%)	

As the polyp grows and symptoms appear, the agent may start noticing these symptoms, and the Andersen Model of Delay is implemented to register the days of delay in seeing the doctor following the process of noticing symptoms (due to various reasons). The agent may start noticing symptoms for the first time with a delay - either because he does not have enough CRC knowledge, or he decides to self-medicate. For instance, people experiencing rectal bleeding (which is a common symptom of CRC) may attribute this symptom to hemorrhoids, therefore, increasing the delay in presenting to a primary care doctor (Bernal et al., 2006).

The risk of getting CRC increases with increasing age, as there is more time for agents to engage in risky behaviors that in turn increases their cancer risk, as shown in Figure 81. The highest risk of contracting CRC occurs later in the life of an agent. During their life cycle, agents can either die of natural death or due to CRC. Once the agents enter the initiation stage, the carcinogenesis processes begin, and if nothing else interferes (such as the agent getting cancer screening), the polyps advance into cancer. If no treatment is performed, the cancer continues to grow and metastasize, leading to cancer-related deaths.

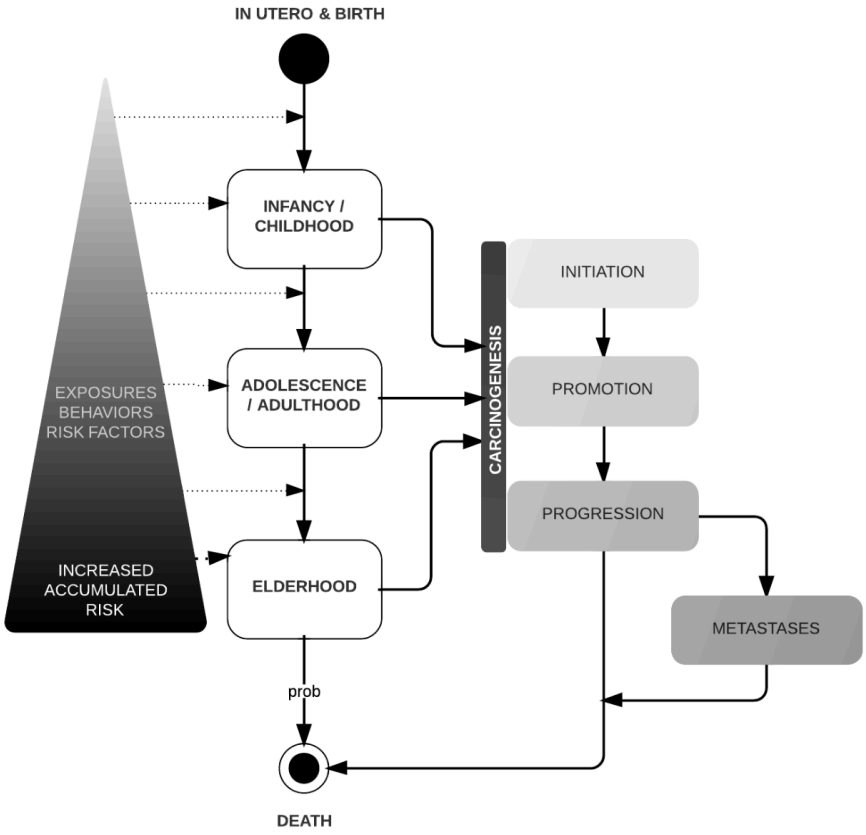


Figure 81: The risk of contracting cancer increases with age

To incorporate the effects of the modifiable and non-modifiable risk factors on the individual cancer risk, cancer index was designed. The cancer index is a construct representing the individual cancer risk at the agent level. As shown in Figure 82, the individual cancer index ranges from 0 to 102 where a cancer index of zero means the agent is healthy with no traces of colorectal polyps or cancer, and at a cancer index of 102 the agent dies. These values were reached to as a result of intensive calibration efforts.

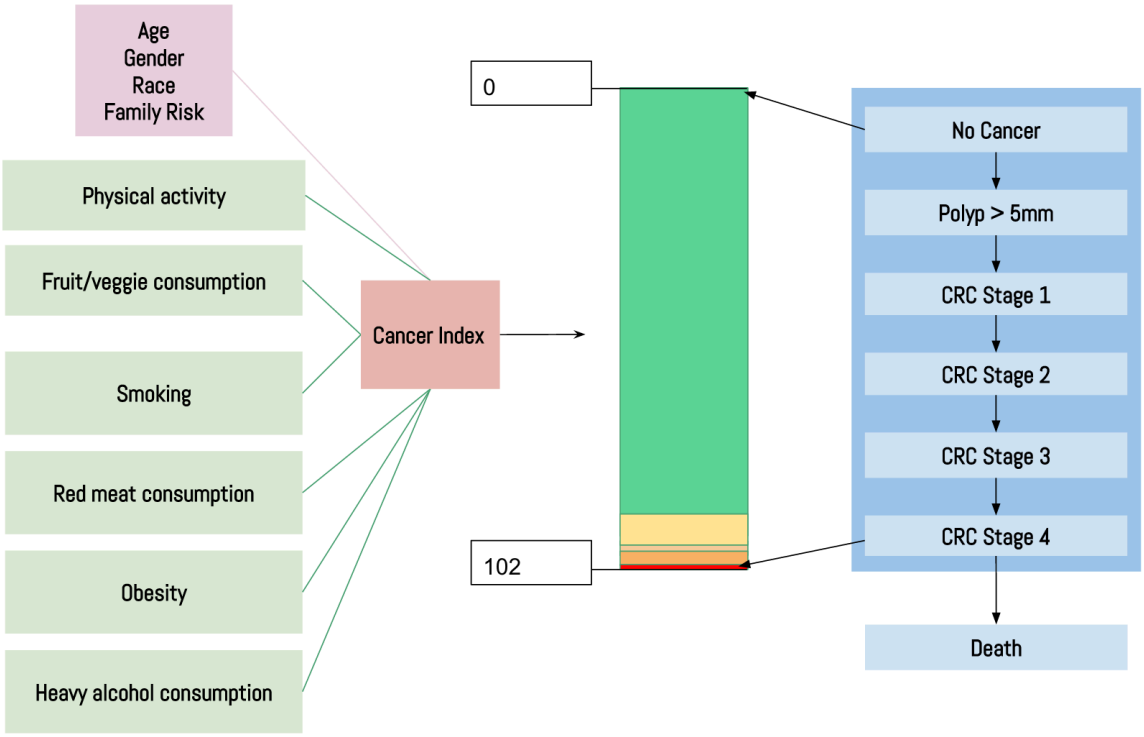


Figure 82: The cancer index diagram

Cancer index is modulated by modifiable and non-modifiable risk factors. Among the non-modifiable factors modulating cancer risk are age, gender, race and family risk. Cancer index also changes every time an agent decides to engage in physical activity, decides to smoke, consume excessive alcohol, fruits and vegetables or red meat, as well as when his BMI level changes. These factors are referred to as modifiable factors. As the cancer index progresses, the individual cancer stage changes as well.

6.7.3.4 Behaviors

Agents decide to engage in behaviors on a regular basis, or whenever these are triggered from the other external FSMs. Agents use the "follow the average" (FTA) rule for making decisions, as described by Hammond & Epstein (2007), as it has been elaborated

in detail in previous sections. In short, agents interact with their social network and with their social networks. When doing so, they evaluate their own status and attributes and decide to behave in a certain way depending on the predominant mean characteristics of their social environment. Social conformity towards an ideal group mean was shown to be an important driver of obesity dynamics, and the FTA rule was shown to be of clear relevance for studying these complex social processes. It is hypothesized that the FTA rule may as well underlie other forms of social adjustment where social conformity could lead to unexpected dynamics (Hammond & Epstein, 2007). This rule may be implicated in processes beyond obesity, and it may well impact the social determinants of cancer, given the similarities between the determinants of the two chronic conditions at the individual level.

6.7.3.5 Environment

The spatial environment of Cancerscape represents the geographic map of Washington, D. C. The shapefiles for constructing the spatial environment involve the primary care facilities, roads and demographic data files for population generation. The description of the spatial environment modeled in Cancerscape and the detailed construction of the environment was described earlier in this chapter.

6.7.3.6 Population Generation

Generating the population of the model began with collecting and handling the data in order to bring it to a standardized form acceptable by the model. The process of collecting the data and preprocessing is described in more detail in Section 6.2. The population is generated after the spatial data is loaded into the model. Starting initially with the method for generating houses and households based on U. S. Census Bureau

(2000) data for Washington, D. C. and distributing them across the road system presented by Wise (2015), agents are subsequently attributed individual characteristics and risk factors obtained from the CDC BRFSS (2000). A high-level diagram of the population generation procedures within Cancerscape is presented in Figure 83:

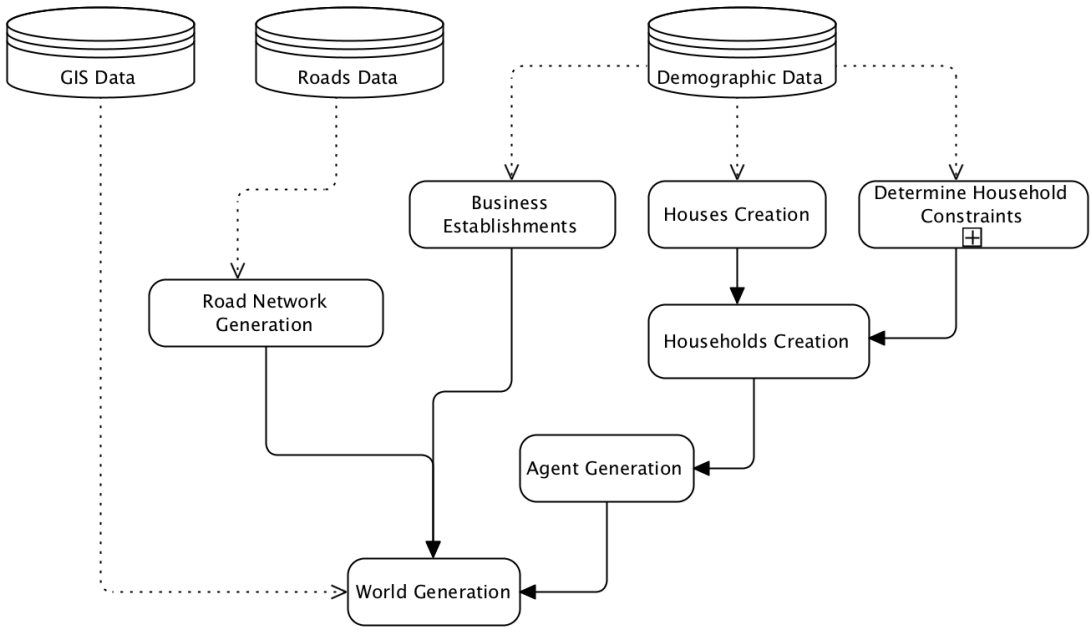


Figure 83: High level diagram of population generation within the Cancerscape world

As part of the world generation, the houses in which Cancerscape households and agents reside are initially generated within the spatial environment of the model. The information about the number of houses per tract has been obtained from the 2000 United States Census Bureau data for Washington, D. C. Having the number of houses for each tract, the houses were generated across the both sides of the residential roads for each individual tract.

The distribution of houses across the roads in Cancerscape is presented in Figure 84. Along with the number of district households, the United States Census Bureau also provides data on individual household constraints, that is the total proportion of the population residing in group housing. This was incorporated into the model as well to achieve a more realistic distribution of the agent population across the Cancerscape environment.

After the housing units have bee distributed across the roads with the corresponding density based on census characteristics, the households have been assigned to each housing units according based on the household constraints data available for each tract.

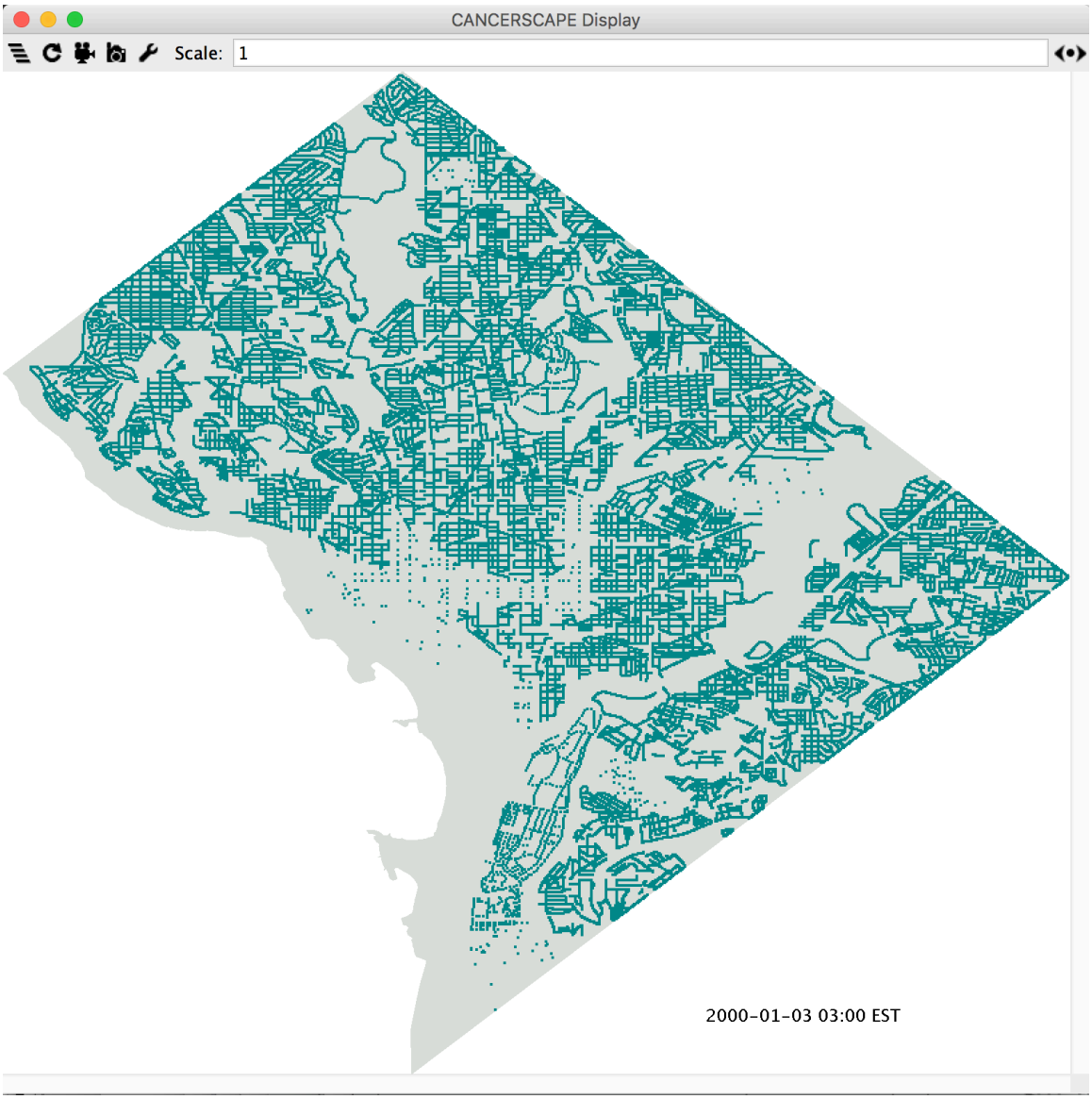


Figure 84: The overall distribution of houses across Cancerscape environment

A close-up of houses generated along the road networks in Cancerscape is presented in Figure 85.

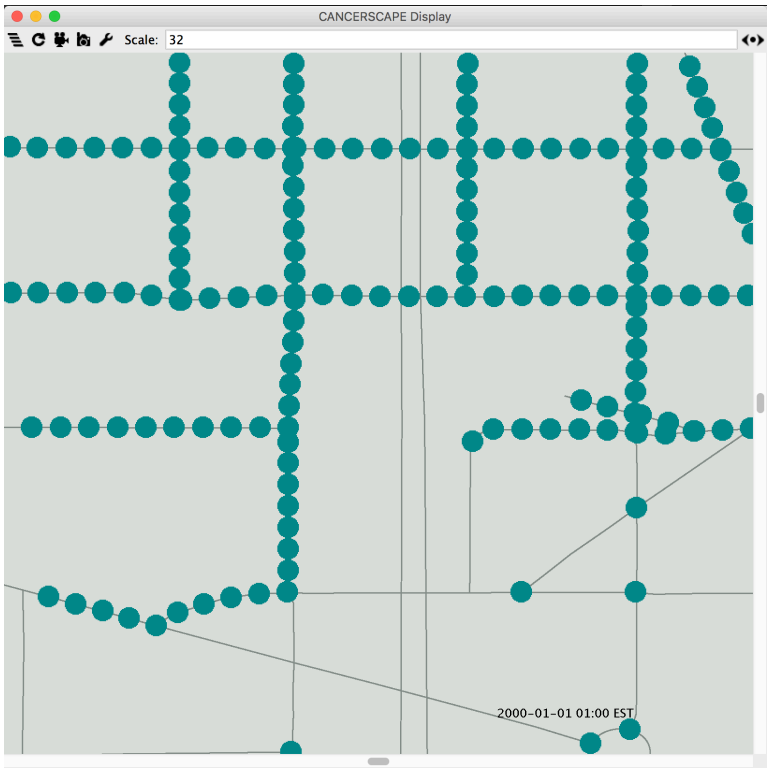


Figure 85: A close-up of houses along the road networks in Cancerscape

The U. S. Census Bureau (2000) data for District of Columbia also includes constraints for individual households, which were used for creating the households for each simulated house. The attribution of agent characteristics is performed as follows: initially, the agent level socio-demographic characteristics are defined, and then the individual level characteristics are assigned. An example showing the process of assigning education based on race, age and gender is presented in the form of pseudocode in Table 16. Because the model is run for five years, births and deaths are not implemented. However, this would be an interesting extension for future versions of the model, given the computational resources to run the model for more years.

Table 16: Pseudocode showing the process of assigning education based on race age and gender

1. Defining the distribution	2. Assigning agent characteristics based on it

Education = {
white: {
female: {
bachelor: 33.04,
graduate: 41.51
...
},
male: {
...
}
...
}
...
}
	population.ageGreaterThan(18)
	.byRace(**WHITE**)
	.byGender(**MALE**)
	.getPercent(Education.male.white.**bachelors**)
	.shuffle()
	.attribute("education", **BACHELORS**);
	
	population.ageGreaterThan(18)
	.byRace(**WHITE**)
	.byGender(**MALE**)
	.getPercent(Education.male.white.**graduate**)
	.shuffle()
	.attribute("education", **GRADUATE**);
	...

Agents are portrayed and color-coded according to their cancer status, as shown in Figure 86. Agents with the "NO CANCER" status are portrayed in light red. Their color changes as the agent progresses from one cancer stage to the next. The darker the hue of red, the more advanced the CRC stage for that specific agent is.

Other visualization types available in Cancerscape, such as visualization by tracts that are color-coded according to the number of agents with CRC within them, will be further discussed in Chapter 7.

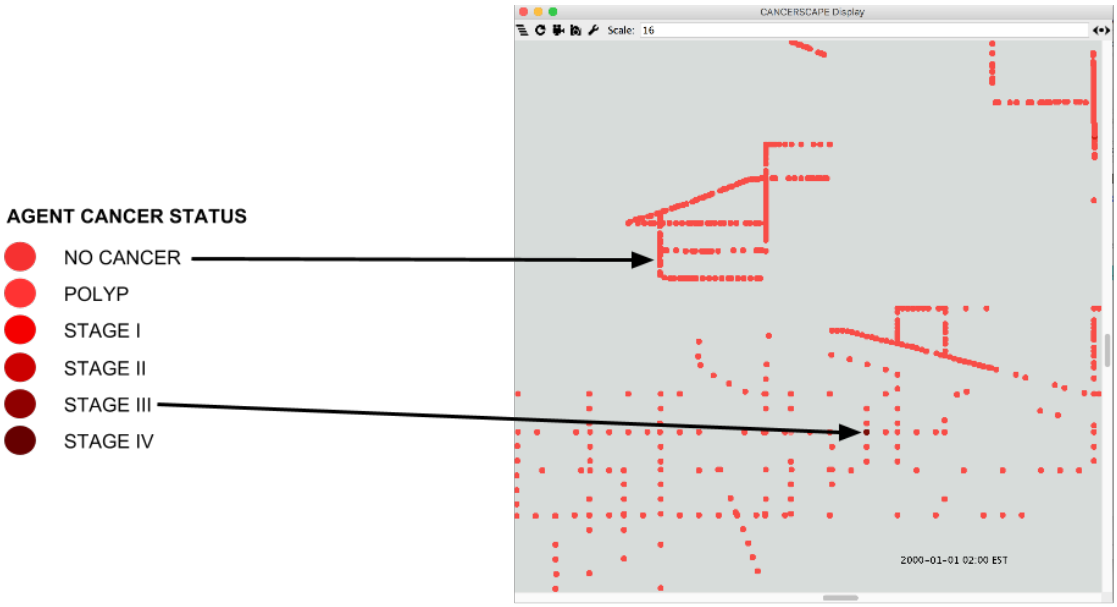


Figure 86: Agents color-coded by their cancer status

The next subsection proceeds by explaining the social networks, their attribution processes as well as the characteristics of the resulting social network in Cancerscape.

6.7.3.7 Social Networks

Social networks influence individual behavior to a much greater extent than individual personal attributes (Valente, 2010). The link between social networks and physical health has long been established. Social networks can influence individual health behaviors in several ways: serving as a source of information regarding health, providing the basis for social support and influencing health-seeking behaviors. Before proceeding, a few basic concepts regarding social networks need to be reviewed. Other network terminology will be introduced as necessary throughout the text.

- A *node* (or vertex) – is an individual unit in the graph (in our case, nodes represent individuals).
- A *network* (system, network) is the collection of units that may be connected to each other.
- An *edge* represents a tie between two individual nodes.
- A *neighborhood* for a node represents the number of other nodes situated in its neighborhood.

Most social network studies approach social networks using either "whole-network" or "egocentric" designs. The "egocentric" approach studies the individual personal network, and its effect on that individual himself. The "whole-network" or socio-centric approach focuses on large groups of people, quantifying their relationships. This method studies how the patterns of interactions within groups affect the community as a whole (Chung et al., 2005; Scott, 2012; Marsden, 2002). Individuals can be part of numerous networks simultaneously. They can be a part of a social network at their jobs, in their families, neighborhoods, within a religious group, an interest group and/or among others.

The individual social network generation in Cancerscape was built with the goal of replicating to an appropriate degree the overall characteristics of a typical real-world social network. Several properties may characterize a typical human social network. Dunbar (1992) explored the size of an individual social network, extrapolating the results of studies on primates and taking into account the average size of the human brain. He found that humans could easily maintain a social network of 150 individuals (ranging

between 100-230 people). This number entered the academic literature as the "Dunbar Number". Dunbar's eponymous number takes into consideration only the close individual relationships (such as immediate family, close friends, extended family, coworkers) and excludes the connections that are trivial or not active at a particular time. McCarty et al. (2001) describe several methods for estimating the size of an individual personal network. Individuals who make up a typical social network range from immediate family, family of the spouse and best friends to coworkers and neighbors, as shown in Table 17.

Table 17: Average number of people and relation types (Source: McCarty et al., 2001)

Relation Types	Average Number
Immediate family	3.5
Other birth family	24.0
Family of spouse or significant other	12.3
Coworkers	35.6
People at work but don't work with directly	62.1
Best friends/confidantes	4.3
People known through hobbies/recreation	12.3
People from religious organization	43.4
People from other organization	17.1
School relations	18.3
Neighbors	12.8
Just friends	22.6
People known through others	22.6
Childhood relations	6.8
People who provide a service	7.7
Other	3.9

Another important aspect of a typical social network is that people tend to befriend people similar to them (be it regarding age, gender, race, education interests or religion, among other characteristics). This phenomenon is generally known as

homophily. Some general properties of the real-world social networks are summarized in Table 18.

Table 18: Properties of a human social network (Source: Wise, 2014)

Parameter	Value	Source
Average Degree	10	Eubank et al., 2004
Network Diameter	6	Eubank et al., 2004
Average Clustering Coefficient	0.480	Eubank et al., 2004
Average Path Length	6	Albert & Barabasi, 2000

To understand how the generated networks compare to the real world social networks – several characteristics should be considered. Real world social networks have a lower density with smaller clustering behaviors (Wellman & Worley, 1990). In network theory, a clustering coefficient is a classical measure of network connectivity looking at the degree to which nodes tend to cluster. The *clustering coefficient* of a node v with k neighbors is calculated as follows:

$$C(v) = \frac{|actual\ edges|}{k(k-1)/2} = \frac{2 * |actual\ edges|}{k(k-1)} \qquad \text{Equation 3}$$

The clustering coefficient in real-word networks is moderate (Watts, 2003). These networks are also characterized by an approximate power-law distribution of node degree (Albert & Barabasi, 2002), as well as by a low average network distance (Szabo et al., 2003).

The average network distance is determined according to the following equation:

$$Average\ network\ size = \frac{\log(m)}{\log\ (n)} \qquad \text{Equation 4}$$

where m is the number of people and n represents the average degree of all people (Bollobás & Riordan, 2003; Watts, 1999).

The social network generation for Cancerscape agents was implemented as an ego network for each single random agent. The large generated social network portrayed by degree is presented in Figure 87. The agent draws its number of friends from a power law distribution. He then proceeds to gather a list of agents to be connected with. The agent initially selects his contacts from the list of friends of friends, and then proceeding to fill in the remaining slots with random individual agents based on their individual characteristics, such as age and gender and geographic proximity between himself and the home location of the other agent.

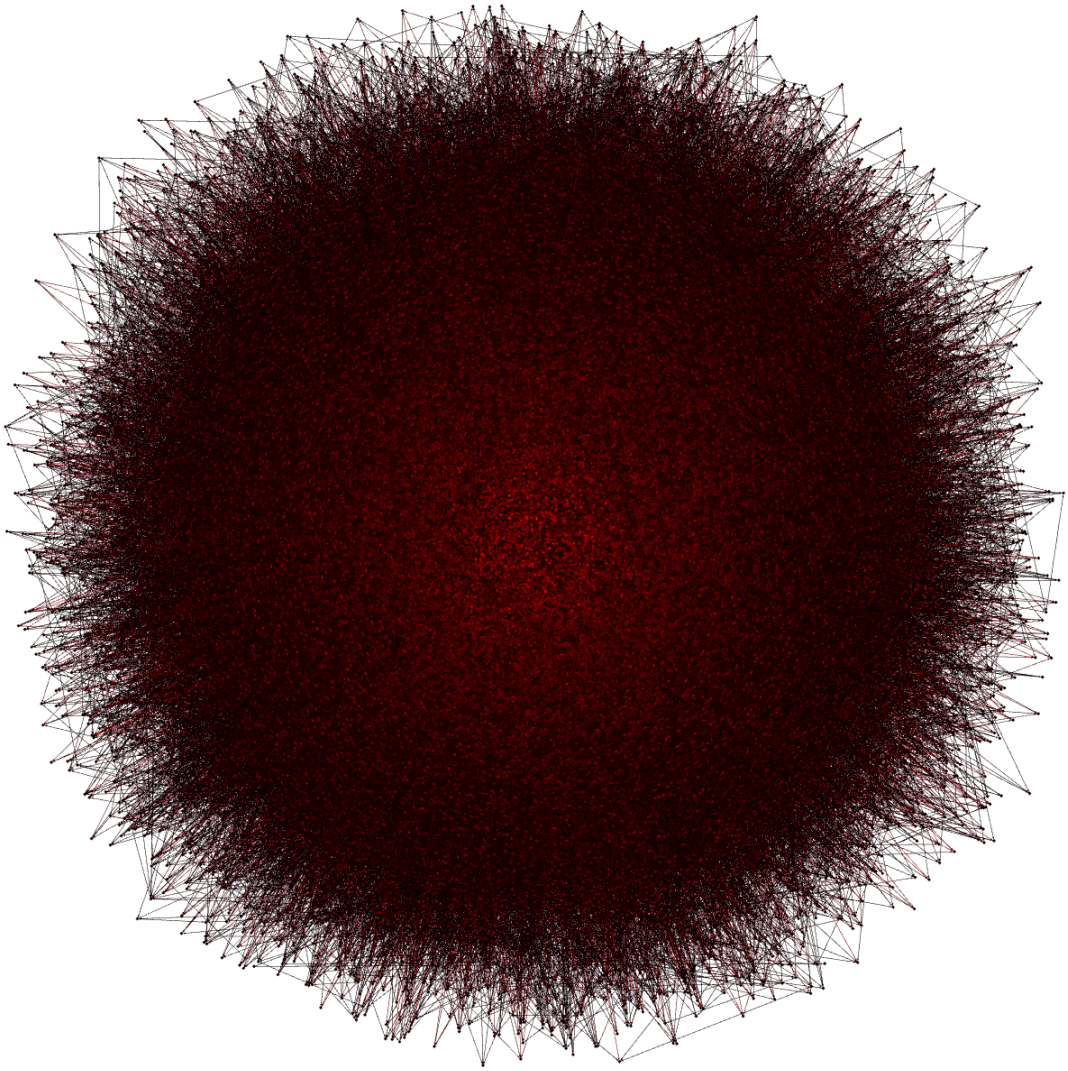


Figure 87: The generated agent social networks in Cancerscape

The resulting egocentric social network is undirected. The agents categorize their contacts based on a measure of social distance, taking into consideration their age, sex and proximity (home location). The degree distribution of the Cancerscape agents network is presented in Figure 88, exhibiting a power law distribution.

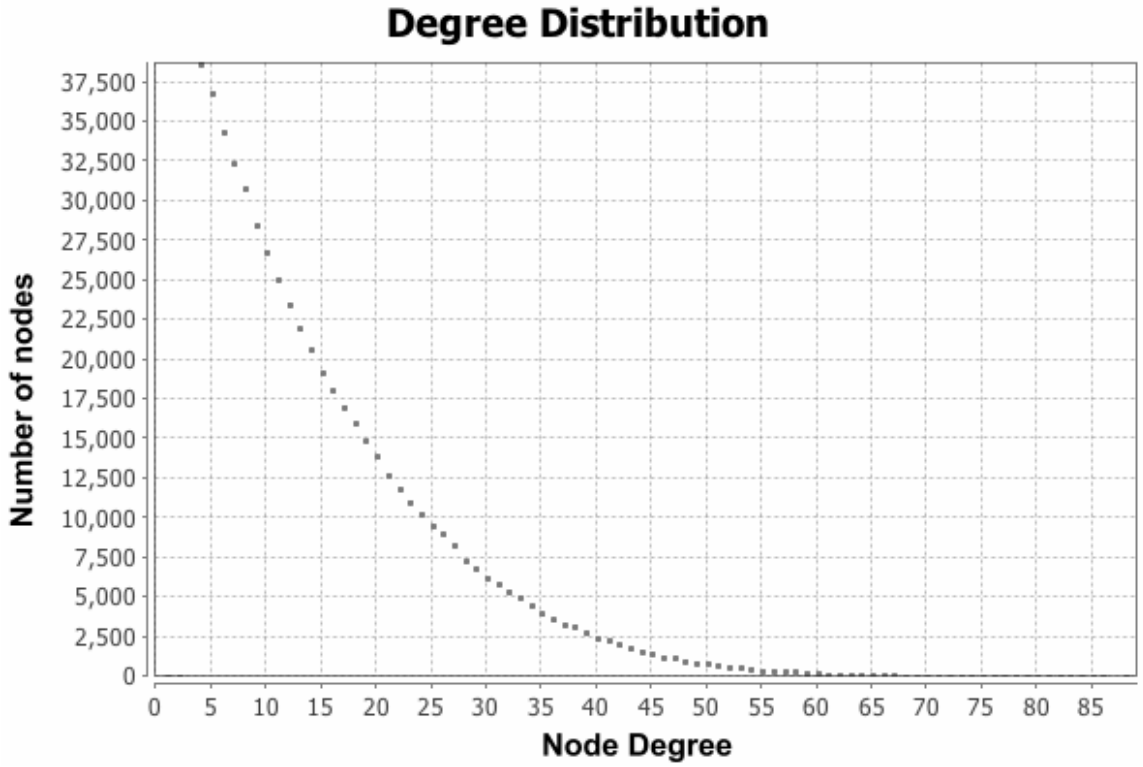


Figure 88: The degree distribution of the generated social network

The power law distribution of the resulting network is in line with current knowledge about real world human social network characteristics (Albert & Barabasi, 2002). The generated network is composed of 572,059 nodes and 9,057,666 edges. The average degree of the network is 15.833 contacts, and is in agreement with previous studies (Watts, 1999). A diameter of 8 for the generated social network largely aligns with previous results presented by Eubank et al. (2004). A summary of these characteristics is presented in Table 18.

6.7.3.8 Screening Submodule

In the context of CRC, screening represents the process of identifying the individuals who are at higher CRC risk among the asymptomatic agent population within the model. Screening has been shown to reduce CRC morbidity as well as mortality in several studies. However, the optimum approach for CRC screening has been the subject of scientific debate for decades.

The available methods for screening and testing for CRC have been described in detail in Chapter 3. To reiterate, the most commonly used CRC screening tests are: FOBT, Sigmoidoscopy, Barium Enema and Colonoscopy. Each screening method varies in price/copay and has its own advantages and disadvantages. In the case when an abnormality is detected after performing a CRC screening test, further actions or investigations are triggered. Is a polyp is found during a colonoscopy procedure – a polypectomy is performed, based on certain agent characteristics (such as whether the agent is covered by health insurance). This is one of the main advantages of the colonoscopy screening – it is also curative – as the polyps found during colonoscopy can be eliminated during the screening procedure itself. The main drawback of colonoscopy is the cost – as it is one of the most expensive CRC screening tests available. Both the advantages and disadvantages of the available CRC screening tests are summarized in Table 19.

Table 19: The pros and cons of selected CRC screening tests (Source: ACS, 2014)

Benefits	Performance & Complexity*	Limitations	Test Time Interval
Structural Exams			
Flexible Sigmoidoscopy			
• Fairly quick • Few complications • Minimal bowel preparation • Does not require sedation or a specialist	**Performance:** High for rectum & lower one-third of the colon **Complexity:** Intermediate	• Views only one-third of colon • Cannot remove large polyps • Small risk of infection or bowel tear • Slightly more effective when combined with annual fecal occult blood testing • Colonoscopy still needed if abnormalities are detected • Limited availability	5 years
Colonoscopy			
• Examines entire colon • Can biopsy and remove polyps • Can diagnose other diseases • Required for abnormal results from all other tests	**Performance:** Highest **Complexity:** Highest	• Full bowel preparation needed • Can be expensive • Sedation of some kind usually needed, necessitating a chaperone to return home • Patient may miss a day of work. • Highest risk of bowel tears or infections compared with other tests	10 years
Double-contrast Barium Enema			
• Can usually view entire colon • Few complications • No sedation needed	**Performance:** High (for large polyps) **Complexity:** High	• Full bowel preparation needed • Some false positive test results • Cannot remove polyps or perform biopsies • Exposure to low-dose radiation • Colonoscopy necessary if abnormalities are detected • Very limited availability	5 years
Computed Tomographic Colonography			
• Examines entire colon • Fairly quick • Few complications • No sedation needed • Noninvasive	**Performance:** High (for large polyps) **Complexity:** Intermediate	• Full bowel preparation needed • Cannot remove polyps or perform biopsies • Exposure to low-dose radiation • Colonoscopy necessary if abnormalities are detected • Not covered by all insurance plans	5 years
Stool Tests (Low-sensitivity stool tests, such as single-sample FOBT done in the doctor's office or toilet bowl tests, are not recommended)			
High-Sensitivity Guaiac-based Fecal Occult Blood Test (FOBT)			
• No bowel preparation • Sampling is done at home • Low cost • Noninvasive	**Performance:** Intermediate for cancer **Complexity:** Low	• Requires multiple stool samples • Will miss most polyps • May produce false-positive test results • Pre-test dietary limitations • Slightly more effective when combined with a flexible sigmoidoscopy every five years • Colonoscopy necessary if abnormalities are detected	Annual

In Cancerscape, agents have the opportunity to undergo screening in two situations: when they go for a yearly routine health exam and are recommended to undergo CRC screening by their PCP, or when they decide to undergo screening based on their knowledge about their own risk status and about CRC. For instance, if the agent has more than 50 years of age, is covered by health insurance and has knowledge about CRC

– he is more likely to know his risk category and has the possibility to take the individual decision to undergo appropriate CRC screening with the corresponding frequency.

As mentioned in Chapter 3, several factors besides CRC knowledge can influence screening rates. Among the most significant barriers to CRC screening are their cost and the general lack of health care access – often due to the lack of health insurance or the geographic location of the screening center (American Cancer Society, 2012).

The slow progress of CRC development provides a unique opportunity for prevention and early detection (Winawer & Zauber, 2002). By screening for CRC, polyps can be found early and removed, and/or cancer can be detected early when the treatment for CRC is most successful. This type of prevention that focuses on the characteristics of individuals has been referred to as "precision prevention" (Gillman & Hammond, 2015). It implies operating "above the skin" and aiming to alter individual behavior as well as community or group behavior.

6.7.3.9 Treatment Submodule

Agents can be diagnosed with CRC in two cases: by regular screening as a prevention measure, or by having their cancer detected as a result of having themselves checked by the doctor due to the presence of severe CRC symptoms. The patient, together with the doctor take decisions regarding treatment, taking into consideration the treatment options for the corresponding stage at diagnosis and the affordability of the patient to financially cover specific treatment regimens. Agents can also refuse to be treated, should they decide so (due to various reasons). Chapter 3 explains the recommended CRC treatment by cancer stage, together with their accompanying success rates. The success rate for each treatment option is also modeled, and there is a possibility for the patient to die

during the procedure. After the agent has been successfully treated, his cancer goes into remission. The agents have a possibility to have a CRC recurrence later in life, in which case the agents run the corresponding schedules, triggering the doctor visit routines. The processes involved during these screening and treatment procedures in Cancerscape are presented in Figure 89.

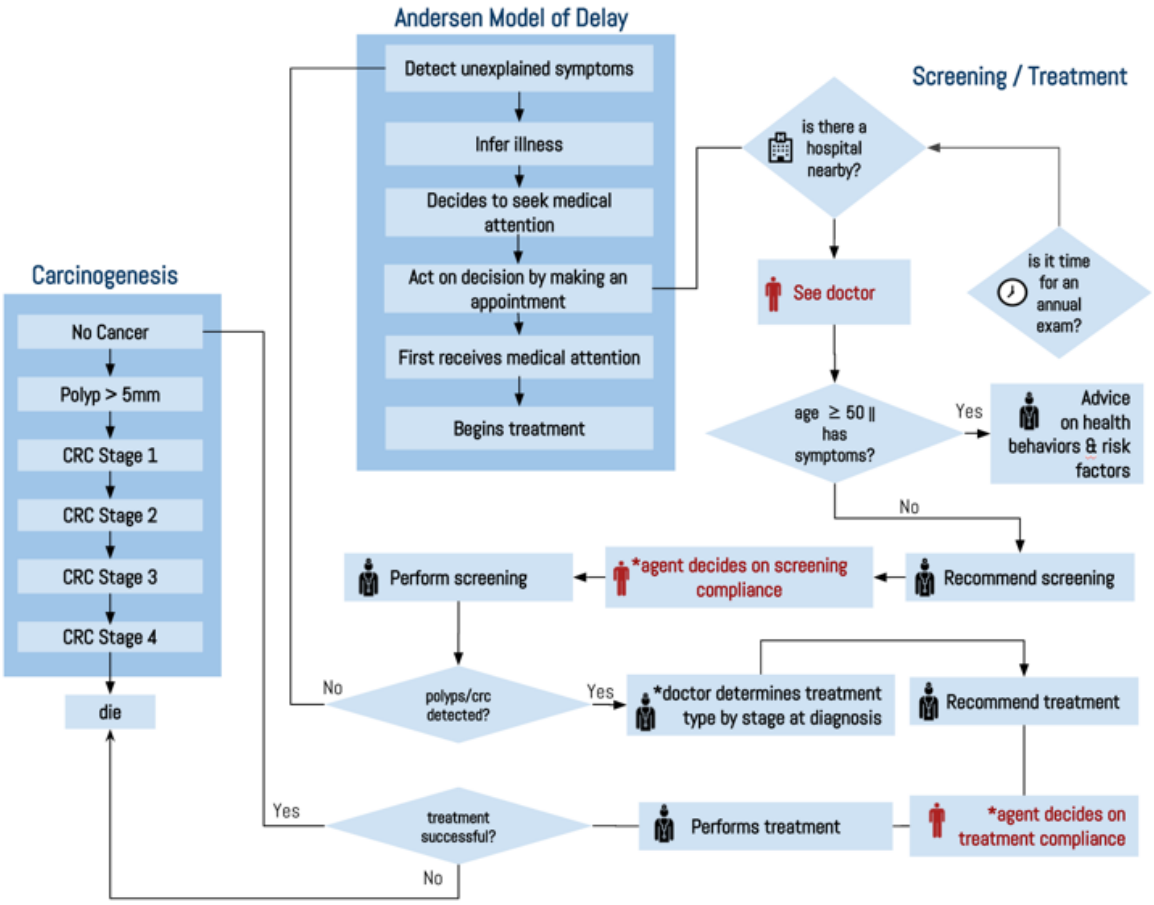


Figure 89: Overview of the screening and treatment processes in Cancerscape

During screening and treatment processes there are several opportunities for delay, that further influences early diagnosis, the CRC incidence as well as mortality

trends. Therefore, careful attention should be given to decreasing the delays at individual, doctor, as well as at system level.

6.7.3.10 User Interface (UI) Submodule

The model includes a UI sub-model, containing the files used to generate the display window is presented in Figure 90.

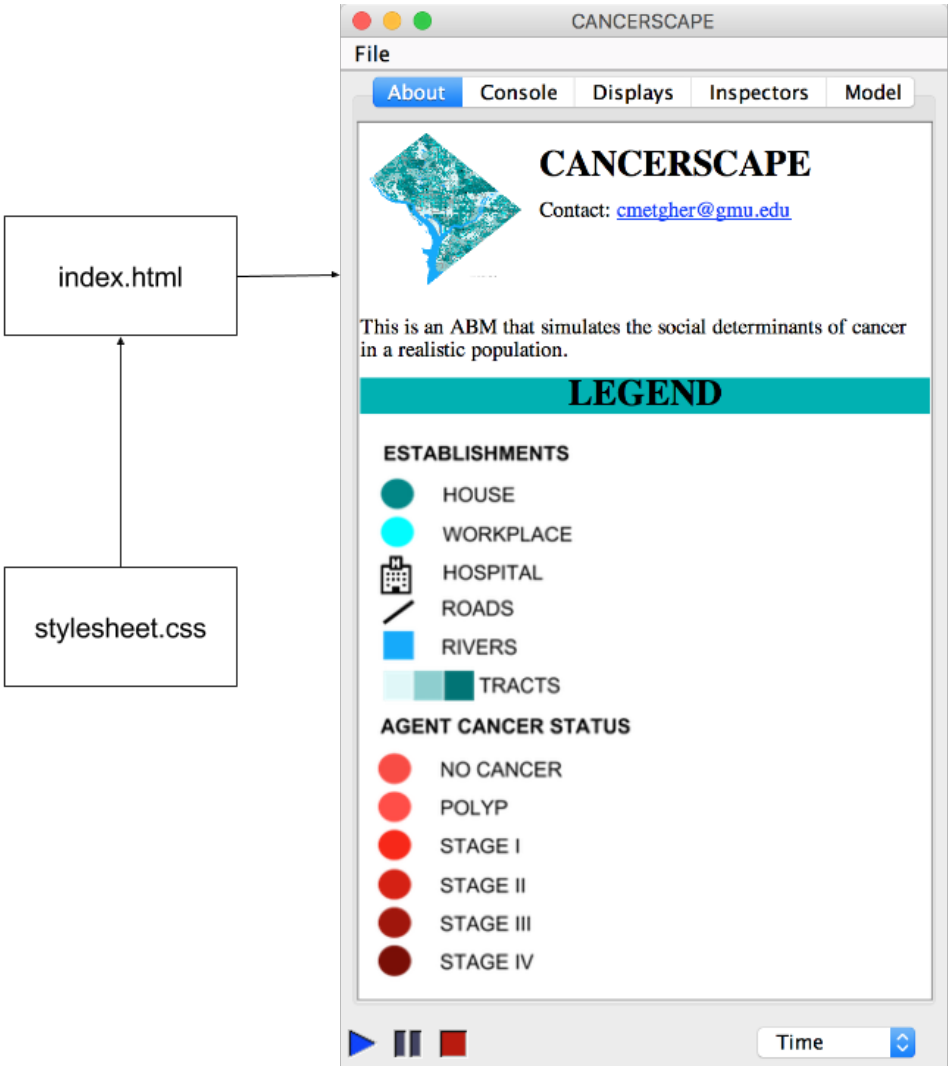


Figure 90: The initial UI console window

MASON has a built-in HTML viewer that allows creating a basic simulation description window. A 'stylesheet' file is added to personalize the simulation display window. The UI Submodule also contains the statistical charts, as well as the legend of the model, and other displays available to "Show" or "Hide" under the Cancerscape "Displays" tab.

6.7.3.11 Model output

Cancerscape model produces both geographic visual output and quantitative output. The results and outputs of the model are thoroughly discussed in the next chapter. A typical model run unfolds a GIS space with agents aging and moving through social systems. They carry out a daily set of activities, and are assigned a vector of risk factor values. As they live their lives within the artificial society, they may develop polyps; that may develop into CRC - as a result of genetic predisposition (presence of first or second-degree relatives with CRC), behaviors or exposures to risk factors. Cancerscape also outputs the CRC incidence, screening rates, survival, and other statistical charts, such as delay in diagnosis and treatment, and other measures that are discussed in more detail in the following chapter.

6.7.3.12 Testing submodule

The testing sub-module comes together with the other modules to ensure that proper model testing and verification was executed. The testing module contains several JUnit (a unit testing framework for Java programming language) tests, as well as mock objects simulating the behavior of real objects for testing purposes. The model testing strategy is elaborated in more depth in the next chapter.

6.8 Summary

In this chapter, Cancerscape model was described in detail, using the ODD protocol (Grimm et al., 2010). The model inputs and assumptions were introduced, and the data types that go into the model were discussed together with the motivation of using state machines for modeling agent behaviors. As part of the "Overview" section, the purpose of the model was discussed, the entities and state variables were presented, and the process overview and scheduling was examined. The following "design concepts" were described as part of the ODD protocol: interaction, emergence, adaptation, stochasticity, collectives and observation. Within the "Details" section, the model initialization details were considered, and the routines for screening and treatment were presented. The next chapter begins with describing the testing strategies for the model and proceeds with outlining the model verification, validation, and results.

CHAPTER 7: VERIFICATION, VALIDATION & RESULTS

Contents

7.1 Introduction 222

7.2 Model Verification and Validation (V&V) 224

7.2.1 Documentation 227

7.2.2 Debugging 228

7.2.3 Code Walkthroughs 229

7.2.4 Unit Testing (Iterative Programmatic Testing) & Mocking 229

7.2.5 Model Profiling 230

7.3 Model Calibration and Sensitivity Analysis 231

7.4 Model Validation 235

7.4.1 Model Validation Considerations 235

7.4.2 Matching Spatial Data 239

7.4.3 General Considerations 242

7.4.4 ABM and GIS Integration within Cancerscape 242

7.4.5 Model Scenario Analysis Overview 248

7.5 Theme I: Exploring CRC Incidence 249

7.5.1 Overall CRC Incidence Results 250

7.5.2 Disparities in CRC Incidence 253

7.6 Theme II: CRC Screening 256

7.7 Theme III: CRC Survival 259
7.8 Theme IV: Delay in CRC Diagnosis and Treatment..... 266
7.9 Scenarios and Discussion of Results 269
7.9.1 Scenario 1: Reducing CRC Screening Age for Average Population..... 269
7.9.2 Scenario 2: Increasing CRC Knowledge for General Population 274
7.9.3 Scenario 3: Increasing the Health Insurance Coverage for Colonoscopies.... 280
7.10 Summary 285

> "Previously, scientists had two pillars of understanding: theory and experiment. Now there is a third pillar: simulation"
>
> (Steven Chu, 2012)

7.1 Introduction

This chapter discusses the process of verification, validation of Cancerscape and presents the model results. Experimenting with the model is performed under several themes, such as CRC incidence, mortality, survival, screening, disparities, as well as delays. The results presented in this chapter refer to the 'base case', that is keeping all the parameters the same and running the model for five years, 30 times. Following the 'base case', three scenarios are proposed that explore CRC incidence and survival (among other measures). The first scenario focuses on lowering the recommended age for CRC screening, the second scenario focuses on increasing the population levels of CRC, and the third one explores the consequences of increasing the healthcare coverage for colonoscopies.

Verification and validation (often abbreviated as V&V) are tremendously important in ABM (Sargent, 2000; Galán et al., 2009). V&V procedures were briefly

introduced in Chapter 4. Verification stands for making sure that the model was build "right", while validation refers to the process of ensuring that the "right model was built" (Balci, 1998).

To select the appropriate combination of V&V strategies, it is helpful to think of models regarding their level of performance, or level of analysis. Axtell & Epstein (1994) suggested four progressive levels that assess the performance of a model, as shown in Table 20. These levels have widely been used as a reference for distinguishing between the performance ranks among ABMs. The four levels proposed by Axtell & Epstein (1994) are gradual, implying that a satisfactory model performance at Level *N* means that it is satisfactory at level *N-i* as well.

Table 20: The four levels of assessing ABM performance and analysis (Source: Axtell & Epstein, 1994)

Level	Description
Level 0	The model is a caricature of reality, as established through the use of simple graphical devices (e.g., allowing visualization of agent motion);
Level 1	The model is in qualitative agreement with empirical macro- structures, as established by plotting, say, distributional properties of the agent population;
Level 2	The model produces quantitative agreement with empirical macrostructures, as established through on-board statistical estimation routines;
Level 3	The model exhibits quantitative agreement with empirical microstructures, as determined from cross-sectional and longitudinal analysis of the agent population.

Various scientific disciplines can benefit from models at different levels to different extents. While Level 0 models can contribute to advancing certain disciplines, most scientific areas probably have few problems to which simulations of this sort can contribute and make significant progress in the field. However, Level 1 models and higher may contribute to addressing a wider range of problems (Axtell & Epstein, 1994).

Due to the fact that to date no other agent-based models have yet been proposed to simulate the social determinants of cancer, Cancerscape aims for a classification target closer to the Level 2, and will be verified and validated accordingly to match the requirements of the performance level it is attempting to attain. For achieving a higher level of validity quantitative agreement with empirical microstructures are necessary. This quantitative agreement would have to be established by cross-sectional and longitudinal analysis of the modeled population. The unavailability of micro level data on cancer development for defining and informing the micro level structures and processes in Cancerscape makes it difficult to aim for a Level 3 validation. However, this could be implemented in future versions of the model.

7.2 Model Verification and Validation (V&V)

Verification and Validation (V&V), together with calibration are often among the most challenging parts of building ABMs (Crooks et al., 2008). The tasks of V&V for any given ABM can get challenging, especially if the model is complex and involves a large parameter space. It would be desirable to refer to model's degree of validity, rather than addressing the model as simply valid or not valid (Law & Kelton, 1991). A detailed diagram explaining the V&V procedures in relation to developing simulation models, in

general, is presented in Figure 91. The V&V paradigm developed by Sargent (2000), and illustrated in Figure 91, shows the process of generating system level theories and simulation models in the context of V&V. The diagram distinguishes between the "Real World" and the "Simulation World" to highlight procedures for developing system theories and implementing simulations that can be verified and validated. Initially, the real world system is observed and an abstraction is developed based on the real observations, in order to generate system theories. V&V involves multiple steps and iterations, and can quickly become a complex and time consuming process, especially if the model is large and complex.

Even when the computer code for the model is written carefully, it is important to make sure that the code does what it supposed to do. As Gilbert (2007) writes: "*You should assume that, no matter how carefully you have designed and built your simulation, it will contain bugs (code that does something different to what you wanted and expected)*". Cancerscape was implemented using the Java programming language, MASON and GeoMason libraries, as well as using several additional Java libraries for testing purposes.

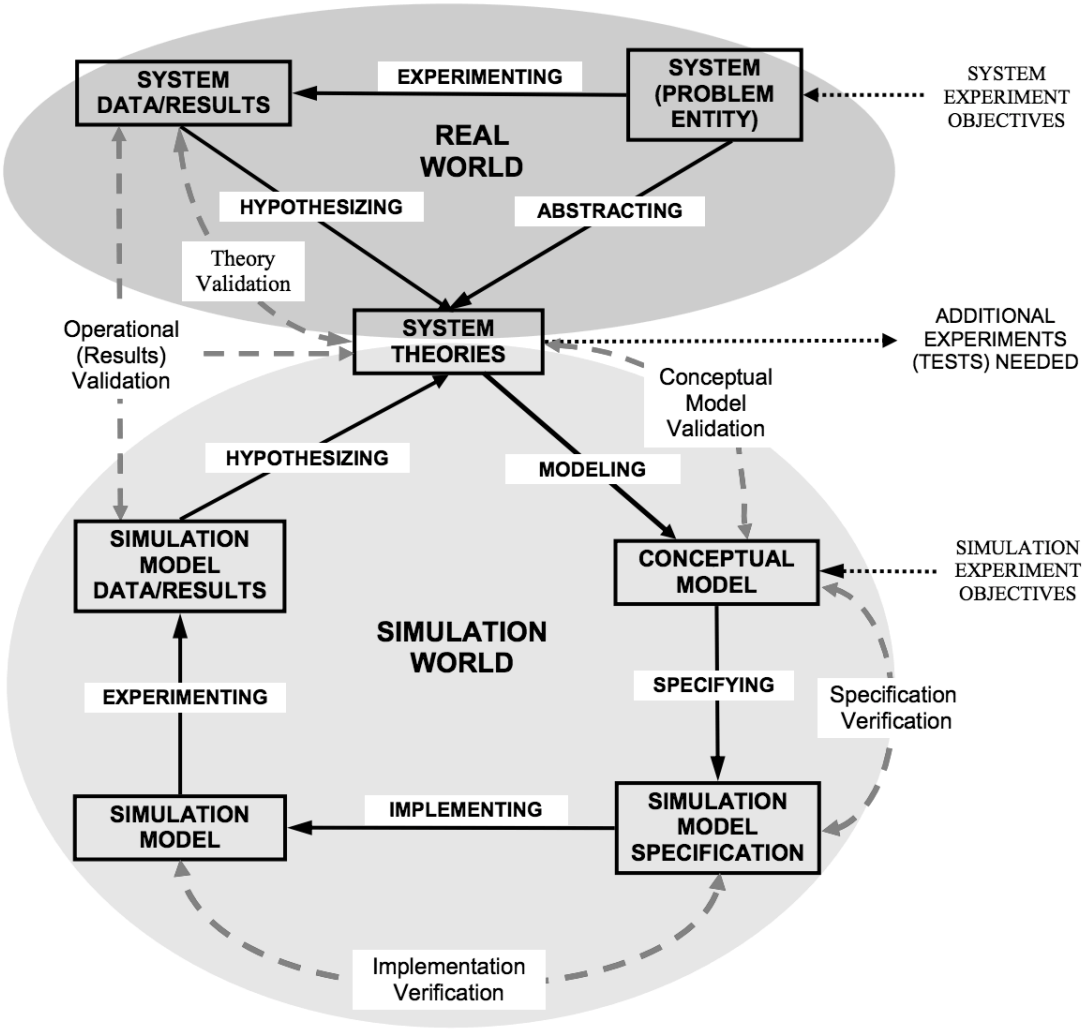


Figure 91: Real world and simulation world relationships with verification and validation (Source: Sargent, 2000)

Practical verification of Cancerscape was performed using the following procedures: code walkthroughs, profiling, and sensitivity analysis. Documentation, programmatic testing and test-case analysis are fundamental for ABM verification (Rand & Rust, 2011). Also, the Cancerscape code is well documented and structured. A test-driven approach benefits model development in several ways, assisting in writing clearer and better-designed code and ensuring continuous quality. IntelliJ IDEA offers a wide

range of tools, such as JUnit or FindBugs to assist large-scale application design, running and testing, that were used throughout Cancerscape implementation and testing.

7.2.1 Documentation

Creating records that explain the conceptual model and its implementation is important not just for ensuring the model is sound, but also for making sure that other researchers and general model users understand each other's creations. The ODD protocol presented in the previous chapter describes the Cancerscape structure explaining the different sub-modules that make up the model. It describes the agents, the environment, the behaviors, and scheduling processes so that the overall model workings can be better understood by the users. Together with a comprehensive ODD framework description, the code has been well documented as well – so that it can readily be understood and utilized by other developers. An example of a documented code snippet is presented in Figure 92.

```
/**
 * this function returns takes as a parameter an arraylist of people and returns
 * the individuals with ages between 50-64 as an arraylist
 */
public ArrayList<Person> getPeople50To64(ArrayList<Person> people) {
   ArrayList<Person> everybody = new ArrayList<Person>();
   for (Person p : people) {
      if (p.getAge() >= 50 && p.getAge() <= 64) {
         everybody.add(p);
      }
   }
   return everybody;
}
```

Figure 92: Example of code documentation

Even in the case where the code is available and well documented – it can get time-consuming (especially in the case of a large-scale model) for anyone other than the programmer to understand it (Gotts & Polhill, 2009). Therefore, making sure proper documentation – both the high-level description of the model and at the code level is an important step of the verification process. It ensures that the model can be understood and extended by others.

7.2.2 Debugging

Debugging is the systematic process of discovering and reducing the number of bugs or errors in a software application. The aim of the debugging process is to make sure the software program behaves as expected. Given that Cancerscape is a computationally intensive large-scale software application – debugging constituted a laborious process of the model development process. Debugging tasks ranged in complexity: from fixing minor errors to performing tedious and extensive tasks of analysis, scheduling, data collection and performance.

Following the addition of each piece of code, the sub-modules were tested, and the bugs were solved as they emerged. Agent activities were plotted using dynamic charts to debug each agent activity depending on the time of the day. Moreover, logging the model results was also employed for verification purposes during debugging. The output log of the agent states and behaviors allowed for a more detailed progressive model analysis. Intellij IDEA 15 provides an excellent built-in debugging environment that was extensively used for model debugging and error fixing.

7.2.3 Code Walkthroughs

Code walkthroughs were performed periodically to ensure the logic flow among the multiple parts of the model. The accuracy of parameter testing, method calling, code commenting and statistical outputs were systematically verified. The code walkthroughs also ensured that the standard code naming conventions (for variables, types and functions) were respected and that the model design was accurately implemented.

7.2.4 Unit Testing (Iterative Programmatic Testing) & Mocking

Unit testing represents a method for testing software applications by which the smallest piece of testable code is separately examined to ensure it is correctly written and functions as expected before the piece is put to use in the bigger context. The goal of unit testing is to make sure that the individual parts of the application are correctly implemented (Huizinga & Kolawa, 2007). Cancerscape makes use of the JUnit Java framework to execute unit testing that is a widely used Java Unit Testing library. JUnit 4 is linked to Cancerscape as a JAR (Java Archive) at compile time. JUnit tests were written to make sure different pieces of code work properly. Even though wiring software tests may seem like a time-consuming task at first – in the long run it ends up saving software verification time.

Mocking is another technique used in Cancerscape that primarily focuses on unit testing and involves creating mock objects that essentially represent simulated objects mimicking the behavior of real objects in a controlled manner. Using mocking, a series of mock agents were created and given various attributes so that they could be tested individually. The subset of generated mock agents was let to interact, their behavior was monitored and their dynamics observed. Different test cases and scenarios were analyzed

using distinctive conditions to make sure the model runs as expected. Using formal testing, logging was used to make sure that the code is accurately written (Dijkstra, 1976).

7.2.5 Model Profiling

Profiling a software program means analyzing it dynamically to measure its complexity, the memory space or time it takes to run, the handling of specific instructions, or the duration of particular function calls. Profiling is usually realized using specific tools called code profilers. These tools are essential for understanding the program behavior as well as ways to optimize it.

Cancerscape profiling was executed as part of model verification process. The model code was analyzed for duration and frequency of particular code units, such as key functions, methods, and classes. The speed of the model as a whole was also examined and decisions were made to minimize the model running time. Model profiling was executed with the JProfiler Intellij 14 plugin, paired with the JProfiler stand-alone external application, as shown in Figure 93.

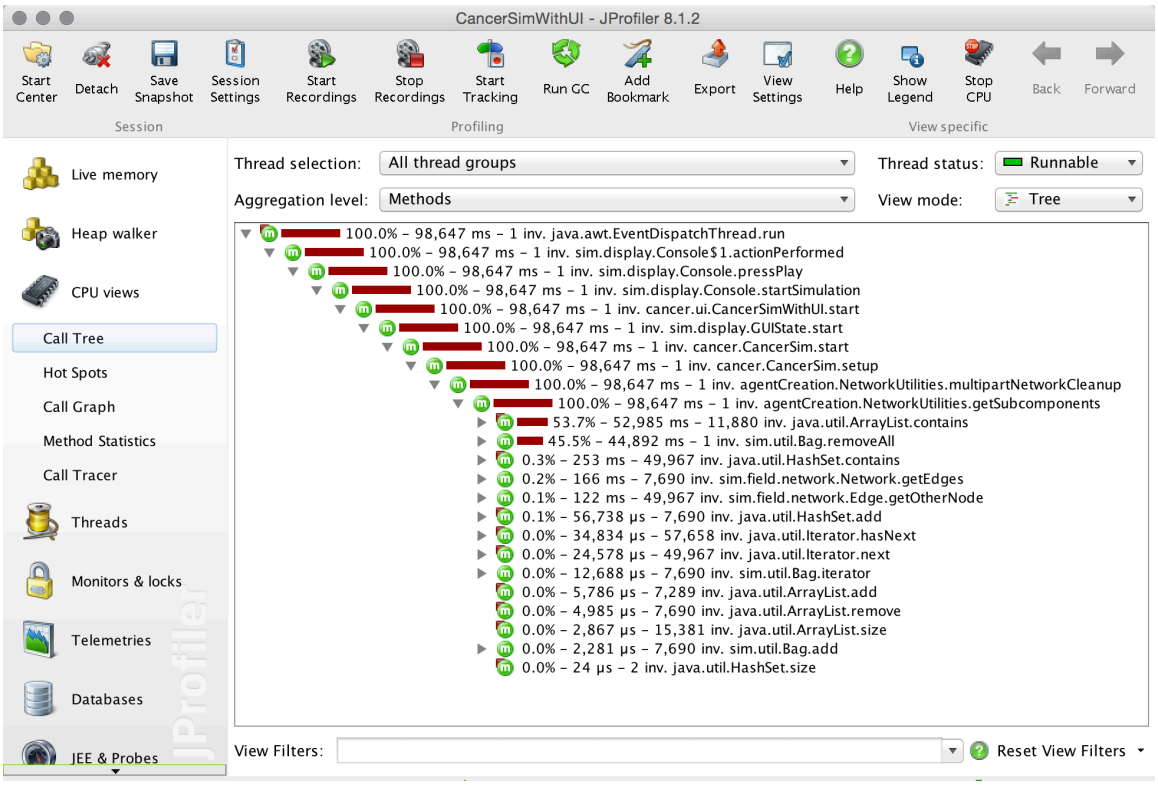


Figure 93: Cancerscape profiling with JProfiler 8.1.2

JProfiler provides an overview and analysis of the memory usage and dynamic allocation leaks, as well as evaluating the CPU and thread conflicts. Executing the model profiling helped reduce the running time of the model and avoid memory overuse issues. Model verification procedures expanded as more tests were written and more errors were found and resolved.

7.3 Model Calibration and Sensitivity Analysis

Only after proper model verification, and ensuring that the code performs correctly as expected is it appropriate to proceed with model calibration and validation. Calibration is an important task in model building, and one of the most challenging steps of building simulations (Crooks et al., 2008). Calibration refers to configuring the ABM parameters

in such a way so that it produces an output that is as similar as possible to the real world system. One way to accomplish this is to compare the model results with expected data and to configure them progressively until the best fit with simulated data is obtained. There are several techniques available for assisting the process of calibration (Heppenstall et al., 2006). Hansen & Heckman (1996) argue that the calibration and verification of models can be posed as econometric estimation and testing problems. However, the use of such techniques on Cancerscape would be impractical because of its large scale and complexity. This is due to multiple reasons, such as the fact that many Cancerscape parameters cannot be translated into appropriate numerical ranges (like the agent's initial locations, for instance), determining the goodness of fit is particularly challenging for models that contain spatial components. Computation time is another reason – Cancerscape being a computationally expensive model, it regards such specific calibration procedures as impossible to run on the available hardware configuration. Consequently, Cancerscape calibration will be executed manually through thorough visual investigation of model results and fine-tuning parameters keeping in mind the theories behind the model dynamics.

Cancerscape is calibrated with data from multiple sources about the demographics and health characteristics of DC residents in 2000. Among the main data sources that have been used for calibrating and validating the model are:

- The 2000 U. S. Census Bureau (2000) demographic data for Washington DC,
- Behavioral Risk Factor Surveillance System (BRFSS) (CDC, 2000),
- National Health Interview Survey (NCHS, 2000),

- National Cancer Institute's Surveillance, Epidemiology, and End Results Program (SEER, 2014),
- DC Department of Health public use data sets (DC DOH, 2000b),
- Secondary sources and other relevant studies.

All scenarios are based on the Washington, D. C. geographic area, and use as the starting point the 2000 DC Census population, which represents 572,059[4] individuals.

The simulated geographic area of Washington D. C. is composed of 188 Census Tracts and divided into eight wards. The 2000 Census tracts serve as the main geographic units for the model. Tracts are also used as the geographic units for generating the agent population within Cancerscape by tract. These county subdivisions contain between 1000 and 8000 people, with most tracts being comprised of about 3000 individuals.

When running the model, the user is presented with the model Console, which is the central object that handles several simulation tasks. The console has "Play/Pause" and "Stop" buttons for controlling the simulation. It also has a "Parameters" tab that holds the global parameters of the model and may be modified for experimentation purposes. During a model run, the user can visualize real-time statistics and observe dynamics of interest in real-time. The Display tab offers a variety of windows that can be hidden or shown depending on the variables of interest to the user. The "Cancerscape Display" window contains the geographic layers and can also be hidden if the user so desires (it is often done to increase the speed of the simulation as the geographical analysis is computationally intensive in itself). The model results are presented and analyzed further

[4] The population of Washington, D. C. on April 1, 2000 was 572 059

in this chapter, and are based on 30 individual model runs over a five-year time period, between 2000 and 2004.

Sensitivity analysis refers to the process of examining how the model parameters affect the outputs. This process in extremely important for ensuring that the simulation is working as expected and for establishing a qualitative agreement with existing empirical data regarding the simulated phenomenon. Cancerscape was carefully analyzed for dependence on the initial conditions, and when strong dependence was observed, the cause was traced and carefully addressed. Such sensitive dependence on the initial conditions is usually indicative of "bugs" in the system, and should be addressed carefully if are to have confidence in the simulation results (Axtell & Epstein, 1994).

As mentioned previously in the opening part of this chapter, Cancerscape seeks a performance classification closer to Level 2 (based on Axtell & Epstein (1994) classification), where the model should be in qualitative agreement with empirical macro-structures as shown by plotting the distributions of interest.

Using sensitivity analysis, the effect of various parameters on specific behaviors and on the overall model outputs was evaluated. Several tests and parameter sweeps were executed to search for cases of high sensitivity among the model parameters and to make sure the agents behave as expected, both spatially and without the explicit spatial environment. Even though executing the model many times with varying parameters can help to assess the robustness of the model (Axtell, 2000) – this task can get computationally expensive and time consuming, especially for ABMs as broad and complex as Cancerscape.

Ideally, it would be desirable to go through the whole range of agent and model parameters and check for highly sensitive parameters. However, due to the large scope and complexity of Cancerscape, and also due to limited computational resources to support such operations, only the parameters suspected to have a significant effect were tested directly. Subsequently, one may argue if parameters known to have little effect on the overall agent or model behavior should be considered at all. While this is a valid concern, Cancerscape was conceptually conceived as a flexible multilevel system able to account for a broad range of influences that characterize the social determinants of cancer.

7.4 Model Validation

7.4.1 Model Validation Considerations

It is common for ABMs to be interdisciplinary in their nature, involving concepts and theories from social, behavioral and complex systems science. Verification works together with validation to remove the barriers in trusting the model, contributing to making it more credible and reliable.

Validation supplements the process of model verification and is used for determining how well the ABM corresponds to the real system it tries to model. That is, it describes the extent to which the model represents the real world system (Casti, 1997), and whether the modeled system accurately reproduces the behaviors characteristic to the system being modeled. A validated model allows us to explore "what if" scenarios, and perform model experiments of interest that are not feasible to perform on the real system. Rather than thinking of models as valid or invalid – it is more appropriate to think in terms of degrees of validity.

As it is the case with model verification, it is practically impossible to fully validate an ABM (North & Macal, 2007). Even though it is considered exceedingly difficult to completely validate and verify a model (Grimm & Railsback, 2005), this is accurate for all models, not just for ABMs. However, models should follow certain V&V standards, if they are to be useful.

Being a large-scale data-driven model, Cancerscape follows the KIDS (Keep It Descriptive Stupid), as opposed to KISS (Keep It Simple Stupid) approach. Edmonds & Moss (2005) distinguish at least two stages of empirical validation for ABMs: *(1) micro-validation* – performed at the level of individual agents, referring to data on individual behavior, and *(2) macro-validation* – usually carried out on the model's aggregate behavior. North & Macal (2007) mentioned several perspectives on model validation:

- *requirements validation* – that implies establishing a clear definition of the real world questions of interest that will be explored in the model;
- *data validation* – referring to the validity of the data used by the model;
- *face validation* – representing a visual validity practice assessing the plausibility of the model assumptions, checking if the model "looks" right;
- *process validation* – refers to checking if the internal steps correspond to similar operations in the real world;
- *model output validation* – verifies if the model output matches the real system being modeled;
- *agent validation* - examining the agent activities and interactions to make sure they correspond to the individuals they attempt to model in the real world.

Darvishi & Ahmadi (2014) divide the existing ABM validation techniques into three categories:

(1) *structural validation* – that has the goal to show the way a simulation system works to produce the observed phenomenon in the real world,

(2) *predictive validation* – referring to the ability of the model to predict a behavior that has not been observed before, and

(3) *replicative validation* – that relies on comparing the output of the model with the real world data.

In the case when the model simulates a complex system, a commonly used approach is structural validation (Galán et al., 2009). Next in this chapter, I will discuss the validation processes that were applied to Cancerscape, as well as three proposed scenarios.

Face validation (sometimes referred to as a verification procedure) is a subjective measure of the extent the model *prima facie* looks like the real system – usually performed at the initial stages of the validation process (Hermann, 1967). The distribution of the population in Cancerscape was plotted by the age groups corresponding to the Census 2000 grouping for Washington DC, as shown in Figure 94.

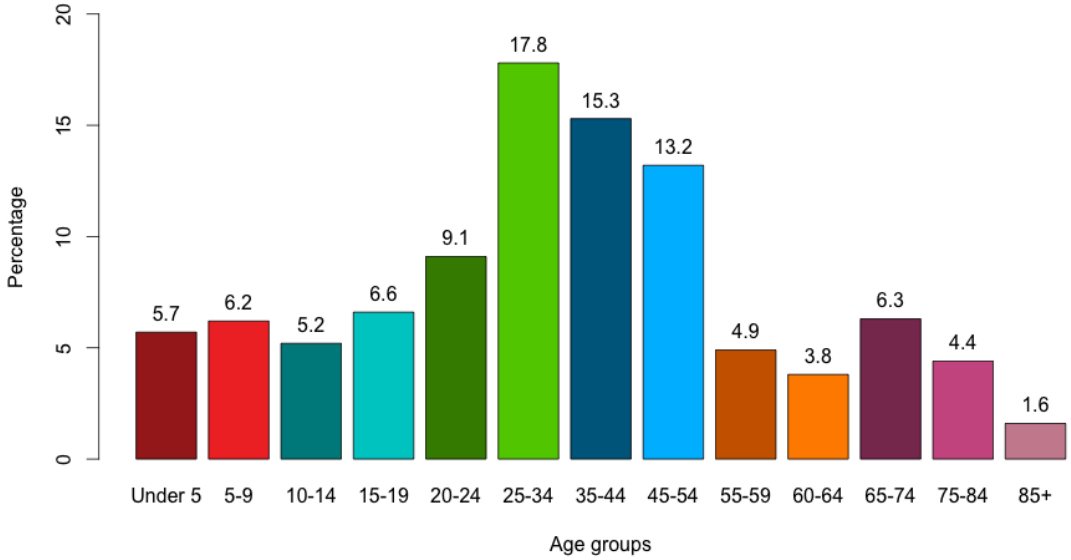


Figure 94: Washington DC 2000 population groups by age

For instance, Figure 94 above and Figure 95 both show the distribution of population groups from the real data, as shown in Figure 94, and after generating the agent population in Cancerscape, as shown in Figure 95.

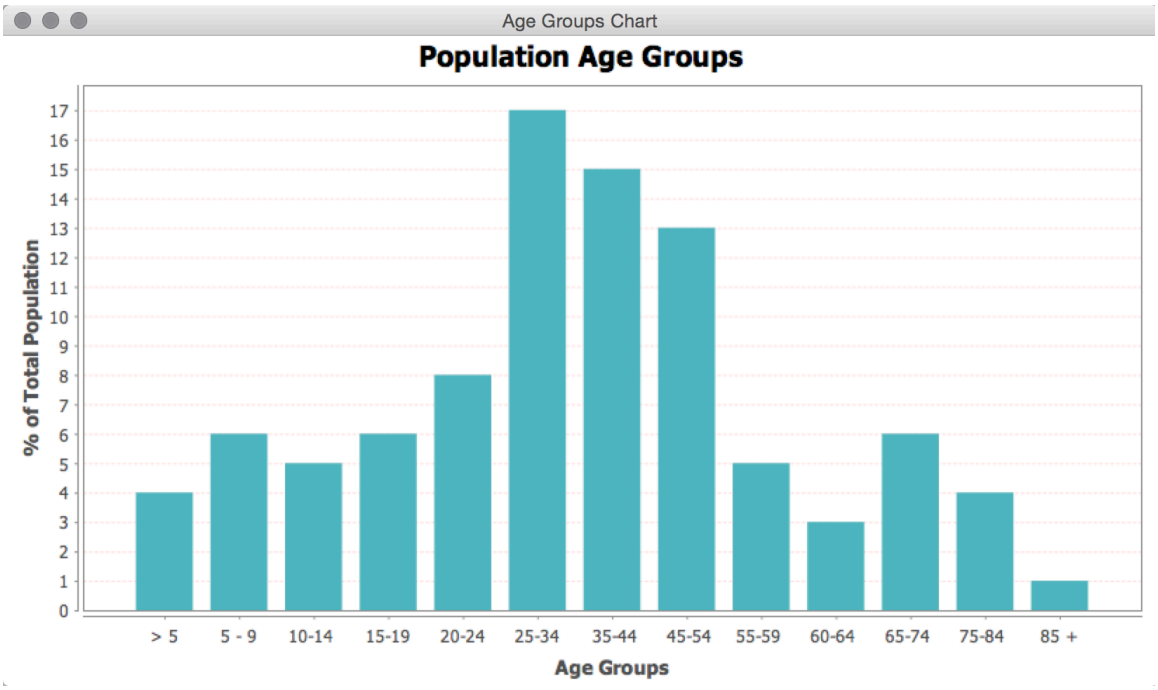


Figure 95: Modeled population groups in Cancerscape

The model produces animations that can be visually assessed in terms of overall system behavior. Each agent can be individually inspected in particular. Along with the model animation, a series of graphs, charts and a time tracker is provided for analyzing in detail the overall model behavior. The model output was evaluated to ensure the results fall within satisfactory ranges of the real world phenomena.

7.4.2 Matching Spatial Data

In order for a model to be in objective qualitative agreement with the real world system that it attempts to simulate, the model has to first "look" accurate (Mandelbrot, 1983). The spatial data used in Cancerscape is provided by authoritative sources that offer a detailed and accurate description of the spatial components of the model. Using

GeoMason, the geospatial data (in shapefiles format) was imported into the model and the spatial objects were distributed across the simulation environment. A *spatially explicit* ABM represents a particular geographical space, such as a real world city or region (Galan et al., 2009). Cancerscape includes a geospatial representation of Washington, D. C., based on the GIS Census data for the year 2000. For a visual comparison of a close-up of a health clinic in DC: Figure 96 shows the Hunt Place Community Health Place on OpenStreetMap (2015), and Figure 97 presents the equivalent of the same institution and its geographic location in Cancerscape.

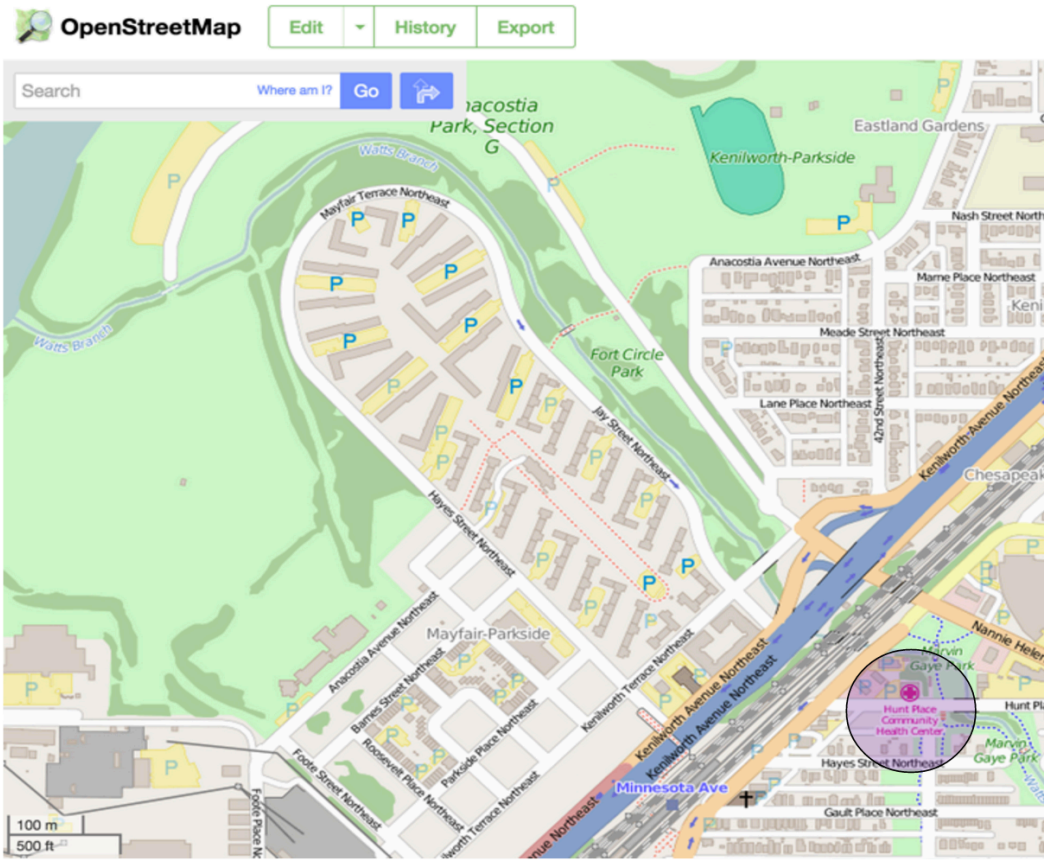


Figure 96: A close-up of the Washington DC map (Source: OpenStreetMap, 2015)

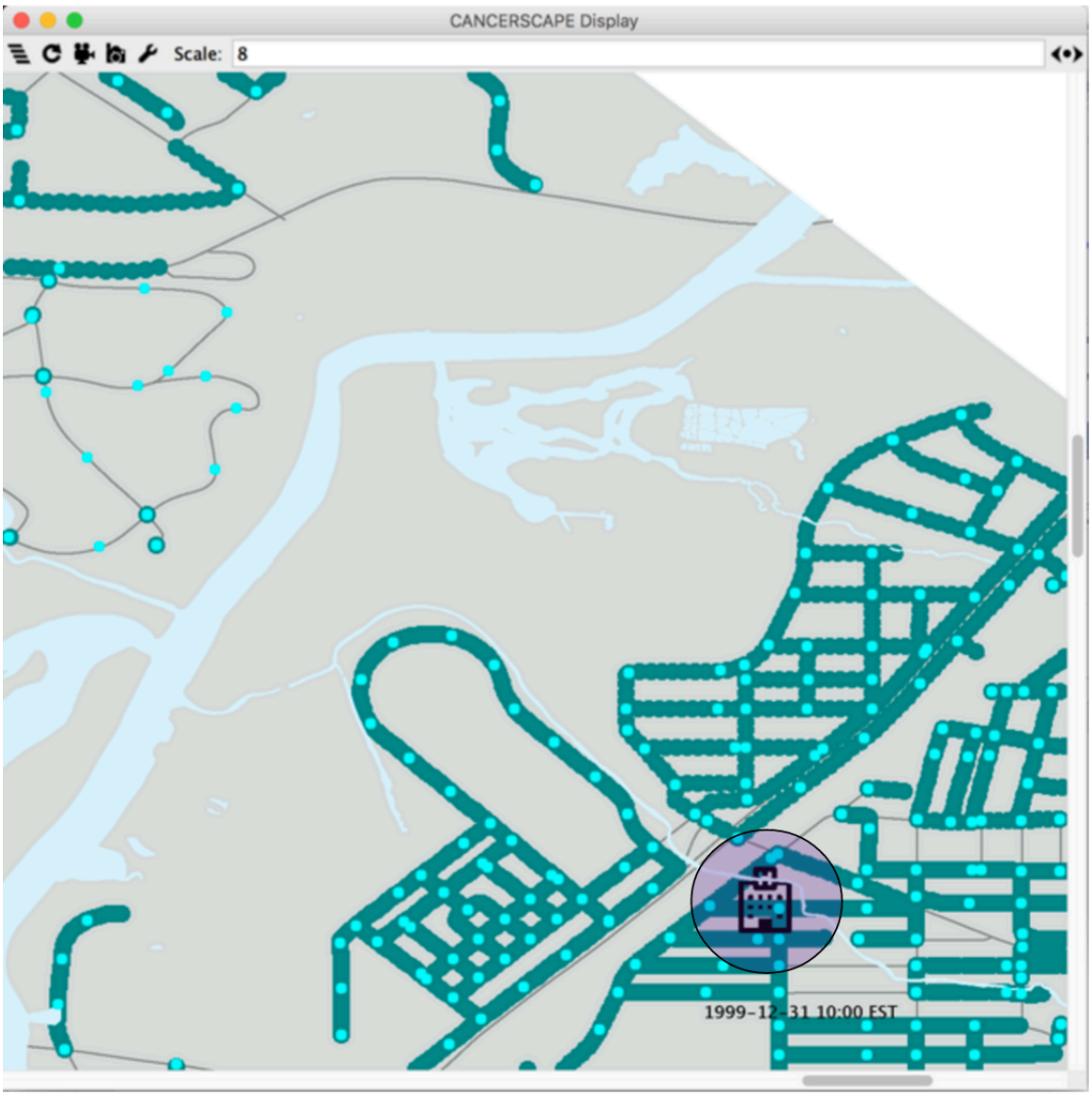


Figure 97: A close-up view of the Cancerscape equivalent of Figure 96

Several similar visual testing procedures were executed throughout the model development process to ensure that the geospatial layers were loaded correctly and that the Cancerscape spatial environment appropriately reproduces the real world geographical space it aims to model.

7.4.3 General Considerations

The model behavior should be evaluated as the simulation progresses, making qualitative judgments of whether the implementation of the behaviors and procedures are implemented correctly (Axelrod, 1997). Object-oriented simulation models tend to be fairly complex, bringing a series of challenges and complexities regarding V&V procedures – making it difficult to assess their correctness (Balci, 1997). Therefore, three experiments will be conducted to explore and evaluate extensively the proposed agent-based model of the social determinants of cancer.

Anticipating the way in which a complex system will evolve is not a trivial endeavor. This can get particularly challenging when the system under study is composed of many individual parts interacting in nonlinear ways across several levels, and whose behavior is not well understood - posing particularly hard challenges. In fact, it is often impossible to predict valid and scientifically sound predictions of the overall behavior of such systems.

The goal of Cancerscape is therefore not to make accurate predictions of the overall system behavior, which would be unreasonable considering the fact that the amount of information about the particular real system it models is relatively scarce. The goal is rather to provide an "artificial laboratory" for rigorous scenario analysis so that insight can be gained regarding the dynamics and behavior of the system.

7.4.4 ABM and GIS Integration within Cancerscape

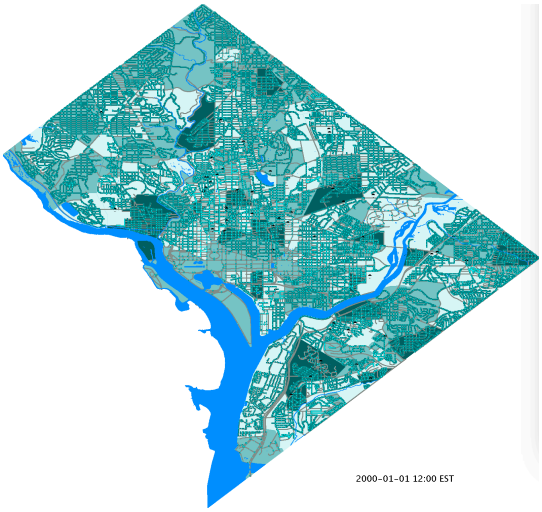
As mentioned in Section 4.2, there are two main ways by which an ABM can be linked with GIS: coupling and integration. Coupling (loose, moderate or tight), that refers to linking two stand-alone systems by data transfer (Crooks & Castle, 2012) while

integration involves embedding the functionality of interest of either ABM or GIS within the dominant system. Depending on this, the final system is referred to as either GIS-centric or ABM-centric. Coupling itself is accomplished using one of three approaches: loose, moderate and tight. Cancerscape adopts a combination of integration and loose coupling to make use of both ABM and GIS methodologies.

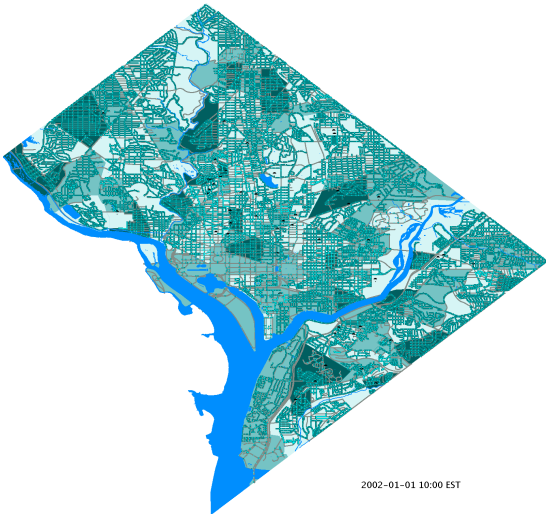
In the case of loose coupling, functions are operated asynchronously within each system, with data exchanged between systems in the form of files. Following this approach, Cancerscape input data files were pre-processed within ArcGIS and later imported into MASON (in the form of shapefiles), where the program execution is performed. However, the geographic data is not sent back to the GIS system for later processing and visualization, but it is rather portrayed and analyzed dynamically within the ABM-centric system. While coupling is commonly the preferred method for linking ABM with GIS, it often results in rather specialized and isolated solutions. Therefore, integration has been combined with loose coupling for optimal linkage of ABM and GIS.

According to Brown et al. (2005), the four strategies for ABM and GIS coupling involve identity, causal, temporal and topological relations. In Cancerscape, identity relations are reflected in the relation between agents and the spatial database that includes the socio-demographic characteristics of the agent population tied to their houses/household location within census tracts. The causal relations in Cancerscape are represented by dynamic changes in spatial features (modification of tract shades) that correspond to the number of agents with CRC. For instance, the more agents with CRC exist within a census tract, the darker teal its shade will become. In Figure 98 (A-E) the

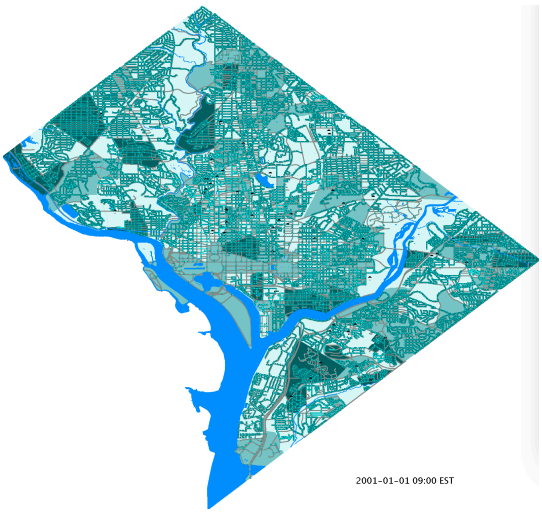
model results for five consecutive years are presented including all layers. Figure 98A starts with the population at generation. The tract color reflects the spatial distribution of agents with CRC within the district at the start when the population (in this particular case, 100,000 agents) was generated and distributed across the spatial environment. Figure 98B presents the CRC population distribution across the environment after one simulation year has passed. Slight differences can be observed in the spatial distribution of the agent population suffering from CRC from to the beginning of the simulation comparing to each incremental year.



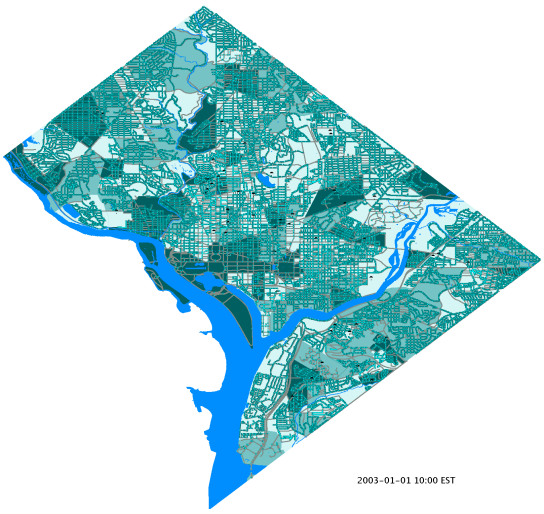
A. Cancerscape model run with all GIS layers, 2000



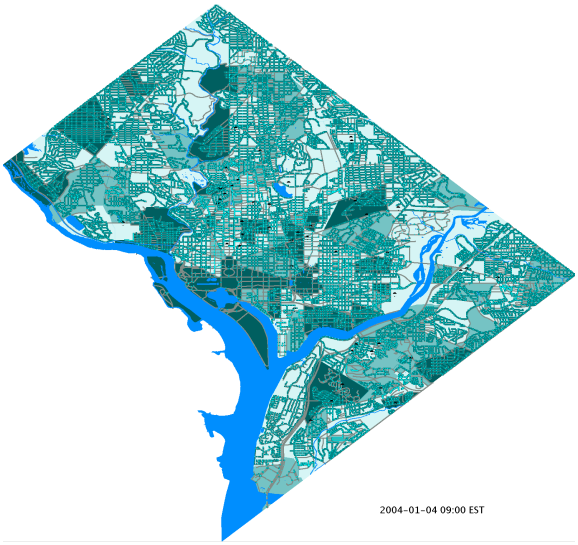
C. Cancerscape model run with all GIS layers, 2002



B. Cancerscape model run with all GIS layers, 2001



D. Cancerscape model run with all GIS layers, 2003



E. Cancerscape model run with all GIS layers, 2004

Figure 98: Cancerscape model runs with all GIS layers for five consecutive years

More agents diagnosed with CRC were located closer towards the center of the district. A one-year simulation run including all GIS layers took about 6 hours of computation time (on a Macintosh Pro machine with a 2.6 GHz Intel Core i7 processor and 16 GB 1600 MHz DDR3 of memory). Therefore, it is advisable to simplify, and leave out the layers that are not of interest for a particular run in order the decrease the model running time.

The color of the agents also changes dynamically, when the individual cancer stage changes. A close up of individual agents, as well as their workplaces and nearby hospitals (excluding home locations) is presented in Figure 99. Agents are assigned nearby workplace location at generation. When they need to go to a hospital, they check their neighborhood within a specific radius and pick a primary care facility that is closest to their home location. If no hospital is located within the specified radius the delay in seeing a primary care doctor for that agent increases.

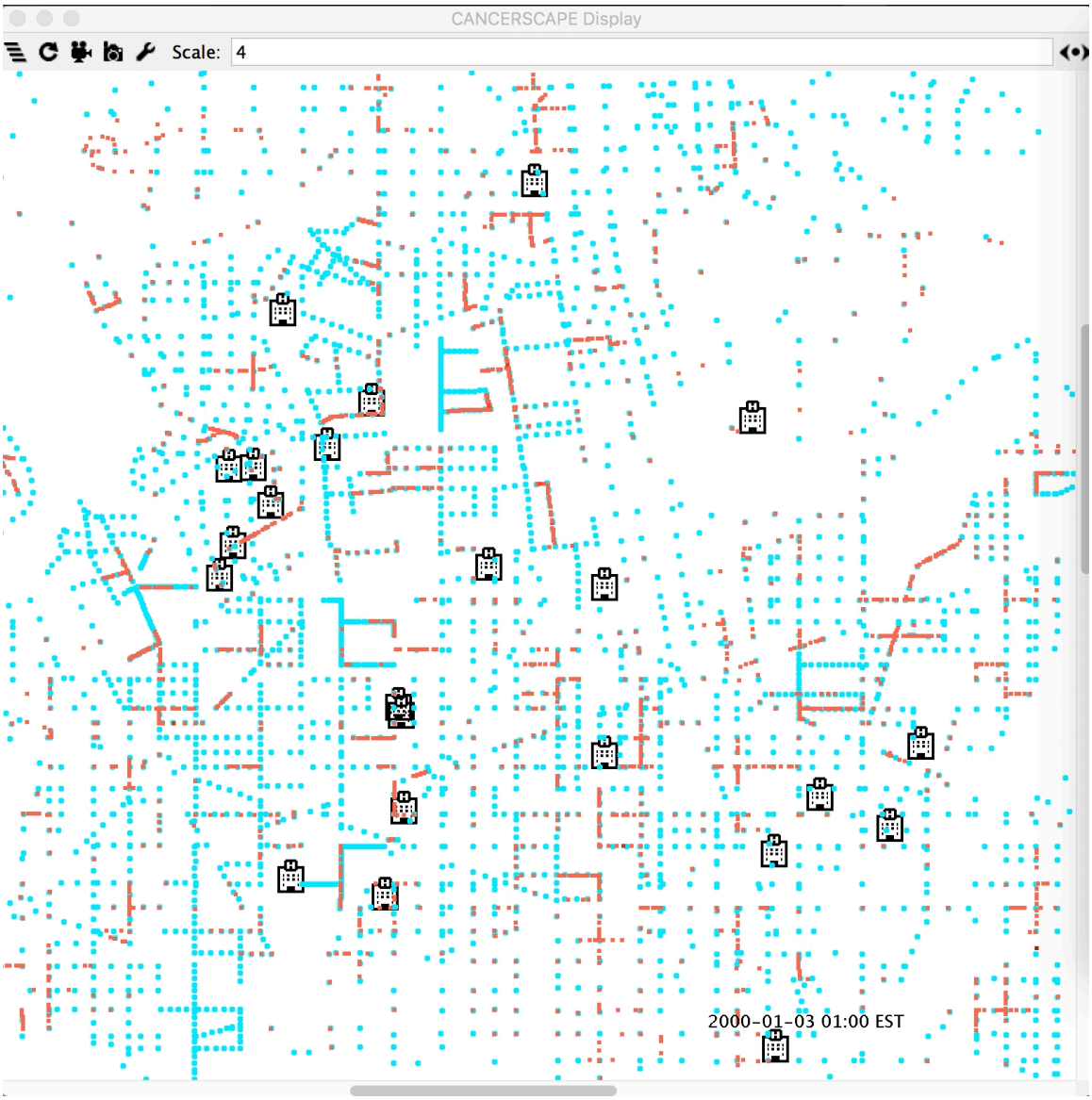


Figure 99: Close up of agents, businesses, and hospitals

Temporal relations are reflected in the way the agents decide to move geographically depending on the time of the day or individual health-related decision-making.

7.4.5 Model Scenario Analysis Overview

Scenario analysis is a widely used exploratory tool in ABM, especially among nonpredictive models (Bankes et al., 2001; Van Asselt, 2000). A scenario can be perceived as a glimpse (that is internally consistent) of what the future may potentially look like – one potential future outcome, rather than a prediction (Porter, 1985). Furthermore, scenario analyses can help policy decision makers make better sense of model results. ABM is particularly suited for modeling systems that mimic real-world scenarios and produce detailed histories about various aspects of the real world system (Axtell, 2000). Scenario analysis is particularly useful for the social science – where most of the experiments are impossible, impractical or extremely difficult or unethical to carry out in the real world (Gilbert, 2004).

Due to the vast scale and complexity of the model, the validation of Cancerscape posed several problems. Firstly, there were gaps in data availability. Cancerscape relies on the CRC data collected by BRFSS for Washington, D. C. The data on screening is not explicitly available for the year 2000. However, in 2002 the BRFSS collected data about the adults 50 or more years of age, who have had an FOBT test within the past two years in Washington DC (basically referring to the period between 2000 and 2002), and regarding the individuals of age 50 or over who have ever had a sigmoidoscopy or colonoscopy.

One of the main goals of Cancerscape is to explore the social determinants of cancer dynamics under multiple conditions. In Cancerscape, several experiments will be performed to explore the social determinants of cancer around three main scenarios: CRC prevention, treatment, and survival. While exploring these scenarios careful attention was

attributed to screening, survival, mortality, treatment and delay in cancer care. The "covering cases" helped to take the model verification and validation a step further (North & Macal, 2007). A model can be considered fully validated when it correctly matches all of the possible real-world cases, which is a task nearly impossible to achieve for most ABMs. Therefore, the model results can be compared with some cases from the real world to achieve consistency. The next section will present the general model output, and the three scenarios: prevention, treatment, and survival.

Up to this point, Cancerscape has been verified, extensively evaluated and tested. In the next section, the cancer-related model outcomes will be presented and discussed. Following this, three scenarios will be provided that explore in detail aspects of CRC prevention, treatment and survival. Having calibrated Cancerscape, it is possible to compare the results of the model with the real world data.

7.5 Theme I: Exploring CRC Incidence

In this section, I explore the results of the model regarding CRC incidence. Incidence rates presented in this section explore the average results of running the model 30 times at full scale (for all 572,059 agents), for five years. One simulation step represents one hour. While the model is not overly sensitive to initial conditions, a slight variation between the different runs was observed, that could be attributed to the stochastic elements of Cancerscape. During the testing phase, the model was also run on a smaller scale as well, to test for robustness. The results were relatively in agreement with the results obtained from running the model at full scale. A comprehensive list of the model input parameters and the default values for each is available in Appendix F.

7.5.1 Overall CRC Incidence Results

The overall CRC incidence trend in the model corresponds with the trend observed in Washington, D. C. during the same period. CRC incidence and mortality increased with increasing age. Similarly, in Cancerscape the incidence of CRC increased with advancing age, and was slightly higher in men compared to women, as shown in Figure 100. Also, the chances of contracting CRC increase considerably after the age of 50.

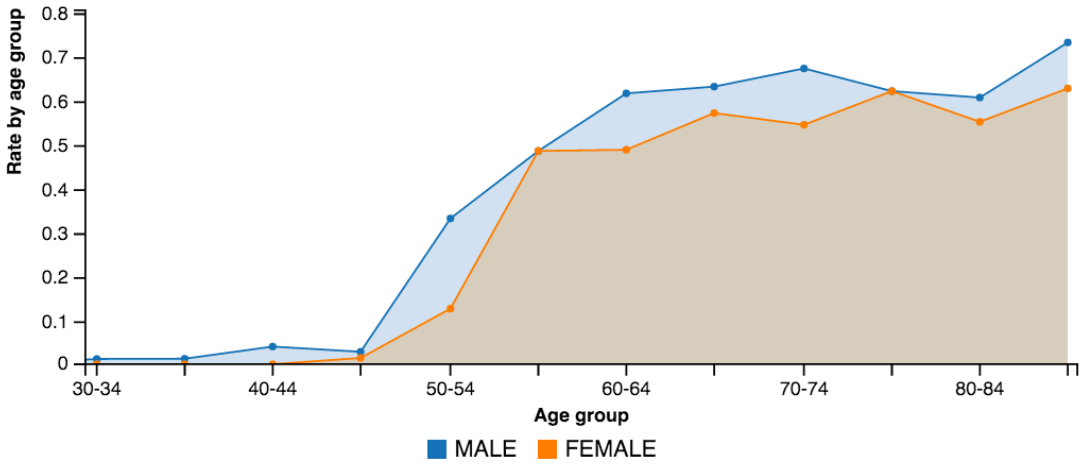


Figure 100: Modeled CRC incidence by gender in DC, 2000-2004

These results are also consistent with the gap between men and women for overall CRC incidence rates across the nation, and the difference in rates was found to be about 35% - 40% higher in men compared to women (ACS, 2012). While the reasons behind these differences are not completely understood, this gap may partially be attributed to hormonal differences but also to gender-related differences in the exposure to several CRC risk factors and behaviors. Among the factors that could explain these gender

differences are obesity and heavy alcohol drinking, with men tend to have higher obesity rates compared to women.

The overall CRC incidence in Cancerscape during the five simulation years is presented in Figure 101, with (A) presenting the CRC incidence rates based on the real data and for comparison, and (B) presenting the modeled CRC incidence rates, generated as the output of the model after 30 full-scale runs. The model was run at full scale, with the starting population as 572,059 agents.

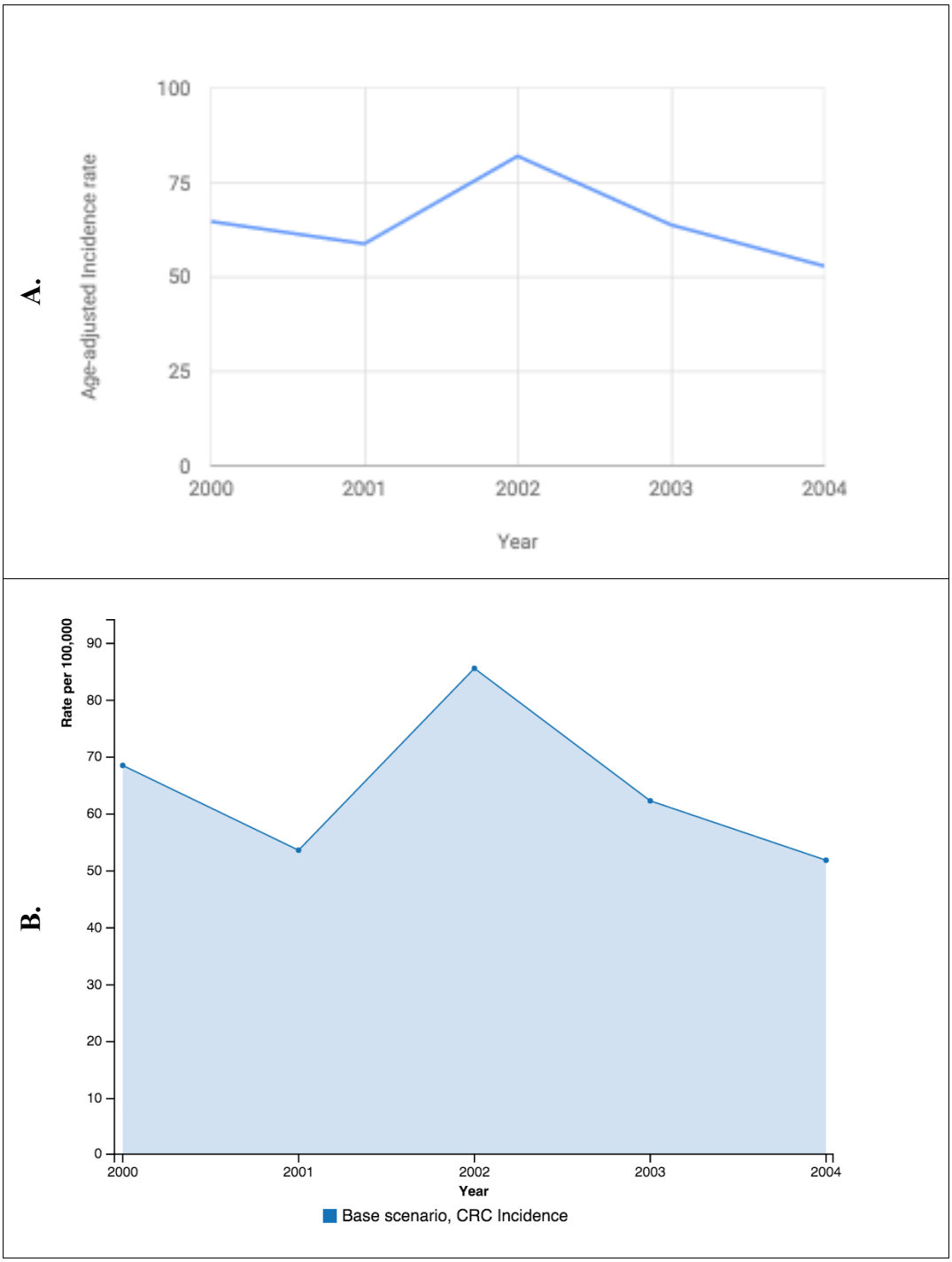


Figure 101: Observed (A) and modeled (B) incidence rates for CRC in DC, 2000-2004

7.5.2 Disparities in CRC Incidence

The simulation model was also designed to explore the disparities persisting in terms of CRC incidence. In particular, disparities by race and gender are investigated. About 59.4% of the Washington D. C. population constituted of African Americans in 2000. Eliminating disparities refers to achieving a reduction in CRC incidence and mortality for a given population. It also refers to achieving an increase in survival among the socially and economically disadvantaged population groups (Ward et al., 2004). As it was expected to be the case, the racial disparities in terms of CRC incidence persisted in Cancerscape as they did in the real world, as shown in Figure 102. The White agent population in Cancerscape experienced considerably lower CRC incidence rates compared to African-Americans.

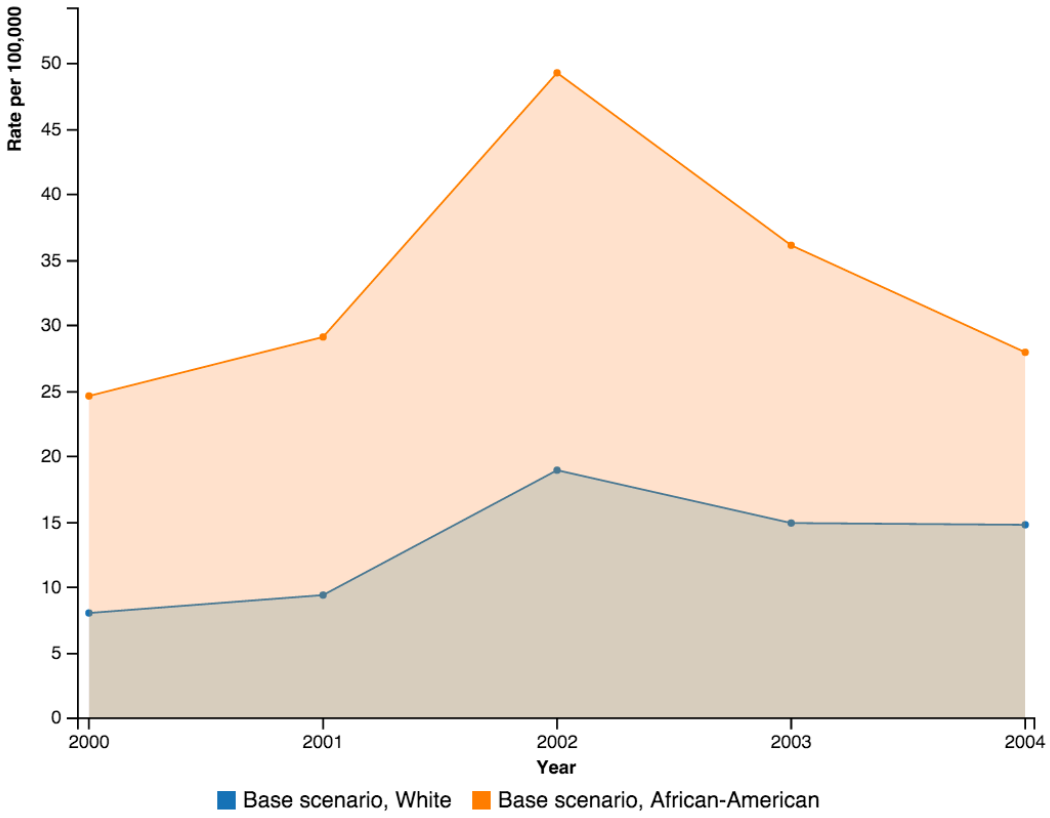


Figure 102: Modeled overall CRC incidence rates by race in DC, 2000-2004

The root causes of racial disparities regarding cancer incidence are multifactorial and complex. Similar to statistics from the real data, CRC incidence rates among the Whites in DC were significantly lower compared to the incidence rates among the African Americans. These disparities persisted during all simulation years, from 2000 to 2004, with the gap between the two races slightly decreasing in the last year of the simulation. The spatial distribution of CRC incidence among African Americans is presented in Appendix C.

The spike observed for the year 2002 is consistent with the increase observed for the same year in the overall CRC incidence and relates to the adoption of the state mandates on CRC screening in DC – as Washington, D. C. implemented the mandate on CRC screening coverage on April 13, 2002 (Cokkinides et al., 2011). Positive associations were found between state mandates and the use of recent CRC screening. Screening mandates may have contributed to more cancers being detected during that period.

The pathways by which cancer affects incidence patterns by race are complex. An increase in risk factor trends was observed among the African Americans of Washington D. C. in 2000. African Americans had higher overweight and obesity rates compared to Whites (Section 5.4.2) and were also more likely to smoke tobacco (Section 5.4.4). In 2000, African Americans in the district were less physically active compared to Whites (Chatterjee et al., 2015). They also had lower healthcare insurance coverage and had less access to care compared to Whites (Section 5.4.6). The proportion of population of age 25 and over with at least a bachelor's degree was about four times higher in Whites compared to African Americans (Section 5.3.4), and the median household income was about twice lower in African American households compared to the White ones (Section 5.3.5). The potential pathways by which these characteristics may translate into the higher CRC incidence rates for African Americans are: less healthcare insurance coverage less access to healthcare and subsequently lower CRC screening rates compared to Whites, fewer years of education and lower levels of knowledge about CRC (Cotunga et al., 1992; Powe, 1993). The more advanced colorectal cancer stage at diagnosis among

African Americans has been associated with lower income levels, lower education levels and higher likelihood of residing in neighborhoods with reduced healthcare access (Schwartz et al., 2003; Ward et al., 2004; Mandelblatt et al., 1996). This may partially attempt to explain the association between the higher CRC incidence rates in African Americans compared to the CRC rates within the White population groups of the district. Understanding the causes of the disparities in terms of cancer incidence is particularly important for reducing the unequal burden of cancer at the population level. One way to achieve this is with proper screening strategies that target the various dissimilarities among the distinctive racial groups of the district.

7.6 Theme II: CRC Screening

The overarching goal of CRC screening is to prevent CRC altogether through detecting and removing precancerous growths (colorectal polyps) that lead to cancer. Screening can reduce the death rates due to CRC, by decreasing the incidence rates and by detecting polyps and tumors early, when the available treatment options are most effective (Lieberman et al., 2008). Two of the screening tests modeled in Cancerscape are Fecal Occult Blood Test (FOBT) and Sigmoidoscopy/Colonoscopy (jointly called Endoscopy tests), that are also the most commonly used test for detecting and/or treating CRC.

The results of the model showing the proportion of population 50 years of age or over who have undergone any CRC screening procedures during the five-year simulation period by gender are presented in Figure 103.

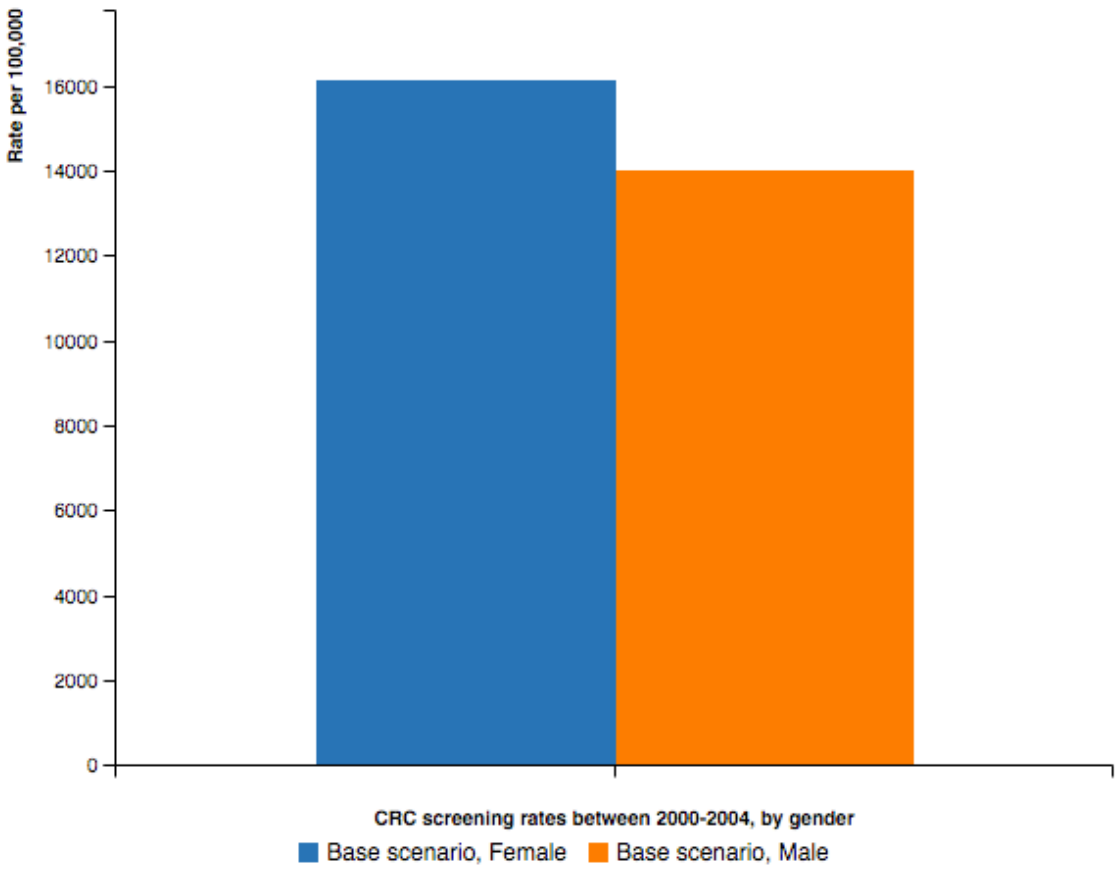


Figure 103: Modeled CRC screening rates by gender, between 2000-2004

According to the data on CRC screening from the Behavioral Risk Factor Surveillance System (BRFSS) of 2004, the percentage of female residents of age 50+ who ever had a sigmoidoscopy or colonoscopy was 65.2%, compared to only 60.2% for males. According to the BRFSS data on CRC screening by gender, the same trend can be observed in the model results, as shown in Figure 104.

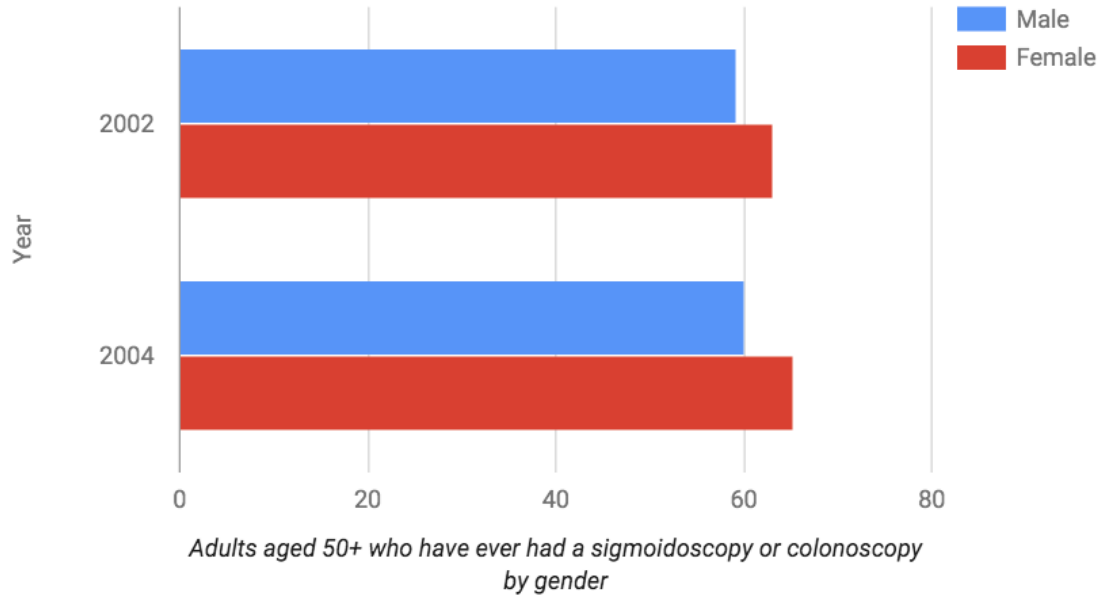


Figure 104: Adults of age 50 years and over who have ever had a sigmoidoscopy or colonoscopy by gender, in 2002 and 2004 (Source: BRFSS, 2015)

Screening in Cancerscape provides an opportunity to highlight one of the benefits of ABM for addressing the issue of missing data. Because CRC screening data for DC has only been collected by the BRFSS every couple of years, using simulated data, we can attempt to offer potential data for filling these gaps.

Disparities are observed in terms of screening as well. Figure 105 shows the screening rates for a period of 5 years for individuals covered by health insurance compared to those without any health insurance coverage.

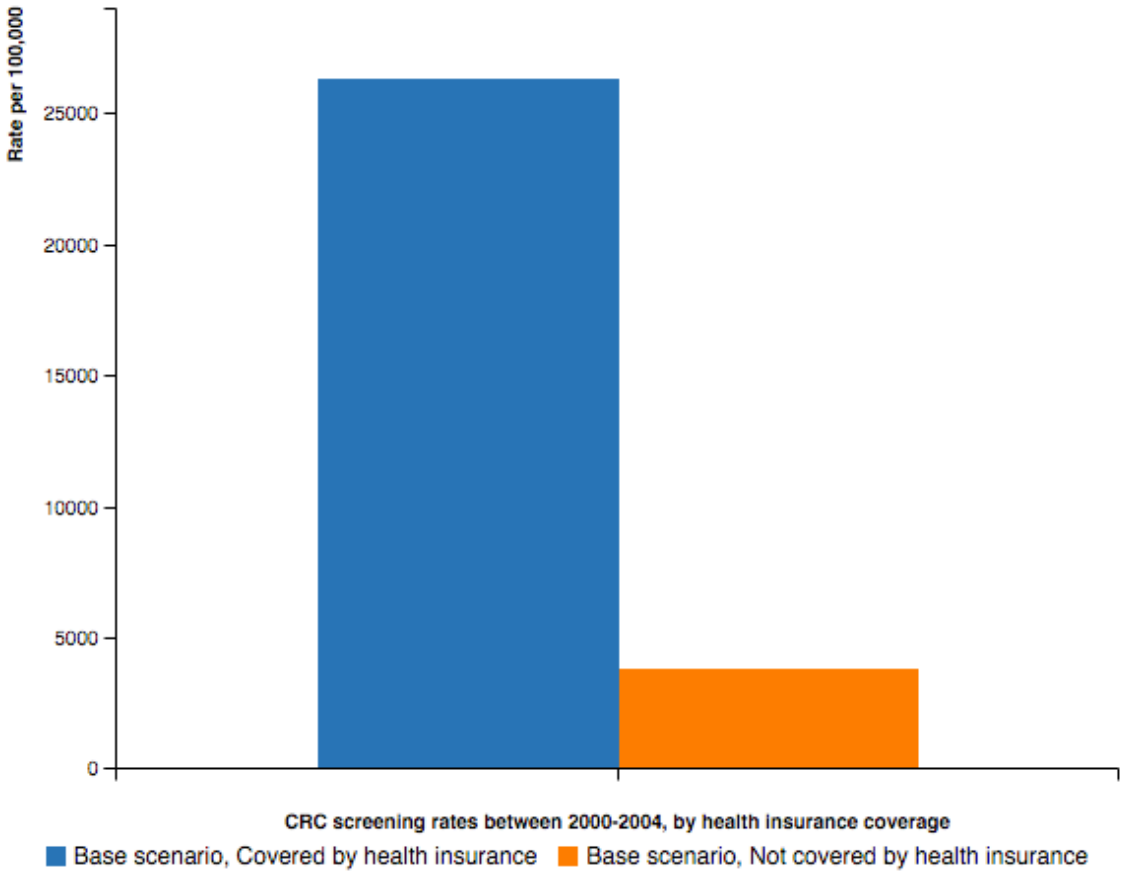


Figure 105: Modeled screening rates for Whites and African Americans by health insurance coverage in DC, 2000-2004

Agents with health insurance were more likely, compared to their counterparts, to receive timely screening during the five simulated years. The difference between the two population groups is large, and lacking health insurance was shown to be an important factor influencing receiving up to date CRC screening, contributing to increased delays.

7.7 Theme III: CRC Survival

In 2000, CRC was a leading cause of cancer deaths in DC, nationwide and worldwide (Greenlee et al., 2000). In DC, CRC mortality rates were above the national average in 2000, with 28.8 and 20.7 deaths per 100,000 individuals, correspondingly.

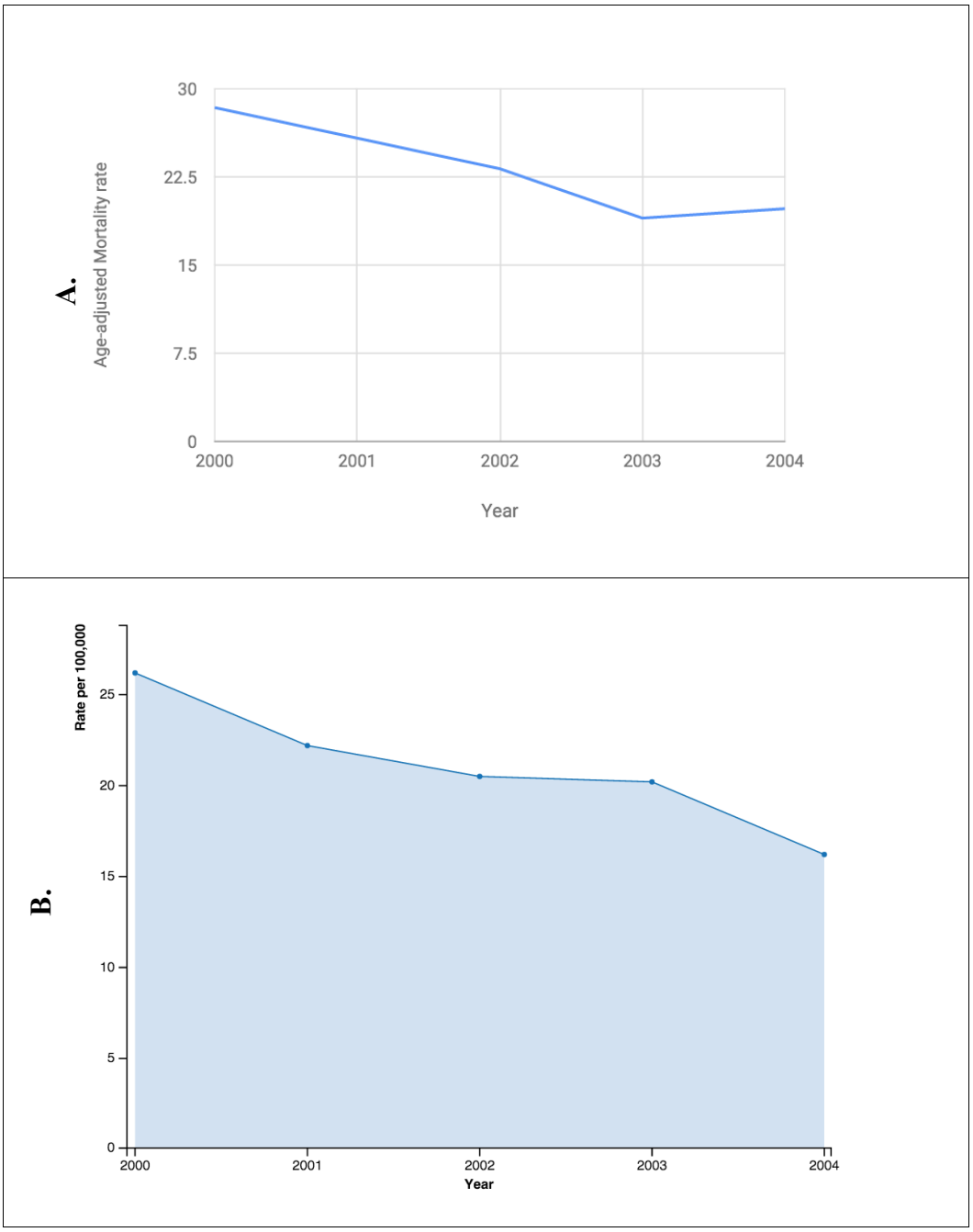


Figure 106: A: Observed (A) and modeled (B) CRC mortality rates in DC, 2000-2004

The CRC mortality rates in Cancerscape compared to the mortality rates in DC between 2000 and 2004 are presented in Figure 106. Similar to the real world data, the

mortality rates in the model continued the downward trend up to the year 2003, followed by a sharper decrease in 2004. The decrease towards the final simulation year may be associated with the higher screening rates in the previous years as well as in the increase of the number of cancers detected early when treatment was most effective. The increase in screening rates during the first years is associated with Washington D. C. implementing the state mandate on CRC screening coverage on April 13, 2002, mandating that all healthcare plans to cover CRC screening procedures (Cokkinides et al., 2011). This was specified as part of input validation of Cancerscape.

Another form of disparities observed in Cancerscape is gender disparities. In Washington, D. C. men had higher CRC incidence rates compared to women (Section 5.6.2). Consequently, the results of the model showing the mortality rates in men compared to women are presented in Figure 107. This gender disparity persisted regarding other CRC measures such as CRC incidence and screening rates and continued to manifest in terms of CRC mortality rates as well.

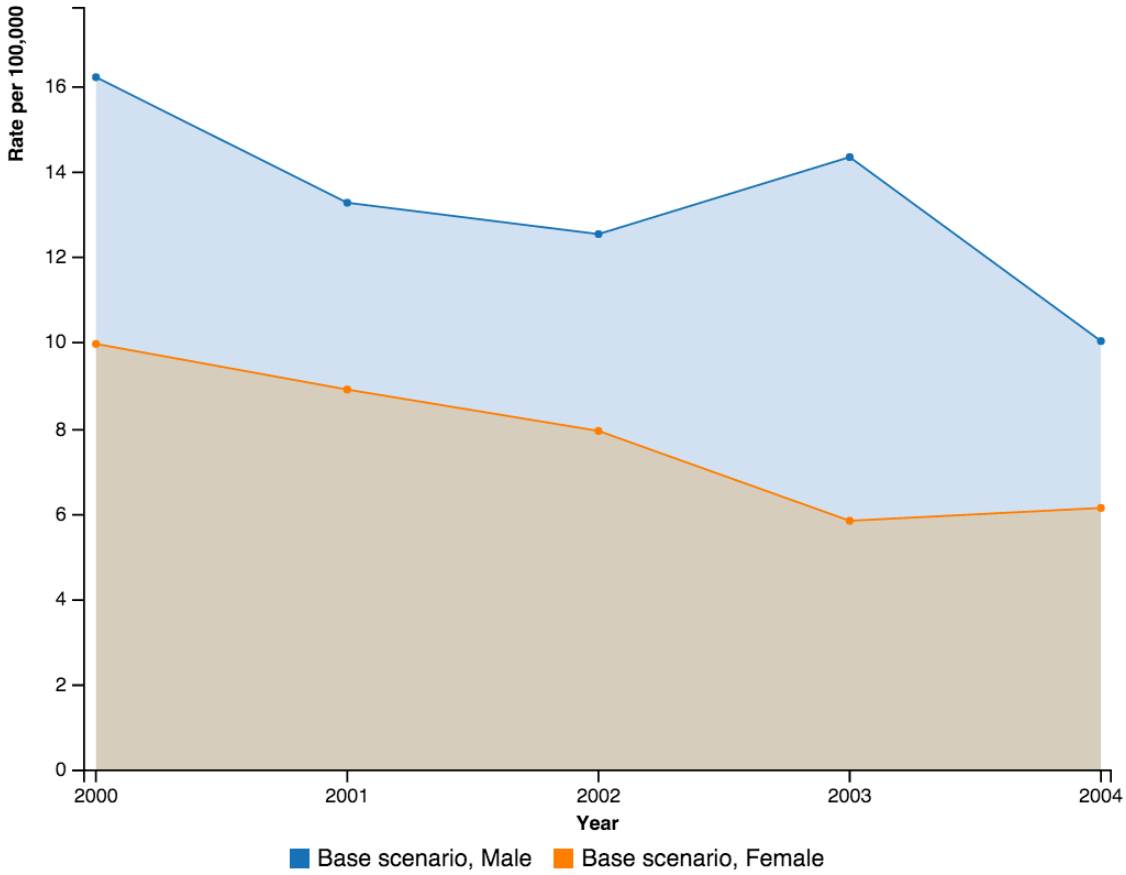


Figure 107: Modeled CRC mortality rates by gender in DC, 2000-2004

A difference in CRC mortality rates across the five simulation years was also noticed for different races/ethnicities, with African Americans leading the gap in this regard as well, as shown in Figure 108.

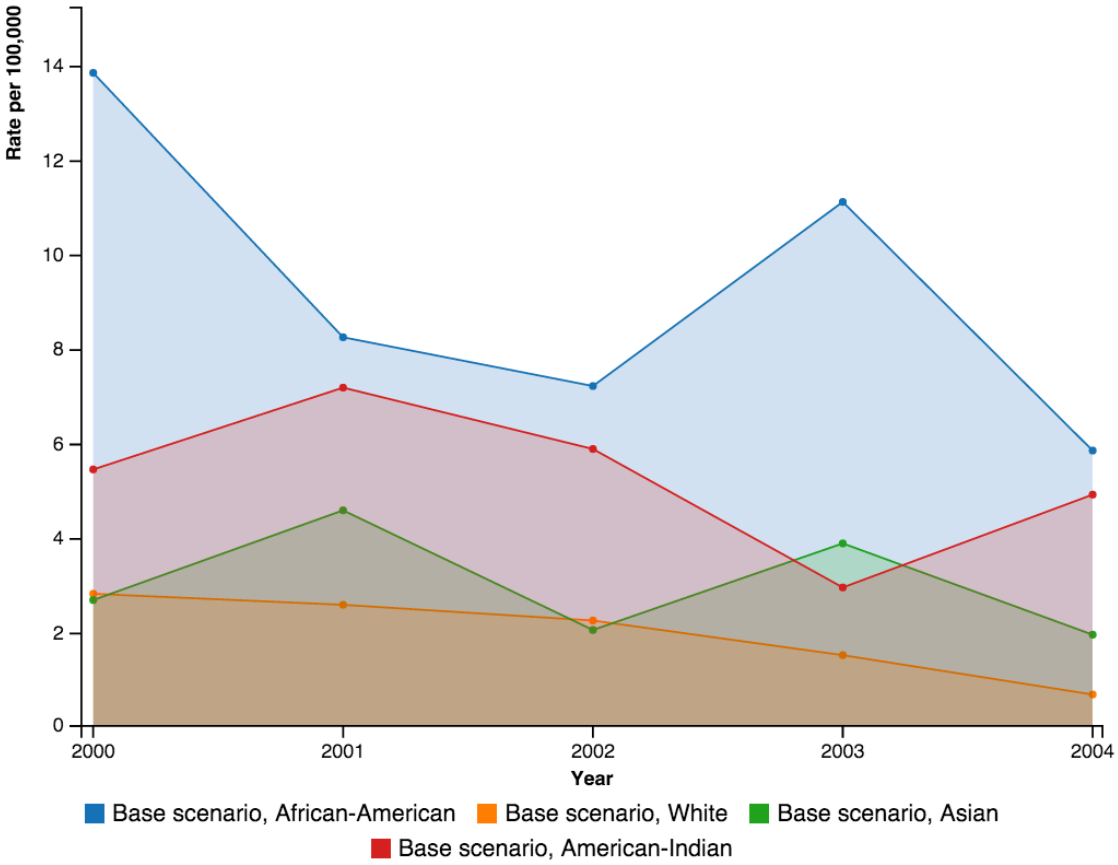


Figure 108: Modeled CRC mortality rates by race in DC, 2000-2004

CRC mortality rates in Cancerscape were higher in African Americans compared to all the other races, with Whites having lower CRC mortality rates compared to African Americans. This is consistent with previous results from real world data showing that African Americans benefited less from timely screening compared to Whites, therefore being more likely to present more often with advanced stages at diagnosis, that considerably influenced their survival.

In Cancerscape, deaths due to CRC depend on several factors with the stage at diagnosis being the most significant. As shown in Figure 109, agents diagnosed at earlier stages survive considerably longer compared to those diagnosed later.

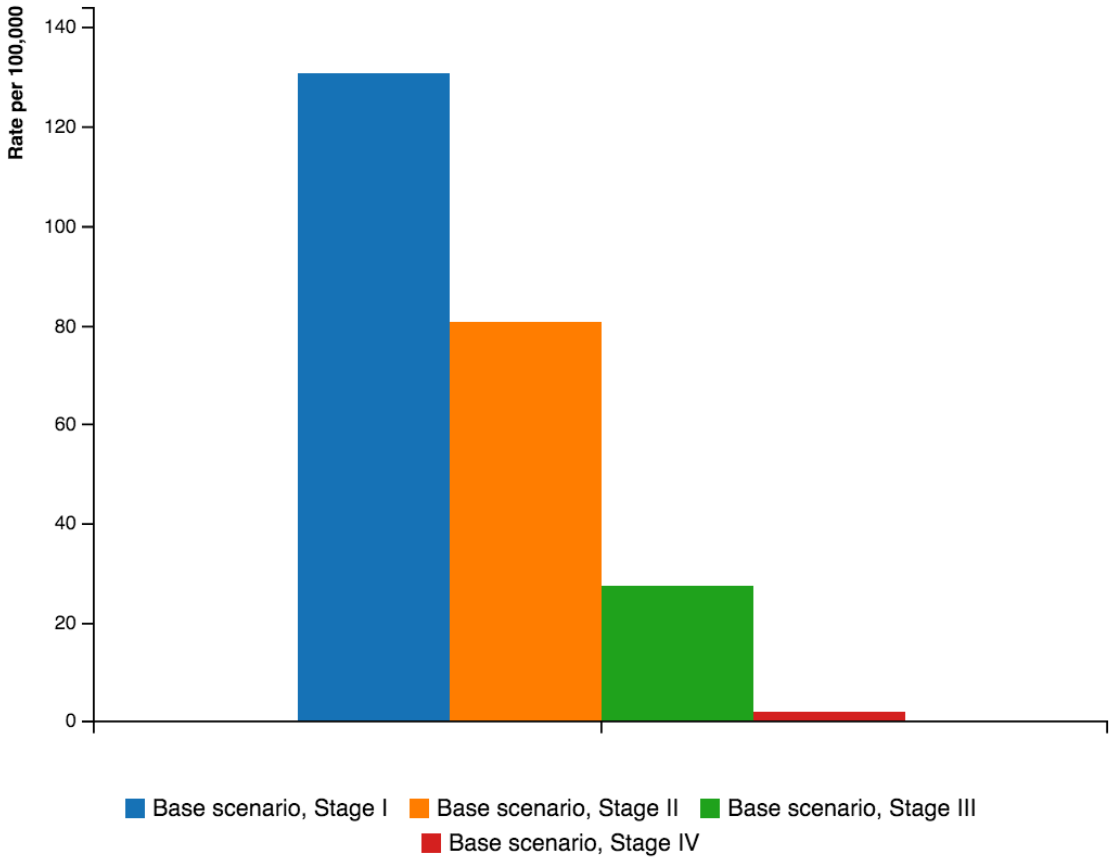


Figure 109: Modeled 5-year CRC survival rate by stage at diagnosis, 2000-2004

The survival rate for agents with Stage 1 at diagnosis was significantly higher compared to all the other stages. Survival rate continued to decrease with advancing cancer stage at diagnosis, being the worst for agents diagnosed with Stage 4 CRC. This is

associated with the proportion of agents who received screening during the simulation period.

Agent survival and mortality also depend on the treatment type, its timeliness, affordability and quality. Treatment for CRC does not always result in successful outcomes. A proportion of agents, albeit small, die annually due to failed treatment or complications during screening and treatment procedures, as shown in Figure 110.

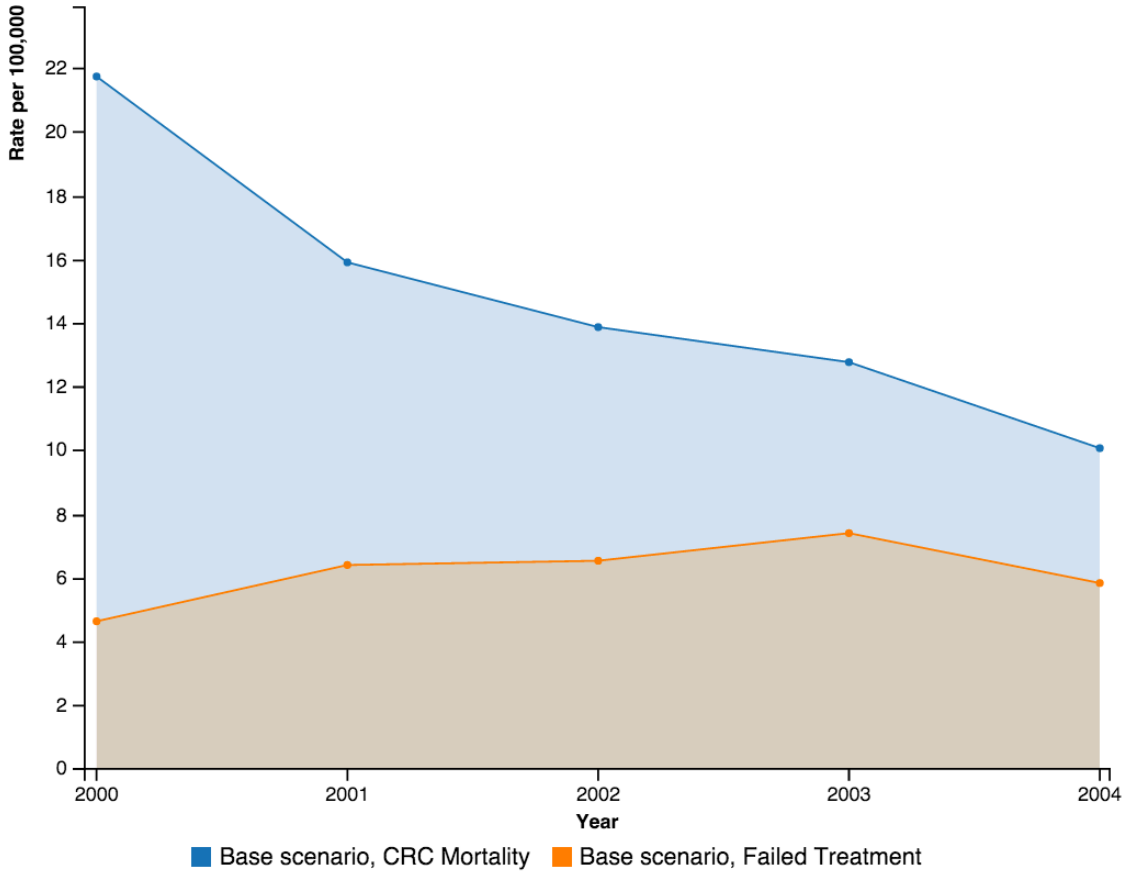


Figure 110: Modeled mortality rates due to CRC and failed treatment, 2000-2004

For instance, the success rate of the colonoscopy varies, as reported by several studies. Nelson et al. (2002) found the success rate of this procedure to be as high as .972. The probability of complications or subsequent death due to this procedure is rather low. Comparatively, the success rate of surgery combined with chemotherapy is about .35. It is considerably lower compared to chemotherapy alone, which is commonly applied to late stage CRC, when cancer is advanced and less responsive to treatment.

7.8 Theme IV: Delay in CRC Diagnosis and Treatment

Because early diagnosis leads to better treatment outcomes, improved rates of survival and sometimes to having CRC cured, time to diagnosis (and delay) has been thoroughly studied. Minimizing the delay of a cancer diagnosis relies heavily on patient factors, but is also influenced by the primary care physician and overall system factors as well. People may delay seeking medical attention for unusual symptoms due to various reasons, such as the lack of knowledge about specific symptoms that are linked to cancer, not perceiving those symptoms as serious and a lack of health insurance coverage, among others.

In order to study CRC-related delay in Cancerscape, the Anderson Model for Total Patient Delay, introduced in Section 3.8, is employed. Andersen et al. (1995) distinguished six main stages of delay in seeking a cancer diagnosis: (1) detecting unexplained symptoms, (2) inferring illness, (3) taking the decision to seek medical attention, (4) acting on that decision by making an appointment, (5) first receiving medical attention, and (6) beginning treatment for illness. In Cancerscape, the stage of taking the decision to seek medical care and acting on that decision by making an

appointment was considered jointly. The distribution of delay in days after running the model for five years is presented in Figure 111.

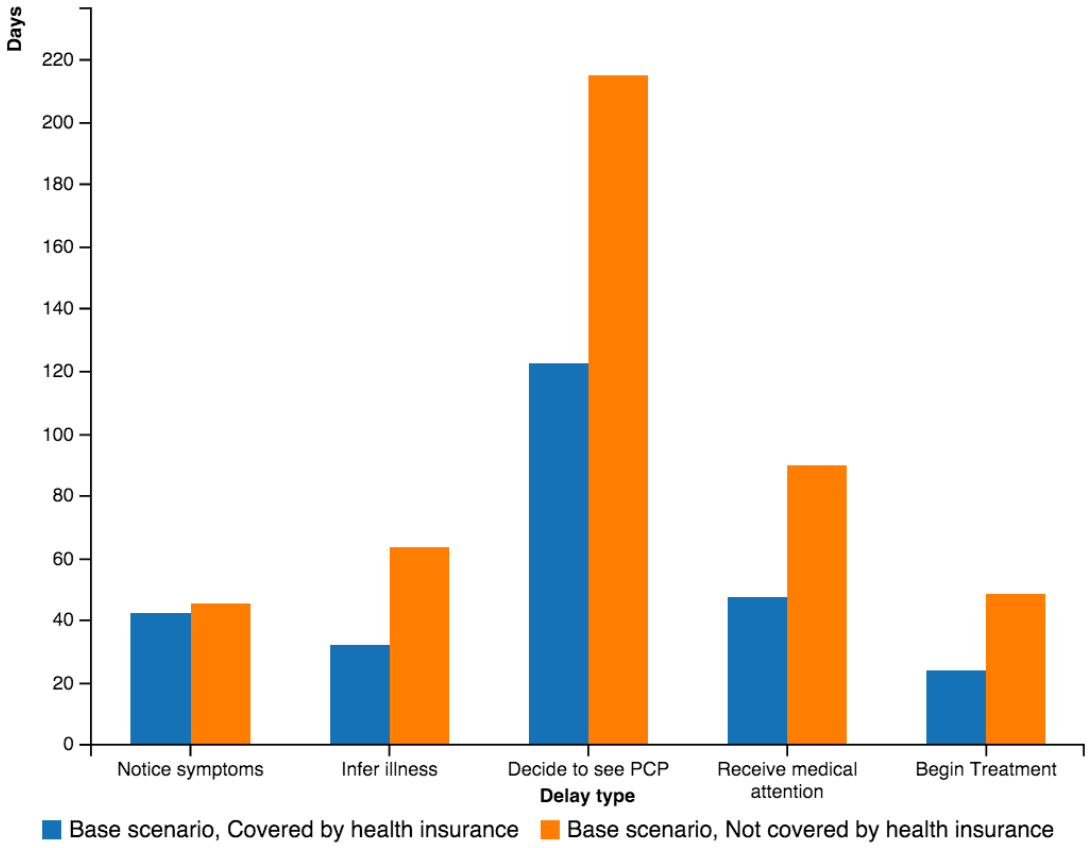


Figure 111: Modeled delay in seeking medical attention and beginning treatment by health insurance (in days)

Agents who had any type of health insurance experience reduced delay throughout all Andersen Model of Delay stages compared to those without any health insurance coverage. The difference was particularly prominent at the deciding to seek medical attention stage. After the agent notices their symptoms and perceives them as

serious, the possession of any type of health insurance becomes an important factor in the decision to seek medical attention.

Being covered by health insurance was a significant factor in influencing the total delay among the African American and White populations, as shown in Figure 112.

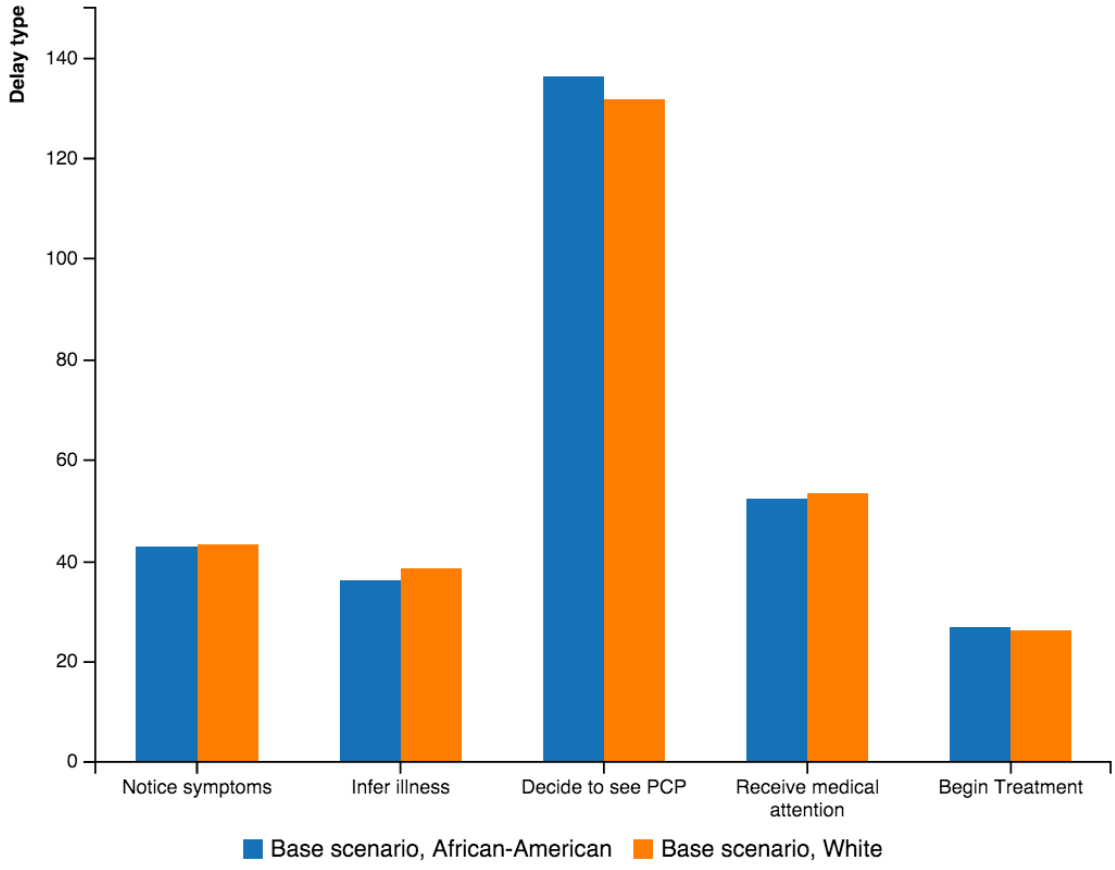


Figure 112: Modeled delay in seeking medical attention and beginning treatment by race (in days)

Both Whites and African Americans experience similar delays across all stages of Andersen Model of Delay. A combination of factors, such as the demographic characteristics of agents, their income, health insurance, perceived significance attributed

to symptoms together with race and ethnicity influenced the delays in getting medical attention and starting treatment for CRC.

Knowledge about CRC and having a family history of CRC were mostly significant within the first two stages of the delay continuum. Taking the decision to get the symptoms checked by a primary care physician was mostly influenced by the possession of health insurance, receiving medical attention and beginning treatment was mostly determined by income, as well as health insurance.

7.9 Scenarios and Discussion of Results

After presenting the model results in terms of incidence, screening, delay and survival, in this section I will discuss the scenarios implemented in the model. Cancerscape supports a variety of types of experiments. For the remainder of this chapter, three scenarios are presented for exploring "what if" situations in Cancerscape to investigate the implications of alternative selections regarding CRC screening, incidence, and survival, followed by an general summary of the chapter. As mentioned earlier in this chapter, a scenario can be referred to as a setting of the analytical experimental parameters that may influence future behaviors of the system (Goto & Takahashi, 2013).

7.9.1 Scenario 1: Reducing CRC Screening Age for Average Population

Several factors may affect screening rates, either directly or indirectly – and age is among the most significant factors. However, the benefits of CRC screening for the certain age groups does not always outweigh the harms. The United States Preventive Services Task Force (USPSTF, 2002) recommends screening for all adults starting the age of 50 and continuing up to age 75. For adults 75 or older the risks and benefits highly vary among the different individuals. Because the risk of developing CRC is higher in individuals

with a family history of CRC, USPSTF recommends that screening for this population group should begin at age 40 or sooner (colonoscopy is recommended ten years before the time of CRC diagnosis of the youngest relative). According to USPSTF (2002), CRC screening is the key to prevention colorectal cancer.

Countries without aggressive CRC screening and population education programs have not experienced the overall decreasing trend in CRC incidence that developed countries had experienced (Center et al., 2009). This highlights the important of screening at state, national and international level. Current screening guidelines recommend starting screening for CRC at the age of 50 for the average population in the United States. However, the CRC stage at diagnosis within this population group is usually advanced, with limited treatment options and worse chances of survival. Therefore, in this section I aim to explore a scenario of lowering the recommended screening age for CRC from 50 to 40 years. While this would be an experiment difficult to carry out in the real world, ABM offers the possibility to explore this on our citywide simulated population, as shown in Figure 113.

The recommended screening age in Cancerscape starts at the age of 50 for the general population and the age of 40 for the high-risk population groups (those with two or more relatives with CRC). Screening for the high-risk individuals starts ten years before the average risk population or 10 years before the earliest case of CRC in the family.

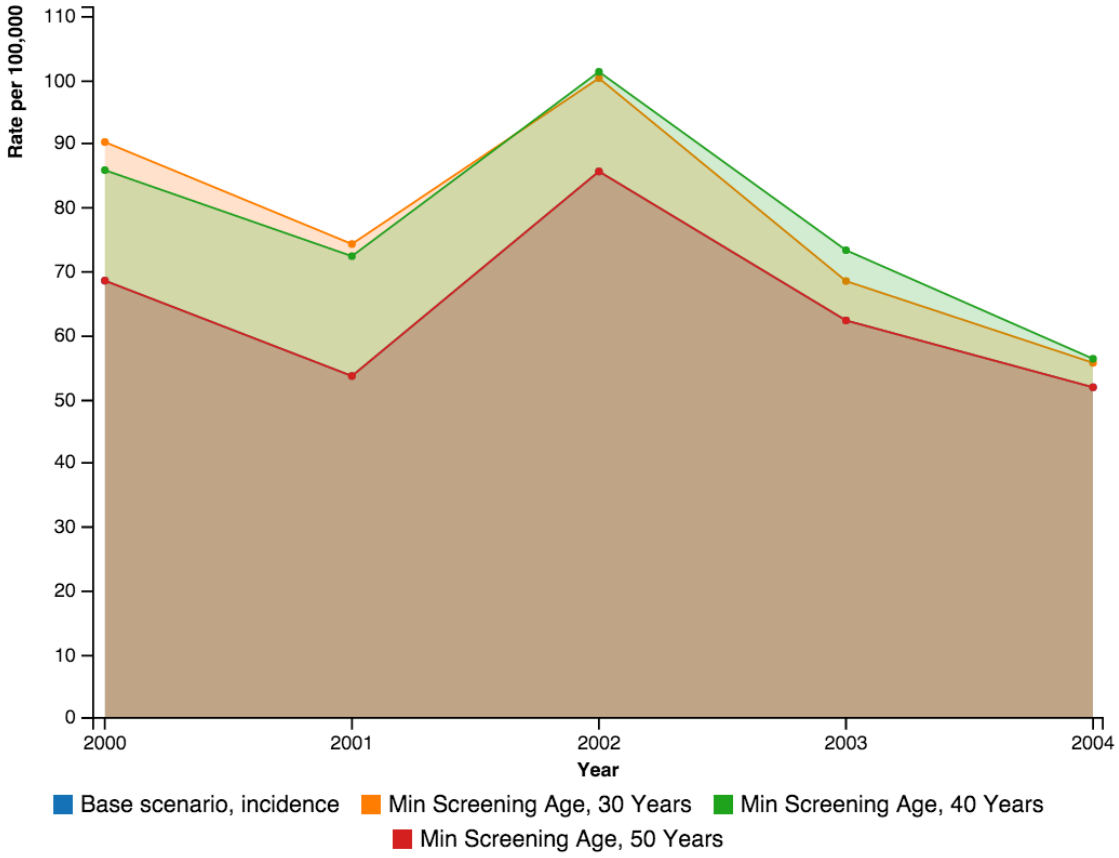


Figure 113: Modeled CRC incidence by recommended screening age starting at 30, 40 and 50 years

Reducing the recommended CRC screening age contributed to increased rates of CRC incidence for all simulated years in Cancerscape, as individuals between the age of 40 and 50 also received screening recommendations from their primary care physicians during annual physical exams. Reducing the recommended screening age further to 30 years resulted in only a small influence on the overall CRC incidence on some years compared to the strategy of starting screening for CRC at the age of 40. The increase in CRC incidence rates for the scenario with recommended screening age at 40 is explained by the fact that this policy helped detect more people with colorectal polyps or tumors

whose diagnosis would otherwise be missed if recommended screening started at 50 years, as shown in Figure 114.

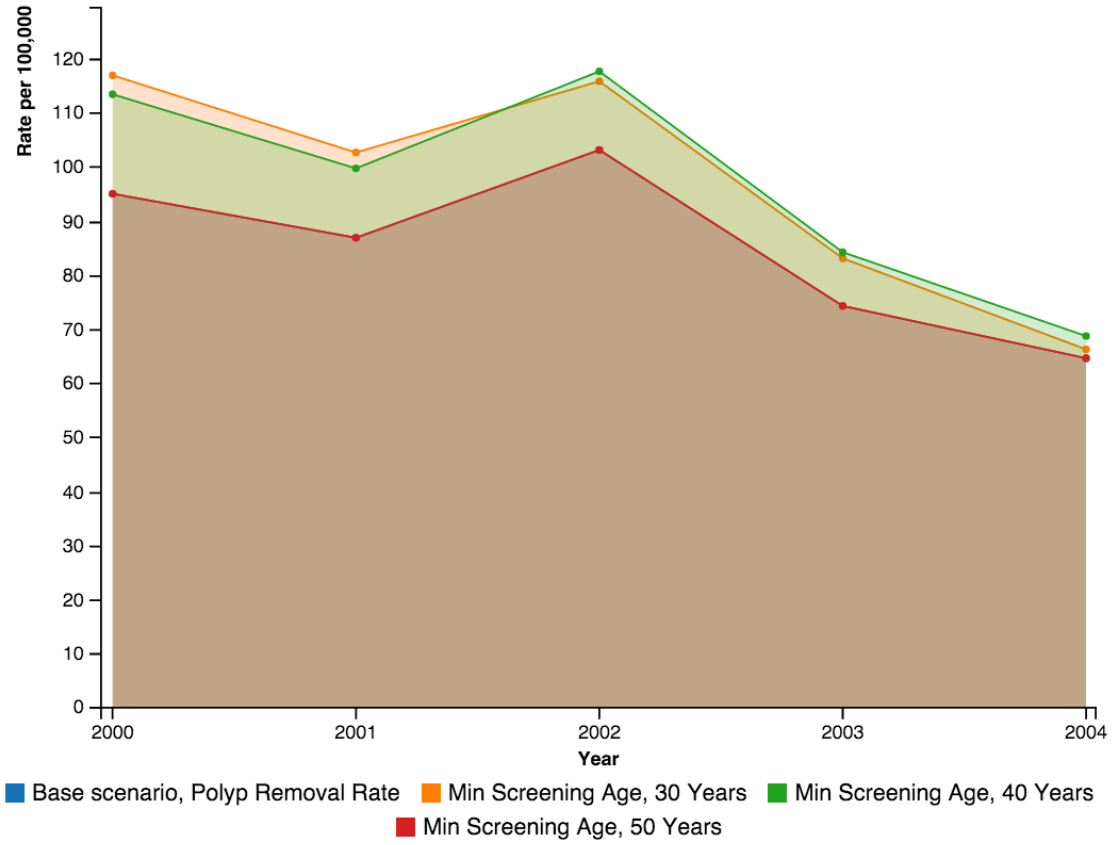


Figure 114: Modeled polyp removal rates by recommended screening age starting at 30, 40 and 50 years

Reducing the recommended CRC screening age resulted in more people having their polyps removed early in the process before they turned into cancer. Starting the CRC screening procedures for the general population beginning at the age of 30 did not result in a similar increase in the overall CRC incidence or polyp removal rates. This was probably due to the lower prevalence of colorectal polyps and tumors within this

population group, which is consistent with previous knowledge about the age distribution of polyps and CRC, as previously discussed in Section 5.6.1.

As a result of the increase in CRC incidence (with fewer agents getting diagnosed early), the policy scenario of reducing the recommended screening age to 40 years also had a positive influence on the overall CRC survival, as shown in Figure 115.

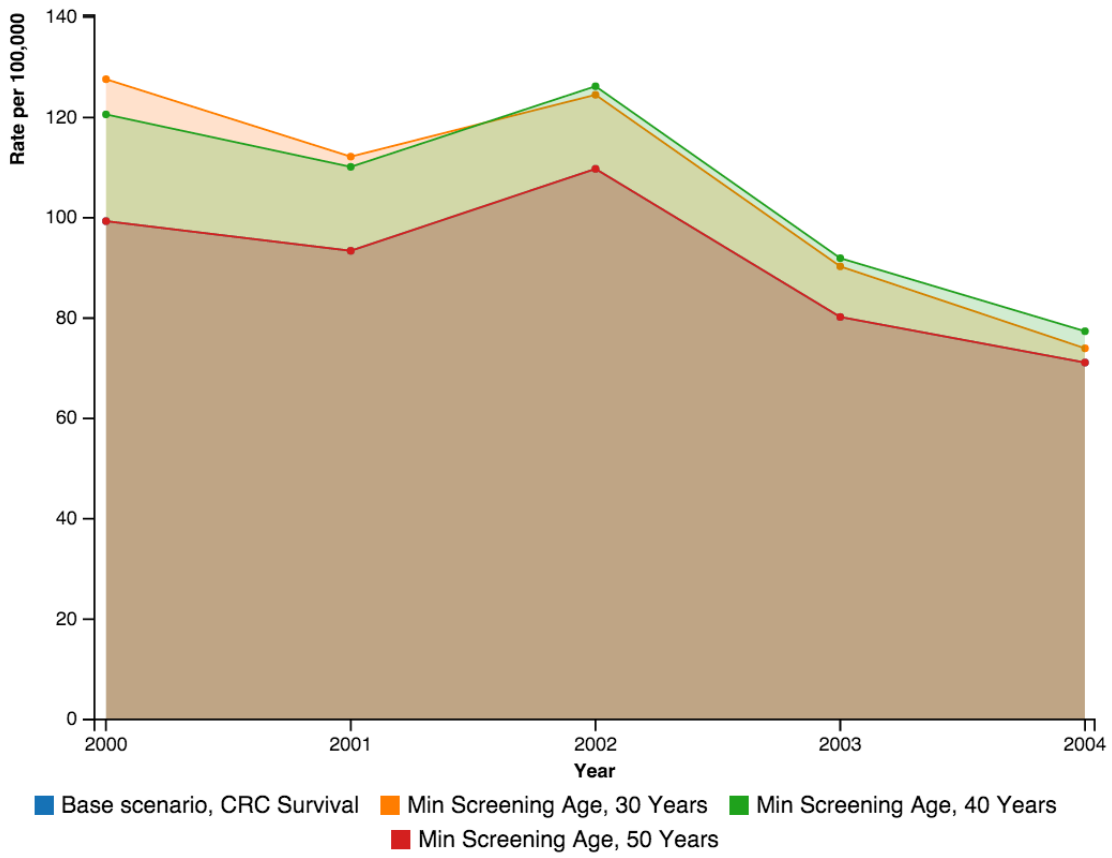


Figure 115: Hours Modeled CRC survival by recommended screening age starting at 30, 40 and 50 years

Compared to the current strategy of starting screening the average risk population at the age of 50, reducing the recommended screening age by ten years resulted in an a

16% total improvement in survival rates for all simulated years. Analogously to what was presented earlier, decreasing the recommended screening age further to 30 did not result in similarly improved CRC survival rates comparing to the scenario with the recommended screening starting at age 40 (Table 21).

Table 21: The change in survival rates by recommended screening age

Scenario	2000	2001	2002	2003	2004	Total (change%)
Base scenario, CRC Survival	99.13	93.23	109.53	80.07	71	452.97
Min Screening Age, 40 Years	120.37 (+21%)	109.93 (+18%)	125.97 (+15%)	91.77 (+15%)	77.23 (+9%)	525.27 (+16%)
Min Screening Age, 30 Years	127.37 (+28%)	111.97 (+20%)	124.27 (+13%)	90.13 (+13%)	73.83 (+4%)	527.57 (+16%)

In the next, section I explore the influence of varying the levels of knowledge (to 30%, 60%, and 90%) about CRC symptoms and CRC risk factors for the general population.

7.9.2 Scenario 2: Increasing CRC Knowledge for General Population

Knowledge about CRC is an important contributor to reducing delays in CRC investigation procedures for agents with CRC. Knowledge about CRC has been shown to influence patient delay, primarily through screening pathways (Gimeno-Garcia et al., 2009; Cockburn et al., 2003). For an agent to consider seeking medical attention upon noticing symptoms, he initially needs to perceive them as serious and attribute these symptoms to a health issue that he has some knowledge about. In the case of CRC, it is common for people to attribute their symptoms to ailments other than CRC, such as diet, stress, hemorrhoids or aging processes (Siminoff et al., 2014). The initial population proportion with knowledge about CRC was set to 47% (Scroggins & Bartley, 1999).

The influence of CRC knowledge on CRC incidence, survival and delays were explored in Cancerscape by varying the population levels of CRC knowledge to 30%, 60%, and 90%, as shown in Figure 116.

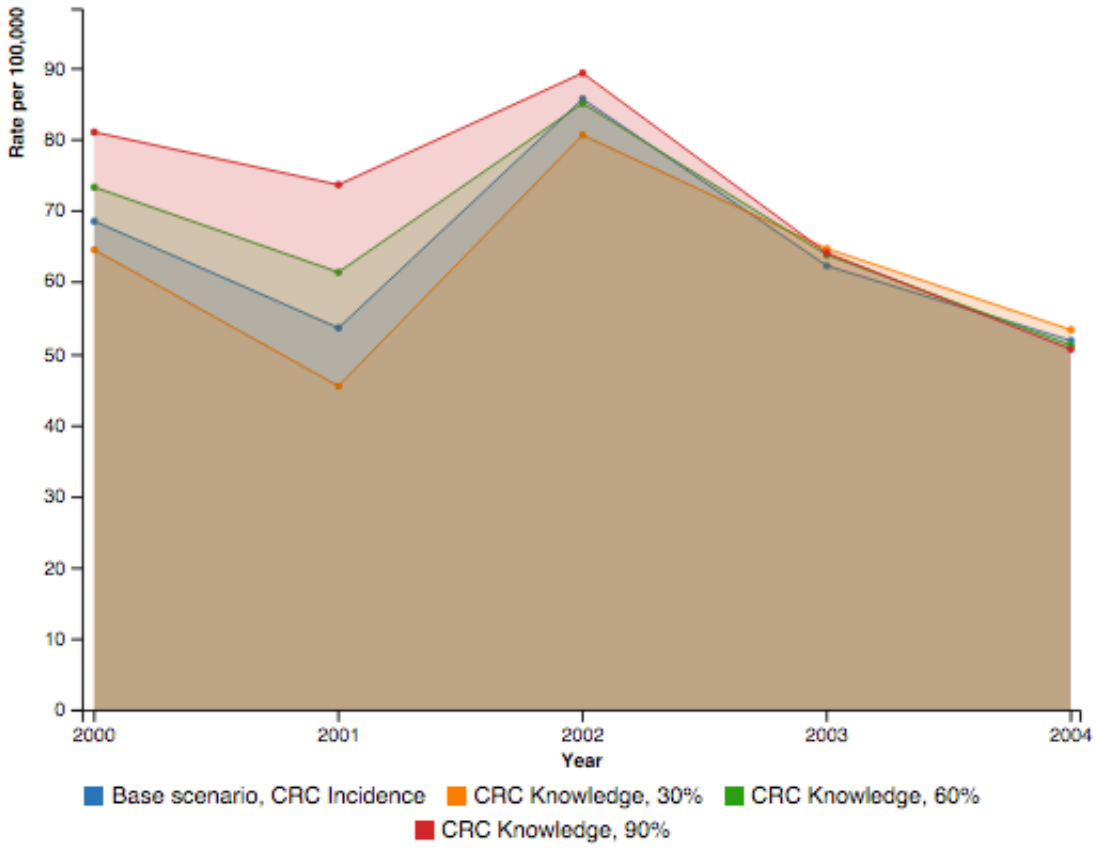


Figure 116: Modeled CRC incidence by CRC knowledge levels at 30%, 60% and 90%

This policy resulted in increased CRC incidence rates during the first couple of simulated years, followed by a small reduction in the subsequent years. As a potential explanation for this pattern may serve the fact that most agents who needed a colonoscopy received it during the first years, and given that the recommended frequency

of this procedure is once every five years, the agents did not need it for the following years, therefore, the decrease in the overall CRC incidence occurred. Higher levels of CRC knowledge were associated with the higher incidence rates, as more agents with symptoms were more likely to attribute them to CRC and proceed with timely seeking medical care for their symptoms.

Underestimation of symptoms severity was shown to increase the patient delay in seeking medical care (Courtney et al., 2012). If the agent has knowledge about CRC or about his family history in particular, he is more likely to perceive the symptoms as serious and requiring medical care. As mentioned previously, for the symptoms typical to CRC it is common for people to attribute their symptoms to other conditions such as diet, stress, hemorrhoids or aging processes (Siminoff et al., 2012), therefore increasing the delay in both diagnosis and treatment.

Furthermore, even when agents have knowledge about CRC and attribute their symptoms to this condition, they may continue experiencing delays due to lack of health insurance. The results in terms of the Andersen Model of Total Patient Delay by health insurance coverage across the various levels of CRC knowledge are presented in Figure 117.

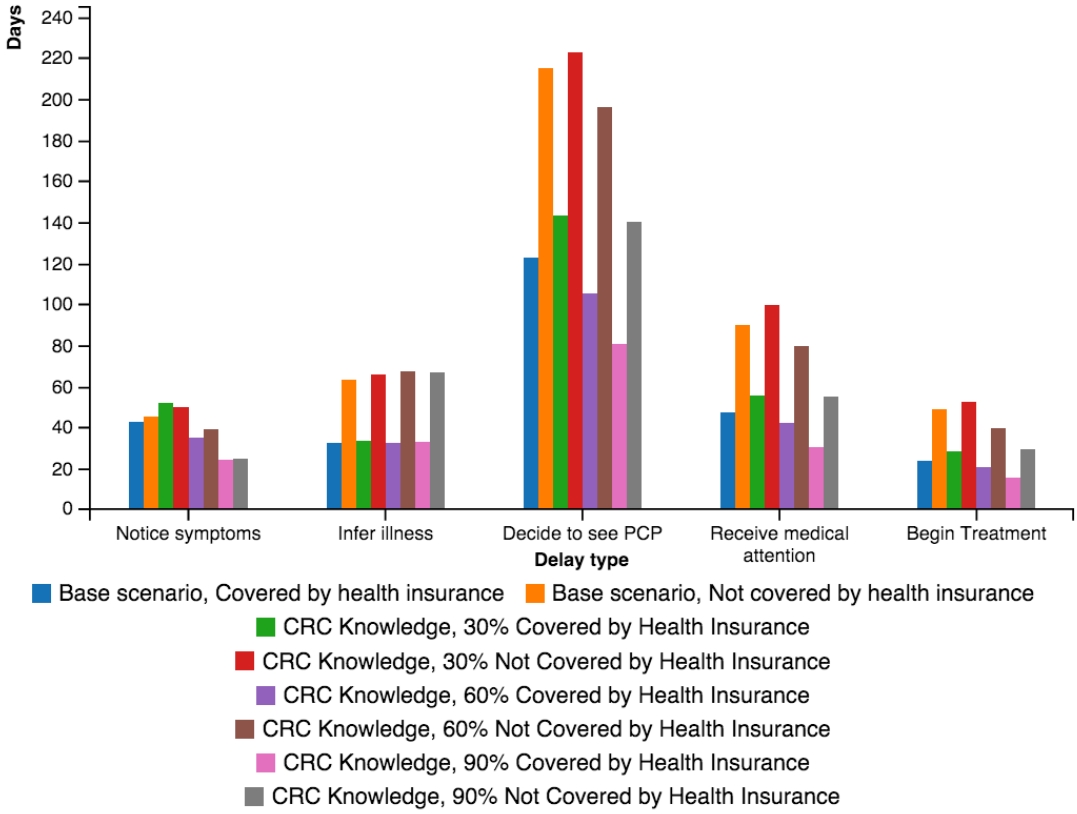


Figure 117: Modeled delays by health insurance coverage for CRC knowledge levels at 30%, 60% and 90%

As expected, the higher the CRC knowledge levels, the lesser the delays across all stages of the Andersen Model of Total Patient Delay. CRC knowledge levels were particularly important in influencing delay at the 'Noticing Symptoms' stage as well as at the 'Deciding to Seek Medical Attention' step.

Higher levels of CRC knowledge resulted in increased rates of polyp removal during the colonoscopy procedures in the model, as shown in Figure 118.

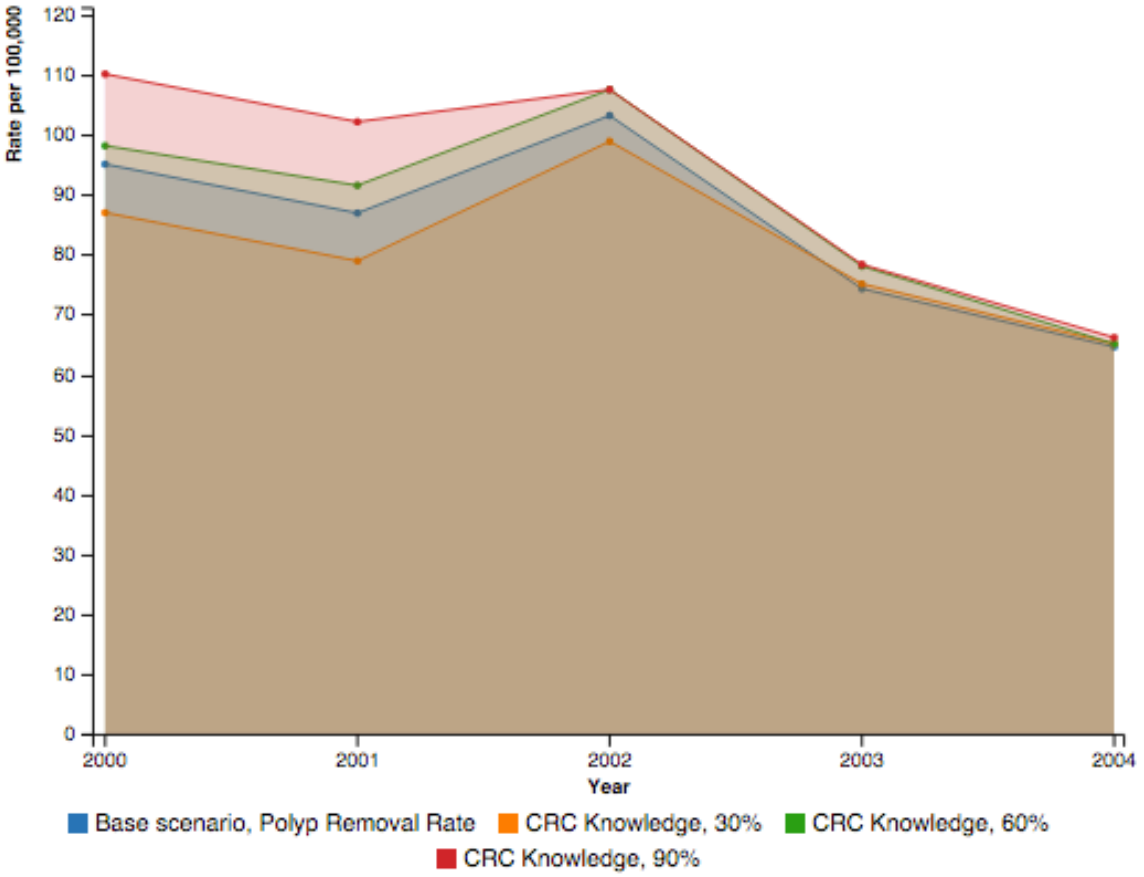


Figure 118: Modeled CRC polyp removal rate by CRC knowledge levels at 30%, 60% and 90%

Similar to the dynamics of CRC incidence for this scenario, the impact of varying the levels of CRC knowledge on polyp removal rates in the model was higher in the first couple of years. This is mainly due to the fact that most agents with existing CRC would associate their symptoms with this particular cancer type and seek medical care for their symptoms.

A similar pattern can be observed for CRC survival, as shown in Figure 119. CRC survival was higher for the setting with the highest CRC knowledge levels and lowest for the situation where agent population knowledge levels about CRC were only 30%.

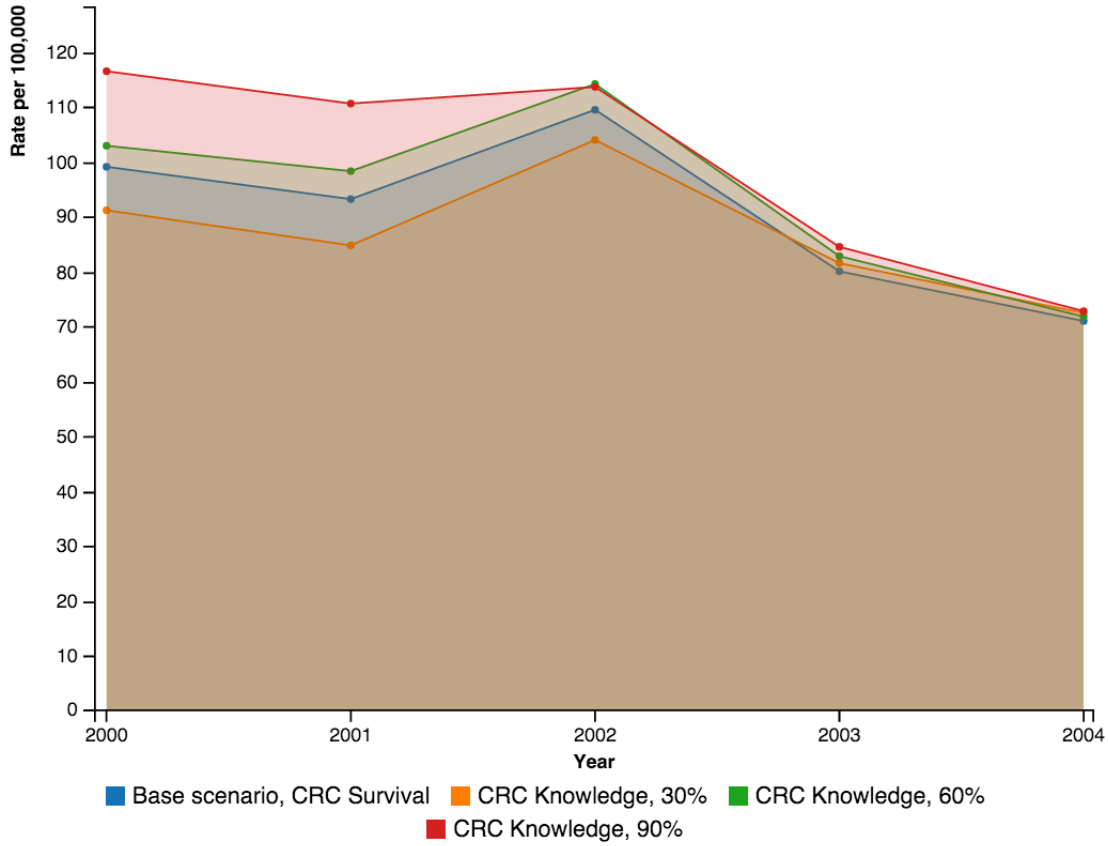


Figure 119: Modeled CRC survival rate by CRC knowledge levels at 30%, 60% and 90%

CRC knowledge is an important factor not only for reducing the delay in CRC diagnosis, treatment, and improving early diagnosis, but also for improving CRC survival rates. Policies should be put in place to make knowledge about CRC symptoms and risk factors accessible to the general population. Declaring the month of March as the CRC Awareness Month, or having popular TV personalities openly discuss the importance of CRC screening are just a few ways by which the overall population knowledge levels about CRC could be increased. Recognizing the CRC symptoms and the importance of knowing one's family history are pathways that could lead to catching CRC earlier when

the CRC death risks are lower and the treatment more efficient. The most desirable scenario (besides primary prevention) is to catch the polyps and remove them before they turn into full-blown CRC.

7.9.3 Scenario 3: Increasing the Health Insurance Coverage for Colonoscopies

Among the most beneficial interventions for decreasing mortality is increasing the health insurance coverage (Milstein et al., 2011; Wilper et al., 2009). Even though the colonoscopy procedures have been shown to considerably reduce CRC related deaths rates, not all health insurance plans cover colonoscopy procedures. As mentioned in Section 3.4, colonoscopy procedures are rather costly and are associated with higher risks compared to the alternative CRC screening methods - such as FOBT or barium enema. Even though these less expensive, they are also less sensitive (have higher false positive rates) and lack the capability to remove polyps during the screening procedure itself, if needed. Therefore, in this section, I explore the effect of different levels of colonoscopy coverage on CRC incidence, polyp removal rates and on the overall CRC mortality.

Varying the levels of coverage for colonoscopy (to 30%, 60%, and 90%) was initially explored in relation to CRC incidence. As expected, increased levels of health insurance coverage for colonoscopies contributed to reducing the overall rates of CRC incidence, as shown in Table 22.

Table 22: The change in incidence rates by health insurance coverage for colonoscopies

Scenario	2000	2001	2002	2003	2004	Total (change%)
Base scenario, CRC Incidence Colonoscopy, 30% (default)	68.47	53.53	85.57	62.23	51.77	321.57
Colonoscopy, 60%	72.63 (+6%)	55.77 (+4%)	85.13 (-1%)	56.8 (-9%)	44.57 (-14%)	314.9 (-2%)
Colonoscopy, 90%	71.83 (+5%)	61.13 (+14%)	83.33 (-3%)	54.53 (-12%)	41.87 (-19%)	312.7 (-3%)

A coverage of 60% for colonoscopies resulted in a slight increase in CRC incidence rates during the first few years followed by a decrease in incidence rates for all the following years. As expected, an increased health care insurance coverage for colonoscopies was associated with an even further decrease in the CRC incidence rates, as shown in Figure 120.

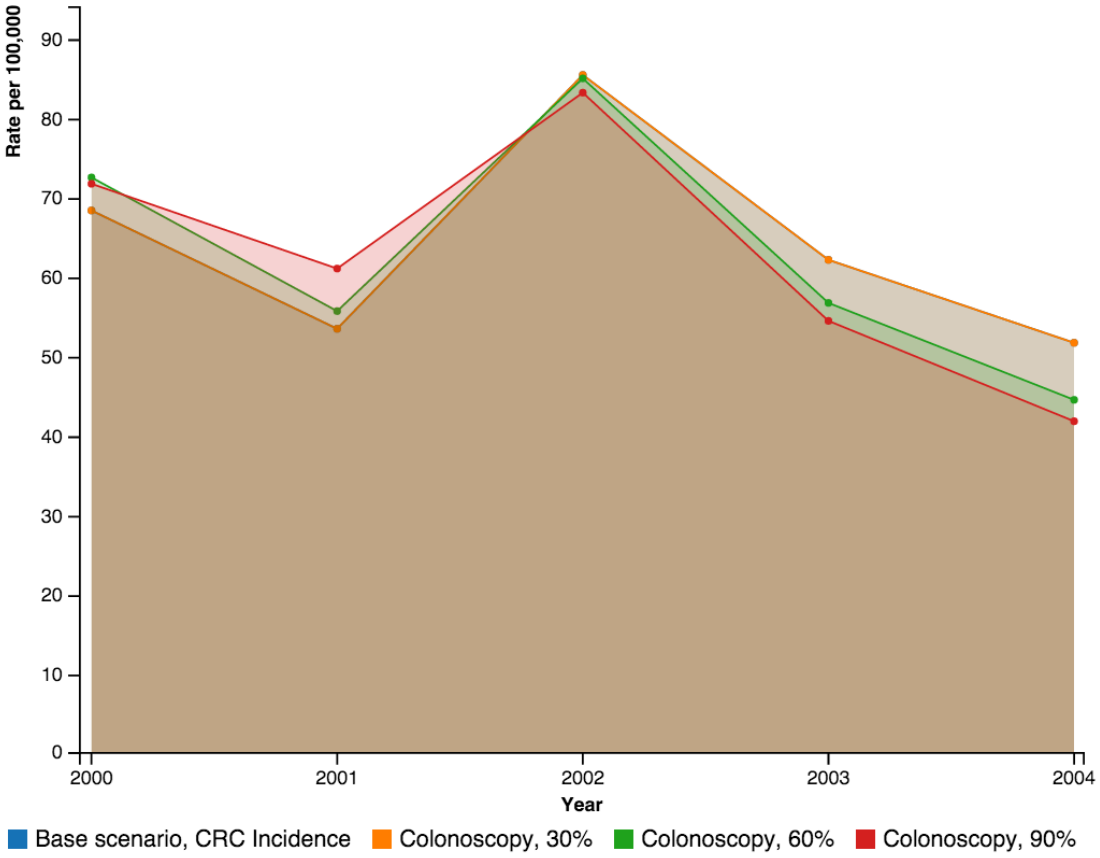


Figure 120: Modeled CRC incidence rate by colonoscopy coverage levels at 30%, 60% and 90%

Correspondingly, increasing colonoscopy coverage rates was also associated with higher polyp removal rates, as shown in Figure 121.

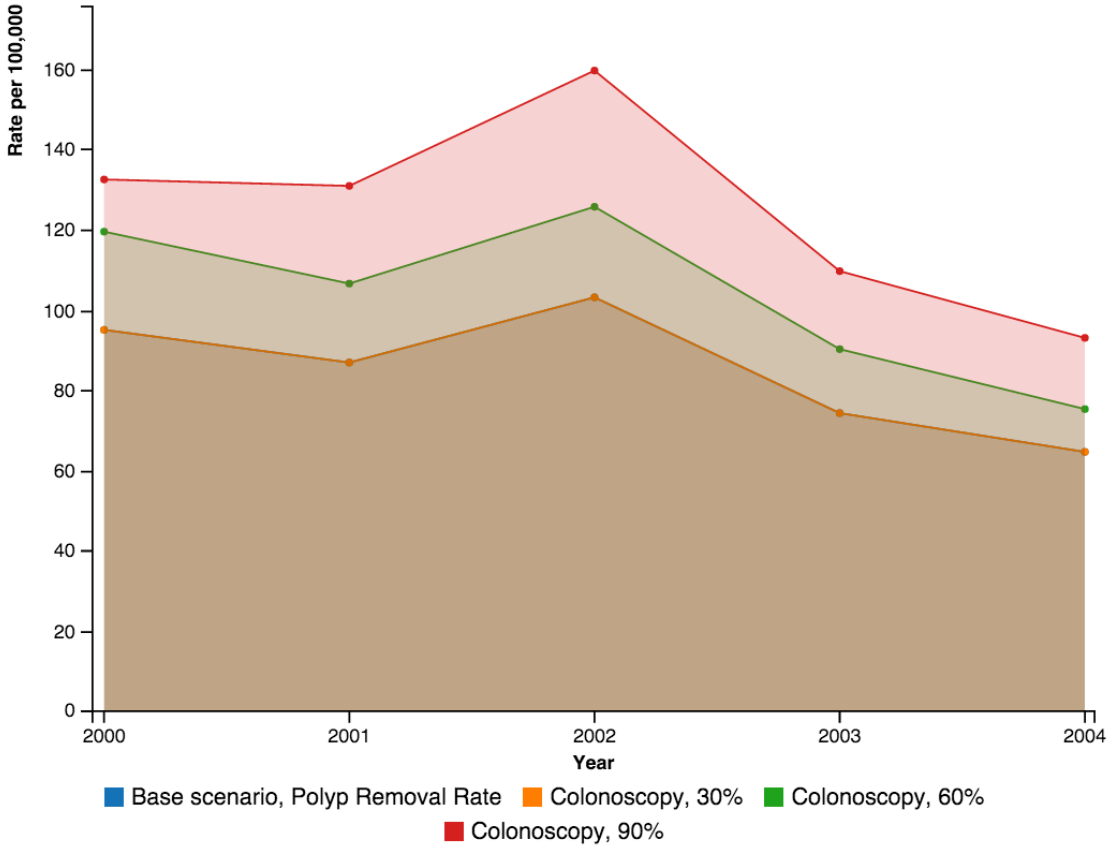


Figure 121: Modeled CRC polyp removal rates by colonoscopy coverage levels at 30%, 60% and 90%

A colonoscopy coverage of 90% led to the highest levels of polyp removal rates. If the primary care physician recommended an agent to undergo colonoscopy, the likelihood is greater that he will undergo the procedure and have his polyps removed (if any) in the case in which he has healthcare coverage that covers this particular procedure. Polyp removal rates increased with increasing colonoscopy coverage levels for this particular procedure. Removing a larger number of polyps, and avoiding having them turn into cancer could prevent premature deaths. Colorectal polyps, that would have

otherwise been missed, were found early and removed, correspondingly influencing both CRC mortality and survival rates in the model.

Accordingly, higher levels of colonoscopy coverage were associated with reduced CRC death rates, as shown in Figure 122.

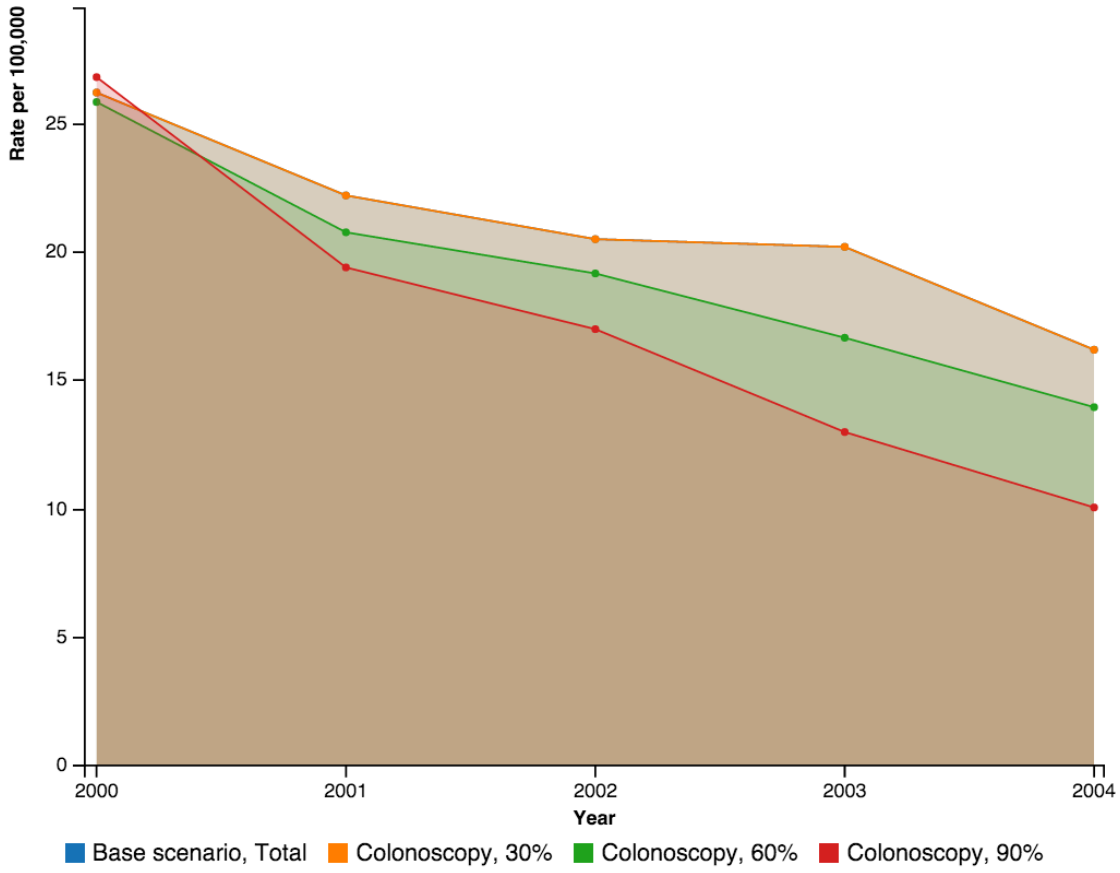


Figure 122: Modeled CRC incidence rate by colonoscopy coverage levels at 30%, 60% and 90%

This pattern could potentially be explained by the fact that individuals who had CRC and had the possibility to undergo the procedure early in the process, avoided early death while contributing to an overall decrease in the CRC mortality rates. Lesser levels

of colonoscopy coverage were associated with lower chances for agents that need this procedure to get it. The polyps were therefore missed, and they continue to develop over time into CRC, causing symptoms. With a more advanced stage at diagnosis, CRC survival rates decrease, and the CRC mortality rates increase. Similarly, a 60% increase in Colonoscopy coverage led to an increase in CRC survival, with the increase continuing when colonoscopy coverage was increased further to 90%, as shown in Figure 123.

Given that higher levels of colonoscopy coverage contributed to increased CRC survival rates, policies, especially regarding state mandates for CRC screening coverage, and particularly the ones covering colonoscopy procedure (or decreasing its costs) should be put in place and further studied. While other tests that screen for CRC are also important, they are not curative. Considering the fact that colonoscopy offers the possibility not only to find but also remove colorectal polyps before they turn into cancer, the possibility to afford colonoscopy, when recommended by a physician, can be life-saving.

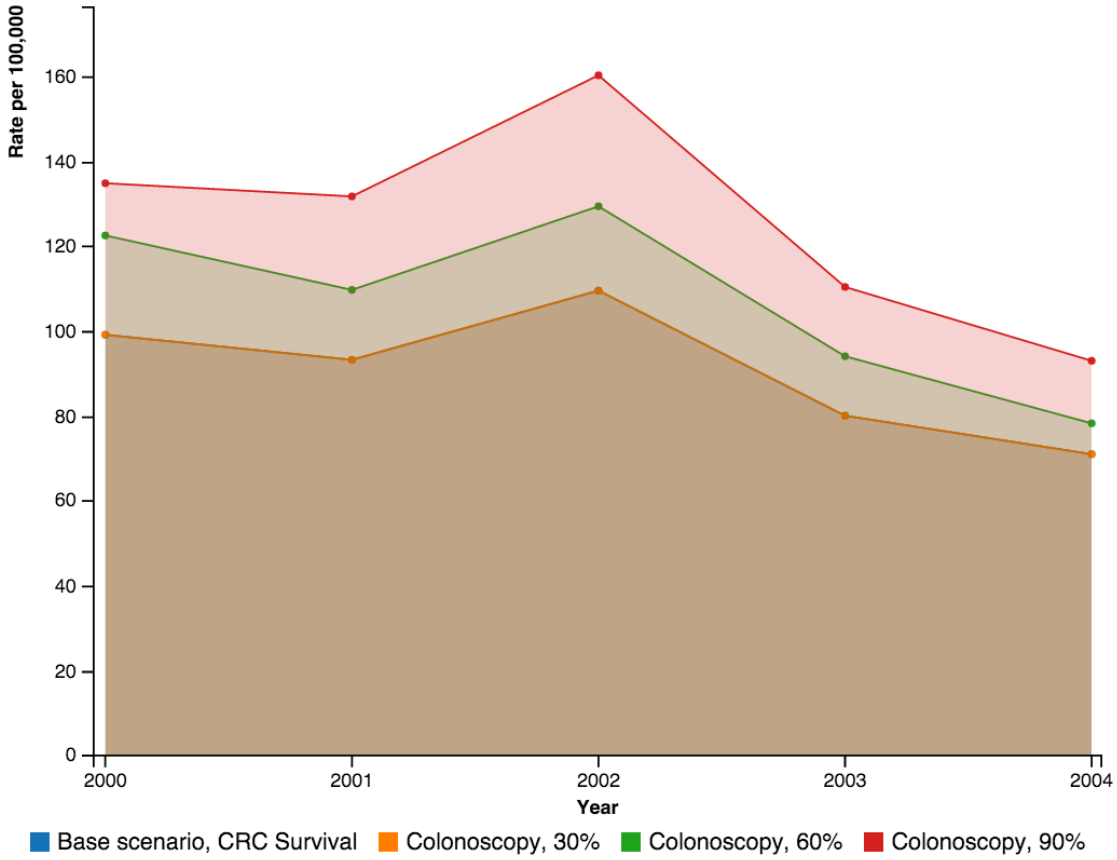


Figure 123: Modeled CRC survival rate by colonoscopy coverage levels at 30%, 60% and 90%

7.10 Summary

Chapter 7 began with presenting the verification, validation and calibration aspects of Cancerscape. Following that, the results of the model were presented, starting with the resulting CRC incidence and mortality rates at the end of five simulation years, with a focus on the distribution of CRC by age, as well as by race, gender and health insurance coverage status.

Disparities were discussed throughout the chapter regarding race, gender and health insurance status. African Americans were more likely to develop CRC in

Cancerscape, and to die because of it compared to Whites. Whites were less likely to have their tumors diagnosed early (when treatment is most effective). Although an overall increase in screening rates has been observed in the district, CRC screening rates were higher in Whites compared to African Americans, which is also consistent with the lower health insurance coverage rates among African Americans. While the fundamental causes of these differences are multifactorial, late stage at diagnosis may explain more than half of the disparities between the White and African American population groups in terms of CRC survival (Mayberry et al., 1995). Disparities concerning gender have also been discussed. The chapter concluded with presenting the analysis and the comparison of the three scenarios explored in Cancerscape, contributing to the design of more informed cancer control policy strategies. Scenario analysis helps to increase the level of validation of ABMs. A comprehensive list of the scenario explorations presented in this section is provided in Appendix G.

A dashboard with the results presented in this section was created and is currently available for interactive exploration at this web address: http://cmetgher.github.io/cancerscape/ (Appendix A). The results at this site are available for the base case and for the three scenarios described in Section 7.9. For each scenario, the results are accessible for 30 individual separate runs. The algorithm for generating the dashboard and the tools used for this purpose are presented in Appendix B.

CHAPTER 8: POLICY IMPLICATIONS

Contents

8.1 Introduction 287

8.2 CRC Legislation 288

8.3 Overall Implications of the Cancerscape Framework 289

8.4 Implications of Scenario Analyses Results 290

8.5 Summary 300

> "Today the network of relationships linking the human race to itself and to the rest of the biosphere is so complex that all aspects affect all others to an extraordinary degree. Someone should be studying the whole system, however crudely that has to be done, because no gluing together of partial studies of a complex nonlinear system can give a good idea of the behavior of the whole."
>
> (Murray Gell-Mann, 1997)

8.1 Introduction

The world is becoming increasingly complex and highly interconnected. The multilevel complexity inherent in the social determinants of cancer calls for new approaches for designing policies and interventions that would alter the dynamics of cancer at the population level.

Complex system methods are powerful tools for planning health policy, and some argue their use will considerably increase in the future (Maglio & Mabry, 2010). ABM,

in particular, is a powerful methodology for combining evidence from multiple sources in a systematic manner in order to generate analytical information and guide policy makers.

Once an ABM has been calibrated and validated, it can potentially be used guiding policy, by using scenario analysis to provide policy-relevant information, such as providing various screening scenarios and assessing their influence on the overall CRC incidence and mortality. In the case of cancer, evaluating scenarios of interest in real life would not only be difficult, but also unethical. Simulation-based models have been successfully used to advice policy and provide recommendations regarding interventions related to different health issues such as mental health (Koizumi et al., 2009), drinking behaviors (Gorman et al., 2006), obesity (Hammond, 2009) and pandemics (Epstein, 2009). Cancerscape integrates diverse data from multiple sources, existing frameworks, theories and knowledge from several disciplines to explore the social determinants of cancer within a unified framework. This chapter will discuss the policy aspect and the overall implications of this work.

8.2 CRC Legislation

Screening is among of the most cost-effective ways to prevent CRC. CRC is one of the merely two cancers that received a grade 'A' recommendation from the USPSTF (a Grade 'A' recommendation for a certain service is given by the USPSTF when there is confidence that the net benefit of offering that service is substantial) (Pignone et al., 2002).

The cost of a colonoscopy procedure (without polyp removal) was estimated to be within the range of $150 (Roge et al., 1994) and $1000 (Neugut & Young, 1996). The cost of colonoscopy including polypectomy (polyp removal) is estimated to be between

$150 (Gyrd-Hansen et al., 1998) and $1500 (Lieberman, 1995). Most states within the United States passed laws that mandate health insurance plans to cover CRC screening procedures, including the endoscopic testing methods. Several studies found a positive association between state mandates and the use of recent CRC screening tests for individuals with health insurance (Rathore et al., 2000; Nelson et al., 2002; Cokkinides et al., 2011). District of Columbia implemented the mandate on CRC screening coverage on April 13, 2002 (Cokkinides et al., 2011). While the mandates at the state level were shown to contribute to an increase in CRC screening test usage, other factors – such as demographic, socioeconomic, or the mere awareness of these benefits at individual, patient or physician level – need to be considered in order to understand the overall uptake in CRC screening test usage.

In the following two sections, I discuss the both the implications of Cancerscape as a tool as well as the implications of the results based on the model scenario analyses. Section 8.3 discusses the general implications of the Cancerscape framework while Section 8.4 focuses on the results of using the framework for exploring the various policy scenarios.

8.3 Overall Implications of the Cancerscape Framework

Among the most notable implications of this work is applying novel, computational social science methods in order to design and build a multi-level ABM of the social determinants of cancer. Cancerscape provides a new way to explore the complex nature of the social determinants of cancer into a dynamic and integrative framework that allows agent heterogeneity (in terms of demographics, socio-economic characteristics, risk

factors and social networks), integration of data and theories within a spatial context and interaction (between agents and among agents and their environment) providing a systems view on the social determinants of cancer. Cancerscape goes beyond the traditional reductionist approaches predominant in the discipline, to accommodate realistic key processes of cancer development as well as interactions and behavioral adaptations (according to peer influence) over time and within a relevant geographic environment. Consequently, cancer incidence and mortality trends emerge as a result of the dynamic interaction among agents and between agents and their environment.

By design, the Cancerscape framework allows extensibility to other cancer types. To accomplish this - spatial, socio-demographic and risk factor data preparation is needed, as well as information about the development and stage progression for the particular cancer type of interest is required. Cancerscape also provides an "artificial laboratory" that can be used as a tool for policy support, by offering the possibility to assess the outcomes of various policy scenarios.

8.4 Implications of Scenario Analyses Results

CRC prevention is one of the most promising avenues in medicine at this time (Sarfaty & Wender, 2007). Screening can both prevent CRC and reduce CRC mortality. While it is one of the most preventable cancers, CRC is the second leading cause of cancer deaths in the district. Most colorectal cancers develop from non-malignant polyps over an extended period, giving a wide window of opportunity to catch CRC early and prevent it from developing. Social determinants of cancer pose a complex public policy challenge, as indicated by the large number of factors involved and the nonlinear interactions between these factors.

The first scenario explored in Cancerscape focused on decreasing the recommended screening age for CRC, as shown in Table 23, and discussed in Section 7.9.1. The current screening age that also served as the default value for the model is the age of 50 for general population (for average risk individuals) and age 40 for those at higher risk, as recommended by the USPSTF (2002).

Table 23: Relative change in CRC incidence by year for Scenario 1

Scenario	2000	2001	2002	2003	2004	Total (change%)
Base scenario, CRC Incidence Min Screening Age, 50 Years	68.47	53.53	85.57	62.23	51.77	321.57
Min Screening Age, 40 Years	85.77 (+25%)	72.3 (+35%)	101.17 (+18%)	73.23 (+18%)	56.23 (+9%)	388.7 (+21%)
Min Screening Age, 30 Years	90.17 (+32%)	74.2 (+39%)	100.17 (+17%)	68.4 (+10%)	55.6 (+7%)	388.53 (+21%)

Decreasing the recommended screening age in Cancerscape to 40 years was associated with a substantial increase in the overall CRC incidence. The result of this policy scenario led to a total increase of 21% in CRC screening rates for the five years in Cancerscape. Further decreasing the recommended screening age to 30 years provided similar results to the previous policy, indicating that decreasing the recommended screening age further would be a less efficient avenue for exploration. This result may be attributed to the lower prevalence of colorectal polyps among the younger population.

Policy experiments with CRC screening in Cancerscape were shown to be the most efficient in reducing both incidence and mortality. The fewer people 50 years of age or older waited to get their symptoms checked by a primary care physician the higher the chances for doctors to find the polyps/tumors early when treatment is most effective,

resulting in better survival rates. The effects of decreasing the minimum recommended screening age were also explored regarding survival, as shown in Table 24.

Table 24: Relative change in CRC survival by year for Scenario 1

Scenario	2000	2001	2002	2003	2004	Total (change%)
Base scenario, CRC Survival	99.13	93.23	109.53	80.07	71	452.97
Min Screening Age, 30 Years	127.37 (+28%)	111.97 (+20%)	124.27 (+13%)	90.13 (+13%)	73.83 (+4%)	527.57 (+16%)
Min Screening Age, 40 Years	120.37 (+21%)	109.93 (+18%)	125.97 (+15%)	91.77 (+15%)	77.23 (+9%)	525.27 (+16%)

Similar to the results regarding incidence, decreasing the recommended screening age to 40 years resulted in improved survival for all years in Cancerscape. While decreasing the screening age further by ten years resulted in slight improvements in 2000 and 2001, the total survival for all years did not increase compared to the strategy of starting screening at the age of 40.

The U. S. Congress designated March of the year 2000 as the first ever CRC awareness month (Shapiro et al., 2001), aiming to increase the awareness among the individuals eligible for screening and contribute to improving the CRC prevention efforts. Cancerscape allows playing out a number of policy scenarios to investigate the change in CRC incidence by exploring parameters such as knowledge about CRC while keeping the initial parameters the same. As shown in Table 25, CRC knowledge levels within the population were varied to explore their effects on CRC incidence.

Lower levels of knowledge about CRC among the population groups resulted in decreased incidence rates.

Table 25: Relative change in CRC incidence by year for Scenario 2

Scenario	2000	2001	2002	2003	2004	Total (change%)
Base scenario, CRC Incidence	68.47	53.53	85.57	62.23	51.77	321.57
CRC Knowledge, 30%	64.47 (-6%)	45.4 (-15%)	80.53 (-6%)	64.63 (+4%)	53.27 (+3%)	308.3 (-4%)
CRC Knowledge, 60%	73.23 (+7%)	61.33 (+15%)	84.97 (-1%)	63.77 (+2%)	51.1 (-1%)	334.4 (+4%)
CRC Knowledge, 90%	80.93 (+18%)	73.57 (+37%)	89.2 (+4%)	64.03 (+3%)	50.57 (-2%)	358.3 (+11%)

Higher levels of CRC knowledge, on the other hand, was associated with increased CRC incidence rates, which came as a result of more agents getting screened and more polyps being removed before they would have eventually developed into cancer. With increased CRC knowledge levels, more opportunities to find cancer opened and more polyps that could have been missed - were removed. This also influenced the rates of CRC mortality. As shown in Table 26, increased CRC knowledge levels were correspondingly associated with increased rates in CRC survival.

Table 26: Relative change in CRC survival by year for Scenario 2

Scenario	2000	2001	2002	2003	2004	Total (change%)
Base scenario, CRC Survival	99.13	93.23	109.53	80.07	71	452.97
CRC Knowledge, 30%	91.2 (-8%)	84.8 (-9%)	104.03 (-5%)	81.57 (+2%)	72.47 (+2%)	434.07 (-4%)
CRC Knowledge, 60%	102.97 (+4%)	98.33 (+5%)	114.23 (+4%)	82.83 (+3%)	71.8 (+1%)	470.17 (+4%)
CRC Knowledge, 90%	116.57 (+18%)	110.67 (+19%)	113.7 (+4%)	84.53 (+6%)	72.8 (+3%)	498.27 (+10%)

Increasing the CRC knowledge to 60% contributed to a total improvement in survival of 4% during the five simulation years in Cancerscape. Increasing the levels of CRC knowledge further to 90% led to a more than double increase in survival. The improvement in CRC incidence rates with increasing levels of CRC knowledge can be attributed to less advanced stages at diagnosis. This indicates the importance of developing policies aimed at targeting an increase in the population levels of CRC knowledge.

CRC awareness month presents one opportunity to accomplish this. Two successful examples of positive results achieved by increasing the levels of CRC knowledge are the "Super Colon" and the "Polyp Man" campaigns, shown in Figure 124 and Figure 125, respectively. The "Super Colon" is a physical model of the human colon, an inflatable installation with a height of 8 feet and a length of 20 feet that replicates the human colon. It has been successfully used on several occasions for the purpose of educating people about CRC risks and symptoms, as well as about the importance of early detection and timely treatment on survival. Several studies have shown these methods to be efficient in educating people about this particular type of cancer and motivating the eligible individuals to get screened as recommended (Finn et al., 2013; Fickle et al., 2015).

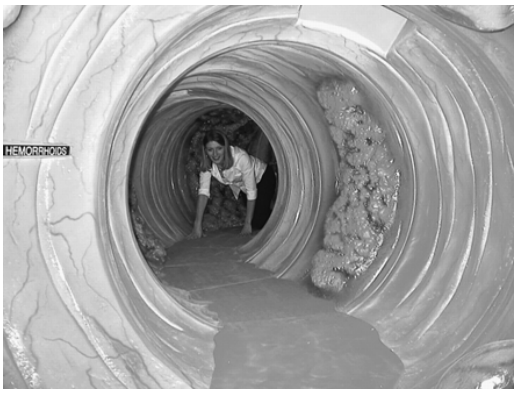


Figure 124: Colossal Colon founder Molly McMaster crawls through the 40-foot-long model of a human colon. The exhibit is on a 20-city tour around the United States (Source: Reynolds, 2003)

Figure 125: In a public service announcement, two doctors apprehend the Polyp Man. The campaign debuted last year during Colon Cancer Awareness Month (Source: Reynolds, 2003)

Another opportunity for increasing the awareness of CRC is indicated by the effects of celebrities and media personalities on CRC screening test use. An example of this is the boost in CRC screening rates associated with the TV personality Katie Couric's televised colonoscopy (Cram et al., 2003). Also, the American Cancer Society with Ad Council created the "Polyp Man" – a "comic-book-style super-villain" (an actor wearing a padded red polyp suit) chased down by doctors in a TV show (Reynolds, 2003). More targeted policies aimed at increasing the knowledge about CRC, its risk factors, demystifying the myths around the screening procedures as well as regarding the importance of screening for early detection are necessary.

Health insurance coverage for colonoscopies was explored in the second scenario and described in Section 7.7. The initial coverage of health care plans for colonoscopy procedures was increased from 30% to 60% and 90%, as shown in Table 27.

Table 27: Relative change in CRC incidence by year for Scenario 3

Scenario	2000	2001	2002	2003	2004	Total (change%)
Base scenario, CRC Incidence Colonoscopy, 30%	68.47	53.53	85.57	62.23	51.77	321.57
Colonoscopy, 60%	72.63 (+6%)	55.77 (+4%)	85.13 (-1%)	56.8 (-9%)	44.57 (-14%)	314.9 (-2%)
Colonoscopy, 90%	71.83 (+5%)	61.13 (+14%)	83.33 (-3%)	54.53 (-12%)	41.87 (-19%)	312.7 (-3%)

Increasing colonoscopy coverage levels contributed to a slight increase in CRC incidence during the first two years, followed by a marked CRC incidence decrease during the following years. The decrease between 2002 and 2004 can be attributed to the fact that more people received screening between 2000 and 2001 and they did not require the procedure for the following five years. The colonoscopy screening procedure and the frequency it should be administered according to the USPSTF are explained in more detail in Sections 3.4.1 and 6.7.3. While these are largely unrealistic cases to explore in the real life given the difficulty of implementation and the ethical and financial toll this would take, this policy can be easily explore in a validated ABM, and it was shown to contribute to a decrease in CRC incidence. One of the pathways by which healthcare coverage may affect CRC incidence is via screening.

Apart from contributing to an overall decrease in CRC incidence rates in Cancerscape, increased colonoscopy coverage was also associated with a significant increase in the polyp removal rates as a result of these procedures. It is assumed that the individuals eligible for screening are in total compliance with seeing a primary care physician on a yearly basis for an annual health exam. The result regarding the amount of

polyps removed by healthcare plans coverage levels for colonoscopies is presented in Table 28. The polyp removal rates increased with increasing levels of coverage.

Table 28: Relative change in polyp removal rates by year for Scenario 3

Scenario	2000	2001	2002	2003	2004	Total (change%)
Base scenario, Polyp Removal Rate Colonoscopy, 30%	95	86.87	103.1	74.27	64.6	423.83
Colonoscopy, 60%	119.47 (+26%)	106.53 (+23%)	125.67 (+22%)	90.2 (+21%)	75.27 (+17%)	517.13 (+22%)
Colonoscopy, 90%	132.47 (+39%)	130.87 (+51%)	159.6 (+55%)	109.63 (+48%)	93 (+44%)	625.57 (+48%)

A 60% level of healthcare plan coverage for colonoscopy procedures was associated with a total polyp removal rate increase of 22% for all years in Cancerscape. The improvement associated with increasing colonoscopy coverage further to 90% was nearly double. This highlights the importance of health care coverage for CRC preventive services, as it gives the opportunity for physicians to provide eligible individuals with screening recommendations, so that they can get screened as necessary, given the fact that they have the procedure covered by health insurance.

DC had the highest mortality rates due to CRC in the United States for both males and females in 2000 (Stewart et al., 2004). Screening for CRC, as well as access to quality treatment, can contribute to considerably decreasing CRC mortality rates. Therefore, the effect of colonoscopy coverage was also explored in terms of CRC

mortality. As expected, increasing levels of coverage for colonoscopies led to decreased levels of CRC mortality, as shown in Table 29.

Table 29: Relative change in CRC mortality by year for Scenario 3

Scenario	2000	2001	2002	2003	2004	Total (change%)
Base scenario, Total Mortality	26.17	22.17	20.47	20.17	16.17	105.13
Colonoscopy, 60%	25.8 (-1%)	20.73 (-6%)	19.13 (-7%)	16.63 (-18%)	13.93 (-14%)	96.23 (-8%)
Colonoscopy, 90%	26.77 (+2%)	19.37 (-13%)	16.97 (-17%)	12.97 (-36%)	10.03 (-38%)	86.1 (-18%)

The total decrease in mortality for all years in the model was 8% for a colonoscopy coverage level of 60% and 18% for 90% colonoscopy coverage. The mortality rates due to CRC would considerably be reduced if more colorectal polyps were found early and removed (during colonoscopy procedures) before they turn into full-blown cancer and when the treatment becomes complicated, expensive or often impossible. If the window of opportunity to prevent cancer through screening was missed, treatment is the next line of defense to avoid CRC mortality. Policies targeting a decrease in the severity of the CRC stage at diagnosis could considerably improve survival. Policies should, therefore, be targeted to lowering the stage at diagnosis as the second line of defense, post primary prevention.

To efficiently accelerate the reduction in CRC mortality, a broader combination of policies aimed at decreasing the prevalence of risk factors, increasing the prevalence of protective factors, increasing the CRC screening rates and providing access to the best available treatments is necessary.

Finding opportunities to reduce the delay is another important factor that policies aimed at decreasing CRC incidence and mortality should consider. The waiting time before having a colorectal tumor diagnosed, and the time patients wait in getting screened or treated need to be reduced. These delays are influenced by factors at individual/agent level (such as knowledge about CRC, attitudes about CRC screening, cultural beliefs, age, gender or marital status), at healthcare provider level (such as doctor attitudes towards CRC screening, doctor gender or cultural background) and at system level – where factors such as CRC screening state mandates, lowering the costs of the screening tests and making them available to the eligible population play a role (delay is discussed in more detail in Section 3.8).

State policies regarding CRC should focus more on ways of preventing CRC, together with the efforts spent on advancing the treatment methods. The most effective way to decrease both incidence and mortality would be targeting CRC screening policies towards educating individuals about risk factors and symptoms of CRC in order to decrease delay before individuals seek medical attention for their symptoms. Using ABMs such as Cancerscape, we can identify areas where data is missing and attempt to fill those gaps using simulation data. Also, this can guide new data collection.

The scenarios explored in Cancerscape are merely for illustration purposes and are not meant to provide any predictive validity. By running the model several times with varying different parameters and noticing their influence on the overall dynamics of CRC, it is possible to develop hypotheses about what would work best and which scenarios are worth a more in-depth examination. However, Cancerscape provides a

framework that could be used for identifying policy interventions that are best suited to address specific concerns about CRC control.

8.5 Summary

Complex system methodologies offer suitable tools and techniques for addressing the complex nature of social determinants of cancer and for informing policy design. ABMs can discipline the policy discussion, making CRC control judgments more informed and reliable. Compared to 'mental' models, that are implicit – ABMs can explicitly be tested and their assumption verified (Epstein, 2006). Using ABMs, policy makers can test several hypotheses and exclude the ones producing system behaviors that are unrealistic and extremely dissimilar from the system being modeled. No model alone can displace human judgment, but ABMs has the advantage of offering transparency about the assumptions that go into the model, as well as in regards to their implementation. Being explicit, ABMs may be easier to understand by policy makers compared to equation-based models that require, is some cases, extensive mathematical elaboration. ABM helps in evaluating the effects of policy interventions before applying them to the real world.

Complex systems pose unprecedented challenges for the design of effective policy interventions. Because the elements of these systems are highly interconnected, policies targeted at one level of the system may influence other levels. Among the most cited examples of a negative influence due to the lack of consideration of the many levels involved in a complex system is the Lake Victoria upheaval (Fuggle, 2001). The introduction of a new fish species (Nile perch, one of the largest species of freshwater fish, reaching over 6ft in length) in Lake Victoria was associated with the near-eradication of a few hundreds of native species, disrupting the whole ecosystem.

This had numerous social and economic implications. New fishing companies were established, and tourism has grown. To preserve the meat of this fish, it needs to be smoked, as drying it under the sun is not an option (a strategy that was used for the previous lake fish species) due to its high-fat content. This led to increased deforestation and soil erosion (among other serious health issues) in the area (Reynolds & Greboval, 1988). While the initial goal of introducing this fish species was to improve the overall welfare of the population in that region, this policy did not consider the other levels of this ecosystem and did not anticipate the response of other system players. The wide range of actors involved, the levels and interactions that make up the social determinants of cancer, need to be systemically considered and accounted for when designing policy interventions.

The policy scenarios in Cancerscape are not specifically meant to predict the exact effects of changes in specific parameters on the overall system (even though that would be possible). It is rather meant to provide a test bed for addressing – qualitatively and quantitatively – the assessment of different policy options under similar initial assumptions. While the model provides the opportunity to test several policy scenarios for exploring various experiments and potential interventions aimed at altering the dynamics of colorectal cancer in terms of screening, incidence, survival, delays and mortality.

Cancerscape largely incorporates the systemic diversity involved in the several levels of the social determinants of colorectal cancer, addressing the heterogeneity of the agents involved in the system (their demographic, social, behavioral, network, tumor and

health-related characteristics among others). A source of diversity can be observed in the decision-making processes the boundedly rational agents take in regard to their health-related behaviors (Simon, 2000).

For future research directions, Cancerscape can be extended to explore more policies and interventions of interest that are not considered in this work. Interesting scenarios to explore may be an in-depth analysis of the diffusion of new treatments on the overall CRC survival, policies targeted on particular risk factor trends – looking at their effect on CRC incidence, environmental aspects or the effect of cancer advocacy groups in shaping CRC screening policy. Another idea would be to implement in the model some of the existing CDC chronic disease state policies for Washington, D. C. (CDC, 2015). Many of these policies may contribute to the risk factors trends in regard to CRC and having the possibility to test them on a dynamic, artificial society resembling the population of DC would also be of interest.

More attention is given to policies that helped individuals that are already sick compared to investing in designing policies that prevent illnesses from occurring in the first place (Heymann, 2005). While CRC is highly curable if found early, it remains to be a significant issue in Washington, D. C. and nationally, both in terms of incidence and mortality. Using and ABM approach we can distinguish system areas that would benefit the most of the specific interventions. These models can serve as 'artificial laboratories' where policy design can take place, the potential effects of certain interventions on the system as a whole can be assessed so that more informed decision can be taken. They can

complement existing methodologies in addressing the social determinants of cancer to offer a more comprehensive approach to controlling cancer.

CHAPTER 9: CONCLUSIONS

Contents

9.1 Introduction 304

9.2 Summary of Research Findings 306

9.3 Scientific Contributions..... 313

9.4 Limitations 316

9.5 Future Research..... 317

9.6 Concluding Remarks 319

> "As individuals and nations, we spend far more time and resources treating illnesses and injuries than we do addressing the conditions that give rise to them"
>
> (Healthy Societies: from Analysis to Action, 2005)

9.1 Introduction

The predominant methodology in social epidemiology is relying on observational data (Kaplan, 2004). SDC's emphasis on a broad range of factors tied to their social context calls for new methods that would address this multilevel complexity. In this dissertation, I attempted to employ novel complex system methodologies, ABM in particular, combined with social networks and GIS, to address this. In this section, I synthesize the

conclusions and findings of this dissertation. I will also discuss the wider scientific and policy implications of this work.

People who are poor live shorter lives and are ill more frequently than the rich ones. This fact suggests the high sensitivity of health to social circumstances (Wilkinson & Marmot, 2003) that are in turn associated with health outcomes. In theory, genes, biology, and health behaviors may explain about 25% of the whole population health, with the social determinants together with medical care accounting for the rest, as shown in Figure 126.

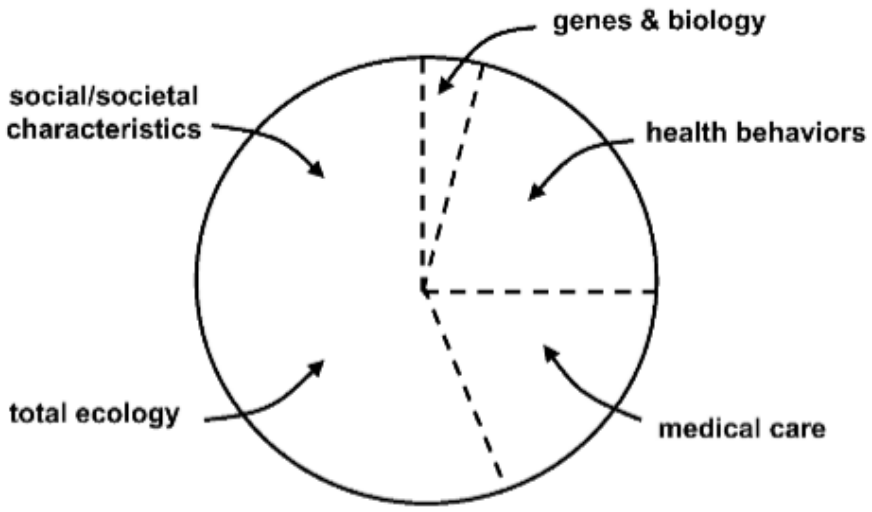


Figure 126: Relative influence of the five major determinant categories of health: rough estimations (Source: Tarlov, 1999)

The exact contribution of each of these factors is not precisely known at this time. Social determinants interact with other factors to influence behaviors that put individuals and communities are higher health risks. The complexity of the systems contributing to cancer incidence and mortality calls for novel, system-oriented approaches that address

the wide range of factors acting at multiple levels. Social determinants play a role at all levels of analysis across the cancer care continuum.

Individual behaviors, such as smoking or engaging in physical activity do not occur in isolation. They are influenced and shaped by a complex mix of biologic, as well as environmental, social and economic factors.

This chapter concludes this dissertation, summarizing the research findings and the broader scientific implications of this work. It discusses the results regarding the objectives highlighted in Section 1.4 and concludes with recommendations for future research and with general comments about what has been accomplished.

As mentioned in Section 1.3 of this dissertation, the overall aim of this work is to apply novel computational social science methodologies to better understand the social determinants of cancer. This dissertation attempted to provide an answer to the following research question: “Can a “bottom-up” agent-based approach help to gain insight into the complexity inherent in the social determinants of cancer?”. The aims and objectives presented in Chapter 1 attempt to address this question. The extent to which these aims have been accomplished is elaborated below, by addressing each aim in particular.

9.2 Summary of Research Findings

This section discusses the extent to which the original objectives of this dissertation have been achieved, while summarizing the main research findings of this work.

1. Review and examine the literature on the social determinants of cancer to distinguish the factors and behaviors that are most relevant to be included in the model.

This objective was accomplished in the first three chapters while laying out the theoretical foundation of this dissertation. These initial chapters described how the

different sections of the model were conceptualized using the available data and existing theories from multiple disciplines. Chapter 2 reviews the relevant scientific literature on the social determinants of cancer. It initially defines and explains what is understood by SDC, and why they are referred to as the "fundamental causes" of health (Link et al., 1998). Several upstream factors, such as demographic, cultural, socioeconomic and environmental factors, among others, that predispose individuals and communities to engage in behaviors that are either protecting or health damaging are discussed (Kaplan 1996; Marmot 1998). I then introduce the NCI's cancer continuum, describing the distinctive aspects of cancer prevention, detection, diagnosis, treatment and survivorship, followed by one of the most cited frameworks in SDC, proposed by Hiatt & Breen (2008). The framework focusing on how social determinants may influence cancer dynamics at the different stages of the cancer continuum is presented in Figure 1.

Chapter 2 proceeds by discussing the existing theories in the area of SDC, followed by a brief review of the importance of social networks to the different aspect of individual and community health. The link between complex systems and cancer is presented and motivated towards the end of the second chapter. Several characteristics of cancer make it a complex system (Agus & Gell-Mann, 2012; Gell-Mann, 2002; Barabási et al., 2011; Galea et al., 2010; Moore et al., 2011) are discussed, calling for new methodologies to address this complexity. The need for new approaches to address the SDC is motivated at the end of Chapter 2, as the current methodologies predominant in this discipline (mainly reductionist correlation-based analytic methods) do not address the full scale and complexity of this issue.

2. Review and examine the relevant literature on colorectal cancer to distinguish the factors and behaviors that are most relevant to be included in the model.

This aim has been accomplished throughout Chapter 3. The motivation behind choosing this cancer site was presented. Even though it is considered as one of the most largely preventable types of cancer, CRC is the third leading cancer in terms of cancer mortality in the world (Shike et al., 1990) and the second leading cancer killer in the United States. District of Columbia has the highest CRC deaths rates in the United States (CDC, 2011), and this one was of the reasons for choosing this area as the case study site for the model.

The process of CRC development is presented, with the assumption that most colorectal cancers begin from a polyp, which develops with the possibility of turning into cancer. Agents with colorectal polyps that are bigger than 1cm may turn into Stage 1 CRC. Therefore, the different CRC staging strategies were discussed. The risk factors that predispose agents to higher risks of developing CRC are discussed: such as smoking, diet, obesity, excessive alcohol consumption, physical activity or having a family history of CRC. All these factors that increase the risk of CRC have been documented and included in the model.

Particular attention was given to CRC prevention, because with regular screening CRC is largely preventable. The methods available for testing and screening for CRC were discussed (e.g., FOBT, colonoscopy), the treatment options and survival, as well as the quality of life, were introduced. Another CRC-related aspect that was addressed in this section and within the model is the delay in CRC diagnosis and treatment and its influence on CRC incidence, mortality and survival trends. Chapter 3 concludes with a

section on CRC disparities and the factors that contribute to the racial and gender gaps in CRC incidence and mortality.

3. Synopsize the existing computational social science methodologies and motivate the use of agent-based modeling as the methodology of choice to address the social determinants of cancer.

To accomplish this aim, Chapter 4 reviewed a few of the most common CSS methodologies, such as system dynamics, microsimulation models, agent-based modeling, geographic information systems (including their linkage/coupling with ABM) and social network analysis. The logic of simulation as a method is presented in Figure 17. The need for a data-driven and bottom-up methodology to address the reductionist view predominant in the study of SDC is presented, with the goal of shedding light into how social systems may influence the dynamics of cancer. General aspects of model calibration, verification and validation are provided. "Understanding our creations" was addressed by discussing UML as a way to describe the model workings, and the ODD protocol, that is commonly used for the purpose of publishing and presenting ABMs.

4. Design and build an agent-based model that captures the complexity inherent in the social determinants of cancer.

After laying the theoretical foundations in Chapters 2, 3 and 4 – the next aim of actually designing and building the model (Cancerscape) was targeted. The following chapter (Chapters 5) describes the study area (Washington, D. C.) and Chapter 6 proceeds with describing the process of designing and building the model.

As mentioned previously, the reason for choosing DC as the study site for the ABM study case is the fact that from 2000 to 2004 it had the highest CRC deaths rates in the United States, ranking sixth in the nation and continuing being the third highest in the

United States for CRC deaths (American Cancer Society, 2008). Because the prevalent methods for addressing the SDC, in general, are mainly reductionist correlation-based analytic methods, these two chapters present a new approach to address this problem. Chapter 5 begins with describing the characteristics of the study area for the study case. It starts with introducing the geography of the nation's capital followed by the socio-demographic characteristics of the district population in 2000. The main focus was on the age distribution, race, ethnicity, marital status, education, income and poverty, employment and occupations. All these characteristics (based on the U. S. Census for DC) were used to synthesize the artificial population. Data from the BRFSS was used for attributing the health characteristics of DC residents. Characteristics such as obesity and overweight, physical activity, smoking and alcohol use, or the coverage by health insurance were included into the model. While the distribution of CRC rates in DC was available for the model simulation period, the data on polyp distribution was not. Therefore, the initial distribution of CRC polyp proportions by age and race among the initial population groups is estimated according to the method provided by Gopalappa et al. (2011). Chapter 5 concludes with data on CRC screening rates in DC, as screening can reduce the mortality due to CRC considerably (Mandel et al., 1993).

Chapter 6 presents the data types that go into the model followed by model description using the ODD protocol. The "Overview" section highlights the purpose of the model presenting a high-level model diagram. It proceeds with discussing the entities, state variables, and scales as well as the process overview and scheduling of Cancerscape. UML diagrams are used to represent the main modeling concepts. The

design concepts included in the model are interaction, emergence, adaptation, stochasticity, collectives and observation. In the 'Details' section, the model initialization, the input and the submodels, such as daily activities, agent behaviors, as well as cancer surveillance are introduced and discussed. At this stage, I also describe how the colorectal carcinogenesis was modeled, explaining how the screening and treatment sub-modules function together. Population generation is also explained close the end of this chapter. After the model was presented and thoroughly described in Chapter 6, the next dissertation aim highlights the processes of verification and validation of Cancerscape, followed by introducing the model results for the five year simulation time and an analysis of the results.

5. Verify, validate and analyze the results of the agent-based model executed during the previous step.

Verification and validation (V&V) are immensely important for ABMs (Sargent, 2000). Verification ensures that the model was build "right", while validation helps make sure that the "right model was built" (Balci, 1998). For a model to be helpful to policy makers, it has to be properly verified and validated. Depending on the performance level the model aims for, certain validation techniques must be performed on the model for it to categorize according to the desired level of performance (Axtell & Epstein, 1994). Because no other agent-based models have been proposed to simulate the social determinants of colorectal cancer to date, Cancerscape aims for a classification target closer to Level 2 of performance. V&V has been executed accordingly to match the requirements of the performance level it attempts to attain. Cancerscape verification was performed using code walkthroughs, profiling, and sensitivity analysis. Validation was

performed from several perspectives. In particular, attention was given to *requirements validation* (having a precise definition of the real world questions of interest), *data validation*, *face validation* (a visual validity practice assessing if the model "looks" right), *process validation* (checking internal model steps to make sure they correspond with their analogous processes in the real world), *model output validation* (checking if the model output matches the actual system being modeled), and *agent validation* - to examine the agent activities and interactions and make sure they correspond to the individuals in the real world. Cancerscape was thoroughly tested as well.

6. Apply the model to different scenarios, to increase the validity of the model, and explore policies that might work best in certain situations.

Building on the previous sections, once the model has been verified and validated, a few scenarios have been explored to assess the extent to which the model replicates the real world phenomenon. The policy scenarios explored in Cancerscape investigate ways to reduce CRC incidence and mortality, as well as to decrease delays in cancer diagnosis and treatment.

As expected, lower stages at diagnosis were associated with increased survival rates and lower mortality rates. If tumors were diagnosed early, the chances of survival are higher compared to the agents with more advanced stages at diagnosis. The effects of delay were also explored as an alternative scenario. Delay significantly impacts the stage at diagnosis and is related to the overall CRC survival. While the goal of Cancerscape is not to explicitly provide precise predictions of CRC dynamics in Washington, D. C., the scenarios explored within the model mimic real world situations rather realistically, providing insights into ways to control CRC within the study area.

7. Assess the potential of agent-based modeling in addressing the social determinants of cancer.

Among the most notable objectives of this dissertation was to evaluate the potential of agent-based modeling to the study of the social determinants of cancer. To attempt to answer the main research question of this dissertation, that is: "Can a "bottom-up" agent-based approach help to gain insight into the complexity inherent in the social determinants of cancer?", a citywide large scale model was built using novel computational social science methodologies such as ABM, SNA and GIS.

This bottom-up approach allowed addressing the research question from a broad perspective, accounting for the wide range of factors contributing to the SDC. Applying the model to a study case and exploring real world scenarios presents a step forward compared to the traditional, equation-based methods predominantly used in addressing the social determinants of cancer issues. The remainder of this chapter will continue answering this question by summarizing the research findings and discussing the broader scientific implications of this work. It will recommend future research directions and will conclude with general remarks.

9.3 Scientific Contributions

One of the previous presidents of Unites States, President Nixon, started the "War on Cancer" in 1971. Since then, cancer continued to remain among the leading causes of death in the nation. This is, in part, due to the reductionist thinking that has been predominant in cancer research for the last decades. This work attempted to address this reductionist thinking by employing novel, computational social science methodologies to the social determinants of cancer, which are often referred to as the "fundamental causes

of disease" or "the causes of the causes" (Link & Phelan, 1995). The work presented in this dissertation is novel. In the academic literature, there are no other published studies that address the social determinants of cancer using agent-based modeling. This dissertation represents a step towards opening new avenues of research in SDC for CSS.

More specifically, the research contribution of this work can be reflected in two aspects. Among the key contributions of this work is integrating social determinants of cancer theories (as mentioned in Section 2.2), models (such as the multistep model of carcinogenesis or Andersen Model of Total Patient Delay) and data from multiple sources into a unified framework. This framework is used to explore the social determinants of cancer using novel, computational social science methodologies. This dissertation may be the first attempt to do so. In contrast with the methods predominantly used to address the social determinants of health (such as regression, equation-based models), the work in this dissertation advances the research on SDC beyond these traditional methods by modeling heterogeneous, boundedly rational agents individually, considering their varied socio-demographic characteristics and interactions across time and space, accounting for the individual risk factors, behaviors and social influences. During the past decades, the focus in cancer research has been on understanding the biological processes of carcinogenesis. The focus was on one individual gene that has gone wrong, one single mutation, and one association. While this direction is certainly meaningful, this reductionist view does not consider the social context of the individual, together with the complexity of the various factors involved in disease causation. These factors, however,

consequently maintain an association with the disease even when intervening mechanisms change (Link & Phelan, 1995).

By integrating different sources of data and theories, this work opens new avenues for social determinants of cancer research by modeling a broad range of factors at different levels, providing an artificial laboratory for exploring the relationships and interactions that influence cancer development and progression. This is particularly important if we are interested in exploring fundamental mechanisms involved the social causation of cancer for providing ways to better control it.

Another way this dissertation contributes to existing research is by providing a test-bed for exploring policy scenarios and for potentially designing more informed policy interventions in the area of SDC. Regression-based models are limiting in their ability to provide inference regarding policy interventions, mainly to the fact that they do not allow realistic representations of the complexity inherent in human interactions and the social systems in the real world. Using agent-based models like Cancerscape we can design and test policy interventions that would be difficult or unethical to test in the real world. In the case of data-driven models, where ABMs represent realistic real world populations, these policy questions can be targeted towards specific populations of interest, so that new policies can be particularly designed for the population they will be further enacted upon. In such a way, the outcomes of implementing these policies can be assessed and modified before they are implemented in the real world.

In the field of social determinants of cancer, this work may contribute to attracting more computational social science applications to the wide range of issues this discipline

faces. ABM has been successfully applied to address several health issues such as obesity, infectious diseases, smoking, and diet among others. The work in this dissertation demonstrates the applicability of these methods to address the social determinants of chronic conditions such as cancer, as well and encourages future efforts to address them in more detail and at larger scale.

9.4 Limitations

The research presented in this dissertation has shown how an agent-based, bottom-up methodology can be used to address the social determinants of cancer. This shed light into how social systems may influence the dynamics of cancer at the population level. When building any model, the modeler faced the challenge of simplifying a real world system to address specific research questions. Because models are mere abstractions of reality, several aspects of the real world system are omitted, phenomena and processes are simplified, and some of the elements of the real world system are left out. Models have limitations, and recognizing them is an important part of the process of documenting the model.

Computational limitations limited the scale and complexity that was initially planned for Cancerscape. Several simplifications had to be made to address these limitations. Simulations were run on a Macintosh Pro with a 2.6 GHz Intel Core i7 processor and 16 GB 1600 MHz DDR3 of memory. Increased computational power would have permitted for a richer specification of agent behaviors, and the employment of a wider range of experiments and scenarios. Also, even though the model has been carefully implemented, extensively tested and thoroughly verified, the potential for software errors to still exist in the code cannot be ruled out.

Another limitation of this work lies in the lack of sufficient data, making it difficult to specify certain agent rules or other processes explicitly. Therefore, in such cases, transition probabilities were used. The main empirical data sources used for the purpose of this model were obtained from surveys. The U. S. Census and BRFSS surveys are the main data sources used for Cancerscape initialization and calibration. By utilizing these representative surveys, the Cancerscape agents were assigned realistic demographic, socioeconomic, and health-related characteristics representative of the DC population (complemented with data from secondary sources). One of the limitations of these datasets is that they use one-time data collection procedures for the measurement of subjects. In the next section, recommendations for future work in presented.

9.5 Future Research

The main focus of this work was to attempt to answer the main research question and explore the social determinants of cancer using a bottom-up, agent-based modeling methodology, combined with SNA and GIS. However, most of the processes and phenomena modeled in this dissertation could benefit from a more in-depth exploration. Complex behaviors that lead to health conditions, like for instance - obesity (that is a CRC risk factor in itself) could benefit from richer and more thorough representation.

This work could be extended in several research directions. An obvious extension of this work would be extending Cancerscape to implement and model other cancer sites, such as prostate or breast cancer. Cancerscape is, by design, relatively easy to extend. Because it is particularly targeted towards the social determinants of cancer, the cancer types that would be best to implement using this framework are the ones for which social determinants have been shown be important drivers of either incidence or mortality.

Another direction that would be interesting to explore in Cancerscape is the influence of migration on cancer dynamics. Migrants tend to embody the cancer patterns of the country to which they migrated (Doll & Peto, 1981). For instance, Japanese people have higher rates of stomach cancer and lower rates of colon cancer, in comparison to the U. S. population. However, these differences have been found to gradually disappear in Japanese families that have moved to the U. S., as the migrants tend to acquire higher colon cancer rates and reduced stomach cancer rates.

Cancer takes a long time to develop. Therefore, it would be interesting to run the model longer than five years. Running it for several decades would allow incorporating early life exposures into the model. This would also permit observing a few generations of agents and better understand the genetic component of CRC, as agents with a history of CRC in a first-degree relative are at increased risk of CRC. If more than one first-degree relative has CRC, the risk continues to increase. Running the model longer would require considerably more computational power, which given the increasing computer power and its availability (as shown by Moore's law) should become easier in the imminent future.

In the presence of cost data, the model could be extended to explore the comparison of the cost-effectiveness of different screening strategies and treatment interventions, as well as the impact of screening on the overall quality-adjusted life years (QALY). Similarly, it would be interesting to extend the model to explore the quality of life and influence of side effects post CRC treatment.

Among the policy scenarios that could additionally be implemented in Cancerscape would be providing workspace incentives for employed individuals to undergo CRC screening. Also, the newly developed CRC treatment methods can be incorporated into the model to explore the combination that is most appropriate for each agent diagnosed with CRC within Cancerscape. Post treatment, it would be interesting to explore what affects the quality of life of agents with CRC, and design policies aimed at increasing their quality of life.

While this would add extra complexity to the already extensive and computationally expensive model, it would be interesting to extend Cancerscape by including CRC tumor-level characteristics. For instance, CRC patients with tumors located on the right side of their colon have a worse prognosis compared to those who have their tumors on the left side (Loupakis et al., 2015).

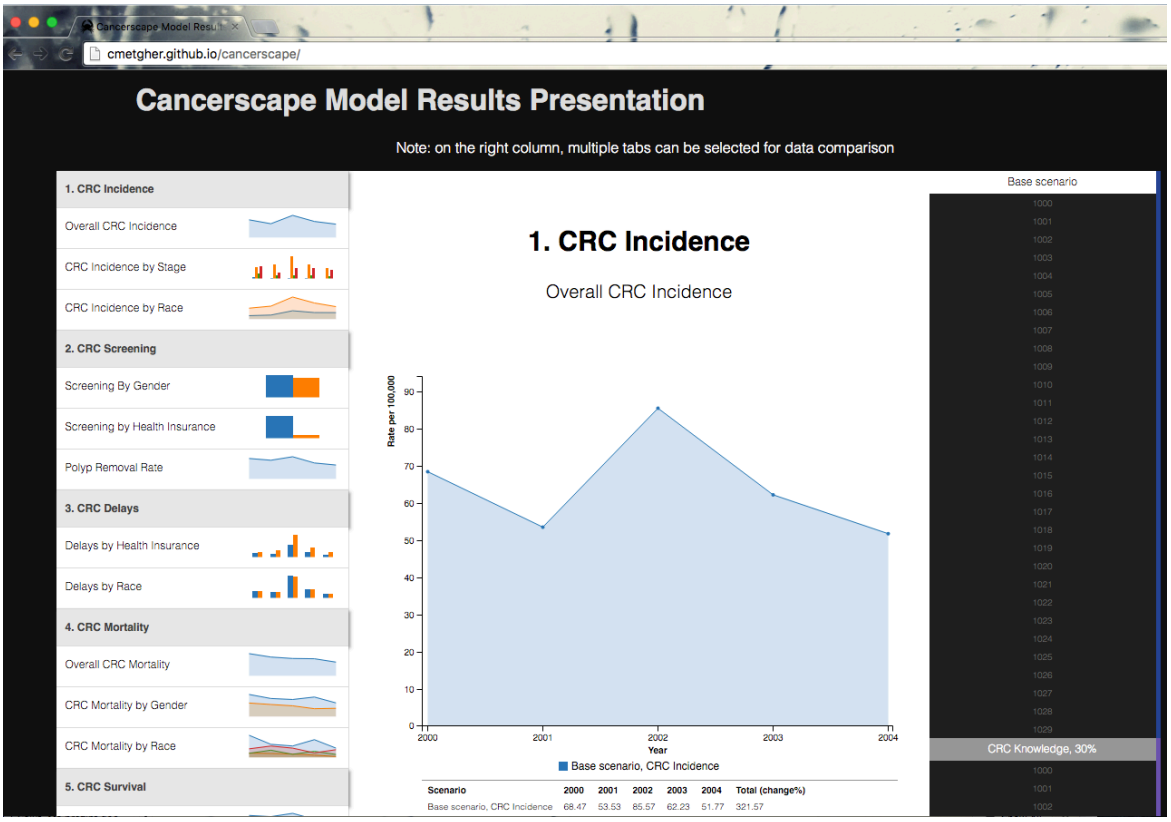
9.6 Concluding Remarks

This dissertation has explored the use of computational social science methods for exploring the social determinants of cancer. By employing a data-driven agent-based approach combined with SNA and GIS, and combining data and theories from multiple sources and disciplines - this research created a unified framework that can account for the multilevel complexity inherent in the SDC and that drives the dynamics of such complex, hard to predict systems. While they certainly have shortcomings, computational social science methods represent the most suitable currently available approaches for modeling systems characterized by a wide range of actors and interactions, which involve many levels of analysis and complex emergent behavior.

The study case of colorectal in Washington, D. C. provided additional validity to the agent-based approach, presenting areas where insights could be gained by employing a CSS approach. Specifically, the large-scale agent-based model presented in this work showed how this methodology could be useful in shedding light into how the SDC influence measures of cancer dynamics such as incidence, screening, delay and mortality. This final chapter presented a summary of the research findings and how they relate to the initial aims and objectives of this dissertation. It discussed the broader scientific contributions of this work. The research limitations were accounted for, and future research and extensions to the existing model were addressed. Finally, concluding remarks addressing the general results of this work were provided.

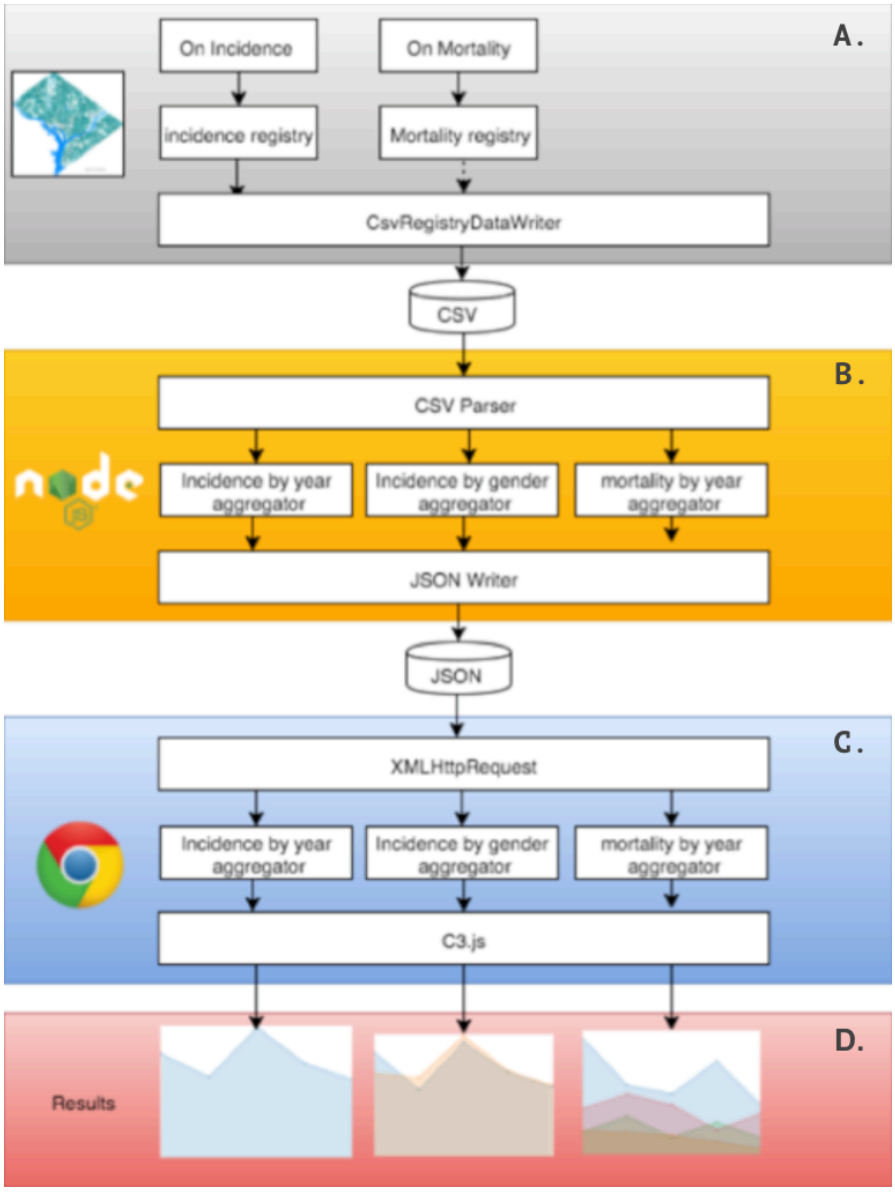
This work lays the foundation for studying the social determinants of cancer dynamically from the bottom-up, using novel, computational social science techniques. While there is space for improvement in potential future versions of Cancerscape, the model – as it is now – represents a significant advance over the predominantly used equation-based regression models in addressing the complexity of the actors, the multiple levels of analysis and the nonlinear interactions and behaviors that characterize the social determinants of cancer. CSS methodologies represent a promising avenue for future research into the social determinants of cancer and for exploring policy scenarios aimed at designing successful cancer control policy interventions.

APPENDIX A: CANCERSCAPE RESULTS PRESENTATION DASHBOARD



The dashboard (available at http://cmetgher.github.io/cancerscape/) was created to facilitate the presentation of the model results and provide an interactive experience with the Cancerscape results that are presented both as the average of 30 separate runs for the five simulated years, as well as for each individual run separately. By selecting several runs from the column on the right, the corresponding results are added on the main chart, for a comparative presentation of the model results.

APPENDIX B: ALGORITHM FOR GENERATING THE RESULTS DASHBOARD



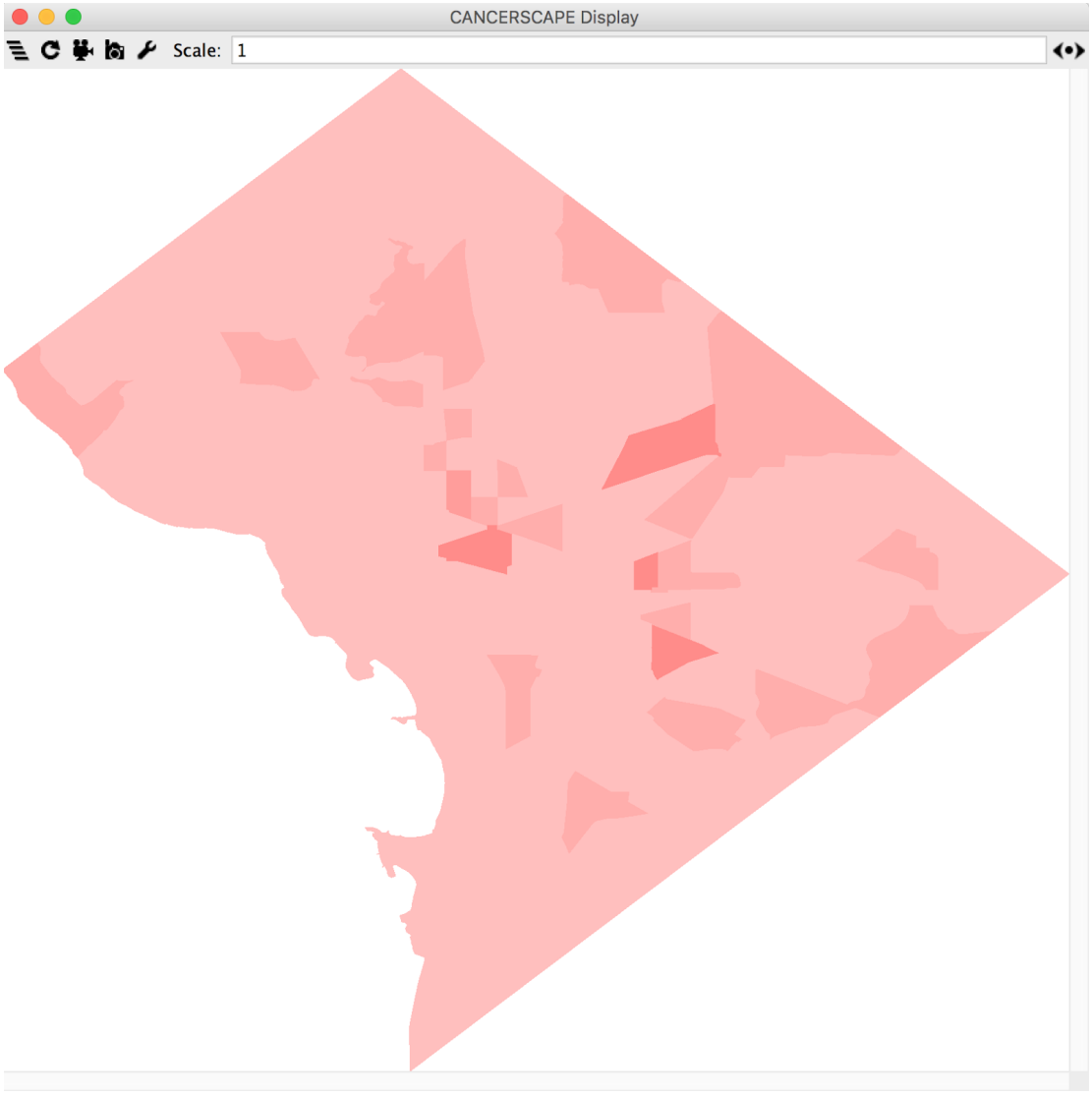
A. When the model is running, events such as incidence, mortality, and screening take place. The data collection registries listen to these events, collect the necessary information and pass them to the data writer. The data writer dumps the information to appropriate CSV files. Following this, the node server preprocesses and aggregates the data.

B. CSV files can contain tens of thousands of lines and be really large. Therefore, the aggregates that are needed for any one chart are really small. This is why the data pre-processing step is necessary so that we can only save in JSON format what is actually needed for display.

C. Following step B, the browser fetches the data from the JSON files. Subsequently, the user interface is built using the D3.js framework. A spot within the UI is prepared for each individual chart.

D. Following the previous steps, the data is finally portrayed for an easier interaction with the model results.

APPENDIX C: SPATIAL DISTRIBUTION OF CRC INCIDENCE FOR AFRICAN-AMERICANS, 2000-2004

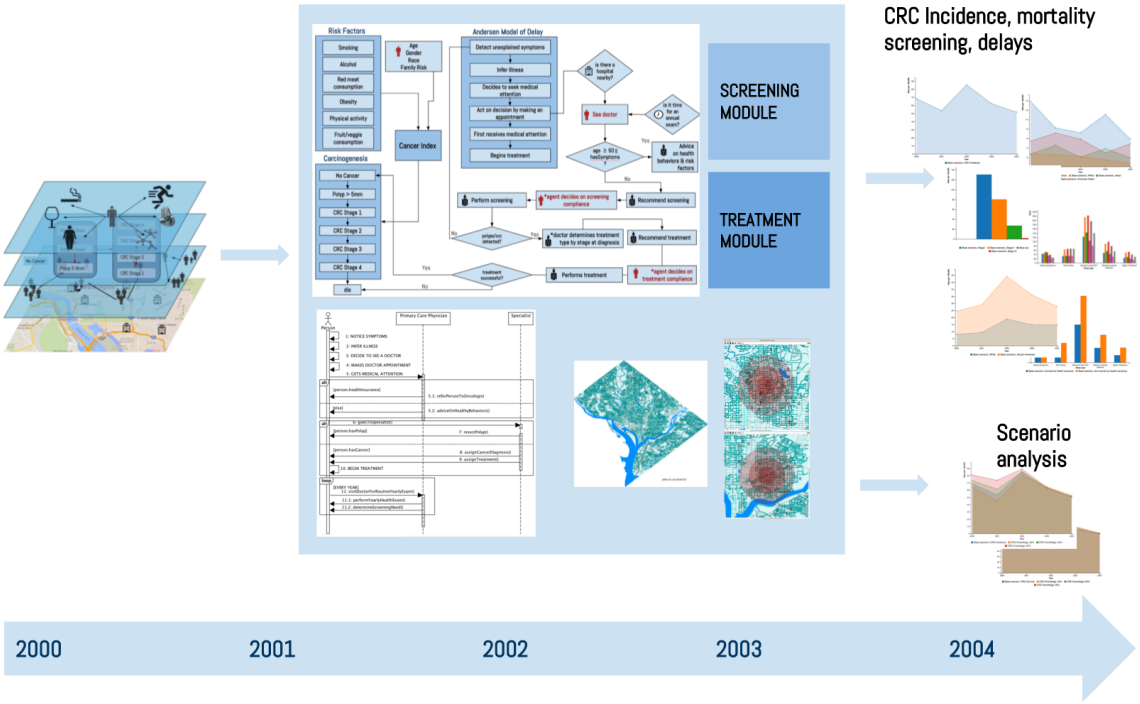


Presented here is the spatial distribution of CRC incidence for African-Americans in Cancerscape, by tract. The different shades of pink for the tracts represent the density of

the African-American agents with CRC within that tract, with tracts in lighter shades of pink containing less African-American CRC agents, and tracts in darker shades of pink containing more African-American agents with CRC. This is consistent with the racial distribution of African-Americans in the Washington DC area, as shown in Figure 26.

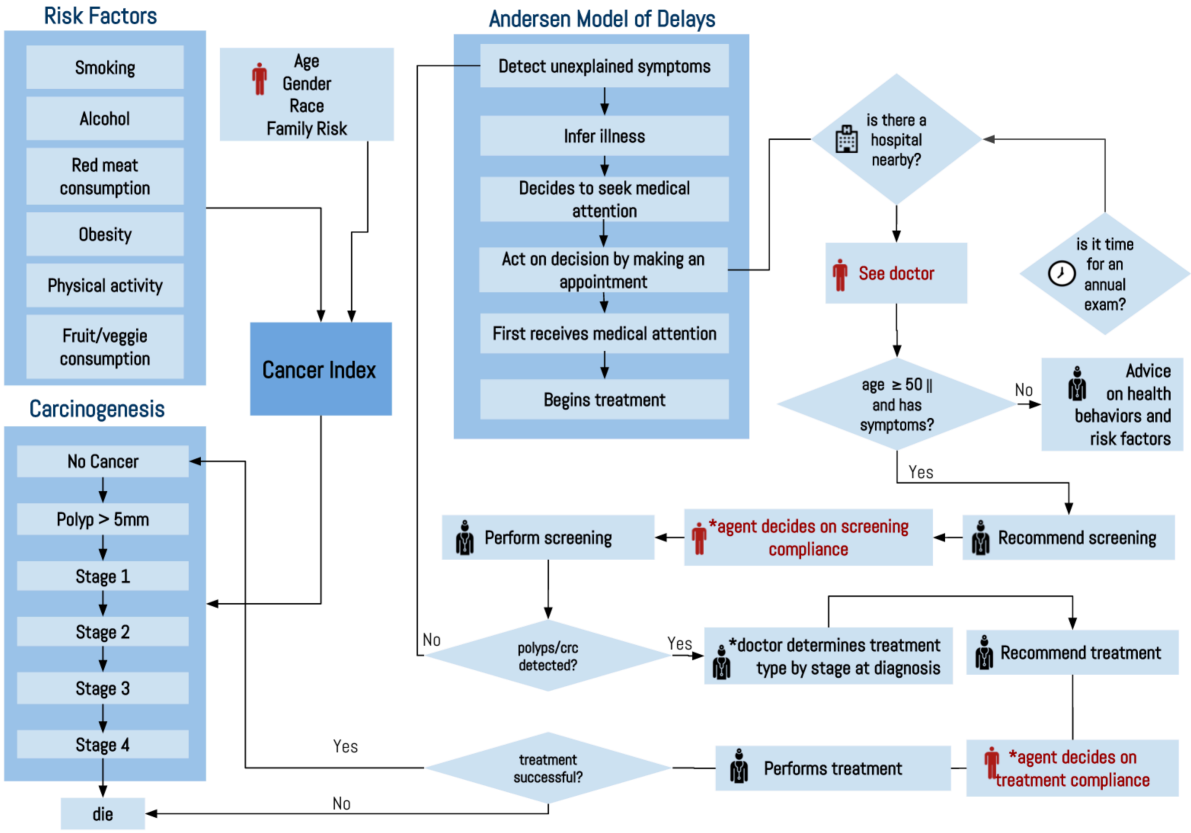
Additionally, as mentioned in Chapter 5, the African Americans in Washington, DC live primarily on the east of 16th street, while Hispanics are concentrated in Silver Spring, Adams Morgan and Arlington/Fairfax, VA. This area of DC contains less primary care facilities and higher poverty rates compared to the other areas of the district, as shown in Figure 32. This highlights the importance of future research to address these disparities.

APPENDIX D: HIGH LEVEL MODEL OVERVIEW



This high-level diagram shows the general model overview. The data types that go into the model are of several types: demographic, social, behavioral, risk factor data, CRC related information, geographic data, and social networks.

APPENDIX E: OVERVIEW OF CANCERSCAPE MODULES INTERACTION



APPENDIX F: SUMMARY OF DEFAULT INPUT PARAMETERS

LEVEL		PARAMETER		DEFAULT		REFERENCE
INDIVIDUAL	DEMO-GRAPHIC	*Age distribution*	0-115	**Age group Percentage** Under 5 5.7 5 to 9 6.2 10 to 14 5.2 15 to 19 6.6 20 to 24 9.1 25 to 34 17.8 35 to 44 15.3 45 to 54 13.2 55 to 59 4.9 60 to 64 3.8 65 to 74 6.3 75 to 84 4.4 85 over 1.6	The distribution of Washington D. C. population by age groups in 2000	USCB, 2000
		Gender Distribution	MALE FEMALE	**Gender Percentage** MEN 47.1 WOMEN 52.9	The proportion of Men and Women in Washington, DC in 2000	USCB, 2000
		Racial distribution	WHITE AFRICAN_AMERICAN AMERICAN_INDIAN ASIAN NATIVE_HAWAIIAN TWO OR MORE RACES SOME_OTHER_RACE	**Race Percentage** WHITE 30.78 AFRICAN_AMERICAN 60.01 AMERICAN_INDIAN 0.3 ASIAN 2.66 NATIVE_HAWAIIAN 0.06 TWO_OR_MORE_RACES 2.88 SOME_OTHER_RACE 3.84	The racial distribution of Washington D. C. population in 2000	USCB, 2000
		Education	LESS_THAN_HIGH_SCHOOL SOME_COLLEGE_OR_ASSOCIATE BACHELORS_DEGREE_OR_HIGHER LESS_THAN_5TH_GRADE FIFTH_TO_8_GRADE NINTH_TO_12_GRADE_NO_DIPLO HIGH_SCHOOL_GRADUATE SOME_COLLEGE_CREDITS ONE_OR_MORE_YRS_OF_COLLEGE ASSOCIATE_DEGREE BACHELORS_DEGREE MASTERS_DEGREE PROFESSIONAL_DEGREE	**Education Level Percentage** Less than 5th grade 3 5th to 8th grade 4.7 9th to 12th grade 14.4 High school graduate 20.6 Some college credit 4.3 1/more years of college 11.1 Associate degree 2.8 Bachelor's degree 18.1 Master's degree 11.9 Professional degree 6.1 Doctorate degree 3 Percent high school graduate/higher	The proportion of people by education level in Washington, DC in 2000	USCB, 2000

		DOCTORATE_DEGREE NA	77.8 Percent bachelor's degree or higher 39.1		
	Marital Status	NEVER_MARRIED MARRIED SEPARATED WIDOWED DIVORCED NA	**Marital Status Percentage** Never Married 48.4 Now Married 29.9 Separated 4.2 Widowed 7.8 Divorced 9.7	The distribution of Washington D. C. population by marital status in 2000	USCB, 2000
	Employment	IN_LABOR_FORCE CIVILIAN_LABOR_FORCE EMPLOYED UNEMPLOYED ARMED_FORCES NOT_IN_LABOR_FORCE	**Employment status Percentage** In labor force 63.6 Civilian labor force 62.9 Employed 56.1 Unemployed 6.8 Armed Forces 0.7 Not in labor force 36.4	The distribution of Washington D. C. population by employment status in 2000	USCB, 2000
	Earnings	$1to$9999orLess, $10000to$14999, $15000to$24999, $25000to$34999, $35000to$49999, $50000to$74999, $75000to$99999, $100000orMore, NA	**Income Percentage** Men Women $1to$9999orLess 2.1 2.6 $10000to$14999 5.1 5.0 $15000to$24999 16.5 17.1 $25000to$34999 20.0 21.9 $35000to$49999 20.7 22.7 $50000to$74999 17.6 17.5 $75000to$99999 8.2 7.2 $100000orMore 9.6 6.0	Individual earnings by sex in Washington D. C. in 2000	USCB, 2000
RISK FACTORS	*Relatives with CRC*	NONE, ONE, TWO_OR_MORE,	**Relatives with CRC Percentage** NONE 80.0 ONE 15.0 TWO_OR_MORE 5.0	Individuals with a familial risk Those who have two or more 1st or 2nd degree relatives (or both) with CRC make up approximately 20 % of all CRC cases	Lynch & Chapelle, 2003
	Smoking Status	true, false	**White Black Hispanic** **Race** Smokes Daily 10 17.1 6.6 SmokesSomeDays 6.1 6.7 8.7 Former Smoker 29.5 19.2 11.9 Never Smoked 54.4 56.9 72.8	Adults who are current smokers by race	CDC BRFSS, 2000
	Physical Activity Status	true, false	**Race Percentage** White 38.6 Black 24.5 Hispanic 26 Others 33	Adults with 30+ minutes of any physical activity five or more days per week	CDC BRFSS, 2000
	Obesity Level	Obese (BMI 30.0 - 99.8) Overweight (BMI 25.0-29.9) Neither overweight nor obese (BMI < 24.9)	**Race Percentage** White 7.4 Black 30.7 Hispanic 20.7 Others 14.9	Distribution of obesity levels by race in Washington D. C. in 2000 Weight classification by Body	CDC BRFSS, 2000

				Mass Index (BMI)	
	Health Insurance Coverage	Not Covered Covered by Private/Government Private: Total Private: Employment based Government: Total Government: Medicaid Government: Medicare Government: Military	**Coverage** **Percentage** Not Covered 12.8 Covered by Private/Government 87.2 Private:Total 69.6 Private: Employment- 62.1 Private: Direct Purchase 7.6 Government: Total 27.4 Government: Medicaid 14.2 Government: Medicare 13.6 Government: Military 2.3	Health insurance coverage in Washington D. C. in 2000	USCB, 2000
	Fruits and Veggie Consumption	Consumed 5 or more times/day Consume less than 5 times/day	**Gender** **Percentage** **Men** **Women** Consumed 5 or more times/day 28.4 34.9 Consume less than 5 times/day 71.6 65.1	Proportion of population who consumed more or less than the recommended 5 servings of fruits and vegetables a day by gender, in Washington, DC in 2000	CDC BRFSS, 2000
	Heavy Drinking Status	true, false	**Age group** Percentage **18-24 25-34 35-44 45-54 55-64 65+** Yes 12.3 8.0 4.6 4.5 3.9 4.6 No 87.7 92.0 95.4 95.5 96.1 95.4	Proportion of heavy drinkers by age group in Washington D. C. in 2000 According to CDC BRFSS (2000) Heavy drinkers are defined as adult men having more than 2 drinks per day and adult women having more than 1 drink per day	CDC BRFSS, 2000
	Red Meat Consumption	0.0..1.0	0.39	Proportion of people by daily consumption of red or processed meat	Cross et al., 2010, Chan et al., 2011
CRC	*Cancer Stage*	NO_CANCER POLYP1_5 STAGE_1 STAGE_2 STAGE_3 STAGE_4	**Gender** **Men** **Women** Stage Race I II III IV(%) I II III IV(%) White 42 35 19 4 40 36 19 5 African American Asian American/ 36 34 25 5 36 34 24 6 **Men** **Women** Pacific Islander American Indian/ 40 37 19 4 38 41 18 3 Alaska Native 36 37 24 3 40 39 17 3 Hispanic 37 36 22 5 37 36 21 5	Distribution of colorectal cancer cases by stage and gender	ACS, 2013

		Cancer Index	0-102	`Assigned proportional to the by stage at diagnosis`	At cancerIndex 0 the agent is healthy. At 102 the agent dies.	
		CRC Knowledge	0.0..1.0	`0.47`	The proportion of population with any knowledge about CRC symptoms and development	Scroggins & Bartley, 1999
		Fear of Cancer	0.0..1.0	`0.59`	The proportion of population with any knowledge about CRC symptoms and development	Vrinten et al., 2015, Bynum, S. et al., 2012.
		Delay Category	SUSPECT_NOTHING NOTICE_SYMPTOMS INFERRING_ILLNESS DECIDE_TO_SEEK_MEDICAL_ATTENTION SCHEDULE_DOCTOR_APPOINTMENT DETERMINE_SCREENING_NEED RECEIVE_SCREENING BEGIN_TREATMENT	**Delay type Probability** DecideToSeekMedicalAttentionBasePr 0.07 setReceiveScreeningBaseProbability 0.2 BeginTreatmentBaseProbability 0.05 NoticeSymptomsBaseProbability 0.1 setInferIllnessBaseProbability 0.2 ReturnToDoctorBaseProbabilily 0.1 ReceiveMedicalAttentionBaseProbability 0.1	The different stages of delay according to Andersen Model of Total Patient Delay	Andersen et al., 1995
	OTHER	*Home Location*	Coordinate	`Coordinates assigned at generation`	Spatial layer with agent home locations	
		Workplace Location	Coordinate	`Coordinates assigned at generation`	Spatial layer with agent workplace locations	
		Nearest Hospital	Coordinate	`Coordinates assigned at generation`	Spatial layer with hospitals	
		Daily Activities	SLEEP HOME WORK HOSPITAL SOCIAL/LEISURE/SPORT	`Activity.SLEEP`	Agent daily activities derived from Bureau of Labor Statistics' American Time Use Survey (ATUS)	USDL, 2015
		Yearly Health Exam	Day	`A random day between 0-365 in 1999 is assigned for each agent as the day of his last annual physical`	Annual physical exams when agents see a doctor for routine check-ups	
	GIS LAYERS	*Tracts Layer*	Shapefile	`188 (tracts)`	Washington, DC tracts layer	
		Houses Layer	int	`274845 (housing units)`	Washington, DC housing units	
		Business Establishments	int	`19655 (businesses)`	Number of business establishments in Washington, DC in 2000	DCG, 2013

GLOBAL		*Primary Care Units*	Shapefile	52 (primary care units)	Number of primary care units in Washington D. C. in 2000	DC GIS, 2015
		Roads Layer	Shapefile	Shapefile	Roads layer	USCB, 2000
	DEMO-GRAPHIC	*Number of Agents*	int	572059	The default number of agents in Cancerscape (population of Washington D. C. in 2000)	USCB, 2000
		Number of Households	int	248,338	Number of households in Washington D. C. in 2000	USCB, 2000
		Household Types	Married Couples With Children Married Couples Without Children Other Family With Children Other Family Without Children Individuals Living Alone	**Household Type** **Percentage** Married Couples With Children 8.4 Married Couples Without Children 14.5 Other Family With Children 11.4 Other Family Without Children 11.7 Individuals Living Alone 43.8	Types of households	USCB, 2000

		Family Risk Coefficient	0-2	if (relativesWithCRC == ONE) {
				familyRiskCoefficient = 1;
				} else if (relativesWithCRC == TWO_OR_MORE) {
				familyRiskCoefficient = 2.0;
				}
					Family risk coefficient (higher in people with more relatives with CRC)	Butterworth et al., 2006,
						Johns & Houlston, 2001

		Age Risk Coefficients	0-2.39	**Age group** **Risk** < 20 .001 20-34 0.13 35-44 0.41 45-54 1.42 55-64 2.12 65-74 2.39 75-84 2.32 > 84 1.21	The distribution of CRC risk by age groups	NCI SEER, 2015
	OTHER	*Screening Sensitivity*	FOBT BARIUM_ENEMA FLEXIBLE_SIGMOIDOSCOPY COLONOSCOPY	**Screening Method** **Success rate** FOBT 0.5 Barium enema 0.67 Flexible sigmoidoscopy 0.67 Colonoscopy 0.95	Sensitivity for screening methods	Winawer et al., 2012
		Treatment Success Probabilities	SURGERY SURGERY_AND_CHEMO CHEMO_PALLIATIVE COLONOSCOPY	**Treatment Method** **Success rate** Surgery 0.65 Surgery/chemotherapy 0.35 Chemo palliative 0.09 Colonoscopy 0.972	Sensitivity for treatment procedures	Nelson et al., 2002
		Minimum Screening Age	int	50	This is the recommended screening age by USPSTF for	USPSTF, 2002

				average risk individuals to start at age 50 for the high risk population group (such as individuals with a family history of CRC) CRC screening should begin at age 40 or sooner (colonoscopy is recommended 10 years before youngest relative)	
	CRC Base Prevalence	int	`1646`	The number of people living with colorectal world in DC in 2000	NCI SEER, 2000
	Mandatory CRC Screening 2002?	true, false	`TRUE`	District of Columbia implemented the mandate on CRC screening coverage on April 13, 2002	Cokkinides et al., 2011
	Time Manager	hourly	`1/1/2000`	A time step (tick) in the model is one hour.	

APPENDIX G: SUMMARY OF MAIN SCENARIO RESULTS

<table>
<tr><th>SCENARIO</th><th>EXPECTED RESULT</th><th>POLICY</th><th>ACTUAL RESULT</th></tr>
<tr>
<td>Scenario I: Reducing recommended CRC screening age</td>
<td>Reducing recommended CRC screening age should increase incidence rates and polyp removal rates.</td>
<td>Ia. Screening starting at age 40 for general population and at 30 for people with family history of CRC</td>
<td>
<table>
<tr><th>Scenario</th><th>2000</th><th>2001</th><th>2002</th><th>2003</th><th>2004</th><th>Total (change%)</th></tr>
<tr><td>Base scenario, Polyp Removal Rate</td><td>95</td><td>86.87</td><td>103.1</td><td>74.27</td><td>64.6</td><td>423.83</td></tr>
<tr><td>Min Screening Age, 40 Years</td><td>113.4 (+19%)</td><td>99.7 (+15%)</td><td>117.63 (+14%)</td><td>84.2 (+13%)</td><td>68.7 (+6%)</td><td>483.63 (+14%)</td></tr>
<tr><td>Base scenario, CRC Incidence</td><td>68.47</td><td>53.53</td><td>85.57</td><td>62.23</td><td>51.77</td><td>321.57</td></tr>
<tr><td>Min Screening Age, 40 Years</td><td>85.77 (+25%)</td><td>72.3 (+35%)</td><td>101.17 (+18%)</td><td>73.23 (+18%)</td><td>56.23 (+9%)</td><td>388.7 (+21%)</td></tr>
</table>
Reducing recommended CRC screening age from 50 to 40 years contributed to more agents getting timely screening and to a subsequent increase in overall CRC incidence. There was a 25% improvement compared to the base model scenario where recommended screening age is 50 years in 2000 that was sustained during the following 4 years. The highest increase in incidence has been observed during the first two years.

Individuals of age 40 or over who went for a yearly checkup received the screening recommendation sooner, which also led to more polyps being detected and removed during the colonoscopy screening procedures. This scenario helps detect more people with colorectal polyps whose diagnosis would be missed if screening started 10 years later (at the currently recommended age of 50).
</td>
</tr>
</table>

<table>
<tr>
<td></td>
<td></td>
<td>Ib. Screening starting at age 30 for general population and at 20 for people with family history of CRC</td>
<td>
<table>
<tr><th>Scenario</th><th>2000</th><th>2001</th><th>2002</th><th>2003</th><th>2004</th><th>Total (change%)</th></tr>
<tr><td>Base scenario, CRC Incidence</td><td>68.47</td><td>53.53</td><td>85.57</td><td>62.23</td><td>51.77</td><td>321.57</td></tr>
<tr><td>Min Screening Age, 30 Years</td><td>90.17 (+32%)</td><td>74.2 (+39%)</td><td>100.17 (+17%)</td><td>68.4 (+10%)</td><td>55.6 (+7%)</td><td>388.53 (+21%)</td></tr>
<tr><td>Base scenario, Polyp Removal Rate</td><td>95</td><td>86.87</td><td>103.1</td><td>74.27</td><td>64.6</td><td>423.83</td></tr>
<tr><td>Min Screening Age, 30 Years</td><td>116.9 (+23%)</td><td>102.63 (+18%)</td><td>115.8 (+12%)</td><td>83.07 (+12%)</td><td>66.23 (+3%)</td><td>484.63 (+14%)</td></tr>
</table>
Further decreasing the recommended CRC screening age resulted in a marginal improvement over the previous policy. This can be due to the overall decreased of the prevalence of colorectal polyps among individuals younger than age 40 (Gopalappa et al., 2011).
</td>
</tr>
<tr>
<td>Scenario II: Increasing CRC knowledge</td>
<td>Increased population levels of knowledge about CRC should lead to increased CRC incidence, increased polyp removal rates and improved survival.</td>
<td>IIa. CRC knowledge at 30%</td>
<td>
<table>
<tr><th>Scenario</th><th>2000</th><th>2001</th><th>2002</th><th>2003</th><th>2004</th><th>Total (change%)</th></tr>
<tr><td>Base scenario, CRC Incidence</td><td>68.47</td><td>53.53</td><td>85.57</td><td>62.23</td><td>51.77</td><td>321.57</td></tr>
<tr><td>CRC Knowledge, 30%</td><td>64.47 (-6%)</td><td>45.4 (-15%)</td><td>80.53 (-6%)</td><td>64.63 (+4%)</td><td>53.27 (+3%)</td><td>308.3 (-4%)</td></tr>
<tr><td>Base scenario, Polyp Removal Rate</td><td>95</td><td>86.87</td><td>103.1</td><td>74.27</td><td>64.6</td><td>423.83</td></tr>
<tr><td>CRC Knowledge, 30%</td><td>86.93 (-8%)</td><td>78.9 (-9%)</td><td>98.8 (-4%)</td><td>75.07 (+1%)</td><td>65.1 (+1%)</td><td>404.8 (-4%)</td></tr>
<tr><td>Base scenario, CRC Survival</td><td>99.13</td><td>93.23</td><td>109.53</td><td>80.07</td><td>71</td><td>452.97</td></tr>
<tr><td>CRC Knowledge, 30%</td><td>91.2 (-8%)</td><td>84.8 (-9%)</td><td>104.03 (-5%)</td><td>81.57 (+2%)</td><td>72.47 (+2%)</td><td>434.07 (-4%)</td></tr>
</table>
The default level of knowledge about colorectal cancer and its symptoms in Cancerscape is 47% (Scroggins & Bartley, 1999). As expected, a decrease in the CRC knowledge within the agent population was associated with a slight decrease in the CRC incidence rates, colorectal polyp removal rates and a subsequent decrease in the overall CRC survival.
</td>
</tr>
</table>

IIb. Increasing CRC knowledge to 60%

Scenario	2000	2001	2002	2003	2004	Total (change%)
Base scenario, **CRC Incidence**	68.47	53.53	85.57	62.23	51.77	321.57
CRC Knowledge, 60%	73.23 (+7%)	61.33 (+15%)	84.97 (-1%)	63.77 (+2%)	51.1 (-1%)	334.4 (+4%)
Base scenario, **Polyp Removal Rate**	95	86.87	103.1	74.27	64.6	423.83
CRC Knowledge, 60%	98.07 (+3%)	91.47 (+5%)	107.43 (+4%)	78.03 (+5%)	65.1 (+1%)	440.1 (+4%)
Base scenario, **CRC Survival**	99.13	93.23	109.53	80.07	71	452.97
CRC Knowledge, 60%	102.97 (+4%)	98.33 (+5%)	114.23 (+4%)	82.83 (+3%)	71.8 (+1%)	470.17 (+4%)

As expected, increasing CRC knowledge led to increased CRC incidence rates. A 60% population level of CRC knowledge was associated with a 7% increase in CRC incidence in 2000 and 15% in 2001. The following years there were some slight fluctuations but the overall increasing CRC incidence trend was sustained.

IIc. Increasing CRC knowledge to 90%

Scenario	2000	2001	2002	2003	2004	Total (change%)
Base scenario, **CRC Incidence**	68.47	53.53	85.57	62.23	51.77	321.57
CRC Knowledge, 90%	80.93 (+18%)	73.57 (+37%)	89.2 (+4%)	64.03 (+3%)	50.57 (-2%)	358.3 (+11%)
Base scenario, **Polyp Removal Rate**	95	86.87	103.1	74.27	64.6	423.83
CRC Knowledge, 90%	110 (+16%)	102.07 (+17%)	107.43 (+4%)	78.3 (+5%)	66.13 (+2%)	463.93 (+9%)
Base scenario, **CRC Survival**	99.13	93.23	109.53	80.07	71	452.97
CRC Knowledge, 90%	116.57 (+18%)	110.67 (+19%)	113.7 (+4%)	84.53 (+6%)	72.8 (+3%)	498.27 (+10%)

A further increase in CRC knowledge was associated with a subsequent increased CRC incidence rates and polyp removal rates. A 90% CRC knowledge level was associated with a total increase in CRC incidence rates of 11%, a total polyp removal rate of 9% and an overall 10% improvement in CRC survival between 2000 and 2004.

<table>
<tr>
<td>Scenario III: Increasing healthcare coverage for colonoscopies</td>
<td>Increased healthcare coverage for colonoscopies should lead to increased polyp removal rates and increased CRC survival rates.</td>
<td>IIIa. Increasing colonoscopy coverage to 60%</td>
<td>
<table>
<tr><th>Scenario</th><th>2000</th><th>2001</th><th>2002</th><th>2003</th><th>2004</th><th>Total (change%)</th></tr>
<tr><td>Base scenario, CRC Incidence</td><td>68.47</td><td>53.53</td><td>85.57</td><td>62.23</td><td>51.77</td><td>321.57</td></tr>
<tr><td>Colonoscopy, 60%</td><td>72.63 (+6%)</td><td>55.77 (+4%)</td><td>85.13 (-1%)</td><td>56.8 (-9%)</td><td>44.57 (-14%)</td><td>314.9 (-2%)</td></tr>
<tr><td>Base scenario, Polyp Removal Rate</td><td>95</td><td>86.87</td><td>103.1</td><td>74.27</td><td>64.6</td><td>423.83</td></tr>
<tr><td>Colonoscopy, 60%</td><td>119.47 (+26%)</td><td>106.53 (+23%)</td><td>125.67 (+22%)</td><td>90.2 (+21%)</td><td>75.27 (+17%)</td><td>517.13 (+22%)</td></tr>
<tr><td>Base scenario, CRC Survival</td><td>99.13</td><td>93.23</td><td>109.53</td><td>80.07</td><td>71</td><td>452.97</td></tr>
<tr><td>Colonoscopy, 60%</td><td>122.53 (+24%)</td><td>109.7 (+18%)</td><td>129.4 (+18%)</td><td>94.07 (+17%)</td><td>78.27 (+10%)</td><td>533.97 (+18%)</td></tr>
</table>
As expected, higher levels of colonoscopy coverage resulted in decreased levels of overall CRC incidence. A 60% increase in colonoscopy coverage (compared to the baseline 30% coverage) was associated with increased polyp removal rates between 2000 and 2004. It also led to increased incidence rates during the first two years and slightly decreased in the following years as most agents got their symptoms checked out by a physician within the first years. This resulted in increased survival rates for all years in Cancerscape.
</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IIIb. Increasing Colonoscopy coverage to 90%</td>
<td>
<table>
<tr><th>Scenario</th><th>2000</th><th>2001</th><th>2002</th><th>2003</th><th>2004</th><th>Total (change%)</th></tr>
<tr><td>Base scenario, CRC Incidence</td><td>68.47</td><td>53.53</td><td>85.57</td><td>62.23</td><td>51.77</td><td>321.57</td></tr>
<tr><td>Colonoscopy, 90%</td><td>71.83 (+5%)</td><td>61.13 (+14%)</td><td>83.33 (-3%)</td><td>54.53 (-12%)</td><td>41.87 (-19%)</td><td>312.7 (-3%)</td></tr>
<tr><td>Base scenario, Polyp Removal Rate</td><td>95</td><td>86.87</td><td>103.1</td><td>74.27</td><td>64.6</td><td>423.83</td></tr>
<tr><td>Colonoscopy, 90%</td><td>132.47 (+39%)</td><td>130.87 (+51%)</td><td>159.6 (+55%)</td><td>109.63 (+48%)</td><td>93 (+44%)</td><td>625.57 (+48%)</td></tr>
<tr><td>Base scenario, CRC Survival</td><td>99.13</td><td>93.23</td><td>109.53</td><td>80.07</td><td>71</td><td>452.97</td></tr>
<tr><td>Colonoscopy, 90%</td><td>134.83 (+36%)</td><td>131.73 (+41%)</td><td>160.27 (+46%)</td><td>110.4 (+38%)</td><td>93 (+31%)</td><td>630.23 (+39%)</td></tr>
</table>
Increasing the colonoscopy coverage even further led to an increase in polyp removal rates and improved CRC survival (over the 5 years period a 48% total improvement in the rates of polyp removed was observed). Also, increasing colonoscopy coverage to 90% let to a total increase in CRC survival of 18%.
</td>
</tr>
</table>

REFERENCES

Agus, D. B. (2012). *The End of Illness*. Simon and Schuster.

Agus, D., & Gell-Mann, M. (2012). Perspective: Meeting of Minds. *Nature*, 491(7425), S61-S61.

Aizer, A. A., Chen, M. H., Mccarthy, E. P., Mendu, M. L., Koo, S., Wilhite, T. J. & Hu, J. C. (2013). Marital Status and Survival in Patients with Cancer. *Journal of Clinical Oncology*, 31(31), 3869-3876.

Albano, J. D., Ward, E., Jemal, A., Anderson, R., Cokkinides, V. E., Murray, T., Thun, M. J. (2007). Cancer Mortality in the United States by Education Level and Race. *Journal of the National Cancer Institute*, 99(18), 1384-1394.

Albert, R., & Barabási, A. L. (2002). Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, 74(1), 47.

Alexander, D. D., Waterbor, J., Hughes, T., Funkhouser, E., Grizzle, W., & Manne, U. (2007). African-American and Caucasian Disparities in Colorectal Cancer Mortality and Survival by Data Source: An Epidemiologic Review. *Cancer Biomarkers: Section a of Disease Markers*, 3(6), 301.

Alfarano, S., Lux, T., & Wagner, F. (2005). Estimation of Agent-Based Models: the Case of An Asymmetric Herding Model. *Computational Economics*, 26(1), 19-49.

Altekruse S, Kosary C, Krapcho M. (2010). *SEER Cancer Statistics Review 1975-2007*, http://seer.cancer.gov/Csr/1975-2007/, Based On November 2009 SEER Data Submission, Posted to the SEER Web Site, April 2010. Bethesda, MD: National Cancer Institute; 2010.

American Academy of Family Physicians (AAFP) (2015). *Primary Care*. Retrieved from: http://www.aafp.org/about/policies/all/primary-care.html

American Cancer Society (ACS) (2013). *Cancer Prevention and Early Detection Facts and Figures*. Retrieved from: http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-037535.pdf

American Cancer Society (ACS) (2014). *Colorectal Cancer Facts & Figures 2014-2016*. Atlanta: American Cancer Society

American Cancer Society (ACS) (2008). *South Atlantic Division Cancer Facts & Figures*. Atlanta: American Cancer Society

American Society of Clinical Oncology (ASCO) (2014). The State of Cancer Care in America, 2014: a Report by the American Society of Clinical Oncology. *Journal of Oncology Practice*, 10(2), 119-142.

Andersen, B. L., Cacioppo, J. T., & Roberts, D. C. (1995). Delay in Seeking a Cancer Diagnosis: Delay Stages and Psychophysiological Comparison Processes. *British Journal of Social Psychology*, 34(1), 33-52.

Andersen, R., & Newman, J. F. (2005). Societal and Individual Determinants of Medical Care Utilization in the United States. *Milbank Quarterly*, 83(4), Online-Only.

Andrews, J. A., Tildesley, E., Hops, H., & Li, F. (2002). The Influence of Peers On Young Adult Substance Use. *Health Psychology*, 21(4), 349.

Andrulis, D. P., & Carrier, B. (1999). *Managed Care and the Inner City: the Uncertain Promise for Providers, Plans, and Communities*. Jossey-Bass Inc Pub.

Andrykowski, M. A., Beacham, A. O., Schmidt, J. E., & Harper, F. W. (2006). Application of the Theory of Planned Behavior To Understand Intentions To Engage in Physical and Psychosocial Health Behaviors After Cancer Diagnosis. *Psycho‐Oncology*, 15(9), 759-771.

Armitage, P., & Doll, R. (1954). The Age Distribution of Cancer and a Multi-Stage Theory of Carcinogenesis. *British Journal of Cancer*, 8(1), 1.

Atkin, W. S., Morson, B. C., & Cuzick, J. (1992). Long-Term Risk of Colorectal Cancer After Excision of Rectosigmoid Adenomas. *New England Journal of Medicine*, 326(10), 658-662.

Auchincloss, A. H., Riolo, R. L., Brown, D. G., Cook, J., & Roux, A. V. D. (2011). An Agent-Based Model of Income Inequalities in Diet in the Context of Residential Segregation. *American Journal of Preventive Medicine*, 40(3), 303-311.

Aune, D., Lau, R., Chan, D. S. M., Vieira, R., Greenwood, D. C., Kampman, E., & Norat, T. (2012). Dairy Products and Colorectal Cancer Risk: a Systematic Review and Meta-Analysis of Cohort Studies. *Annals of Oncology*, 23(1), 37-45.

Aune, D., Lau, R., Chan, D. S., Vieira, R., Greenwood, D. C., Kampman, E., & Norat, T. (2011). Nonlinear Reduction in Risk for Colorectal Cancer by Fruit and Vegetable Intake Based On Meta-Analysis of Prospective Studies. *Gastroenterology, 141,*1, 106-18.

Axelrod, R. (1986). An Evolutionary Approach to Norms. *American Political Science Review*, 80(04), 1095-1111.

Axelrod, R. (1997). Advancing the Art of Simulation in the Social Sciences. In *Simulating Social Phenomena* (pp. 21-40). Springer Berlin Heidelberg.

Axtell, R. L. (2000). Why Agents? On the Varied Motivations for Agent Computing in the Social Sciences. *Proceedings of the Workshop On Agent Simulation: Applications, Models, and*

Tools. C. M. Macal and D. Sallach. Chicago, Illinois, Argonne National Laboratory: 3-24.

Axtell, R. L., & Epstein, J. M. (1994). Agent-Based Modeling: Understanding Our Creations. *the Bulletin of the Santa Fe Institute*, 9(2), 28-32.

Axtell, R., Axelrod, R., Epstein, J. M., & Cohen, M. D. (1996). Aligning Simulation Models: a Case Study and Results. *Computational & Mathematical Organization Theory*, 1(2), 123-141.

Ayanian, J. Z., Zaslavsky, A. M., Fuchs, C. S., Guadagnoli, E., Creech, C. M., Cress, R. D., & Wright, W. E. (2003). Use of Adjuvant Chemotherapy and Radiation Therapy for Colorectal Cancer in a Population-Based Cohort. *Journal of Clinical Oncology*, 21(7), 1293-1300.

Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., Bouvard, V., Altieri, A., Cogliano, V. & WHO (2007). International Agency for Research On Cancer Monograph Working Group. Carcinogenicity of Alcoholic Beverages. *The Lancet Oncology*, 8(4), 292-293.

Bach, P. B., Schrag, D., Brawley, O. W., Galaznik, A., Yakren, S., & Begg, C. B. (2002). Survival of Blacks and Whites After a Cancer Diagnosis. *JAMA*, 287(16), 2106-2113.

Balci, O. (1997). Verification Validation and Accreditation of Simulation Models. In *Proceedings of the 29th Conference On Winter Simulation* (pp. 135-141). IEEE Computer Society.

Balci, O. 1998. Verification, Validation, and Testing. In *The Handbook of Simulation*, ed. J. Banks. New York: John Wiley.

Baldwin, L. M., Cai, Y., Larson, E. H., Dobie, S. A., Wright, G. E., Goodman, D. C., Matthews, B., Hart, L. G. (2008). Access to Cancer Services for Rural Colorectal Cancer Patients. *The Journal of Rural Health*, 24(4), 390-399.

Bankes, S. C., Lempert, R. J., & Popper, S. W. (2001). Computer-Assisted Reasoning. *Computing in Science & Engineering*, 3(2), 71-77.

Barabási, A. L., Gulbahce, N., & Loscalzo, J. (2011). Network Medicine: a Network-Based Approach to Human Disease. *Nature Reviews Genetics*, 12(1), 56-68.

Barger, S. D. (2013). Social Integration, Social Support and Mortality in the US National Health Interview Survey. *Psychosomatic Medicine*, 75(5), 510-517.

Bass, S. B., Gordon, T. F., Ruzek, S. B., Wolak, C., Ward, S., Paranjape, A., Lin, K., Meyer, B., & Ruggieri, D. G. (2011). Perceptions of Colorectal Cancer Screening in Urban African American Clinic Patients: Differences by Gender and Screening Status. *Journal of Cancer Education*, 26(1), 121-128.

Becker, G. (2004). Deadly Inequality in the Health Care "Safety Net": Uninsured Ethnic Minorities' Struggle to Live with Life-Threatening Illnesses. *Medical Anthropology Quarterly*, 18(2), 258-275.

Berger, P. L., & Luckmann, T. (1991). *The Social Construction of Reality: a Treatise in the Sociology of Knowledge* (No. 10). Penguin UK.

Berger, T., & Parker, D. C. (2001). Examples of Specific Research: Introduction To Standard Descriptions of Projects'. In D. C. Parker, T. Berger, & S. M. Manson (Eds.), Meeting the Challenge of Complexity (pp. 48–54). *Proceedings of a Special Workshop On Land-Use/Land-Cover Change*, Irvine.

Berkman, L. F., & Glass, T. (2000). Social Integration, Social Networks, Social Support, and Health. *Social Epidemiology*, 1, 137-173.

Berkman, L. F., & Syme, S. L. (1979). Social Networks, Host Resistance, and Mortality: a Nine-Year Follow-Up Study of Alameda County Residents. *American Journal of Epidemiology*, 109(2), 186-204.

Berkman, L. F., Kawachi, I. (2000). *Social Epidemiology*. Oxford University Press.

Berkman, L. F., Leo-Summers, L., & Horwitz, R. I. (1992). Emotional Support and Survival After Myocardial Infarction: a Prospective, Population-Based Study of the Elderly. *Annals of Internal Medicine*, 117(12), 1003-1009.

Bernal, I., Domènech, E., Garcia-Planella, E., Marín, L., Mañosa, M., Navarro, M., & Gassull, M. A. (2006). Medication-Taking Behavior in a Cohort of Patients With Inflammatory Bowel Disease. *Digestive Diseases and Sciences*, 51(12), 2165-2169.

Bernstein, C. N., Blanchard, J. F., Kliewer, E., & Wajda, A. (2001). Cancer Risk in Patients With Inflammatory Bowel Disease. *Cancer*, 91(4), 854-862.

Berry, J., Bumpers, K., Ogunlade, V., Glover, R., Davis, S., Counts-Spriggs, M., Kauh, J., & Flowers, C. (2009). Examining Racial Disparities in Colorectal Cancer Care. *Journal of Psychosocial Oncology*, 27(1), 59-83.

Bersini, H. (2012). UML for ABM. *Journal of Artificial Societies and Social Simulation*, *15*(1), 9.

Berube, A., Prince, R., Smith, H., Brookings Institution (U.S.). (2003). *Washington, DC in Focus: a Profile from Census 2000*. Washington, D.C: Brookings Institution Center On Urban and Metropolitan Policy.

Beydoun, H. A., & Beydoun, M. A. (2008). Predictors of Colorectal Cancer Screening Behaviors Among Average-Risk Older Adults in the United States. *Cancer Causes & Control*, 19(4), 339-359.

Bianchini, F., Kaaks, R., & Vainio, H. (2002). Overweight, Obesity, and Cancer Risk. *The Lancet Oncology*, 3(9), 565-574.

Birkin, M., & Wu, B. (2012). A Review of Microsimulation and Hybrid Agent-Based Approaches. In *Agent-Based Models of Geographical Systems* (pp. 51-68). Springer Netherlands.

Blanchard, C. M., Stein, K. D., Baker, F., Dent, M. F., Denniston, M. M., Courneya, K. S., & Nehl, E. (2004). Association Between Current Lifestyle Behaviors and Health-Related Quality of Life in Breast, Colorectal, and Prostate Cancer Survivors. *Psychology & Health*, 19(1), 1-13.

Blausen Medical Communications (2016). *Anatomy of the Lower Digestive System*. Retrieved from: http://blausen.com/

Blume, L. E., & Easley, D. (2008). Rationality. *The New Palgrave Dictionary of Economics*. Palgrave Macmilla

Blumer, H. (1986). *Symbolic Interactionism: Perspective and Method*. University of California Press.

Bollobás, B., & Riordan, O. M. (2003). Mathematical Results On Scale-Free Random Graphs. *Handbook of Graphs and Networks: From the Genome to the Internet*, 1-34.

Bond, J. H., & Practice Parameters Committee of the American College of Gastroenterology. (2000). Polyp Guideline: Diagnosis, Treatment, and Surveillance for Patients with Colorectal Polyps. *The American Journal of Gastroenterology*, 95(11), 3053-3063.

Booch, G. (2006). *Object Oriented Analysis & Design with Application*. Pearson Education India.

Borshchev, A., & Filippov, A. (2004). From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools. In *Proceedings of the 22nd International Conference of the System Dynamics Society* (No. 22).

Botteri, E., Iodice, S., Bagnardi, V., Raimondi, S., Lowenfels, A. B., & Maisonneuve, P. (2008). Smoking and Colorectal Cancer: a Meta-Analysis. *JAMA*, 300(23), 2765-2778.

Bourisly, A. K. (2013). An Obesity Agent Based Model: a New Decision Support System for the Obesity Epidemic. In *Asiasim* 2013 (pp. 37-48). Springer Berlin Heidelberg.

Boyle, T., Keegel, T., Bull, F., Heyworth, J., & Fritschi, L. (2012). Physical Activity and Risks of Proximal and Distal Colon Cancers: a Systematic Review and Meta-Analysis. *Journal of the National Cancer Institute*, Djs354.

Breen, N., Wagener, D. K., Brown, M. L., Davis, W. W., & Ballard-Barbash, R. (2001). Progress in Cancer Screening Over a Decade: Results of Cancer Screening from the 1987, 1992, and 1998 National Health Interview Surveys. *Journal of the National Cancer Institute*, 93(22), 1704-1713.

Brenner, H., Hoffmeister, M., Stegmaier, C., Brenner, G., Altenhofen, L., & Haug, U. (2007). Risk of Progression of Advanced Adenomas to Colorectal Cancer by Age and Sex: Estimates Based On 840 149 Screening Colonoscopies. *Gut*, 56(11), 1585-1589.

Brenner, M. H., & Mooney, A. (1983). Unemployment and Health in the Context of Economic Change. *Social Science & Medicine*, 17(16), 1125-1138.

Brody, J. G., Aschengrau, A., Mckelvey, W., Rudel, R. A., Swartz, C. H., & Kennedy, T. (2004). Breast Cancer Risk and Historical Exposure to Pesticides from Wide-Area Applications Assessed with GIS. *Environmental Health Perspectives*, 112(8), 889.

Brown, D. G., Riolo, R., Robinson, D. T., North, M., & Rand, W. (2005). Spatial Process and Data Models: Toward Integration of Agent-Based Models and GIS. *Journal of Geographical Systems*, 7(1), 25-47.

Brown, M. L., & Potosky, A. L. (1990). The Presidential Effect: the Public Health Response to Media Coverage About Ronald Reagan's Colon Cancer Episode. *Public Opinion Quarterly*, 54(3), 317-329.

Brown, S. T., Tai, J. H., Bailey, R. R., Cooley, P. C., Wheaton, W. D., Potter, M. A., Mcglone, S. M. (2011). Would School Closure for the 2009 H1N1 Influenza Epidemic Have Been Worth the Cost? A Computational Simulation of Pennsylvania. *BMC Public Health*, 11(1), 353.

Buffart, L. M., Thong, M. S., Schep, G., Chinapaw, M. J., Brug, J., & Van De Poll-Franse, L. V. (2012). Self-Reported Physical Activity: Its Correlates and Relationship with Health-Related Quality of Life in a Large Cohort of Colorectal Cancer Survivors. *Plos One*, 7(5), E36164.

Burr, V. (2015). *Social Constructionism*. Routledge. New York, NY.

Bury, M. R. (1986). Social Constructionism and the Development of Medical Sociology. *Sociology of Health & Illness*, 8(2), 137-169.

Butterworth, A. S., Higgins, J. P., & Pharoah, P. (2006). Relative and Absolute Risk of Colorectal Cancer for Individuals with a Family History: a Meta-Analysis. *European Journal of Cancer*, 42(2), 216-227.

Byles, J. E., Redman, S., Hennrikus, D., Sanson-Fisher, R. W., & Dickinson, J. (1992). Delay in Consulting a Medical Practitioner About Rectal Bleeding. *Journal of Epidemiology and Community Health*, 46(3), 241-244.

Cacioppo, J. T., & Petty, R. E. (1982). A Biosocial Model of Attitude Change. In J. T. Cacioppo & R. E. Petty (eds.) *Perspectives in Cardiovascular Psychophysiology*. New York: Guilford.

Cacioppo, J. T., Andersen, B. L., Turnquist, D. C., & Tassinary, L. G. (1989). Psychophysiological Comparison Theory: On the Experience, Description, and Assessment of Signs and Symptoms. *Patient Education and Counseling*, 13(3), 257-270.

Cacioppo, J. T., Petty, R. E., & Andersen, B. L. (1988). Social Psychophysiology as a Paradigm. *Social Psychophysiology and Emotion: Theory and Clinical Applications*, 173-191.

Cairns, J. *Cancer: Science and Society* (1978). San Francisco: WH Freeman.

Campbell, P. T., Deka, A., Jacobs, E. J., Newton, C. C., Hildebrand, J. S., Mccullough, M. L., Limburg, P. J., Gapstur, S. M. (2010). Prospective Study Reveals Associations Between Colorectal Cancer and Type 2 Diabetes Mellitus or Insulin Use in Men. *Gastroenterology*, 139(4), 1138-1146.

Capra, F., & Luisi, P. L. (2014). *The Systems View of Life: a Unifying Vision*. Cambridge University Press.

Carley, K. M., Fridsma, D. B., Casman, E., Yahja, A., Altman, N., Chen, L. C., Kaminsky, B., Nave, D. (2006). Biowar: Scalable Agent-Based Model of Bioattacks. *Systems, Man and Cybernetics, Part A: Systems and Humans,* 36(2), 252-265.

Cassel, J. (1976). The Contribution of the Social Environment to Host Resistance. *American Journal of Epidemiology*, 104(2), 107-123.

Casti, J. L. (1998). *Would-be worlds: How Simulation is Changing the Frontiers of Science*. John Wiley & Sons, Inc.

Center, M. M., Jemal, A., & Ward, E. (2009). International Trends in Colorectal Cancer Incidence Rates. *Cancer Epidemiology Biomarkers & Prevention*,18(6), 1688-1694.

Centers for Disease Control and Prevention (CDC) (2013). Vital Signs: Colorectal Cancer Screening Test Use-- United States, 2012. MMWR. *Morbidity and Mortality Weekly Report*, 62(44), 881.

Centers for Disease Control and Prevention (CDC) (2011). Vital Signs: Colorectal Cancer Screening, Incidence, and Mortality--United States, 2002-2010. MMWR. *Morbidity and Mortality Weekly Report*, 60(26), 884.

Cattell, V. (2001). Poor People, Poor Places, and Poor Health: the Mediating Role of Social Networks and Social Capital. *Social Science & Medicine*, 52(10), 1501-1516.

Centers for Disease Control and Prevention (CDC). (2008). Use of Colorectal Cancer Tests--United States, 2002, 2004, and 2006. MMWR. *Morbidity and Mortality Weekly Report*, 57(10), 253.

Centers for Disease Control and Prevention (CDC) (2000). *Behavioral Risk Factor Surveillance System Survey (BRFSS) Data*. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Washington DC, 2000.

Centers for Disease Control and Prevention (CDC) (2015). *Chronic Disease State Policy Tracking System*. Retrieved from: http://nccd.cdc.gov/cdphppolicysearch/default.aspx

Cialdini R. B. & Trost M. R. (1998). Social Influence: Social Norms, Conformity, and Compliance." In Gilbert D. T., Fiske S. T., Lindzey G. (Eds.), *the Handbook of Social Psychology*, 4th Ed.: 151–192. New York: Oxford University Press.

Cioffi-Revilla, C., Rogers, J. D., & Hailegiorgis, A. (2011). Geographic Information Systems and Spatial Agent-Based Model Simulations for Sustainable Development. *ACM Transactions On Intelligent Systems and Technology (TIST)*, 3(1), 10.

Chandra, A., Blanchard, J. C., & Ruder, T. (2013). *District of Columbia Community Health Needs Assessment*. RAND.

Chao, A., Connell, C. J., Jacobs, E. J., Mccullough, M. L., Patel, A. V., Calle, E. E., Cokkinides, V. E., Thun, M. J. (2004). Amount, Type, and Timing of Recreational Physical Activity in Relation to Colon and Rectal Cancer in Older Adults: the Cancer Prevention Study II Nutrition Cohort. *Cancer Epidemiology Biomarkers & Prevention*, 13(12), 2187-2195.

Chao, A., Thun, M. J., Jacobs, E. J., Henley, S. J., Rodriguez, C., & Calle, E. E. (2000). Cigarette Smoking and Colorectal Cancer Mortality in the Cancer Prevention Study II. *Journal of the National Cancer Institute*, 92(23), 1888-1896.

Chao, A., Thun, M. J., Connell, C. J., Mccullough, M. L., Jacobs, E. J., Flanders, W. D., Rodriguez, C., Sinha, R., Calle, E. E. (2005). Meat Consumption and Risk of Colorectal Cancer. *JAMA*, 293(2), 172-182.

Chatterjee, S., Chattopadhyay, A., & Levine, P. H. (2015). Between-Ward Disparities in Colorectal Cancer Incidence and Screening in Washington DC. *Journal of Epidemiology and Global Health*, 5(4), S1-S9.

Chirikos, T. N., Russell-Jacobs, A., & Jacobsen, P. B. (2002). Functional Impairment and the Economic Consequences of Female Breast Cancer. *Women & Health*, 36(1), 1-20.

Cho, E., Smith-Warner, S. A., Ritz, J., Van Den Brandt, P. A., Colditz, G. A., Folsom, A. R., Freudenheim, J.L., Giovannucci, E., Goldbohm, R.A., Graham, S., Holmberg, L. (2004). Alcohol Intake and Colorectal Cancer: a Pooled Analysis of 8 Cohort Studies. *Annals of Internal Medicine*, 140(8), 603-613.

Christakis, N. A. (2004). Social Networks and Collateral Health Effects. *BMJ*, 329(7459), 184-185.

Chung, K. K., Hossain, L., & Davis, J. (2005). Exploring Sociocentric and Egocentric Approaches for Social Network Analysis. In *Proceedings of the 2nd International Conference On Knowledge Management in Asia Pacific* (pp. 1-8).

Clegg, L. X., Reichman, M. E., Miller, B. A., Hankey, B. F., Singh, G. K., Lin, Y. D., Goodman, M. T., Lynch, C. F., Schwartz, S. M., Chen, V. W., Bernstein, L. (2009). Impact of Socioeconomic Status On Cancer Incidence and Stage at Diagnosis: Selected Findings from the Surveillance, Epidemiology, and End Results: National Longitudinal Mortality Study. *Cancer Causes & Control*, 20(4), 417-435.

Cnattingius, S. (2004). The Epidemiology of Smoking During Pregnancy: Smoking Prevalence, Maternal Characteristics, and Pregnancy Outcomes. *Nicotine & Tobacco Research*, 6 (Suppl 2), S125-S140.

Cobb, S. (1976). Social Support as a Moderator of Life Stress. *Psychosomatic Medicine*, 38(5), 300-314.

Cockburn, J., Paul, C., Tzelepis, F., Mcelduff, P., & Byles, J. (2003). Delay in Seeking Advice for Symptoms That Potentially Indicate Bowel Cancer. *American Journal of Health Behavior*, 27(4), 401-407.

Cohen, S. (2004). Social Relationships and Health. *American Psychologist*, 59(8), 676.

Cokkinides, V., Bandi, P., Shah, M., Virgo, K., & Ward, E. (2011). The Association Between State Mandates of Colorectal Cancer Screening Coverage and Colorectal Cancer Screening Utilization Among US Adults Aged 50 To 64 Years with Health Insurance. *BMC Health Services Research*, 11(1), 19.

Collins, J. F., Lieberman, D. A., Durbin, T. E., & Weiss, D. G. (2005). Accuracy of Screening for Fecal Occult Blood On a Single Stool Sample Obtained by Digital Rectal Examination: a Comparison with Recommended Sampling Practice. *Annals of Internal Medicine*, 142(2), 81-85.

Compton, C. C., & Greene, F. L. (2004). The Staging of Colorectal Cancer: 2004 and Beyond. *CA: A Cancer Journal for Clinicians*, 54(6), 295-308.

Conrad, P. (2005). *The Sociology of Health & Illness: Critical Perspectives*. New York: Worth Publishers.

Conrad, P., & Barker, K. K. (2010). The Social Construction of Illness Key Insights and Policy Implications. *Journal of Health and Social Behavior*, 51(1 Suppl), S67-S79.

Cook, K. S., Cheshire, C., Rice, E. R., & Nakagawa, S. (2013). Social Exchange Theory. In *Handbook of Social Psychology*, Handbooks of Sociology and Social Research, J. Delamater (ed.), Springer US (pp. 61–88).

Cotugna, N., Subar, A. F., Heimendinger, J., & Kahle, L. (1992). Nutrition and Cancer Prevention Knowledge, Beliefs, Attitudes, and Practices: the 1987 National Health Interview Survey. *Journal of the American Dietetic Association*, 92(8), 963-968.

Courtney, R. J., Paul, C. L., Sanson-Fisher, R. W., Macrae, F., Attia, J., & Mcevoy, M. (2012). Current State of Medical-Advice-Seeking Behaviour for Symptoms of Colorectal Cancer: Determinants of Failure and Delay in Medical Consultation. *Colorectal Disease*, 14(5), E222-E229.

Cram, P., Fendrick, A. M., Inadomi, J., Cowen, M. E., Carpenter, D., & Vijan, S. (2003). The Impact of a Celebrity Promotional Campaign On the Use of Colon Cancer Screening: the Katie Couric Effect. *Archives of Internal Medicine*, 163(13), 1601-1605.

Crawford, R. (1977). You Are Dangerous to Your Health: the Ideology and Politics of Victim Blaming. *International Journal of Health Services*, 7(4), 663-680.

Christakis, N. A., & Fowler, J. H. (2009). Social Network Visualization in Epidemiology. *Norsk Epidemiologi*, 19(1), 5.

Crooks, A., Castle, C., & Batty, M. (2008). Key Challenges in Agent-Based Modelling for Geo-Spatial Simulation. *Computers, Environment and Urban Systems*, 32(6), 417-430.

Crooks, A. T., & Castle, C. J. (2012). The Integration of Agent-Based Modelling and Geographical Information for Geospatial Simulation. In *Agent-Based Models of Geographical Systems* (pp. 219-251). Springer Netherlands.

Crooks, A. T., & Heppenstall, A. J. (2012). Agent-Based Models of Geographical Systems. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), Agent-based Models of Geographical Systems (Vol. 164, pp. 85–108). New York, NY: Springer.

Cross, A. J., Ferrucci, L. M., Risch, A., Graubard, B. I., Ward, M. H., Park, Y. & Sinha, R. (2010). A Large Prospective Study of Meat Consumption and Colorectal Cancer Risk: An Investigation of Potential Mechanisms Underlying This Association. *Cancer Research*, 70(6), 2406-2414.

Cutler, D. M., & Lleras-Muney, A. (2006). *Education and Health: Evaluating Theories and Evidence* (No. W12352). National Bureau of Economic Research.

Dai, Z., Xu, Y. C., & Niu, L. (2007). Obesity and Colorectal Cancer Risk: a Meta-Analysis of Cohort Studies. *World Journal of Gastroenterology*, 13(31), 4199.

Danaei, G., Ding, E. L., Mozaffarian, D., Taylor, B., Rehm, J., Murray, C. J., & Ezzati, M. (2009). the Preventable Causes of Death in the United States: Comparative Risk Assessment of Dietary, Lifestyle, and Metabolic Risk Factors. *Plos Medicine*, 6(4), E1000058.

Dancik, G. M., Jones, D. E., & Dorman, K. S. (2010). Parameter Estimation and Sensitivity Analysis in an Agent-Based Model of Leishmania Major Infection. *Journal of Theoretical Biology*, 262(3), 398-412.

Darvishi, M., & Ahmadi, G. (2014). Validation Techniques of Agent Based Modelling for Geospatial Simulations. ISPRS-International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences*, 1, 91-95.

Davey Smith, G., Gunnell, D., & Ben-Shlomo, Y. (2001). Life-Course Approaches to Socio-Economic Differentials in Cause-Specific Adult Mortality. *Poverty, Inequality and Health: An International Perspective*, Oxford University Press, New York.

Davis, R. M., Genel, M., Howe, J. P., Karlan, M. S., Kennedy, W. R., Numann, P. J., O'Neill, J. F. (1996). Health Care Needs of Gay Men and Lesbians in the United States. *JAMA*, 275(17), 1354-1359.

Decosse, J. J., & Cennerazzo, W. J. (1997). Quality of Life Management of Patients with Colorectal Cancer. *CA: A Cancer Journal for Clinicians*, 47(4), 198-206.

Demark-Wahnefried, W., Aziz, N. M., Rowland, J. H., & Pinto, B. M. (2005). Riding the Crest of the Teachable Moment: Promoting Long-Term Health After the Diagnosis of Cancer. *Journal of Clinical Oncology*, 23(24), 5814-5830.

De Nooijer, J., Lechner, L., & De Vries, H. (2001). A Qualitative Study On Detecting Cancer Symptoms and Seeking Medical Help; An Application of Andersen's Model of Total Patient Delay. *Patient Education and Counseling*, 42(2), 145-157.

Diaz, R. W. (2004). Unequal Access: the Crisis of Health Care Inequality for Low-Income African American Residents of the District of Columbia. *Journal Health Care Law & Policy*, 7, 120.

Dijkstra, E. W. (1976). *A Discipline of Programming* (Vol. 1). Englewood Cliffs: Prentice-Hall.

District of Columbia (DC) Department of Health (2000a). *Patient Safety Reporting Program: District of Columbia, Annual Report*. Washington, DC: Department of Health.

District of Columbia (DC) Department of Health (DOH) (2000b). *District of Columbia Cancer Registry Program*. Washington, DC: Department of Health.

District of Columbia (DC) Department of Health (2001). *Addiction Prevention and Recovery Administration*. Washington, DC: Department of Health.

District of Columbia (DC) Department of Health (2002). *State Center for Health Statistics Administration*. Washington, DC: Department of Health.

District of Columbia Government (DCG) (2013). *District of Columbia Data Book: Revenue and Economy*. Retrieved from: http://cfo.dc.gov/sites/default/files/dc/sites/ocfo/publication/attachments/dc%2520data%2520book%2520september%25202013%5b1%5d.pdf

District of Columbia Geographic Information System (DC GIS) Program (2015). *Open Data Catalog*. Retrieved from: http://data.octo.dc.gov/

Doll, R., & Peto, R. (1981). The Causes of Cancer: Quantitative Estimates of Avoidable Risks of Cancer in the United States Today. *Journal of the National Cancer Institute*, 66(6), 1192-1308.

Doubeni, C. A., Laiyemo, A. O., Young, A. C., Klabunde, C. N., Reed, G., Field, T. S., & Fletcher, R. H. (2010). Primary Care, Economic Barriers to Health Care, and Use of Colorectal Cancer Screening Tests Among Medicare Enrollees Over Time. *The Annals of Family Medicine*, 8(4), 299-307.

Doyal, L. (1995). What Makes Women Sick? Gender and the Political Economy of Health. *BMJ*, 311, 577.

Dréau, D., Stanimirov, D., Carmichael, T., & Hadzikadic, M. (2009). An Agent-Based Model of Solid Tumor Progression. In *Bioinformatics and Computational Biology* (pp. 187-198). Springer Berlin Heidelberg.

Dunbar, R. I. (1992). Neocortex Size as a Constraint On Group Size in Primates. *Journal of Human Evolution*, 22(6), 469-493.

Dunbar, R. I. (1995). Neocortex Size and Group Size in Primates: a Test of the Hypothesis. *Journal of Human Evolution*, 28(3), 287-296.

Durkheim, E. (1951). *Suicide: a Study in Sociology* [1897]. Translated by JA Spaulding and G. Simpson (Glencoe, Illinois: the Free Press, 1951).

Durrett, R. (2007). *Random Graph Dynamics* (Vol. 200, No. 7). Cambridge: Cambridge University Press.

Earman, J. (Ed.). (1983). Testing Scientific Theories. University of Minnesota Press.

Edge, S. B., & Compton, C. C. (2010). The American Joint Committee On Cancer: the 7th Edition of the AJCC Cancer Staging Manual and the Future of TNM. *Annals of Surgical Oncology*, *17*(6), 1471-1474.

Edmonds, B., & Moss, S. (2005). *From KISS to KIDS–An 'Anti-Simplistic' Modelling Approach* (pp. 130-144). Springer Berlin Heidelberg.

Edwards, B. K., Noone, A. M., Mariotto, A. B., Simard, E. P., Boscoe, F. P., Henley, S. J., Jemal, A., Cho, H., Anderson, R. N., Kohler, B. A., Eheman, C. R. (2014). Annual Report to the Nation On the Status of Cancer, 1975-2010, Featuring Prevalence of Comorbidity and Impact On Survival Among Persons with Lung, Colorectal, Breast, Or Prostate Cancer. *Cancer*, 120(9), 1290-1314.

Ejima, K., Aihara, K., & Nishiura, H. (2013). Modeling the Obesity Epidemic: Social Contagion and Its Implications for Control. *Theoretical Biology and Medical Modelling*, 10(1), 17.

El-Sayed, A. M., Scarborough, P., Seemann, L., & Galea, S. (2012). Social Network Analysis and Agent-Based Modeling in Social Epidemiology. *Epidemiologic Perspectives & Innovations*, 9(1), 1.

Elstad, J. I. (1998). The Psycho-Social Perspective On Social Inequalities in Health. *Sociology of Health & Illness*, 20(5), 598-618.

Epstein, J. M. (2006). Remarks On the Foundations of Agent-Based Generative Social Science. In K. Judd and L. Tesfatsion, (eds). *Handbook of Computational Economics*, 2, 1585-1604. North Holland Press.

Epstein, J. M. (2009). Modelling to Contain Pandemics. *Nature*, 460(7256), 687-687.

Epstein, J. M., & Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Brookings Institution Press.

Epstein, J. M. (2001). Learning to Be Thoughtless: Social Norms and Individual Computation. *Computational Economics*, 18(1), 9-24.

Ermentrout, G. B., & Edelstein-Keshet, L. (1993). Cellular Automata Approaches to Biological Modeling. *Journal of Theoretical Biology*, 160(1), 97-133.

Ernster, V. L., Sacks, S. T., Selvin, S., & Petrakis, N. L. (1979). Cancer Incidence by Marital Status: US Third National Cancer Survey. *Journal of the National Cancer Institute*, 63(3), 567-585.

Eubank, S. G., Guclu, H., Kumar, V. S. A., & Marathe, M. V. (2004). Modelling Disease Outbreaks in Realistic Urban Social Networks. *Nature*, 429(6988), 180–184.

Evans, A. (2012). Uncertainty and Error. In *Agent-Based Models of Geographical Systems* (pp. 309-346). Springer Netherlands.

Eyer, J., & Sterling, P. (1977). Stress-Related Mortality and Social Organization. *Review of Radical Political Economics*, 9(1), 1-44.

Ezzati, M., & Lopez, A. D. (2003). Estimates of Global Mortality Attributable to Smoking in 2000. T*he Lancet*, 362(9387), 847-852.

Fagiolo, G., Moneta, A., & Windrum, P. (2007). A Critical Guide to Empirical Validation of Agent-Based Models in Economics: Methodologies, Procedures, and Open Problems. *Computational Economics*, 30(3), 195-226.

Fearon, E. R., & Vogelstein, B. (1990). A Genetic Model for Colorectal Tumorigenesis. *Cell*, 61(5), 759-767.

Fedra, K. (1996) Distributed Models and Embedded GIS: Integration Strategies and Case Studies. In: Goodchild, Steyeart, and Parks (Eds.) *GIS and Environmental Modeling: Progress and Research Issues*, (pp. 413-417). Fort Collins, GIS World Books.

Feinstein, L., Sabates, R., Anderson, T. M., Sorhaindo, A., Hammond, C. What Are the Effects of Education On Health? In: Desjardins R., Schuller T., Editors. *Measuring the Effects of Education On Health and Civic Engagement: Proceedings of the Copenhagen Symposium*. Paris: OECD; 2006. P. 171–313.

Fernback, J. (2007). Beyond the Diluted Community Concept: a Symbolic Interactionist Perspective On Online Social Relations. *New Media & Society*, 9(1), 49-69.

Ferrari, P., Jenab, M., Norat, T., Moskal, A., Slimani, N., Olsen, A., Tjønneland, A., Overvad, K., Jensen, M. K., Boutron-Ruault, M. C, Clavel-Chapelon, F., (2007). Lifetime and Baseline Alcohol Intake and Risk of Colon and Rectal Cancers in the European Prospective Investigation into Cancer and Nutrition (EPIC). *International Journal of Cancer*, 121(9), 2065-2072.

Fevang, B. T., Fevang, J., Stangeland, L., Søreide, O., Svanes, K., & Viste, A. (2000). Complications and Death After Surgical Treatment of Small Bowel Obstruction: a 35-Year Institutional Experience. *Annals of Surgery*, 231(4), 529-537.

Fickle, D. K., Randall, K. E., Jordan, A. L., Torres, N. A., & Paskett, E. D. (2015). Abstract A06: Guided Tours of a "Super Colon" Increase Knowledge and Willingness to Discuss Colon Cancer Among Residents of Appalachian Ohio. *Cancer Epidemiology Biomarkers & Prevention*, 24(10 Supplement), A06-A06.

Finn, K., Jacobsen, B., O'Brien, M., & Roy, H. (2013). *Colorectal Cancer Screening at Boston Medical Center: 2013 Public Report*. Boston Medical Center.

Ford, P. J., & Farah, C. S. (2013). Early Detection and Diagnosis of Oral Cancer: Strategies for Improvement. *Journal of Cancer Policy*, 1(1), E2-E7.

Forsberg, C., Björvell, H., & Cedermark, B. (1996). Well-Being and Its Relation to Coping Ability in Patients with Colorectal and Gastric Cancer Before and After. *Scandinavian Journal of Caring Sciences*, 10(1), 35-44.

Foucault, M. (2012). *The Birth of the Clinic*. Routledge.

Franceschi, S., & La Vecchia, C. (1994). Alcohol and the Risk of Cancers of the Stomach and Colon-Rectum. *Digestive Diseases,* 12(5), 276-289.

Frankenberg, R. (1974). Functionalism and After? Theory and Developments in Social Science Applied To the Health Field. *International Journal of Health Services*, 4(3), 411-427.

Fraumeni Jr, J. F. (1988). Etiologic Insights from Cancer Mapping. Unusual Occurrences as Clues to Cancer Etiology. *Japan Sci Soc Press*, Tokyo, 13-25.

Freeman, H. P. (2004). Poverty, Culture, and Social Injustice: Determinants of Cancer Disparities. *CA: A Cancer Journal for Clinicians*, 54(2), 72-77.

Fuchs, C. S., Giovannucci, E. L., Colditz, G. A., Hunter, D. J., Stampfer, M. J., Rosner, B., Speizer, F. E., Willett, W. C. (1999). Dietary Fiber and the Risk of Colorectal Cancer and Adenoma in Women. *New England Journal of Medicine*, 340(3), 169-176.

Fuggle, R. F. (2001). *Lake Victoria: a Case Study of Complex Interrelationships*. United Nations Environment Programme, Nairobi.

Galán, J. M., Izquierdo, L. R., Izquierdo, S. S., Santos, J. I., Del Olmo, R., López-Paredes, A., Edmonds, B. (2009). Errors and Artefacts in Agent-Based Modelling. *Journal of Artificial Societies and Social Simulation*, 12(1), 1.

Galán, J. M., López-Paredes, A., & Del Olmo, R. (2009). An Agent-Based Model for Domestic Water Management in Valladolid Metropolitan Area. *Water Resources Research*, 45(5).

Galdston, I. (Ed.). (1949). *Social Medicine: Its Derivations and Objectives.* Commonwealth Fund.

Galea, S., Riddle, M., & Kaplan, G. A. (2010). Causal Thinking and Complex System Approaches in Epidemiology. *International Journal of Epidemiology*, 39(1), 97-106.

Galea, S., Tracy, M., Hoggatt, K. J., Dimaggio, C., & Karpati, A. (2011). Estimated Deaths Attributable to Social Factors in the United States. *American Journal of Public Health*, 101(8), 1456-1465.

Galiatsatos, P., & Foulkes, W. D. (2006). Familial Adenomatous Polyposis. *The American Journal of Gastroenterology*, 101(2), 385-398.

Ganster, D. C., & Schaubroeck, J. (1991). Work Stress and Employee Health. *Journal of Management*, 17(2), 235-271.

Garabrant, D. H., Peters, J. M., Mack, T. M., & Bernstein, L. (1984). Job Activity and Colon Cancer Risk. *American Journal of Epidemiology*, 119(6), 1005-1014.

Gell-Mann, M. (2002). What Is Complexity? In *Complexity and Industrial Clusters* (pp. 13-24). Physica-Verlag HD.

Gerhardsson, M., Norell, S. E., Kiviranta, H., Pedersen, N. L., & Ahlbom, A. (1986). Sedentary Jobs and Colon Cancer. *American Journal of Epidemiology*, 123(5), 775-780.

Gilbert, N. (2004). Agent-Based Social Simulation: Dealing with Complexity. *The Complex Systems Network of Excellence*, 9(25), 1-14.

Gilbert, N. (2007). *Agent-Based Models* (Quantitative Applications in the Social Sciences) Gilbert N., Sage Publications.

Gilbert, N., & Bankes, S. (2002). Platforms and Methods for Agent-Based Modeling. *Proceedings of the National Academy of Sciences*, 99(Suppl 3), 7197-7198.

Gilbert, N., & Terna, P. (2000). How to Build and Use Agent-Based Models in Social Science. *Mind & Society*, 1(1), 57-72.

Gilbert, N., & Troitzsch, K. (2005). *Simulation for the Social Scientist*. McGraw-Hill International.

Gilli, M., & Winker, P. (2003). A Global Optimization Heuristic for Estimating Agent Based Models. *Computational Statistics & Data Analysis*, 42(3), 299-312.

Gillman, M. W., & Hammond, R. A. (2015). Precision Treatment and Precision Prevention: Integrating "Below and Above the Skin". *JAMA Pediatrics*, 1-2.

Gimblett, H. R. (2002) *Integrating Geographic Information Systems and Agent-Based Modeling Techniques for Simulating Social and Ecological Processes*. Vol. 168. New York: Oxford University Press

Gimeno-García, A. Z., Quintero, E., Nicolás-Pérez, D., Parra-Blanco, A., & Jiménez, A. (2009). Colorectal Cancer Screening in a Spanish Population. *Medicina Clínica*, 133(19), 736-740.

Giovannucci, E., Rimm, E. B., Stampfer, M. J., Colditz, G. A., Ascherio, A., Kearney, J., & Willett, W. C. (1994). A Prospective Study of Cigarette Smoking and Risk of Colorectal Adenoma and Colorectal Cancer in US Men. *Journal of the National Cancer Institute*, 86(3), 183-191.

Glade, M. J. (2008). World Cancer Research Fund/American Institute for Cancer Research Food, Nutrition, Physical Activity and the Prevention of Cancer: a Global Perspective. American Institute for Cancer Research. Washington, D. C. ISBN: 978-0-9722522-2-5. *Nutrition*, 24(4), 393-398.

Gnilomedov, I., & Nikolenko, S. (2010). Agent-Based Economic Modeling with Finite State Machines. In *Combined Proceedings of the International Symposium On Social Network Analysis and Norms for MAS* (pp. 28-33).

Gonzalez-Hermoso, F., Perez-Palma, J., Marchena-Gomez, J., Lorenzo-Rocha, N., & Medina-Arana, V. (2004). Can Early Diagnosis of Symptomatic Colorectal Cancer Improve the Prognosis? *World Journal of Surgery*, 28(7), 716-720.

Google Maps. (2016). *Washington, DC*. Retrieved from: https://www.google.com/maps/place/washington,+dc/

Gopalappa, C., Aydogan-Cremaschi, S., Das, T. K., & Orcun, S. (2011). Probability Model for Estimating Colorectal Polyp Progression Rates. *Health Care Management Science*, 14(1), 1-21.

Gorman, D. M., Mezic, J., Mezic, I., & Gruenewald, P. J. (2006). Agent-Based Modeling of Drinking Behavior: a Preliminary Model and Potential Applications to Theory and Practice. *American Journal of Public Health*, 96(11), 2055-2060.

Goto, Y., & Takahashi, S. (2013). Agent-Based Modeling and Simulation Validation by Scenario Analysis. In *Agent-Based Approaches in Economic and Social Complex Systems VII* (pp. 3-16). Springer Japan.

Gottlieb, B. H. (1981). *Social Networks and Social Support* (Vol. 4). Sage Publications, Inc.

Gotts, N. M., & Polhill, J. G. (2009). Narrative Scenarios, Mediating Formalisms, and the Agent-Based Simulation of Land Use Change. In *Epistemological Aspects of Computer Simulation in the Social Sciences* (pp. 99-116). Springer Berlin Heidelberg.

Greenlee, R. T., Murray, T., Bolden, S., & Wingo, P. A. (2000). Cancer Statistics, 2000. *CA: A Cancer Journal for Clinicians*, 50(1), 7-33.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H. H., Weiner, J., Wiegand, T., Deangelis, D. L. (2005). Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. *Science*, 310(5750), 987-991.

Grimm, V., & Railsback, S. F. (2005). *Individual-Based Modeling and Ecology.* Princeton Series in Theoretical and Computational Biology.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., Deangelis, D. L. (2006). A Standard Protocol for Describing Individual-Based and Agent-Based Models. *Ecological Modelling*, 198(1), 115-126.

Grimm, V., Berger, U., Deangelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD Protocol: a Review and First Update. *Ecological Modelling*, 221(23), 2760-2768.

Gupta, S., Halm, E. A., Rockey, D. C., Hammons, M., Koch, M., Carter, E., Valdez, L., Tong, L., Ahn, C., Kashner, M., Argenbright, K., (2013). Comparative Effectiveness of Fecal Immunochemical Test Outreach, Colonoscopy Outreach, and Usual Care for Boosting

Colorectal Cancer Screening Among the Underserved: a Randomized Clinical Trial. *JAMA Internal Medicine*, 173(18), 1725-1732.

Grizzi, F., & Chiriva-Internati, M. (2006). Cancer: Looking for Simplicity and Finding Complexity. *Cancer Cell International*, 6(1), 4.

Gyrd-Hansen, D., Søggard, J., & Kronborg, O. (1998). Colorectal Cancer Screening: Efficiency and Effectiveness. *Health Economics*, 7(1), 9-20.

Hacking, I. (1999). *The Social Construction of What?* Harvard University Press.

Hailegiorgis, A., & Crooks, A. T. (2012). Agent-Based Modeling for Humanitarian Issues: Disease and Refugee Camps. In *the Computational Social Science Society of America Conference*, Santa Fe, NM.

Hammond, R. A. (2009). Complex Systems Modeling for Obesity Research. *Preventing Chronic Disease*, 6(3).

Hammond, R. A., & Epstein, J. M. (2007). *Exploring Price-Independent Mechanisms in the Obesity Epidemic*. Brookings Institution Center On Social and Economic Dynamics Working Paper.

Hansen, L. P., & Heckman, J. J. (1996). The Empirical Foundations of Calibration. *The Journal of Economic Perspectives*, 87-104.

Hansen, H. J., Morsel-Carlsen, L., & Bülow, S. (1997). Patients' Perception of Symptoms in Colorectal Cancer. a Cause of Delay in Diagnosis and Treatment. *Ugeskrift for Laeger*, 159(13), 1941-1944.

Hansen, R. P., Olesen, F., Sørensen, H. T., Sokolowski, I., & Søndergaard, J. (2008). Socioeconomic Patient Characteristics Predict Delay in Cancer Diagnosis: a Danish Cohort Study. *BMC Health Services Research*, 8(1), 1.

Hassan, S., Pavon, J., & Gilbert, N. (2008). *Injecting Data into Simulation: Can Agent-Based Modelling Learn from Microsimulation*. In World Congress of Social Simulation.

Human Services. (2000). *Healthy People 2010* (Vol. 2). US Department of Health and Human Services.

Heaney, C. A., & Israel, B. A. (2008). Social Networks and Social Support. *Health Behavior and Health Education: Theory, Research, and Practice*, 4, 189-210.

Heckbert, S., Straton, A., Smajgl, A., & Ward, J. (2007, May). Agent-Based Modelling Calibrated from Experimental Economics: Bidding Behaviour in a Water Market. In *Proceedings of the 18th IASTED International Conference: Modelling and Simulation* (pp. 648-652). ACTA Press.

Heng, H. H., Bremer, S. W., Stevens, J. B., Ye, K. J., Liu, G., & Ye, C. J. (2009). Genetic and Epigenetic Heterogeneity in Cancer: a Genome-Centric Perspective. *Journal of Cellular Physiology*, 220(3), 538-547.

Heppenstall, A. J., Evans, A. J., & Birkin, M. H. (2006). Application of Multi-Agent Systems to Modelling a Dynamic, Locally Interacting Retail Market. *Journal of Artificial Societies and Social Simulation*, 9(3), 2.

Hermann, C. F. (1967). Validation Problems in Games and Simulations with Special Reference to Models of International Politics. *Behavioral Science*, 12(3), 216-231.

Hetzel, L., Smith, A. (2001). *The 65 Years and Over Population, 2000*. Washington, DC: U.S. Dept. of Commerce, Economics and Statistics Administration, U.S. Census Bureau.

Hewitt, M., Breen, N., & Devesa, S. (1999). Cancer Prevalence and Survivorship Issues: Analyses of the 1992 National Health Interview Survey. *Journal of the National Cancer Institute*, 91(17), 1480-1486.

Hewitt, M., Rowland, J. H., & Yancik, R. (2003). Cancer Survivors in the United States: Age, Health, and Disability. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 58(1), M82-M91.

Heymann, J., Hertzman, C., Barer, M. L., & Evans, R. G. (Eds.). (2005). *Healthier Societies: From Analysis to Action*. Oxford University Press.

Hiatt, R. A., & Breen, N. (2008). The Social Determinants of Cancer: a Challenge for Transdisciplinary Science. *American Journal of Preventive Medicine*, 35(2), S141-S150.

Hiatt, R. A., & Rimer, B. K. (1999). A New Strategy for Cancer Control Research. *Cancer Epidemiology Biomarkers & Prevention*, 8(11), 957-964.

Hill, A. B. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*, 58(5), 295.

Holland, J. H. (2012). *Signals and Boundaries: Building Blocks for Complex Adaptive Systems*. MIT Press.

Homans, G. C. (1958). Social Behavior as Exchange. *American Journal of Sociology*, 597-606.

Homer, J. B., & Hirsch, G. B. (2006). System Dynamics Modeling for Public Health: Background and Opportunities. *American Journal of Public Health*, 96(3), 452-458.

House, J. S., Landis, K. R., & Umberson, D. (1988). Social Relationships and Health. *Science*, 241(4865), 540-545.

House, J. S., Robbins, C., & Metzner, H. L. (1982). The Association of Social Relationships and Activities with Mortality: Prospective Evidence from the Tecumseh Community Health Study. *American Journal of Epidemiology*, 116(1), 123-140.

Howlader, N., Noone, A. M., & Krapcho, M. (2014). *SEER Cancer Statistics Review, 1975-2011*. SEER Website. National Cancer Institute, USA.

Huizinga, D., & Kolawa, A. (2007). *Automated Defect Prevention: Best Practices in Software Management*. John Wiley & Sons.

Imperiale, T. F., Ransohoff, D. F., Itzkowitz, S. H., Levin, T. R., Lavin, P., Lidgard, G. P., Ahlquist, D. A., Berger, B. M. (2014). Multitarget Stool DNA Testing for Colorectal-Cancer Screening. *New England Journal of Medicine*, 370(14), 1287-1297.

Ionescu, M. V., Carey, F., Tait, I. S., & Steele, R. J. (1998). Socioeconomic Status and Stage at Presentation of Colorectal Cancer. *Lancet, 352,*9138.)

Jemal, A., Clegg, L. X., Ward, E., Ries, L. A., Wu, X., Jamison, P. M., Wingo, P. A., Howe, H. L., Anderson, R. N., Edwards, B. K. (2004). Annual Report to the Nation On the Status of Cancer, 1975–2001, With a Special Feature Regarding Survival. *Cancer*, 101(1), 3-27.

Jernigan, J. C., Trauth, J. M., Neal-Ferguson, D., & Cartier-Ulrich, C. (2001). Factors That Influence Cancer Screening in Older African American Men and Women: Focus Group Findings. *Family & Community Health*, 24(3), 27-33.

Jones, P. A., & Baylin, S. B. (2002). The Fundamental Role of Epigenetic Events in Cancer. *Nature Reviews Genetics*, 3(6), 415-428.

Johns, L. E., & Houlston, R. S. (2001). A Systematic Review and Meta-Analysis of Familial Colorectal Cancer Risk. *The American Journal of Gastroenterology*, 96(10), 2992-3003.

Jones, N. A., Smith, A. S., Black, A. A., Hawaiian, N., & Indian, A. (2001). *The Two Or More Races Population, 2000*. (Vol. 8, No. 2). US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Kaplan, G. A. (1996). People and Places: Contrasting Perspectives On the Association Between Social Class and Health. *International Journal of Health Services*, 26(3), 507-519.

Kaplan, G. A. (2004). What's Wrong with Social Epidemiology, And How Can We Make It Better? *Epidemiologic Reviews*, 26(1), 124-135.

Kasaie, P., Dowdy, D. W., & Kelton, W. D. (2013). An Agent-Based Simulation of a Tuberculosis Epidemic: Understanding the Timing of Transmission. In *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World* (pp. 2227-2238). IEEE Press.

Kerssens, J. J., Bensing, J. M., & Andela, M. G. (1997). Patient Preference for Genders of Health Professionals. *Social Science & Medicine*, 44(10), 1531-1540.

Kessler, D. A., & Pines, W. L. (1990). The Federal Regulation of Prescription Drug Advertising and Promotion. *JAMA*, 264(18), 2409-2415.

Koizumi, N., Rothbard, A. B., & Kuno, E. (2009). Distance Matters in Choice of Mental Health Program: Policy Implications for Reducing Racial Disparities in Public Mental Health Care. *Administration and Policy in Mental Health and Mental Health Services Research*, 36(6), 424-431.

Kouklakis, G. & Bampali, A. (2013). *Colorectal Cancer, Colonoscopy and Colorectal Cancer Screening - Future Directions*, Dr. Marco Bustamante (Ed.), ISBN: 978-953-51-0949-5, Intech, DOI: 10.5772/53524. Retrieved from:

http://www.intechopen.com/books/colonoscopy-and-colorectal-cancer-screening-future-directions/colorectal-cancer

Krieger, N. (1987). Shades of Difference: Theoretical Underpinnings of the Medical Controversy On Black/White Differences in the United States, 1830-1870. *International Journal of Health Services*, 17(2), 259-278.

Krieger, N. (1994). Epidemiology and the Web of Causation: Has Anyone Seen the Spider? *Social Science & Medicine*, 39(7), 887-903.

Krieger, N. (2001). A Glossary for Social Epidemiology. *Journal of Epidemiology and Community Health*, 55(10), 693-700.

Krieger, N. (2001a). Theories for Social Epidemiology in the 21st Century: An Ecosocial Perspective. *International Journal of Epidemiology*, 30(4), 668-677.

Krieger, N. (2005). Defining and Investigating Social Disparities in Cancer: Critical Issues. *Cancer Causes & Control*, 16(1), 5-14.

Krieger, N., Quesenberry Jr, C., Peng, T., Horn-Ross, P., Stewart, S., Brown, S., Swallen, K., Guillermo, T., Suh, D., Alvarez-Martinez, L., Ward, F. (1999). Social Class, Race/Ethnicity, and Incidence of Breast, Cervix, Colon, Lung, and Prostate Cancer Among Asian, Black, Hispanic, and White Residents of the San Francisco Bay Area, 1988–92 (United States). *Cancer Causes & Control*, 10(6), 525-537.

Kuh, D., & Ben-Shlomo, Y. (1997). *A Life Course Approach to Chronic Disease Epidemiology* (No. Ed. 2). Oxford University Press.

Kulldorff, M., Athas, W. F., Feurer, E. J., Miller, B. A., & Key, C. R. (1998). Evaluating Cluster Alarms: a Space-Time Scan Statistic and Brain Cancer in Los Alamos, New Mexico. *American Journal of Public Health*, 88(9), 1377-1380.

Kune, G. A., & Vitetta, L. (1992). Alcohol Consumption and the Etiology of Colorectal Cancer: a Review of the Scientific Evidence from 1957 To 1991. *Nutrition and Cancer*, 18(2), 97-111.

Landis, S. H., Murray, T., Bolden, S., & Wingo, P. A. (1999). Cancer Statistics, 1999. *CA: A Cancer Journal for Clinicians*, 49(1), 8-31.

Lane, D. C., Monefeldt, C., & Rosenhead, J. V. (2000). Looking in the Wrong Place for Healthcare Improvements: a System Dynamics Study of an Accident and Emergency Department. *Journal of the Operational Research Society*, 518-531.

Langenbach, M. R., Schmidt, J., Neumann, J., & Zirngibl, H. (2003). Delay in Treatment of Colorectal Cancer: Multifactorial Problem. *World Journal of Surgery*, 27(3), 304-308.

Lansdorp-Vogelaar, I., Van Ballegooijen, M., Zauber, A. G., Habbema, J. D. F., & Kuipers, E. J. (2009). Effect of Rising Chemotherapy Costs On the Cost Savings of Colorectal Cancer Screening. *Journal of the National Cancer Institute*, 101(20), 1412-1422.

Larsson, S. C. & Wolk, A. (2007). Obesity and Colon and Rectal Cancer Risk: a Meta-Analysis of Prospective Studies. *The American Journal of Clinical Nutrition*, 86(3), 556-565.

Law, A. M. & Kelton, W. D. (1991). *Simulation Modeling and Analysis*. New York: Mcgraw-Hill.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M. and Jebara, T. (2009). Life in the Network: the Coming Age of Computational Social Science. *Science* (New York, NY), 323(5915), 721.

Le, H., Ziogas, A., Lipkin, S. M., & Zell, J. A. (2008). Effects of Socioeconomic Status and Treatment Disparities in Colorectal Cancer Survival. *Cancer Epidemiology Biomarkers & Prevention*, 17(8), 1950-1962.

Leck, I. (1976). Atlas of Cancer Mortality for US Counties: 1950-1969. *British Journal of Cancer*, 33(2), 236.

Lee, J. E., & Chan, A. T. (2011). Fruit, Vegetables, and Folate: Cultivating the Evidence for Cancer Prevention. *Gastroenterology*, 141(1), 16.

Lempert, R. (2002). Agent-Based Modeling as Organizational and Public Policy Simulators. *Proceedings of the National Academy of Sciences,* 99(Suppl 3), 7195-7196.

Levin, B., Lieberman, D. A., Mcfarland, B., Smith, R. A., Brooks, D., Andrews, K. S., Dash, C., Giardiello, F. M., Glick, S., Levin, T. R. and Pickhardt, P., Winawer, S. J. (2008). Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous Polyps, 2008: A Joint Guideline from the American Cancer Society, the US Multi-Society Task Force On Colorectal Cancer, and the American College of Radiology*†. *CA: A Cancer Journal for Clinicians*, 58(3), 130-160.

Lewin-Epstein, N. (1991). Determinants of Regular Source of Health Care in Black, Mexican, Puerto Rican, and Non-Hispanic White Populations. *Medical Care*, 29(6), 543-557.

Lewis, S. F., & Jensen, N. M. (1996). Screening Sigmoidoscopy. *Journal of General Internal Medicine*, 11(9), 542-544.

Le, H., Ziogas, A., Lipkin, S. M., & Zell, J. A. (2008). Effects of Socioeconomic Status and Treatment Disparities in Colorectal Cancer Survival. *Cancer Epidemiology Biomarkers & Prevention*, 17(8), 1950-1962.

Levin, G., Roberts, E. B., & Hirsch, G. B. (1975). *The Persistent Poppy: a Computer-Aided Search for Heroin Policy*. Cambridge, Massachusetts: Ballinger Publishing Company.

Lewis, D. (2008). *Convention: a Philosophical Study*. John Wiley & Sons.

Lieberman, D. A. (1995). Cost-Effectiveness Model for Colon Cancer Screening. *Gastroenterology*, 109(6), 1781-1790.

Lillard, L. A., & Waite, L. J. (1995). 'Til Death Do Us Part: Marital Disruption and Mortality. *American Journal of Sociology*, 1131-1156.

Lillie-Blanton, M. (2003). *The Kaiser Family Foundation DC Health Care Access Survey Highlights and Chartpack.* Kaiser Family Foundation, Oakland CA.

Lillie-Blanton, M. D., Rushing, O. E., & Ruiz, S. (2003). *Key Facts: Race, Ethnicity & Medical Care*. Henry J. Kaiser Family Foundation, Oakland CA.

Ling, B. S., Moskowitz, M. A., Wachs, D., Pearson, B., & Schroy, P. C. (2001). Attitudes Toward Colorectal Cancer Screening Tests. *Journal of General Internal Medicine*, 16(12), 822-830.

Link, B. G., Northridge, M. E., Phelan, J. C., & Ganz, M. L. (1998). Social Epidemiology and the Fundamental Cause Concept: On the Structuring of Effective Cancer Screens by Socioeconomic Status. *Milbank Quarterly*, 76(3), 375-402.

Link, B. G., & Phelan, J. (1995). Social Conditions as Fundamental Causes of Disease. *Journal of Health and Social Behavior*, 80-94.

López-Paredes, A., Saurí, D., & Galán, J. M. (2005). Urban Water Management with Artificial Societies of Agents: The FIRMABAR Simulator. *Simulation*, 81(3), 189-199.

Loupakis, F., Yang, D., Yau, L., Feng, S., Cremolini, C., Zhang, W., Maus, M. K., Antoniotti, C., Langer, C., Scherer, S. J., Müller, T. (2015). Primary Tumor Location as a Prognostic Factor in Metastatic Colorectal Cancer. *Journal of the National Cancer Institute*, 107(3), Dju427.

Lynch, H. T., & De La Chapelle, A. (2003). Hereditary Colorectal Cancer. *New England Journal of Medicine*, 348(10), 919-932.

Luke, S., Cioffi-Revilla, C., Panait, L., Sullivan, K., & Balan, G. (2005). Mason: a Multiagent Simulation Environment. *Simulation*, 81(7), 517-527.

Lundy, J. J., Coons, S. J., Wendel, C., Hornbrook, M. C., Herrinton, L., Grant, M., & Krouse, R. S. (2009). Exploring Household Income as a Predictor of Psychological Well-Being Among Long-Term Colorectal Cancer Survivors. *Quality of Life Research*, 18(2), 157-161.

Macklin, P., Edgerton, M. E., Thompson, A. M., & Cristini, V. (2012). Patient-Calibrated Agent-Based Modelling of Ductal Carcinoma in Situ (DCIS): From Microscopic Measurements to Macroscopic Predictions of Clinical Progression. *Journal of Theoretical Biology*, 301, 122-140.

Maglio, P. P., & Mabry, P. L. (2011). Agent-Based Models and Systems Science Approaches to Public Health. *American Journal of Preventive Medicine*, 40(3), 392-394.

Maguire, D. J. (2005). Towards a GIS Platform for Spatial Analysis and Modeling. *GIS, Spatial Analysis and Modelling*, ESRI Press, Redlands, California, CA, 19-39.

Mandel, J. S., Bond, J. H., Church, T. R., Snover, D. C., Bradley, G. M., Schuman, L. M., & Ederer, F. (1993). Reducing Mortality from Colorectal Cancer by Screening for Fecal Occult Blood. *New England Journal of Medicine*, 328(19), 1365-1371.

Mandelblatt, J., Andrews, H., Kao, R., Wallace, R., & Kerner, J. (1996). The Late-Stage Diagnosis of Colorectal Cancer: Demographic and Socioeconomic Factors. *American Journal of Public Health*, 86(12), 1794-1797.

Mandelbrot, B. B. (1983). *The Fractal Geometry of Nature*. Macmillan.

Manne, S. L., Chung, D. C., Weinberg, D. S., Vig, H. S., Catts, Z., Cabral, M. K., Shannon, K., Meropol, N. J. (2007). Knowledge and Attitudes About Microsatellite Instability Testing Among High-Risk Individuals Diagnosed with Colorectal Cancer. *Cancer Epidemiology Biomarkers & Prevention*, 16(10), 2110-2117.

Manson, S. M. (2003). Validation and Verification of Multi-Agent Models for Ecosystem Management. *Complexity and Ecosystem Management: the Theory and Practice of Multi-Agent Approaches*, 63-74.

Manson, S. M., Sun, S., & Bonsal, D. (2012). Agent-Based Modeling and Complexity. In *Agent-Based Models of Geographical Systems* (pp. 125-139). Springer Netherlands.

Marmot, M. G. (1998). Improvement of Social Environment to Improve Health. *The Lancet*, 351(9095), 57-60.

Marmot, M. (2002). The Influence of Income On Health: Views of an Epidemiologist. *Health Affairs*, 21(2), 31-46.

Marmot, M., Friel, S., Bell, R., Houweling, T. A., Taylor, S., (2008). Closing the Gap in a Generation: Health Equity Through Action On the Social Determinants of Health. *The Lancet*, 372 (9650), 1661-1669.

Marmot, M., & Wilkinson, R. (Eds.). (2005). *Social Determinants of Health*. OUP Oxford.

Marsden, P. V. (2002). Egocentric and Sociocentric Measures of Network Centrality. *Social Networks*, 24(4), 407-422.

Maskarinec, G., Murphy, S., Shumay, D. M., & Kakai, H. (2001). Dietary Changes Among Cancer Survivors. *European Journal of Cancer Care*, 10(1), 12-20.

Mayberry, R. M., Coates, R. J., Hill, H. A., Click, L. A., Chen, V. W., Austin, D. F., Redmond, C. K., Fenoglio-Preiser, C. M., Hunter, C. P., Haynes, M. A., Muss, H. B. (1995). Determinants of Black/White Differences in Colon Cancer Survival. *Journal of the National Cancer Institute,* 87(22), 1686-1693.

Meyerhardt, J. A., Heseltine, D., Niedzwiecki, D., Hollis, D., Saltz, L. B., Mayer, R. J., Thomas, J., Nelson, H., Whittom, R., Hantel, A., Schilsky, R. L. (2006). Impact of Physical Activity On Cancer Recurrence and Survival in Patients with Stage III Colon Cancer: Findings from CALGB 89803. *Journal of Clinical Oncology*, 24(22), 3535-3541.

Mayo Clinic (2016). *Colon Cancer Stages* [Online Image] Retrieved from: http://www.mayoclinic.org/colon-cancer-stages/img-20007084

Mccarty, C., Killworth, P. D., Bernard, H. R., Johnsen, E. C., & Shelley, G. A. (2001). Comparing Two Methods for Estimating Network Size. *Human Organization*, 60(1), 28-39.

Mcewen, B. S., & Seeman, T. (1999). Protective and Damaging Effects of Mediators of Stress: Elaborating and Testing the Concepts of Allostasis and Allostatic Load. *Annals of the New York Academy of Sciences*, 896(1), 30-47.

Mckinlay, J. B. (1979). A Case for Refocusing Upstream: the Political Economy of Illness. *Patients, Physicians and Illness: a Sourcebook in Behavioral Science and Health*, 9-25.

Michels, K. B., Giovannucci, E., Joshipura, K. J., Rosner, B. A., Stampfer, M. J., Fuchs, C. S., Colditz, G. A., Speizer, F. E., Willett, W. C. (2000). Prospective Study of Fruit and Vegetable Consumption and Incidence of Colon and Rectal Cancers. *Journal of the National Cancer Institute*, 92(21), 1740-1752.

Mctiernan, A. (2008). Mechanisms Linking Physical Activity with Cancer. *Nature Reviews Cancer*, 8(3), 205-211.

Mctiernan, A., Ulrich, C., Slate, S., & Potter, J. (1998). Physical Activity and Cancer Etiology: Associations and Mechanisms. *Cancer Causes & Control*, 9(5), 487-509.

Miller, J. H., & Page, S. E. (2009). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press.

Milstein, B., Homer, J., Briss, P., Burton, D., & Pechacek, T. (2011). Why Behavioral and Environmental Interventions Are Needed to Improve Health at Lower Cost. *Health Affairs*, 30(5), 823-832.

Moghaddam, A. A., Woodward, M., & Huxley, R. (2007). Obesity and Risk of Colorectal Cancer: a Meta-Analysis of 31 Studies with 70,000 Events. *Cancer Epidemiology Biomarkers & Prevention*, 16(12), 2533-2547.

Moore, N. M., Kuhn, N. Z., Hanlon, S. E., Lee, J. S. H., & Nagahara, L. A. (2011). De-Convoluting Cancer's Complexity: Using a 'Physical Sciences Lens' To Provide a Different (Clearer) Perspective of Cancer. *Physical Biology, 8,* 1, 10302.

Mokdad, A. H., Serdula, M. K., Dietz, W. H., Bowman, B. A., Marks, J. S., & Koplan, J. P. (1999). The Spread of the Obesity Epidemic in the United States, 1991-1998. *JAMA*, 282(16), 1519-1522.

Mor, V., Masterson-Allen, S., Goldberg, R., Guadagnoli, E., & Wool, M. S. (1990). Pre-Diagnostic Symptom Recognition and Help Seeking Among Cancer Patients. *Journal of Community Health*, 15(4), 253-266.

Moss, S. (2008). Alternative Approaches to the Empirical Validation of Agent-Based Models. *Journal of Artificial Societies and Social Simulation*, 11(1), 5.

Müller, B., F. Bohn, G. Dressler, J. Groeneveld, C. Klassert, M. Schlüter, J. Schulze, H. Weise, and N. Schwarz, 2013. Describing Human Decisions in Agent-Based Models - ODD+D, An Extension of the ODD Protocol. *Environmental Modelling & Software* 48, 37-48.

Murabito, J. M., Evans, J. C., Larson, M. G., Kreger, B. E., Splansky, G. L., Freund, K. M., Moskowitz, M. A., Wilson, P. W. (2001). Family Breast Cancer History and Mammography: Framingham offspring Study. *American Journal of Epidemiology*, 154(10), 916-923.

Murray, C. J., Kulkarni, S. C., Michaud, C., Tomijima, N., Bulzacchelli, M. T., Iandiorio, T. J., & Ezzati, M. (2006). Eight Americas: Investigating Mortality Disparities Across Races, Counties, and Race-Counties in the United States. *Plos Medicine*, 3(9), E260.

Muto, T., Bussey, H. J. R., & Morson, B. C. (1975). The Evolution of Cancer of the Colon and Rectum. *Cancer*, 36(6), 2251-2270.

Myers, R. E., Sifri, R., Hyslop, T., Rosenthal, M., Vernon, S. W., Cocroft, J., Wolf, T., Andrel, J., Wender, R. (2007). A Randomized Controlled Trial of the Impact of Targeted and Tailored Interventions On Colorectal Cancer Screening. *Cancer*, 110(9), 2083-2091.

Nadel, M. R., Shapiro, J. A., Klabunde, C. N., Seeff, L. C., Uhler, R., Smith, R. A., & Ransohoff, D. F. (2005). a National Survey of Primary Care Physicians' Methods for Screening for Fecal Occult Blood. *Annals of Internal Medicine*, 142(2), 86-94.

National Cancer Institute: PDQ® *Colorectal Cancer Prevention*. (2016) National Cancer Institute. Last Modified On 02/11/2016. Retrieved from: http://www.cancer.gov/types/colorectal/hp/colorectal-prevention-pdq.

National Cancer Institute: PDQ® *Colon Cancer Treatment*. (2016) Bethesda, MD: National Cancer Institute. Last Modified On 07/22/2015. Retrieved from: http://www.cancer.gov/types/colorectal/patient/colon-treatment-pdq.

National Center for Health Statistics (NCHS). (2000). *National Health Interview Survey: Public Use Datasets*. Retrieved from: ftp://ftp.cdc.gov/pub/health_statistics/nchs/program_code/nhis/2000/readme.txt

National Center for Health Statistics (NCHS) (2015). *United States Cancer Statistics Data: All Data Combined*. Public-Use Data File and Documentation. Retrieved from: http://www.cdc.gov/cancer/npcr/uscs/download_data.htm

National Cancer Institute (NCI) (2015) *The Cancer Control Continuum*. Retrieved from: http://cancercontrol.cancer.gov/od/continuum.html.

National Cancer Institute (NCI) (2000) *State Cancer Profiles*. Retrieved from: http://cancercontrol.cancer.gov

NeighborhoodInfo, D. C. (2016). *Neighborhood Profiles, Washington DC 2000 Census Tracts*. Retrieved from: http://www.neighborhoodinfodc.org/censustract/census.html

Nelson, A. (2002). Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. *Journal of the National Medical Association*, 94(8), 666.

Nelson, D. B., Mcquaid, K. R., Bond, J. H., Lieberman, D. A., Weiss, D. G., & Johnston, T. K. (2002). Procedural Success and Complications of Large-Scale Screening Colonoscopy. *Gastrointestinal Endoscopy*, 55(3), 307-314.

Nelson, D. E., Bland, S., Powell-Griner, E., Klein, R., Wells, H. E., Hogelin, G., & Marks, J. S. (2002). State Trends in Health Risk Factors and Receipt of Clinical Preventive Services Among US Adults During the 1990s. *JAMA*, 287(20), 2659-2667.

Neugut AI, Young GP. Screening for Colorectal Cancer: An Overview. In: Young GP, Rozen P, Levin B, Editors. *Prevention and Early Detection of Colorectal Cancer*. London (U.K.): Saunders; 1996. P. 357–67.

Ngo, T. A., & See, L. (2012). Calibration and Validation of Agent-Based Models of Land Cover Change. In *Agent-Based Models of Geographical Systems* (pp. 181-197). Springer Netherlands.

Nishi, A., & Christakis, N. A. (2015). Human Behavior Under Economic Inequality Shapes Inequality. *Proceedings of the National Academy of Sciences*, 112(52), 15781-15782.

Norat, T., Bingham, S., Ferrari, P., Slimani, N., Jenab, M., Mazuir, M., Overvad, K., Olsen, A., Tjønneland, A., Clavel, F., Boutron-Ruault, M. C. (2005). Meat, Fish, and Colorectal Cancer Risk: the European Prospective Investigation into Cancer and Nutrition. *Journal of the National Cancer Institute*, 97(12), 906-916.

North, M. J., & Macal, C. M. (2007). *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*. Oxford University Press.

Oakes, J. M., & Kaufman, J. S. (Eds.). (2006). *Methods in Social Epidemiology* (Vol. 1). John Wiley & Sons.

Olesen, F., Hansen, R. P., & Vedsted, P. (2009). Delay in Diagnosis: the Experience in Denmark. *British Journal of Cancer*, 101, S5-S8.

O'Brien, M. E. R., Borthwick, A., Rigg, A., Leary, A., Assersohn, L., Last, K., & Smith, I. E. (2006). Mortality Within 30 Days of Chemotherapy: a Clinical Governance Benchmarking Issue for Oncology Patients. *British Journal of Cancer*, 95(12), 1632-1636.

O'Carroll, R. E., Smith, K. B., Grubb, N. R., Fox, K. A. A., & Masterton, G. (2001). Psychological Factors Associated with Delay in Attending Hospital Following a Myocardial Infarction. *Journal of Psychosomatic Research*, 51(4), 611-614.

O'Malley, A. S., Beaton, E., Yabroff, K. R., Abramson, R., & Mandelblatt, J. (2004). Patient and Provider Barriers to Colorectal Cancer Screening in the Primary Care Safety-Net. *Preventive Medicine*, 39(1), 56-63.

Openstreetmap. (2015). *Washington, D.C.* Retrieved from: http://www.openstreetmap.org/relation/5396194

Optenberg, S. A., Thompson, I. M., Friedrichs, P., Wojcik, B., Stein, C. R., & Kramer, B. (1995). Race, Treatment, and Long-Term Survival from Prostate Cancer in an Equal-Access Medical Care Delivery System. *JAMA*, 274(20), 1599-1605.

Orcutt, G. H. (1957). A New Type of Socio-Economic System. *the Review of Economics and Statistics, 39,* 2, 116-123.

Ormond, B. A. (1999). *Health Care for Low-Income People in the District of Columbia*. Urban Institute, Washington, D.C.

Panish, J. F. (1979). Management of Patients with Polypoid Lesions of the Colon--Current Concepts and Controversies. *The American Journal of Gastroenterology*,71(3), 315-324.

Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and Using Sensitivity, Specificity and Predictive Values. *Indian Journal of Ophthalmology*, 56(1), 45.

Parker, J., & Epstein, J. M. (2011). A Distributed Platform for Global-Scale Agent-Based Models of Disease Transmission. *ACM Transactions On Modeling and Computer Simulation* (TOMACS), 22(1), 2.

Parkin D. M., Muir C. S. (1992). *Cancer Incidence in Five Continents*. Lyon, France: International Agency for Research On Cancer Scientific Publications.

Parunak, H. V. D., Savit, R., & Riolo, R. L. (1998). Agent-Based Modeling Vs. Equation-Based Modeling: a Case Study and Users' Guide. In *Multi-Agent Systems and Agent-Based Simulation* (pp. 10-25). Springer Berlin Heidelberg.

Patel, A., Crooks, A., & Koizumi, N. (2012). Slumulation: An Agent-Based Modeling Approach To Slum Formations. *Journal of Artificial Societies and Social Simulation*, 15(4), 2.

Patterson, R. E., Neuhouser, M. L., Hedderson, M. M., Schwartz, S. M., Standish, L. J., & Bowen, D. J. (2003). Changes in Diet, Physical Activity, and Supplement Use Among Adults Diagnosed with Cancer. *Journal of the American Dietetic Association*, 103(3), 323-328.

Pedersen, A. F., Hansen, R. P., & Vedsted, P. (2013). Patient Delay in Colorectal Cancer Patients: Associations with Rectal Bleeding and Thoughts About Cancer. *Plos One*, 8(7), E69700.

Perez, L., & Dragicevic, S. (2009). International Journal of Health Geographics. *International Journal of Health Geographics*, 8, 50.

Peters-Golden, H. (1982). Breast Cancer: Varied Perceptions of Social Support in the Illness Experience. *Social Science & Medicine*, 16(4), 483-491.

Pignone, M., Rich, M., Teutsch, S. M., Berg, A. O., & Lohr, K. N. (2002). Screening for Colorectal Cancer in Adults at Average Risk: a Summary of the Evidence for the US Preventive Services Task Force. *Annals of Internal Medicine*, 137(2), 132-141.

Phipps, E., Braitman, L. E., Stites, S., & Leighton, J. C. (2008). Quality of Life and Symptom Attribution in Long-Term Colon Cancer Survivors. *Journal of Evaluation in Clinical Practice*, 14(2), 254-258.

Pickle, L. W., Szczur, M., Lewis, D. R., & Stinchcomb, D. G. (2006). The Crossroads of GIS and Health Information: a Workshop On Developing a Research Agenda to Improve Cancer Control. *International Journal of Health Geographics*, 5(1), 1.

Pinquart, M., & Duberstein, P. R. (2010). Associations of Social Networks with Cancer Mortality: a Meta-Analysis. *Critical Reviews in Oncology/Hematology*, 75(2), 122-137.

Pinto, B. M., Eakin, E., & Maruyama, N. C. (2000). Health Behavior Changes After a Cancer Diagnosis: What Do We Know and Where Do We Go from Here? *Annals of Behavioral Medicine*, 22(1), 38-52.

Polhill, G., Izquierdo, L., & Gotts, N. (2004). The Ghost in the Model (and Other Effects of Floating Point Arithmetic). *Journal of Artificial Societies and Social Simulation*, 8(1).

Polhill, J. G., Parker, D., Brown, D., & Grimm, V. (2008). Using the ODD Protocol for Describing Three Agent-Based Social Simulation Models of Land-Use Change. *Journal of Artificial Societies and Social Simulation*, 11(2), 3.

Porter, M. E. (1985). *Competitive Advantage: Creating and Sustaining Superior Performance*. New York: Free Press

Potter, J. D. (1996). Nutrition and Colorectal Cancer. *Cancer Causes & Control*, 7(1), 127-146.

Powe, B. D. (1993). Perceptions of Cancer Fatalism Among African Americans: the Influence of Education, Income, and Cancer Knowledge. *Journal of National Black Nurses' Association*: *JNBNA*, 7(2), 41-48.

Price, R. A., Blanchard, J. C., Harris, R., Ruder, T., & Gresenz, C. R. (2012). *Monitoring Cancer Outcomes Across the Continuum*. RAND Corporation

Qi, A. S., Zheng, X., Du, C. Y., & An, B. S. (1993). A Cellular Automaton Model of Cancerous Growth. *Journal of Theoretical Biology*, 161(1), 1-12.

Quaranta, V., Weaver, A. M., Cummings, P. T., & Anderson, A. R. (2005). Mathematical Modeling of Cancer: the Future of Prognosis and Treatment. *Clinica Chimica Acta*, 357(2), 173-179.

Rejeski, W. J., Brawley, L. R., & Shumaker, S. A. (1996). Physical Activity and Health-Related Quality of Life. *Exercise and Sport Sciences Reviews*, 24(1), 71-108.

Ramos, M., Esteva, M., Cabeza, E., Campillo, C., Llobera, J., & Aguiló, A. (2007). Relationship of Diagnostic and Therapeutic Delay with Survival in Colorectal Cancer: a Review. *European Journal of Cancer*, 43(17), 2467-2478.

Ramsey, S. D., Berry, K., Moinpour, C., Giedzinska, A., & Andersen, M. R. (2002). Quality of Life in Long Term Survivors of Colorectal Cancer. *The American Journal of Gastroenterology*, 97(5), 1228-1234.

Rand, W. (2012). Business Applications and Research Questions Using Spatial Agent-Based Models. In *Agent-Based Models of Geographical Systems* (pp. 463-480). Springer Netherlands.

Rand, W., & Rust, R. T. (2011). Agent-Based Modeling in Marketing: Guidelines for Rigor. *International Journal of Research in Marketin*g, 28(3), 181-193.

Renehan, A. G., Tyson, M., Egger, M., Heller, R. F., & Zwahlen, M. (2008). Body-Mass Index and Incidence of Cancer: a Systematic Review and Meta-Analysis of Prospective Observational Studies. *The Lancet*, 371(9612), 569-578.

Reynolds, T. (2003). Colossal Colon, Polyp Man Aimed at Increasing Screening. *Journal of the National Cancer Institute*, 95(8), 575-575.

Reynolds, J. E., & Greboval, D. F. (1988). *Socio-Economic Effects of the Evolution of Nile Perch Fisheries in Lake Victoria: a Review* (No. 17). Food & Agriculture Org.

Reynolds, P., & Kaplan, G. A. (1990). Social Connections and Risk for Cancer: Prospective Evidence from the Alameda County Study. *Behavioral Medicine*,16(3), 101-110.

Rathore, S. S., Mcgreevey, J. D., Schulman, K. A., & Atkins, D. (2000). Mandated Coverage for Cancer-Screening Services: Whose Guidelines Do States Follow? *American Journal of Preventive Medicine*, 19(2), 71-78.

Ribba, B., Alarcón, T., Marron, K., Maini, P. K., & Agur, Z. (2004). The Use of Hybrid Cellular Automaton Models for Improving Cancer Therapy. In *Cellular Automata* (pp. 444-453). Springer Berlin Heidelberg.

Richards, M. A. (2009). The Size of the Prize for Earlier Diagnosis of Cancer in England. *British Journal of Cancer*, 101, S125-S129.

Ries, L. A. G., Eisner, M. P., Kosary, C. L., Hankey, B. F., Miller, B. A., Clegg, L., Edwards, B. (2007). *SEER Cancer Statistics Review, 1975-2004*. National Cancer Institute; Bethesda, MD: 2007. Retrieved from: seer.cancer.gov/csr/1975-2001.

Robards, J., Evandrou, M., Falkingham, J., & Vlachantoni, A. (2012). Marital Status, Health and Mortality. *Maturitas*, 73(4), 295-299.

Robinson, S. (2014). *Simulation: the Practice of Model Development and Use*. Palgrave Macmillan.

Robinson D. T., Brown D. G., Parker D. C., Schreinemachers P., Janssen M. A., Huigen M., Wittmer H., Gotts N., Promburom P., Irwin E., Berger T., Gatzweiler F., Barnaud C. (2007) Comparison of Empirical Methods for Building Agent Based Models in Land Use Science. *Journal of Land Use Science*, 2(1), 31-55.

Roge, J. D., Elmore, M. F., Mahoney, S. J., Brown, E. D., Troiano, F. P., Wagner, D. R., Black, D. J., Pound, D. C. (1994). Low-Cost, Office-Based, Screening Colonoscopy. *The American Journal of Gastroenterology*, 89(10), 1775-1780.

Royston, G., Dost, A., Townshend, J., & Turner, H. (1999). Using System Dynamics to Help Develop and Implement Policies and Programs in Health Care in England. *System Dynamics Review*, 15(3), 293.

Rosenblueth, A., & Wiener, N. (1945). The Role of Models in Science. *Philosophy of Science*, 12(4), 316-321.

Rosenstock, I. M. (1974). Historical Origins of the Health Belief Model. *Health Education & Behavior*, 2(4), 328-335.

Ross, C. E., & Wu, C. L. (1995). The Links Between Education and Health. *American Sociological Review*, 719-745.

Rowland, J. H. (2008). Cancer Survivorship: Rethinking the Cancer Control Continuum. In *Seminars in Oncology Nursing* (Vol. 24, No. 3, pp. 145-152). WB Saunders.

Rubin, M. (2002). *2000 Census Numbers Reveal Higher Poverty Numbers in the District by Ward and Neighborhood Cluster*. A DC Agenda Neighborhood Information Services Research Paper, DC Agenda, Washington, DC.

Rushton, G., Peleg, I., Banerjee, A., Smith, G., & West, M. (2004). Analyzing Geographic Patterns of Disease Incidence: Rates of Late-Stage Colorectal Cancer in Iowa. *Journal of Medical Systems*, 28(3), 223-236.

Ruth A. Wallace, & Alison Wolf. (1999). *Contemporary Sociological Theory: Expanding the Classical Tradition*. Pearson College Division.

Ryle, J. A. (1994). *Changing Disciplines: Lectures On the History, Method, and Motives of Social Pathology*. Transaction Publishers.

Sabatino, S. A., Coates, R. J., Uhler, R. J., Alley, L. G., & Pollack, L. A. (2006). Health Insurance Coverage and Cost Barriers to Needed Medical Care Among US Adult Cancer Survivors Age < 65 Years. *Cancer*, 106(11), 2466-2475.

Safer, M. A., Tharps, Q. J., Jackson, T. C., & Levknthal, H. (1979). Determinants of Three Stages of Delay in Seeking Care at a Medical Clinic. *Medical Care*, 17(1), 11-29.

Samad, A. K. A., Taylor, R. S., Marshall, T., & Chapman, M. A. (2005). A Meta‐Analysis of the Association of Physical Activity with Reduced Risk of Colorectal Cancer. *Colorectal Disease*, 7(3), 204-213.

Sanjoaquin, M. A., Allen, N., Couto, E., Roddam, A. W., & Key, T. J. (2005). Folate Intake and Colorectal Cancer Risk: a Meta-Analytical Approach. *International Journal of Cancer*, 113(5), 825-828.

Sakamoto, T., Matsuda, T., Nakajima, T., & Saito, Y. (2013). Clinico-Pathological Features of Colorectal Polyps: Evaluation of the 'Predict, Resect and Discard' Strategies. *Colorectal Disease*, 15(6), E295-E300.

Sakellariou, I. (2012). Agent Based Modelling and Simulation Using State Machines. In *SIMULTECH* (pp. 270-279).

Sapp, A. L., Trentham-Dietz, A., Newcomb, P. A., Hampton, J. M., Moinpour, C. M., & Remington, P. L. (2003). Social Networks and Quality of Life Among Female Long-Term Colorectal Cancer Survivors. *Cancer*, 98(8), 1749-1758.

Sarfaty, M., & Wender, R. (2007). How to Increase Colorectal Cancer Screening Rates in Practice. *CA: A Cancer Journal for Clinicians*, 57(6), 354-366.

Sargent, R. G. (2005). Verification and Validation of Simulation Models. In *Proceedings of the 37th Conference On Winter Simulation* (pp. 130-143). Winter Simulation Conference.

Satia, J. A., Campbell, M. K., Galanko, J. A., James, A., Carr, C., & Sandler, R. S. (2004). Longitudinal Changes in Lifestyle Behaviors and Health Status in Colon Cancer Survivors. *Cancer Epidemiology Biomarkers & Prevention*, 13(6), 1022-1031.

Scanlan, D. A. (1989). *Structured Flowcharts Outperform Pseudocode: An Experimental Comparison*. Software, IEEE, 6(5), 28-36.

Schachter, S., & Singer, J. (1962). Cognitive, Social, and Physiological Determinants of Emotional State. *Psychological Review*, 69(5), 379.

Schelling, T. C. (1969). Models of Segregation. *The American Economic Review*, 59(2), 488-493.

Schaefer, C., Quesenberry, C. P., & Wi, S. (1995). Mortality Following Conjugal Bereavement and the Effects of a Shared Environment. *American Journal of Epidemiology*, 141(12), 1142-1152.

Schoen, R. E., Pinsky, P. F., Weissfeld, J. L., Yokochi, L. A., Church, T., Laiyemo, A. O., Bresalier, R., Andriole, G. L., Buys, S. S., Crawford, E. D., Fouad, M. N. (2012). Colorectal-Cancer Incidence and Mortality with Screening Flexible Sigmoidoscopy. *New England Journal of Medicine*, 366(25), 2345-2357.

Schroevers, M., Ranchor, A. V., & Sanderman, R. (2006). Adjustment to Cancer in the 8 Years Following Diagnosis: a Longitudinal Study Comparing Cancer Survivors with Healthy Individuals. *Social Science & Medicine*, 63(3), 598-610.

Schulz, R., & Beach, S. R. (1999). Caregiving as a Risk Factor for Mortality: the Caregiver Health Effects Study. *JAMA*, 282(23), 2215-2219.

Schwartz, K. L., Crossley-May, H., Vigneau, F. D., Brown, K., & Banerjee, M. (2003). Race, Socioeconomic Status and Stage at Diagnosis for Five Common Malignancies. *Cancer Causes & Control*, 14(8), 761-766.

Scott, J. (2012). *Social Network Analysis*. Sage.

Scroggins Jr, T. G., & Bartley, T. K. (1999). Cancer Knowledge, Attitudes and Beliefs Among African Americans. *The Ochsner Journal*, 1(2), 52-57.

Shakya, H. B., Christakis, N. A., & Fowler, J. H. (2015). Self-Comparisons as Motivators for Healthy Behavior. *Obesity*, 23(12), 2477-2484.

Shapiro, J. A., Seeff, L. C., & Nadel, M. R. (2001). Colorectal Cancer-Screening Tests and Associated Health Behaviors. *American Journal of Preventive Medicine*, 21(2), 132-137.

Shike, M., Winawer, S. J., Greenwald, P. H., Bloch, A., Hill, M. J., & Swaroop, S. V. (1990). Primary Prevention of Colorectal Cancer. The WHO Collaborating Centre for the Prevention of Colorectal Cancer. *Bulletin of the World Health Organization*, 68(3), 377.

Shotter, J. (1993). C*ultural Politics of Everyday Life: Social Constructionism, Rhetoric and Knowing of the Third Kind*. University of Toronto Press.

Secretan, B., Straif, K., Baan, R., Grosse, Y., El Ghissassi, F., Bouvard, V., WHO International Agency for Research On Cancer Monograph Working Group. (2009). A Review of Human Carcinogens - Part E: Tobacco, Areca Nut, Alcohol, Coal Smoke, and Salted Fish. *The Lancet Oncology*, 10(11), 1033-1034.

Seidman, C. S. (2006). An Introduction to Prostate Cancer and Geographic Information Systems. *American Journal of Preventive Medicine*, 30(2), S1-S2.

Siegel, R. L., Jemal, A., & Ward, E. M. (2009). Increase in Incidence of Colorectal Cancer Among Young Men and Women in the United States. *Cancer Epidemiology Biomarkers & Prevention*, 18(6), 1695-1698.

Siegel, R., Ma, J., Zou, Z., & Jemal, A. (2014). Cancer Statistics, 2014. *CA: A Cancer Journal for Clinician*s, 64(1), 9-29.

Siegel, R. L., Miller, K. D., & Jemal, A. (2016). Cancer Statistics, 2016. *CA: A Cancer Journal for Clinicians*, 66(1), 7-30.

Siminoff, L., Thomson, M., & Dumenci, L. (2014). Factors Associated with Delayed Patient Appraisal of Colorectal Cancer Symptoms. *Psycho-Oncology*, 23(9), 981-988.

Simon, H. A. (1996). *The Sciences of the Artificial*. MIT Press.

Simon, H. A. (2000). Bounded Rationality in Social Science: Today and Tomorrow. *Mind & Society*, 1(1), 25-39.

Singer, A. (2004). *The Living Cities Census Series*. Washington, DC: Brookings Institution.

Singh, H., Khan, R., Giardina, T.D., Daci, K., Gould, M., & El-Serag, H. (2012). Post-Referral Colonoscopy Delays in Diagnosis of Colorectal Cancer: a Mixed-Methods Analysis. *Quality Management in Health Care*, 21(4), 252.

Smith, K. P., & Christakis, N. A. (2008). Social Networks and Health. *Annual Review of Sociology*, 34, 405-429.

Smith, L. K., Pope, C., & Botha, J. L. (2005). Patients' Help-Seeking Experiences and Delay in Cancer Presentation: a Qualitative Synthesis. *The Lancet*, 366(9488), 825-831.

Smith, R. A., Caleffi, M., Albert, U. S., Chen, T. H., Duffy, S. W., Franceschi, D., & Nyström, L. (2006). Breast Cancer in Limited-Resource Countries: Early Detection and Access to Care. *The Breast Journal*, 12(S1), S16-S26.

Sontag, S. (2001). *Illness as Metaphor and AIDS and Its Metaphors*. Macmillan.

Stefanek, M., Mcdonald, P. G., & Hess, S. A. (2005). Religion, Spirituality and Cancer: Current Status and Methodological Challenges. *Psycho‐Oncology*, 14(6), 450-463.

Stewart, S. L., King, J. B., Thompson, T. D., Friedman, C., & Wingo, P. A. (2004). Cancer Mortality Surveillance - United States, 1990–2000. *MMWR Surveillance Summaries*, 53(3), 1-108.

Suarez, L., Ramirez, A. G., Villarreal, R., Marti, J., Mcalister, A., Talavera, G. A., Trapido, E., Perez-Stable, E. J. (2000). Social Networks and Cancer Screening in Four US Hispanic Groups. *American Journal of Preventive Medicine*, 19(1), 47-52.

Sullivan, K., Coletti, M., & Luke, S. (2010). *GeoMason: Geospatial Support for MASON*. Department of Computer Science, George Mason University.

Sullivan, M. C., Roman, S. A., & Sosa, J. A. (2012). Does Chemotherapy Prior to Emergency Surgery Affect Patient Outcomes? Examination of 1912 Patients. *Annals of Surgical Oncology*, 19(1), 11-18.

Sun, X., Zhang, L., Tan, H., Bao, J., Strouthos, C., & Zhou, X. (2012). Multi-Scale Agent-Based Brain Cancer Modeling and Prediction of TKI Treatment Response: Incorporating EGFR Signaling Pathway and Angiogenesis. *BMC Bioinformatics*, 13(1), 218.

Surveillance, Epidemiology, and End Results (SEER) Program (2016) (www.seer.cancer.gov) *SEER Stat Fact Sheets: Colon and Rectum Cancer*.

Surveillance, Epidemiology, and End Results (SEER) Program (2015) (www.seer.cancer.gov) *SEER 18 2008-2012*, All Races, Both Sexes. Retrieved from: http://seer.cancer.gov/statfacts/html/colorect.html

Surveillance, Epidemiology, and End Results (SEER) Program (2014) (www.seer.cancer.gov) *Research Data (1973-2012)*, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch.

Sweeney, C., Schmitz, K. H., Lazovich, D., Virnig, B. A., Wallace, R. B., & Folsom, A. R. (2006). Functional Limitations in Elderly Female Cancer Survivors. *Journal of the National Cancer Institute*, 98(8), 521-529.

Szabó, G., Alava, M., & Kertész, J. (2003). Structural Transitions in Scale-Free Networks. *Physical Review* E, 67(5), 056102.

Szeto, K. L., & Devlin, N. J. (1996). The Cost-Effectiveness of Mammography Screening: Evidence from a Microsimulation Model for New Zealand. *Health Policy*, 38(2), 101-115.

Talley, N. J., & Jones, M. (1998). Self-Reported Rectal Bleeding in a United States Community: Prevalence, Risk Factors, and Health Care Seeking. *The American Journal of Gastroenterology*, 93(11), 2179-2183.

Tarlov, A. R. (1999). Public Policy Frameworks for Improving Population Health. *Annals of the New York Academy of Sciences,* 896(1), 281-293.

Thomas, W., White, C. M., Mah, J., Geisser, M. S., Church, T. R., & Mandel, J. S. (1995). Longitudinal Compliance with Annual Screening for Fecal Occult Blood. *American Journal of Epidemiology*, 142(2), 176-182.

Toma, J., Paszat, L. F., Gunraj, N., & Rabeneck, L. (2008). Rates of New or Missed Colorectal Cancer After Barium Enema and Their Risk Factors: a Population-Based Study. *The American Journal of Gastroenterology*, 103(12), 3142-3148.

Tomlinson, C., Wong, C., Au, H. J., & Schiller, D. (2012). Factors Associated with Delays to Medical Assessment and Diagnosis for Patients with Colorectal Cancer. Canadian Family Physician, 58(9), E495-E501.

Torrens, P., Li, X., & Griffin, W. A. (2011). Building Agent‐Based Walking Models by Machine‐Learning On Diverse Databases of Space‐Time Trajectory Samples. *Transactions in GIS*, 15(S1), 67-94.

Troitzsch, K. G. (2004). Validating Simulation Models. In *18th European Simulation MultiConference. Networked Simulations and Simulation Networks* (pp. 265-270).

United Nations (UN). (1988). Department of International Economic and Social Affairs. Social and Economic Differentials in Mortality in Developed Countries. In: *World Population Trends and Policies: 1987 Monitoring Report.* New York: United Nations, 1988.

U. S. Department of Health & Human Services (DHHS). (1996). *Physical Activity and Health: a Report of the Surgeon General*. DIANE Publishing.

U. S. Department of Labor (USDL). (2015). *Bureau of Labor Statistics*. American Time Use Survey (ATUS).

U. S. Preventive Services Task Force (USPSTF). (2002). Screening for Colorectal Cancer: Recommendation and Rationale. *Annals of Internal Medicine*, 137(2), 129.

U. S. Census Bureau. (2000). Census 2000: *District of Columbia Profile*. Retrieved from: http://www2.census.gov/geo/maps/special/profile2k/dc_2k_profile.pdf

U. S. Census Bureau. (2000). *American Factfinder Fact Sheet: District of Columbia* Retrieved from: http://factfinder.census.gov/

U. S. Census Bureau. (2000). T*iger Mapping Service: Washington, D.C*. Retrieved from: http://tiger.census.gov/.

U. S. Cancer Statistics Working Group (USCSWG). (2015). *United States Cancer Statistics: 1999–2012 Incidence and Mortality Web-Based Report*. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2015. Retrieved from: https://nccd.cdc.gov/dcpc_inca/.

Valente, T. W. (2010). *Social Networks and Health: Models, Methods, and Applications*. Oxford University Press.

Valente, T. W. (2012). Network Interventions. *Science*, 337(6090), 49-53.

Vaira, M., Cioppa, T., D'Amico, S., De Marco, G., d'Alessandro, M., Fiorentini, G., & De Simone, M. (2010). Treatment of Peritoneal Carcinomatosis from Colonic Cancer by Cytoreduction, Peritonectomy and Hyperthermic Intraperitoneal Chemotherapy (HIPEC). Experience of Ten Years. *In Vivo*, 24(1), 79-84.

Van Asselt, M. B., & Rotmans, J. (2002). Uncertainty in Integrated Assessment Modelling. *Climatic Change*, 54(1-2), 75-105.

Verbrugge, L. M. (1979). Marital Status and Health. *Journal of Marriage and the Family*, 267-285.

Verburg, P. H., Kok, K., Pontius Jr, R. G., & Veldkamp, A. (2006). Modeling Land-Use and Land-Cover Change. In *Land-Use and Land-Cover Change* (pp. 117-135). Springer Berlin Heidelberg.

Vogelstein, B., & Kinzler, K. W. (1993). The Multistep Nature of Cancer. *Trends in Genetics*, 9(4), 138-141.

Vrinten, C., Waller, J., Von Wagner, C., & Wardle, J. (2015). Cancer Fear: Facilitator and Deterrent to Participation in Colorectal Cancer Screening. *Cancer Epidemiology Biomarkers & Prevention*, 24(2), 400-405.

Wahls, T. L., & Peleg, I. (2009). Patient-and System-Related Barriers for the Earlier Diagnosis of Colorectal Cancer. *BMC Family Practice*, 10(1), 65.

Walter, F., Webster, A., Scott, S., & Emery, J. (2012). The Andersen Model of Total Patient Delay: a Systematic Review of Its Application in Cancer Diagnosis. *Journal of Health Services Research & Policy*, 17(2), 110-118.

Wang, Y., Xue, H., Chen, H. J., & Igusa, T. (2014). Examining Social Norm Impacts On Obesity and Eating Behaviors Among US School Children Based On Agent-Based Model. *BMC Public Health*, 14(1), 923.

Ward, E., Jemal, A., Cokkinides, V., Singh, G. K., Cardinez, C., Ghafoor, A., & Thun, M. (2004). Cancer Disparities by Race/Ethnicity and Socioeconomic Status. *CA: A Cancer Journal for Clinicians*, 54(2), 78-93.

Ward, E., Halpern, M., Schrag, N., Cokkinides, V., Desantis, C., Bandi, P., Siegel, R., Stewart, A., Jemal, A. (2008). Association of Insurance with Cancer Care Utilization and Outcomes. *CA: A Cancer Journal for Clinicians*, 58(1), 9-31.

Watts, D. J. (1999). *Small Worlds: the Dynamics of Networks Between Order and Randomness*. Princeton University Press.

Watts, D. J. (2004). *Six Degrees: the Science of a Connected Age*. WW Norton & Company.

Watts, D. J. (2013). Computational Social Science: Exciting Progress and Future Directions. *The Bridge On Frontiers of Engineering*, 43(4), 5-10.

Waye, J. D., Lewis, B. S., & Yessayan, S. (1992). Colonoscopy: a Prospective Report of Complications. *Journal of Clinical Gastroenterology*,15(4), 347-351.

Weaver, K. E., Rowland, J. H., Bellizzi, K. M., & Aziz, N. M. (2010). Forgoing Medical Care Because of Cost. *Cancer*, 116(14), 3493-3504.

Wee, C. C., Mccarthy, E. P., & Phillips, R. S. (2005). Factors Associated with Colon Cancer Screening: the Role of Patient Factors and Physician Counseling. *Preventive Medicine*, 41(1), 23-29.

Wei, R., Anderson, R. N., Curtin, L. R., & Arias, E. (2012). *US Decennial Life Tables for 1999-2001: State Life Tables*. National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System, 60(9), 1-66.

Weller, D., Vedsted, P., Rubin, G., Walter, F. M., Emery, J., Scott, S., Campbell, C., Andersen, R.S., Hamilton, W., Olesen, F., Rose, P. (2012). The Aarhus Statement: Improving Design and Reporting of Studies On Early Cancer Diagnosis. *British Journal of Cancer*, 106(7), 1262-1267.

Wellman, B., & Wortley, S. (1990). Different Strokes from Different Folks: Community Ties and Social Support. *American Journal of Sociology*, 558-588.

Weisz, G. (Ed.). (1998). *Greater Than the Parts: Holism in Biomedicine*, 1920-1950. Oxford University Press.

White, A. J., Teitelbaum, S. L., Wolff, M. S., Stellman, S. D., Neugut, A. I., & Gammon, M. D. (2013). Exposure to Fogger Trucks and Breast Cancer Incidence in the Long Island Breast Cancer Study Project: a Case–Control Study. *Environmental Health*, 12(1), 24.

Whitson, H. E., Hastings, S. N., Landerman, L. R., Fillenbaum, G. G., Cohen, H. J., & Johnson, K. S. (2011). Black–White Disparity in Disability: the Role of Medical Conditions. *Journal of the American Geriatrics Society*, 59(5), 844-850.

Wilkinson, R. G., & Marmot, M. G. (2003). *Social Determinants of Health: the Solid Facts*. World Health Organization.

Willems, S., De Maesschalck, S., Deveugele, M., Derese, A., & De Maeseneer, J. (2005). Socio-Economic Status of the Patient and Doctor–Patient Communication: Does It Make a Difference? *Patient Education and Counseling*, 56(2), 139-146.

Wilper, A. P., Woolhandler, S., Lasser, K. E., Mccormick, D., Bor, D. H., & Himmelstein, D. U. (2009). Health Insurance and Mortality in US Adults. *American Journal of Public Health*, 99(12), 2289-2295.

Winawer, S. J., Flehinger, B. J., Schottenfeld, D., & Miller, D. G. (1993). Screening for Colorectal Cancer with Fecal Occult Blood Testing and Sigmoidoscopy. *Journal of the National Cancer Institute*, 85(16), 1311-1318.

Winawer, S. J., Zauber, A. G., Ho, M. N., O'Brien, M. J., Gottlieb, L. S., Sternberg, S. S., Waye, J. D., Schapiro, M., Bond, J. H., Panish, J. F. and Ackroyd, F. (1993). Prevention of Colorectal Cancer by Colonoscopic Polypectomy. *New England Journal of Medicine*, 329(27), 1977-1981.

Winawer, S. J., Fletcher, R. H., Miller, L. A. U. R. A., Godlee, F., Stolar, M. H., Mulrow, C. D., Woolf, S. H., Glick, S. N., Ganiats, T. G., Bond, J. H., Rosen, L. (1997). Colorectal Cancer Screening: Clinical Guidelines and Rationale. *Gastroenterology*, 112(2), 594-642.

Winawer, S., Classen, M., Lambert, R., Fried, M., Dite, P., Goh, K.L., Guarner, F., Lieberman, D., Eliakim, R., Levin, B. and Saenz, R., 2008. Colorectal Cancer Screening: World Gastroenterology Organisation/International Digestive Cancer Alliance Practice Guidelines. *South African Gastroenterology Review*, 6(1), pp.13-20.

Winawer, S. J., & Zauber, A. G. (2002). The Advanced Adenoma as the Primary Target of Screening. *Gastrointestinal Endoscopy Clinics of North America*, 12(1), 1-9.

Winker, P., Gilli, M., & Jeleskovic, V. (2007). An Objective Function for Simulation Based Inference On Exchange Rate Data. *Journal of Economic Interaction and Coordination*, 2(2), 125-145.

Wise, S. (2014). *Using Social Media Content to Inform Agent-Based Models for Humanitarian Crisis Response* (Doctoral Dissertation, George Mason University).

Wu, H., Dai, Q., Shrubsole, M. J., Ness, R. M., Schlundt, D., Smalley, W. E., Chen, H., Li, M., Shyr, Y., Zheng, W. (2009). Fruit and Vegetable Intakes Are Associated with Lower Risk of Colorectal Adenomas. *The Journal of Nutrition*, 139(2), 340-344.

Wolfram, S. (1984). Cellular Automata as Models of Complexity. *Nature*, 311(5985), 419-424.

World Health Organization. (2014). *Global Status Report On Alcohol and Health*.

World Health Organization. (2008). *Social Determinants of Health. Key Concepts*. World Health Organization. Retrieved from: http://www.who.int/social_determinants/thecommission/finalreport/key_concepts/en/.

Zhang, L., Wang, Z., Sagotsky, J. A., & Deisboeck, T. S. (2009). Multiscale Agent-Based Cancer Modeling. *Journal of Mathematical Biology*, 58(4-5), 545-559.

Zapka, J. G., Taplin, S. H., Solberg, L. I., & Manos, M. M. (2003). A Framework for Improving the Quality of Cancer Care the Case of Breast and Cervical Cancer Screening. *Cancer Epidemiology Biomarkers & Prevention*, 12(1), 4-13.

BIOGRAPHY

Cristina Metgher is originally from Vorniceni (Strășeni), Republic of Moldova. She received her Bachelor of Science degree from the faculty of Computers, Informatics and Microelectronics of the Technical University of Moldova in 2008, followed by a Master in Science degree in Applied Informatics from the same university in 2010. Between 2008 and 2010 she worked in industry as an engineer. Cristina entered the PhD program in Computational Social Science at George Mason University in 2010, and was awarded a fellowship from the Open Society Institute that covered most tuition and living costs.

During her graduate studies, Cristina received several awards, such as a travel and tuition Scholarship Award from the University of Washington for attending the Network Modeling for Epidemics summer school, a travel and tuition award for attending the "Agent-based and System Dynamics Models: New Tools for Understanding Obesity" workshop from Johns Hopkins University, George Mason University Travel Fund Award (in 2012 and 2014), travel and tuition award for attending 2011 Santa Fe Institute's Complex Systems Summer School - and more recently – a GMU Provost Dissertation Completion Grant.

Apart from attending and presenting at numerous conferences, Cristina served as a peer reviewer for many academic journals, and since 2015 is the Peer Review Editor for the Journal of Mason Graduate Research (JMGR). She is fluent in Romanian (native), Russian, and English.

Cristina Metgher is a candidate for the PhD degree in Computational Social Science from George Mason University in May 2016.