An Investigation of the Transmission Dynamics of *M. tuberculosis*

Emilia Vynnycky

Department of Epidemiology and Population Sciences

London School of Hygiene & Tropical Medicine

Thesis submitted for the degree of Doctor of Philosophy

Faculty of Medicine, London University

May, 1996

Abstract

We develop an age-structured deterministic model (TBDYN3) of the transmission dynamics of *M. tuberculosis* in England and Wales since 1900. The model estimates the age-specific risks of developing 'primary', 'endogenous' and 'exogenous' disease and analyzes the incubation period, serial interval, lifetime risks and basic and net reproduction numbers for tuberculosis.

Best-fitting disease risk estimates predict peaks in both morbidity and mortality rates among young adults for successive birth cohorts during the pre-chemotherapy era, which fit observed patterns. Model predictions attribute much of the adult morbidity until the 1940s to exogenous reinfection. Chemotherapy, introduced after 1950, accelerated the decline in the risk of infection and the overall morbidity for all age groups.

Age and calendar year at infection are shown to determine the lifetime risks of developing tuberculosis, the incubation period and the serial interval. Twenty year olds faced higher lifetime risks of respiratory tuberculosis following infection (e.g. 26% and 17% in 1900 and 1950 respectively) than did any other age group. Estimates suggest that about 3% and over 25% of diseased 20 year olds in 1900 experienced initial infections in 1900 and during infancy respectively. Analogous estimates are provided for other age groups and calendar years.

Although tuberculous incidence and mortality declined since the 19th century, the estimated net reproduction number slightly exceeded one from 1900 until 1930 and declined thereafter. We discuss the implications of this paradox for diseases for which epidemiological parameters can change appreciably during a serial interval. Inferences about the basic reproduction number for tuberculosis are drawn by simulating the introduction of a sputum-positive case into an uninfected population identical in demography to that in England and Wales since 1900. A 'founder' case would have led to about 3 secondary cases in 1900, to about 2 by 1950 and to less than 1 after 1960.

Contents

Abstract 2

Acknowledgements 20

Introduction 21

1 Background 25

1.1 The natural history of tuberculosis 25

1.2 Review of related work 39

1.2.1 Studies of the risks of developing disease 39

1.2.2 Studies of the transmission of *M. tuberculosis* 49

1.3 Models for tuberculosis 55

1.3.1 Transmission models for tuberculosis 56

1.3.2 Models for parameter estimation 63

1.3.3 Conclusions 67

2 Epidemiology of tuberculosis in England and Wales — data sources and trends 68

2.1 Mortality rates 68

2.2 Notifications of Tuberculosis 74

2.2.1 Data sources and case definitions 74

2.2.2 Trends in notifications 75

2.2.3 Discussion 81

2.3 Annual risk of infection 84

2.3.1 Pre-chemotherapy era 84

2.3.2 Chemotherapy era — after 1950 87

2.4 BCG vaccination data 90

3 Development and evaluation of a model for tuberculosis — TBDYN3 94

3.1 Formulation of TBDYN3 94

3.1.1 Definitions of 'primary' and 'post-primary' disease 94

3.1.2 General structure of TBDYN3 96

3.1.3 Transitions between disease states 98

3.1.4 Incidence of disease 104

3.2 Methods and input data 106

3.2.1 Input data 106

3.2.2 Estimating the age-specific risks of developing disease 107

3.2.3 Sensitivity analyses of parameter estimates 109

3.2.4 Application to Dutch data 110

3.3 Estimates of the risks of developing disease 112

3.3.1 Results 112

3.3.2 Sensitivity analyses 114

3.3.3 Effect of simplifications on the disease risk estimates obtained 117

3.3.4 Disease risk estimates for the Netherlands 124

3.4 Discussion and implications for past trends in the age-specific incidence of disease 129

3.4.1 Discussion 129

3.4.2 Implications for age-specific patterns in the disease incidence predicted by TBDYN3 132

4 The incubation period, serial interval and lifetime risks of developing tuberculosis 141

4.1 Definitions and relevance 141

4.2 Methods 144

4.2.1 Definition of a 'causal infection/reinfection' event 144

4.2.2 Retrospective analyses of the incubation period/serial interval 145

4.2.3 Prospective analyses of the incubation period and serial interval . . 148

4.3 Results 152

4.3.1 Retrospective analyses of the incubation period/serial interval 152

4.3.2 Prospective analyses of the incubation period and serial interval . . 162
4.3.3 Effects of simplifications on the distributions obtained 180
4.4 Discussion 182
4.5 Brief review of key findings 186

5 The basic and net reproduction numbers for tuberculosis 187

5.1 Introduction 187
5.1.1 Relevance and definitions of the net and basic reproduction numbers for infectious diseases 187
5.1.2 Definitions of the basic and net reproduction numbers for tuberculosis 190
5.2 Method 193
5.2.1 Derivation of an effective contact number 193
5.2.2 Analyses of the basic and 'founder case' reproduction numbers . . . 196
5.2.3 Analyses of the net reproduction number 200
5.3 Results 203
5.3.1 Analyses of the basic and founder case reproduction numbers 203
5.3.2 The net reproduction number 218
5.3.3 Effect of simplifying assumptions on the results obtained 226
5.4 Discussion 227

6 Discussion and conclusions 232

6.1 Insights obtained 233
6.1.1 The natural history of tuberculosis 233
6.1.2 The incubation period, serial interval and lifetime risks of developing tuberculosis 236
6.1.3 The basic and net reproduction numbers 238
6.1.4 Concluding remark 243
6.2 Directions for future work 243

Summary 246

A Formulation and results from TBDYN3 250

A.1 PDEs describing the model formulation for TBDYN3 250

A.2 Approximations for the partial derivatives of the age and time-specific disease incidence with respect to fitted parameter estimates 252
A.3 Methods used for sensitivity analyses of disease risk estimates 255
A.4 Results from applying TBDYN3 to Dutch data 257

B Methods and results for the incubation period and serial interval for tuberculosis 260

B.1 PDEs illustrating the methods used to analyze the incubation period/serial interval using TBDYN3 260
B.2 Further results from analyses of the incubation period and serial interval . . 262
B.2.1 Retrospective analyses 262
B.2.2 Prospective analyses 266

C The basic and net reproduction numbers for tuberculosis 278

C.1 Relationship between the basic and net reproduction numbers for tuberculosis under hypothetical situations 278
C.2 The relationship between $\tilde{R}_0(t)$ and $\tilde{R}_n(t)$ for tuberculosis 280

Bibliography 282

List of Tables

1.1 Relative contribution of smear and culture-positive disease to notifications of pulmonary tuberculosis in Saskatchewan during the period 1960–1973 . . . 35

1.2 The ten-year incidence of tuberculosis following 'conversion' from tuberculin negativity (0–4mm to 100TU) to positivity (5^{+}mm to 3 TU), according to age at 'conversion', as found by the UK MRC BCG trial during the 1950s . 41

1.3 Results from the US Public Health Service chemoprophylaxis trial during the 1950s 42

1.4 Summary of the age-specific cumulative risks of developing pulmonary tuberculosis following conversion to tuberculin-positivity, as found by the follow-up study of patients during the period 1929–1944 through the Oslo Public Health Services 44

2.1 Summary of notification rates of all forms, respiratory and non-respiratory forms of tuberculosis in the general population in England and Wales during the period 1965–1988, together with the relative contribution from the ISC population 79

2.2 Crude annual notification rates (per 100,000) of all forms of tuberculosis in the white, Indian and Pakistani ethnic groups found during the national surveys of notifications between 1965 and 1988 in England and Wales . . . 80

2.3 Average annual decline in the notifications of all forms of tuberculosis by age in white males in England and Wales between 1953 and 1988, as estimated by Springett *et al* 81

2.4 Summary of the main events in the history of BCG vaccination in England and Wales 91

2.5 Summary of BCG vaccination policy in 201 districts in England and Wales in 1984 92

3.1 Definitions of class variables used in TBDYN3 99
3.2 Definitions of transition variables used in TBDYN3 100
3.3 Best-estimates of the risks of developing disease, as derived using TBDYN3 by fitting to notifications in white ethnic males in England and Wales . . . 112
3.4 Summary of the disease risk estimates obtained by repeating the fitting procedure on 50 simulated notification data sets, derived by sampling the numbers of notifications from the national tuberculosis surveys from the Poisson distribution 115
3.5 Comparison between the best-fitting values for the risks of developing the first primary episode and endogenous disease derived assuming individuals cannot be reinfected, and the values obtained by including reinfection (original TBDYN3 model) 117
3.6 Comparison between the best-fitting values for the risks of developing the first primary episode and endogenous disease derived assuming vaccination confers no protection against infection, and the values obtained by assuming 77% of individuals are protected from infection for life (original TBDYN3 model) 123
3.7 Comparison between the risks of developing the first primary episode, endogenous and exogenous disease derived with TBDYN3 for the Netherlands, and the corresponding risks derived by Sutherland *et al* 125

4.1 Follow-up population and disease categories used to analyze the incubation period and serial interval for tuberculosis prospectively 151

List of Figures

1.1 Flowchart showing the natural history of tuberculosis 27

1.2 Age-specific mortality rates from respiratory tuberculosis in males in England and Wales during the period 1881–1950, plotted by year and birth cohort . 30

1.3 Distribution of time interval between 'conversion' and clinical onset of tuberculosis for 243 individuals who developed disease, as found in the UK MRC BCG trial during 1950s 31

1.4 Age-specific notifications of pulmonary and 'infectious' (sputum-positive) forms of tuberculosis in males in Norway during the period 1951–1969 . . . 33

1.5 Relative contribution of sputum-positive disease to age-specific notifications of pulmonary tuberculosis in males in Norway during the period 1951–1969 34

1.6 Cumulative risks of developing tuberculosis for 733 children of white sputum-positive cases, according to age at first 'exposure', as found by the Williamson County tuberculosis study during the period 1931–1954 46

1.7 Schematic diagram showing the structure of simple models of the transmission dynamics for *M. tuberculosis* 55

1.8 Schematic diagram showing the structure of the model produced by Chorba and Sanders in 1971 58

1.9 Schematic diagram showing the structure of the model produced by Joesoef in 1989 59

1.10 Schematic diagram showing the structure of the model produced by Brogger in 1967 61

1.11 Schematic diagram showing the structure of the age-structured model produced by Waaler in 1968 62

2.1 Mortality rates from respiratory tuberculosis in males in England and Wales during the period 1847–1990 69
2.2 Age-specific mortality rates from respiratory tuberculosis in males in England and Wales during the period 1850–1990 71
2.3 Age-specific mortality rates from respiratory tuberculosis in males in England and Wales during the period 1850–1990, plotted on a *log* scale 72
2.4 Age-specific mortality rates from respiratory tuberculosis in males in England and Wales during the period 1881–1950, plotted by year and birth cohort . 73
2.5 Crude annual mortality and notification rates (non-age standardized) of respiratory tuberculosis in England and Wales between 1901 and 1990 76
2.6 Published age-specific notification and mortality rates of respiratory tuberculosis in males in England and Wales during the period 1930–1989 77
2.7 Comparison between age-specific notification rates of respiratory tuberculosis in males in the general population in England and Wales and those estimated for the white ethnic group 82
2.8 Age-specific prevalence of tuberculin-positive males in England and Wales, as found by the national tuberculin survey during the period 1949–50 . . . 85
2.9 Mortality rates from tuberculous meningitis in 0–4 year old males in England and Wales during the period 1860–1957 86
2.10 Comparison between mortality rates of tuberculous meningitis in 0–4 year old males in England and Wales until 1957 and the corresponding notification rates from 1954 88
2.11 Prevalence of tuberculin sensitivity among school children and students tested as part of the national BCG vaccination scheme 93

3.1 Schematic diagram showing the overall structure of the TBDYN3 model . . 97
3.2 Approximate risks of developing the first primary disease episode, relative to those experienced during the first year after infection, estimated from the distribution of the time interval between 'conversion' and disease onset of those who were tuberculin-negative at the start of the UK MRC BCG trial 101
3.3 Relationship between the risk of developing the first primary disease episode and the age at infection assumed in TBDYN3 102

3.4 Summary of the relative contribution of sputum-positive disease to age-specific notifications of pulmonary tuberculosis in males in Norway during the period 1951–1969, together with the corresponding values assumed by TBDYN3 104
3.5 Estimated age-specific notification rates of respiratory tuberculosis in white ethnic males used to derive the risks of developing disease by TBDYN3 . . 108
3.6 Comparison between the incidence of respiratory tuberculosis in 0–74 year old males estimated using TBDYN3 for England and Wales, and the notifications used in the fitting process 113
3.7 Comparison between the incidence of respiratory tuberculosis in 0–74 year old males, derived using the best-fitting parameters assuming individuals could not be reinfected and experienced *constant* risks of developing the first primary and endogenous disease after age 20 years, and the notifications used in the fitting process 118
3.8 Comparison between the expected age-specific prevalence of infection in England and Wales in 1949, as used in TBDYN3 and the prevalence of tuberculin sensitivity in south urban, north urban and rural England, as found during the national tuberculin survey during the period 1949–1950 120
3.9 Comparison between the incidence of respiratory tuberculosis in 0–74 year old males estimated using TBDYN3 for England and Wales, assuming vaccination conferred no protection against infection, and the notifications used in the fitting process 122
3.10 Comparison between the estimated annual risk of infection in the Netherlands and that in England and Wales during the period 1901–1990 126
3.11 Comparison between notification rates of respiratory tuberculosis in males in the Netherlands during the period 1951–1970 and those estimated from survey data among white ethnic males in England and Wales 127
3.12 Comparison between tuberculous meningitis and respiratory tuberculosis mortality rates among 0–4 year old males in England and Wales during the period 1860–1957 130

3.13 Comparison between tuberculous meningitis and respiratory tuberculosis mortality rates among 0–4 year old males in England and Wales until 1957, and the corresponding notification rates in the general population from 1953 . . 131
3.14 Estimated age-specific incidence of respiratory tuberculosis since 1900 in England and Wales attributable to first primary episodes, exogenous and endogenous disease, as predicted using TBDYN3 using the best-fitting parameter values 133
3.15 Estimated age-specific incidence of respiratory tuberculosis since 1950 in England and Wales attributable to first primary episodes, exogenous and endogenous disease, as predicted using TBDYN3 using the best-fitting parameter values 134
3.16 Comparison between the incidence of respiratory tuberculosis attributable to first primary episodes, exogenous and endogenous disease for males during the period 1901–1950 in England and Wales, as predicted using TBDYN3 and the corresponding male mortality rates 136
3.17 Comparison between the mortality rates and incidence rates of respiratory and sputum-positive tuberculosis for given years during the period 1901–1950, as predicted using TBDYN3, and the actual mortality rates from respiratory tuberculosis in the general population 137
3.18 Comparison between the incidence of respiratory tuberculosis attributable to first primary episodes, exogenous and endogenous disease for males during the period 1955–1980 in England and Wales, as predicted using TBDYN3, and the corresponding notification rates for white ethnic males 138
3.19 Comparison between the mortality rates among male cohorts born since 1880 in England and Wales and the corresponding incidence and mortality rates from respiratory tuberculosis among cohorts, as predicted using TBDYN3 using the best-fitting parameter values 139

4.1 Distribution of the time interval *since first infection* of individuals who developed respiratory tuberculosis when aged 20 years during the period 1900–1990, as estimated using TBDYN3 153

4.2 Distribution of the time interval since first infection of *individuals who developed respiratory tuberculosis when aged 20 years* during the period 1900–1990, as estimated using TBDYN3 155
4.3 Distribution of the time interval since first infection of *individuals who developed respiratory tuberculosis when aged 70 years* during the period 1900–1990, as estimated using TBDYN3 157
4.4 Distribution of the time interval *since the infection or reinfection event* causing the current (sputum-positive) episode among *individuals developing disease when aged 20 years* during the period 1900–1990, as estimated using TBDYN3 158
4.5 Distribution of the time interval *since the infection or reinfection event* causing the current (sputum-positive) episode among *individuals developing disease when aged 50 years* during the period 1900–1990, as estimated using TBDYN3 160
4.6 Distribution of the time interval *since the infection or reinfection event* causing the current (sputum-positive) episode among *individuals developing disease when aged 70 years* during the period 1900–1990, as estimated using TBDYN3 161
4.7 Lifetime risk of developing *respiratory* tuberculosis for individuals infected at different ages during 1900–1990 in England and Wales, as derived using TBDYN3 162
4.8 Comparison between the age-specific and overall lifetime risks (weighted by the number of individuals experiencing initial infection at each age) of developing *respiratory* tuberculosis for individuals infected during the period 1900–1990 in England and Wales, as derived using TBDYN3 165
4.9 Distribution of the time interval between first infection and the first episode of respiratory disease for individuals *infected in their first year of life* during the period 1900–1990, as estimated using TBDYN3 166
4.10 Distribution of the time interval between first infection and the first episode of respiratory disease for individuals *infected at different ages in 1900*, as estimated using TBDYN3 167

4.11 Lifetime risk of developing *sputum-positive* tuberculosis for individuals infected at different ages during the period 1900–1990 in England and Wales, as derived using TBDYN3 170
4.12 Comparison between the age-specific and overall lifetime risks (weighted by the number of individuals experiencing initial infection at each age) of developing *sputum-positive* tuberculosis for individuals infected during the period 1900–1990 in England and Wales, as derived using TBDYN3 171
4.13 Distribution of the time interval between first infection and the first episode of sputum-positive disease for individuals *infected in their first year of life* during the period 1900–1990, as estimated using TBDYN3 172
4.14 Distribution of the time interval between first infection and the first episode of sputum-positive disease for individuals *infected at different ages in 1900*, as estimated using TBDYN3 173
4.15 Lifetime risk of developing respiratory tuberculosis *attributable to initial infection* for individuals infected at different ages during 1900–1990 in England and Wales, as derived using TBDYN3 175
4.16 Lifetime risk of developing sputum-positive tuberculosis *attributable to initial infection* for individuals infected when aged 0–70 years during 1900–1990 in England and Wales, as derived using TBDYN3 176
4.17 Distribution of the time interval between first infection and the first episode of respiratory disease *attributable to the initial infection* for individuals infected in their first year of life during the period 1900–1990, as estimated using TBDYN3 178
4.18 Distribution of the time interval between first infection and the first episode of *sputum-positive disease attributable to the initial infection* for individuals infected in their first year of life during the period 1900–1990, as estimated using TBDYN3 179

5.1 Crude incidence and prevalence of sputum-positive tuberculosis per 100,000 males aged 0–100 years in England and Wales during the period 1901–1990, as estimated using TBDYN3 195

5.2 Annual effective contact number between sputum-positive and other individuals in the population in England and Wales during the period 1901–1990, as estimated using TBDYN3 195
5.3 Number of secondary sputum-positive cases resulting from a founder case who was infectious for one or two years, introduced into an uninfected population similar in structure to that in England and Wales during the period 1900–1990, as estimated using TBDYN3 203
5.4 Number of incident and prevalent sputum-positive cases following the introduction of one sputum-positive individual into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TBDYN3. This assumes that the sputum-positive case was infectious for two years 205
5.5 Total number of incident and prevalent sputum-positive cases following the introduction of one sputum-positive individual into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TBDYN3. This assumes that the sputum-positive case was infectious for two years 206
5.6 Predicted crude incidence rate of sputum-positive disease after the introduction of a founder case, who was infectious for either one or two years, into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TBDYN3 208
5.7 Predicted crude prevalence of infection after the introduction of a founder case into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TBDYN3. This assumes that the founder case was infectious for two years 209
5.8 Relative contribution of '1st primary', endogenous and exogenous disease to the predicted incidence (per 100,000 general male population) of sputum-positive tuberculosis after the introduction of one founder case into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TBDYN3. This assumes that the founder case was infectious for two years 210

5.9 Predicted annual risk of infection after the introduction of one sputum-positive individual into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TB-DYN3. This assumes that the sputum-positive case was infectious for two years 211

5.10 Number of secondary sputum-positive cases resulting from a founder case who was infectious for one or two years, introduced into an uninfected population similar in structure to that in England and Wales during the period 1900–1990, as estimated using TBDYN3, assuming that there was *no decline in the effective contact number* subsequent to his/her introduction 213

5.11 Comparison between the basic and founder case reproduction numbers for tuberculosis during the period 1900–1990, as estimated using TBDYN3 . . 214

5.12 Comparison between the annual risk of infection subsequent to the introduction of a founder case into a population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, assuming that the effective contact number did and did not subsequently decline 215

5.13 Relative contribution of '1st primary', endogenous and exogenous disease to the predicted incidence (per 100,000 general male population) of sputum-positive tuberculosis after the introduction of one founder case into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TBDYN3. This assumes that the founder case was infectious for two years, and that *there was no subsequent change in the case-fatality rate and the effective contact number* 216

5.14 Comparison between the estimated effective contact number and estimated number of individuals infected for the first time or reinfected in England and Wales during the period 1900–1990 219

5.15 Percentage composition (by age) of *uninfected* individuals in the population in England and Wales between 1900 and 1980, as estimated using TBDYN3. 220

5.16 Percentage composition (by age) of individuals in the '*latent*' class in the population in England and Wales between 1900 and 1980, as estimated using TBDYN3. 220

5.17 Percentage of individuals infected for the first time during the period 1900–1990 when aged 0–100 years *who develop sputum-positive disease attributable to their infection episode*, as estimated using TBDYN3. 222
5.18 Percentage of individuals reinfected during the period 1900–1990 when aged 0–100 years *who develop sputum-positive disease attributable to this reinfection episode*, as estimated using TBDYN3. 222
5.19 Estimates of the 'net reproduction number' for tuberculosis, as derived using TBDYN3 223
5.20 Comparison between the estimated net, basic and founder case reproduction numbers for tuberculosis, as derived using TBDYN3 224
5.21 Age distribution of the male population in England and Wales during the period 1900–1990 in England and Wales 229

A.1 Summary of the 95% range of the age-specific notification rates of respiratory tuberculosis during the survey years (for 0–54 year olds) and for all years (for 55–74 year olds) used in the 50 simulated data sets to analyze the sensitivity of the disease risk estimates 256
A.2 Comparison between the incidence of respiratory tuberculosis in 0–39 year old males estimated using TBDYN3 for the *Netherlands* and the notifications used in the fitting process 258
A.3 Comparison between the incidence of respiratory tuberculosis in 40–69 year old males estimated using TBDYN3 for the *Netherlands* and the notifications used in the fitting process 259

B.1 Distribution of the time interval *since first infection* of individuals who developed respiratory tuberculosis when aged 50 years during the period 1900–1990, as estimated using TBDYN3. 263
B.2 Distribution of the time interval *since first infection* of individuals who developed respiratory tuberculosis when aged 70 years during the period 1900–1990, as estimated using TBDYN3. 264
B.3 Distribution of the time interval since first infection of individuals who developed respiratory tuberculosis when aged 50 years during the period 1900–1990, as estimated using TBDYN3 265

B.4 Distribution of the time interval between first infection and the first episode of respiratory disease for individuals infected *at different ages in 1930*, as estimated using TBDYN3 266
B.5 Distribution of the time interval between first infection and the first episode of respiratory disease for individuals infected *at different ages in 1960*, as estimated using TBDYN3 267
B.6 Distribution of the time interval between first infection and the first episode of respiratory disease for individuals infected *at different ages in 1990*, as estimated using TBDYN3 268
B.7 Distribution of the time interval between first infection and the first episode of sputum-positive disease for individuals infected *at different ages in 1930*, as estimated using TBDYN3 269
B.8 Distribution of the time interval between first infection and the first episode of sputum-positive disease for individuals infected *at different ages in 1960*, as estimated using TBDYN3 270
B.9 Distribution of the time interval between first infection and the first episode of sputum-positive disease for individuals infected *at different ages in 1990*, as estimated using TBDYN3 271
B.10 Distribution of the time interval between first infection and the first episode of respiratory disease *attributable to initial infection* for individuals infected at different ages in 1900, as estimated using TBDYN3 272
B.11 Distribution of the time interval between first infection and the first episode of respiratory disease *attributable to initial infection* for individuals infected at different ages in 1960, as estimated using TBDYN3 273
B.12 Distribution of the time interval between first infection and the first episode of respiratory disease *attributable to initial infection* for individuals infected at different ages in 1900, as estimated using TBDYN3 274
B.13 Distribution of the time interval between first infection and the first episode of sputum-positive disease *attributable to initial infection* for individuals infected at different ages in 1900, as estimated using TBDYN3 275

B.14 Distribution of the time interval between first infection and the first episode of sputum-positive disease *attributable to initial infection* for individuals infected at different ages in 1930, as estimated using TBDYN3 276
B.15 Distribution of the time interval between first infection and the first episode of sputum-positive disease *attributable to the initial infection* for individuals infected at different ages in 1990, as estimated using TBDYN3 277

Acknowledgements

First, I would like thank my supervisor, Professor Paul Fine for his contribution to the development of this thesis. I have benefitted immensely from his supervision and his enthusiasm for the work is greatly appreciated. It has been a privilege and a pleasure to have had the opportunity learn from his enormous experience.

I also wish to thank the Medical Research Council for their financial support, without which this thesis would not have been possible.

Thanks are also due to Dr Styblo (International Union Against Tuberculosis), who provided data on the notifications of tuberculosis in the Netherlands. I am also grateful to Dr Bjartveit (Tuberculosis Screening Service, Norway) for supplying data on sputum-positive notifications from Norway, and to Dr Gunnar Bjune for his help in securing these data. I also appreciate the co-operation of the UK Government Actuary's Department in providing data on the age and time-specific mortality rates within cohorts born in England and Wales since 1841, and thank Dr John Watson (CDSC) for providing data from recent national tuberculosis surveys in England and Wales.

Discussions with a number of individuals have also helped in the development of this thesis. I am particularly grateful to Dr Ian Sutherland (MRC Biostatistics Unit, Cambridge) for his comments on several aspects of the work. I have also benefitted from discussions with Nigel Gay (CDSC), who provided some relevant insights into modelling, and wish to thank Jonathan Sterne for his help in some of the statistical aspects included in this thesis. Thanks are especially due to Daan Muulder, who through discussions, helped to shape some of the assumptions incorporated in the model. I am also grateful to Dr John Watson (CDSC), Chris Dye and Christian Lienhardt for their comments on the work described at the time of upgrading. Judith Glynn has also provided useful suggestions.

Thanks are also due to Michael Jones, who, as a fellow PhD student and long-suffering office companion, witnessed much of the development of this thesis, and provided a good source of answers to statistical queries.

Ultimately, I would like to thank my parents and brother for providing back-up support and encouragement throughout the development of this thesis.

Introduction

Tuberculosis has been declining in most developed countries at least since the beginning of this century. Since the late 1980s, the decline in notifications of tuberculosis has slowed down, and it has even reversed in several Western countries [1]. This has been attributed to various factors, including the emergence of HIV infection, changing patterns of immigration from high-risk populations, multi-drug resistant strains of *M. tuberculosis*, the increasing problem of homelessness, and changes in tuberculosis control programmes.

The basic natural history of tuberculosis is complicated even without these factors. Individuals can develop disease either soon after infection ('primary' disease) or many years thereafter ('post-primary' disease) either through 'endogenous' reactivation or after 'exogenous' reinfection. The risks of developing disease are age-dependent, and are higher in young adult life than among infants and young children, as indicated by birth cohort analyses of mortality rates from respiratory tuberculosis in developed countries from the pre-chemotherapy era [2–5]. The relative contribution of 'primary', 'endogenous' and 'exogenous' forms to the disease incidence in any population is therefore a function both of age and of the risk of infection, which has been declining in developed countries throughout this century [6].

Hitherto, no studies have attempted to quantify the *age-specific* risks of developing 'primary', 'endogenous' and 'exogenous' disease and the effect of the decline in the risk of infection on the relative contribution of these forms to the decline in disease incidence in given age groups. In fact, only one study, that of Sutherland *et al* [7], which focussed on the disease incidence among Dutch *adult* males during the period 1951–1970, has estimated the risk of developing disease subsequent to reinfection in a developed country.

The incubation period, serial interval and basic and net reproduction numbers of tuberculosis have also received little attention in the past, even though they are the most

important parameters underlying the dynamics of any infectious disease. The incubation period provides information on when, given infection, an individual is likely to develop disease, the serial interval indicates when that individual is likely to infect others, and the basic and net reproduction numbers describe the number of secondary infectious cases which will result from one infectious case under given circumstances.

The facts that the risks of developing disease are age-dependent, and that the risk of infection declined over time suggest that the incubation periods and serial intervals must differ by age group and by calendar year. Similar considerations apply to the overall lifetime risk of developing disease after infection, which is often stated to be 5–10% [8]. The fact that reinfection can occur further complicates the definitions of these parameters and of the basic and net reproduction numbers, since both infected and *reinfected* individuals who subsequently develop infectious pulmonary tuberculosis are secondary cases of some infectious source. The only published study of the basic reproduction number for tuberculosis [9] considered a hypothetical population without age-structure, and assumed that reinfection cannot occur. There has been no published study of the net reproduction number for tuberculosis.

Overall we see that the dynamics underlying the decline in tuberculosis in developed countries during the last century are complicated. Intuitively, these dynamics should be understood further before we assess the likely effect of other factors, such as the HIV epidemic, on the *current* tuberculosis situation. This has been the rationale for the work described in this thesis.

Disease dynamics are difficult to study using traditional epidemiological tools such as case control or cohort studies, and are more easily analyzed using theoretical modelling techniques. Such modelling studies in turn require reliable data on the long-term tuberculosis situation in a given population, which are not available for many countries. England and Wales are unique in this respect, as age and sex-specific mortality data have been published since 1847. Hence our analyses of the transmission dynamics of *M. tuberculosis* are restricted mainly to the tuberculosis situation in England and Wales during this century. The age-structured model developed (TBDYN3) focuses on the following aspects of tuberculosis epidemiology:

1. The basic natural history of tuberculosis. The aims are to:

(a) estimate the *age-specific* risks of developing 'primary', 'endogenous' and 'exogenous' disease, and

(b) assess the effect of the decline in the risk of infection on the relative contribution of these different disease forms to the age and time-specific disease incidence and mortality in England and Wales.

2. The incubation period, serial interval and lifetime risk of developing disease, with the aims of:

(a) assessing the definitions of the incubation period and serial interval for infectious diseases and discuss their relevance for tuberculosis and

(b) deriving estimates of these parameters according to age and calendar year.

3. The basic and net reproduction numbers. Our aims are to:

(a) assess their relevance for tuberculosis, and

(b) estimate these parameters in England and Wales since 1900, and to relate their magnitude to the decline observed in tuberculosis during the last century.

TBDYN3 extends the work of Sutherland *et al* [7], who are the only workers to have modelled the disease incidence in a developed country as a function of the different risks of developing 'primary', 'endogenous' and 'exogenous' disease. While acknowledging this foundation, the present work goes much further in being the first to consider the incubation period, serial interval and lifetime risks of developing disease, and the implications of the complex natural history of tuberculosis for the definitions of the basic and net reproduction numbers.

Chapter 1 reviews current knowledge of the natural history of tuberculosis, and studies of the risks of developing disease and the transmission of *M. tuberculosis* in outbreak situations. Section 1.3 reviews models of the transmission dynamics of *M. tuberculosis* published by other workers in the past. Sections 1.1 and 1.2.1 provide the basis for the assumptions incorporated into TBDYN3.

Chapter 2 describes the epidemiology of tuberculosis in England and Wales, together with the data used in TBDYN3.

Chapter 3 describes the development of TBDYN3, and estimates for the age-dependent risks of developing disease are presented in section 3.3.1. The implications of these estimates

for the relative contribution of exogenous and endogenous forms to past trends in disease incidence in England and Wales are presented in section 3.4.2.

The assumptions incorporated into TBDYN3 provide the basis for the analyses in Chapter 4 of the incubation period, serial interval and lifetime risks of developing tuberculosis. Section 4.1 assesses the implications of the complex natural history for the definitions and interpretation of the incubation period and serial interval in section 4.1, and distributions of these parameters and estimates of the lifetime risks of developing disease according to age and calendar year are derived in section 4.3.

The penultimate chapter reviews the definitions of the basic and net reproduction numbers as applied to infectious diseases and discusses their relevance for tuberculosis on the basis of estimates derived using TBDYN3. We conclude in Chapter 6 with a summary of the main insights obtained by this modelling work, and outline possible directions for future work.

Chapter 1

Background

1.1 The natural history of tuberculosis

Tuberculosis is defined as the disease attributable to infection with *Mycobacterium tuberculosis* (the vast majority of cases), *M. bovis* or *M. africanum*. The tubercle bacillus, *M. tuberculosis*, generally affects the lungs, although any other organ may become involved. On the basis of the site affected, tuberculosis is described as either 'pulmonary' (respiratory), or 'extrapulmonary'.

The tubercle bacillus is transmitted most commonly through coughing by an individual with infectious or 'open' pulmonary tuberculosis. *M. bovis* is typically transmitted from a tuberculous cow via contaminated milk. This has become rare in developed countries since the introduction of routine pasteurization of milk. Infection with *M. bovis* is associated classically with cervical lymphadenitis (scrofula) [10].

The implantation of tubercle bacilli into tissue constitutes the first stage in the natural history of tuberculosis. Bacilli are then disseminated via the lymphatics to the regional lymph nodes, which may lead to the development of a 'primary complex', comprising a lesion at the implantation site (usually in the lungs) and in the associated lymph nodes. Cell-mediated immunity to *M. tuberculosis* antigens develops within 2–6 weeks after first infection, and typically leads to the formation of granulomas and calcification around the primary lesions. These ultimately contain the infection, although it may take up to 3 years for the primary complex to resolve completely [10]. Until this occurs, bacilli may escape from these lesions into the blood to initiate foci elsewhere in the body e.g. in other parts of the lung, the kidneys etc. During the first three months after the development of cell-

mediated immunity, individuals are at greatest risk of developing serious systemic forms of tuberculosis, such as tuberculous meningitis and miliary tuberculosis [11]. These forms are particularly common in infants and young children. Lesions in the bones, joints and kidneys, on the other hand, may take up to three years to appear.

Once contained in the calcified lesions, the infection generally lies dormant until the individual's immune response weakens, leading to proliferation of bacilli and to clinical disease months or years after first infection. These calcified lesions are sometimes detectable by X-ray. As the sensitivity and specificity of X-rays are both considerably less than 100%, they do not provide conclusive evidence of infection with the tubercle bacillus.

The tuberculin skin test is conventionally used to detect tuberculous infection. The standard test is by the 'Mantoux' technique, whereby individuals are injected intradermally with tuberculin, a soluble protein derivative of the tubercle bacillus. An induration typically appears at the inoculated site after 48–72 hours in individuals whose T-cells have been sensitized to *M. tuberculosis* antigens. The distribution of induration sizes in the population under study to a given tuberculin test determines the definition of a positive result. This varies between populations, since individuals who have been exposed to certain environmental mycobacteria or who have received BCG vaccine may also be sensitive to tuberculin.

A further complication of the tuberculin test is the fact that repeated testing of uninfected individuals can itself induce sensitivity [12, 13]. In addition, individuals who have exhibited a reaction to tuberculin may lose their sensitivity with time or age [14]. Diagnosing a tuberculous infection is particularly problematical in individuals with HIV infection and AIDS. Such individuals lose their cellular immunity and hence their responsiveness to tuberculin. Problems regarding the use of the tuberculin test in determining the prevalence of infection in a population have recently been reviewed by Rieder [15].

Disease occurring less than 5 years after first infection is conventionally defined as "primary disease", whilst disease occurring thereafter is defined as "post-primary disease" [16]. These definitions are discussed in greater detail in section 3.1.1. Post-primary disease is believed to occur either through the reactivation of the earlier infection ("endogenous reactivation"), or as a consequence of reinfection with the tubercle bacillus ("exogenous reinfection"). On this basis, the basic natural history of tuberculosis can be summarized schematically as shown in Figure 1.1.

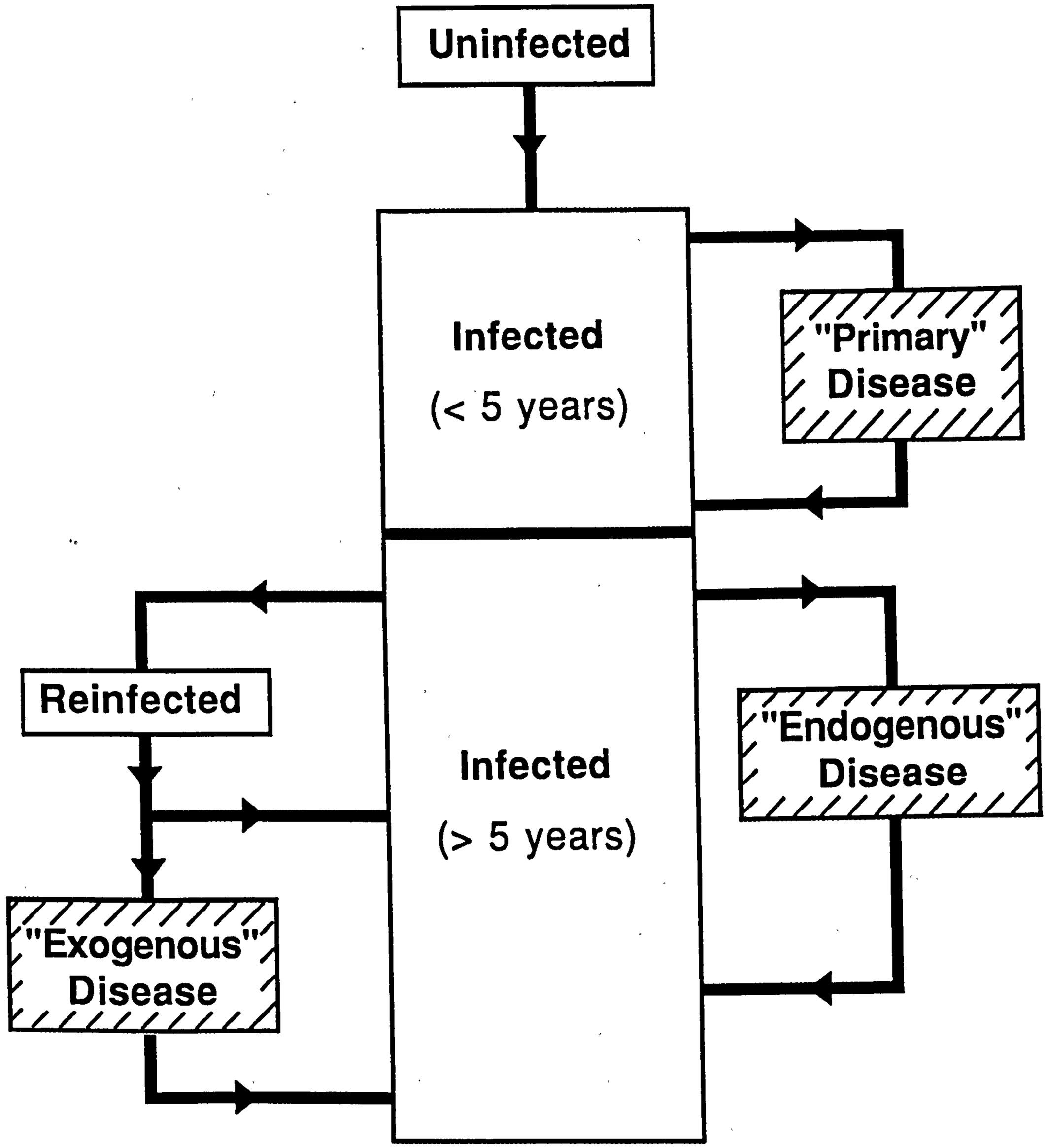


Figure 1.1: Flowchart showing the natural history of tuberculosis.

The relative frequency of disease through endogenous reactivation and exogenous reinfection has been disputed.

The earliest evidence for the role of exogenous disease comes from anatomical studies carried out during the period 1930–1950 of individuals who died from diseases other than tuberculosis [17]. These studies revealed what were considered to be post-primary lesions in 20–60% of adults and the number of such lesions increased with the age of the individual examined. Canetti [17] describes results from other studies, in which bacilli were taken from primary complexes at various stages in development and inoculated into guinea pigs. About 85% of the calcified lesions did not yield live bacilli. On such evidence, Canetti inferred that bacilli may not survive indefinitely in primary complexes and therefore that disease through endogenous reactivation was rare.

Further evidence for the role of exogenous disease is drawn from the tuberculosis situation in Alaska during the 1950s [18]. This was considered to be worse than that in any other region in the world, with an estimated risk of infection of about 25%, and as many as 95% of children tuberculin-positive (induration of 8mm or more to 1TU PPD-S) by age 13–14 years. After the introduction of chemoprophylaxis, the risk of infection declined together with the incidence of disease across all ages. This decline, especially among older individuals, is attributed in large part to a reduction in the incidence of exogenous disease [18].

With the advent of chemotherapy, evidence for the existence of exogenous disease has been provided by changes in drug-susceptibility patterns in patients with previously treated tuberculosis, as illustrated by individual case reports. For example, two sons developed disease with the same drug-resistance pattern as their mother. The sons had had prior treatment for tuberculosis with different drugs from the ones to which they were subsequently resistant [19], and hence could have been reinfected by their mother.

With the refinement of phage typing and, more recently, DNA fingerprinting, it has become possible to ascertain whether the strain of tubercle bacillus with which an individual is currently infected is identical to one with which they were infected many years previously i.e. whether the current episode is attributable to endogenous reactivation. In one survey of tuberculosis in veterans in the US Armed Forces in 1975, 9 out of 26 relapsed cases had different phage types from those causing their previous disease episode [20].

Phage typing and DNA fingerprinting have also been of particular use in verifying transmission in outbreak settings [21–24]. In a famous outbreak of 42 cases associated with a

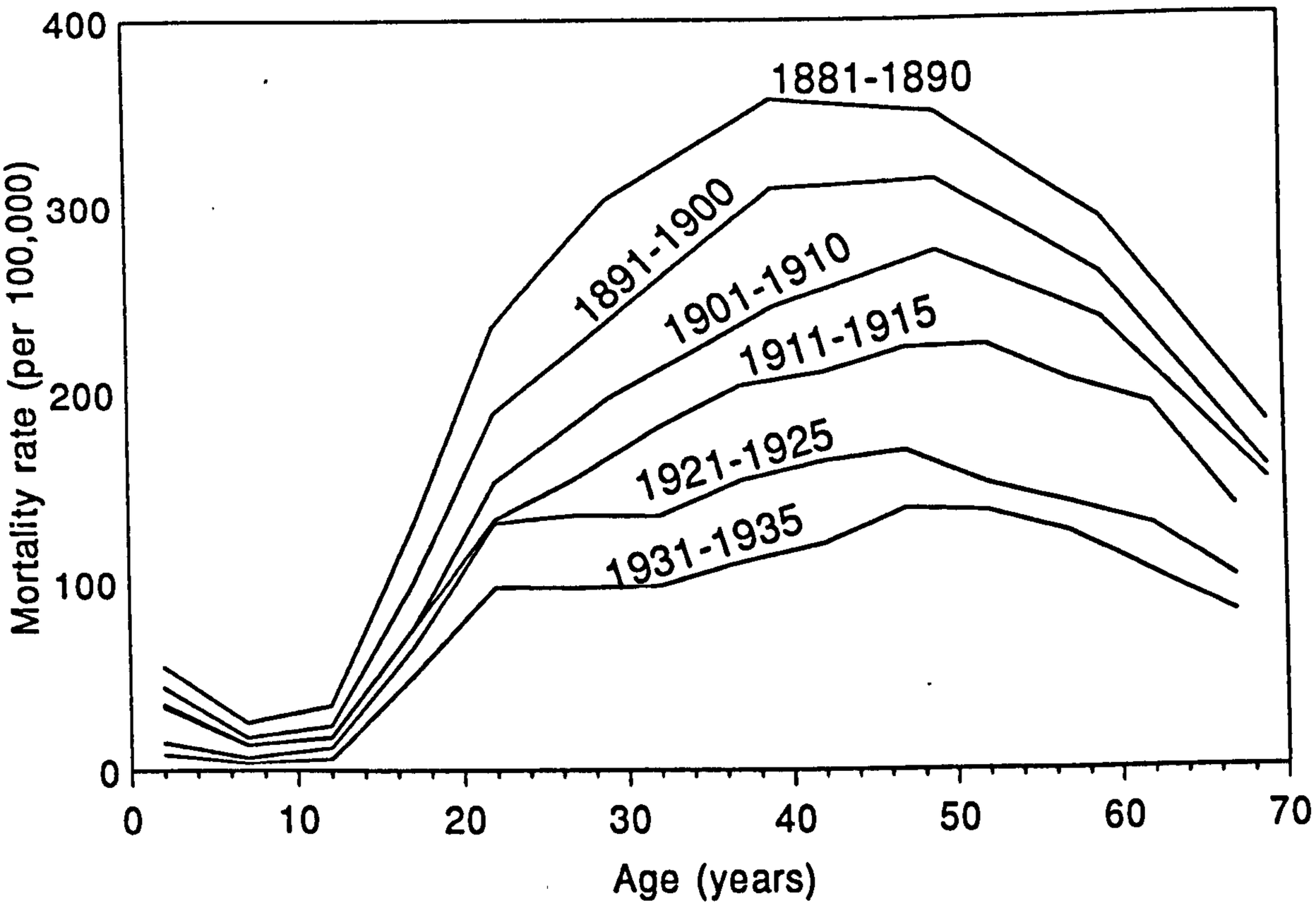
homeless shelter in Boston during the 1980s [21], for example, 22 cases were found to have identical phage types and drug resistance patterns. Moreover, the disease was attributable to exogenous reinfection in 7 of these cases, as indicated by previous tuberculin-positivity and a difference in drug susceptibility between the current and prior disease episode.

With the low risk of infection in developed countries today, it is likely that only a small proportion of disease is exogenous in origin. The situation in developing countries is unclear, however. Studies in Bangalore during the period 1961–1966 [25] and in Madras during the 1970s [26], for example, found much higher incidences of disease in 'infected' (i.e. tuberculin-positive), as compared with uninfected individuals. Given the high risks of infection in both areas, many of the 'infected' individuals could have experienced disease through exogenous reinfection.

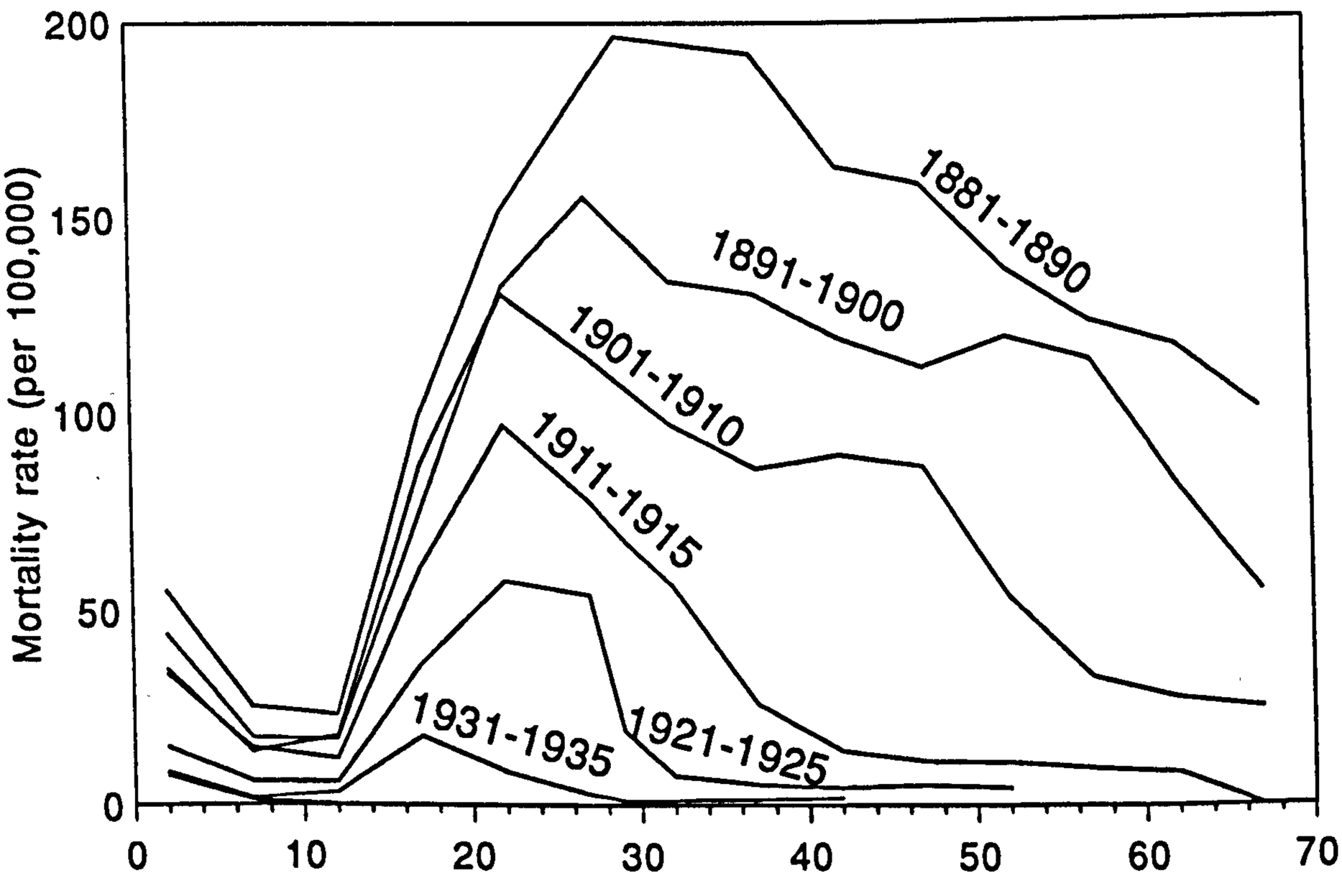
The risks of developing disease are age-dependent. This is illustrated by age-specific tuberculosis mortality data in developed countries from the pre-chemotherapy era, which were a reliable indicator of the actual disease incidence. Figure 1.2(a) presents the age-specific mortality rates from respiratory tuberculosis in males in England and Wales between 1880 and 1950. This shows that the mortality rates in a given year increased between adolescence and middle age, and declined thereafter, and that the peak age generally increased over time. The corresponding mortality rates plotted by birth cohort, as shown in Figure 1.2(b), illustrate that after an initial high mortality rate during infancy, the greatest mortality rate, reflecting the greatest risk of developing disease, was experienced during adolescence and early adult life [3].

Similar analyses by sex suggest that females experience greater risks of developing disease during adolescence, as compared to males, and lower risks during old age. Similar patterns in age-specific mortality rates from all forms of tuberculosis were also found in Norway [5], the US [27] and in New Zealand [4].

Further evidence for a high risk of developing disease during adolescence in developed countries is provided by a study of 82,269 tuberculin-positive (6mm or greater induration to 1 or 10TU PPD) children aged 0–19 years in Puerto Rico, followed up for 18–20 years after 1949 [28]. This found that the case-rate, according to age at diagnosis, followed a similar pattern to that observed for mortality rates by cohort. However, this study did not describe the age at infection of these individuals, and thus it is unclear whether the cases experienced primary, endogenous or exogenous disease.



(a) Mortality rates by year.



(b) Mortality rates by birth-cohort.

Figure 1.2: Age-specific mortality rates from respiratory tuberculosis in males in England and Wales during the period 1881–1950, plotted by year and birth cohort.
Data source: Adelstein [3]

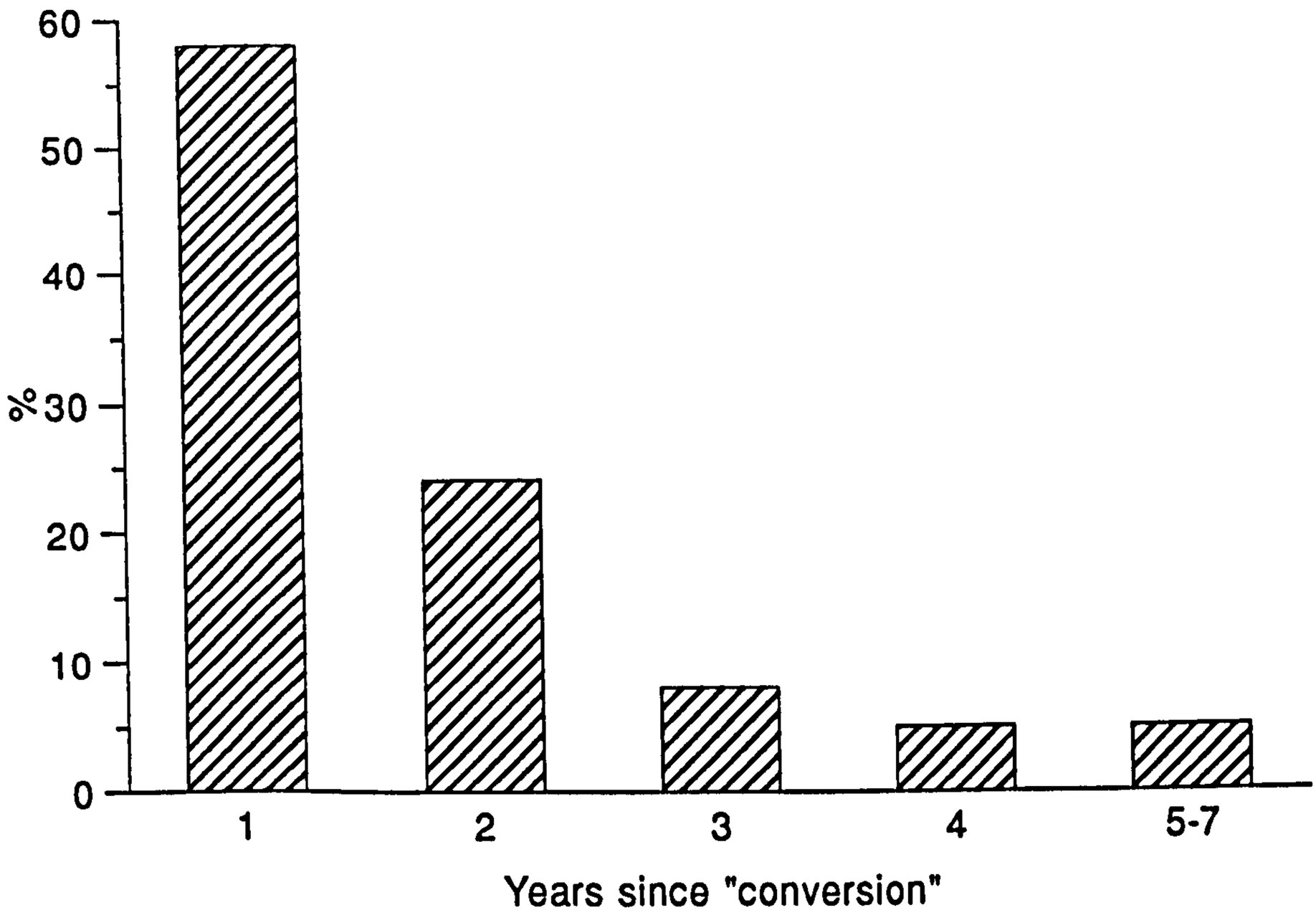


Figure 1.3: Distribution of time interval between 'conversion' and clinical onset of tuberculosis for 243 individuals who developed disease, as found in the UK MRC BCG trial during 1950s [29].

The MRC BCG trial in England and Wales during the 1950s provides the most detailed available information on the risk of developing disease according to time since first infection [29]. Among the 12,867 tuberculin-negative (0–4mm induration to 100TU) 14–15 year olds chosen at random and left unvaccinated, 243 individuals developed tuberculosis during the following 10 years. The distribution of time intervals between 'conversion' to tuberculin-positivity and disease onset of these individuals is shown in Figure 1.3. This shows that of those who developed disease, almost 60% did so during the first year after 'conversion'. These data are discussed in greater detail in section 1.2.1.

The type of disease individuals develop is also age-dependent. About 60% of disease in children is extrapulmonary, for example, as compared with less than 40% in adults. Furthermore, not all forms of tuberculosis are infectious. Extrapulmonary tuberculosis is rarely infectious, and the infectiousness of pulmonary cases depends on the quantity of tubercle bacilli present in their sputum. If such bacilli are detectable by microscopy, then an individual is said to be "smear-positive". Bacilli in the sputum of smear-negative individuals are sometimes detectable by culture: in this case, the individual is said to be "culture-positive" and "smear-negative". It usually takes 4–6 weeks for tubercle bacilli to

grow in culture, which means that there is a delay before a smear-negative tuberculous case receives treatment. In the case of drug-resistant tuberculosis, for which the time interval between onset of disease and subsequent death may be short, this delay until diagnosis is particularly important [30].

Smear-positive individuals are usually more infectious than those who are only smear-negative and culture-positive. For example, a study in Rotterdam during the period 1967–9, indicated that roughly 50% of 0–14 year old household contacts of smear-positive cases were tuberculin-positive, compared with about 1% of individuals in this age group in the general population and about 6% of contacts of culture-positive, but smear-negative individuals [31].

A detailed survey of morbidity among more than 8,000 intimate and 11,000 casual contacts of smear-positive, culture-positive and culture-negative white and American-Indian cases was carried out during the period 1966-71 in British Columbia and Saskatchewan [32]. This found that in any age group, intimate contacts of smear-positive cases had a higher prevalence of tuberculin sensitivity than those of culture-positive but smear-negative cases. The proportion of 0–14 year old contacts of smear-positive cases found to be tuberculin-positive was also lower than that for Rotterdam (e.g. 29%, 36% and 40% for 0–4, 5–9 and 10–14 year olds respectively). The differences in the results from the two studies could have been a consequence of differences in methods of contact tracing or tuberculin-testing. The Canadian study found that a greater proportion of tuberculin-positive contacts of smear-positive cases also developed disease, as compared with those of culture-positive cases (15.0% vs 3.3%).

There are few countries with reliable data on the relative contribution of smear-positive disease to the incidence of pulmonary tuberculosis. Notifications of pulmonary tuberculosis collected in Norway since 1942 have distinguished between 'non-infectious' and 'infectious' cases. The latter are defined as individuals with tubercle bacilli demonstrable in the sputum, or cases with a cavity and/or a tendency to progressive ('phthisic') tuberculosis. Cases with bacilli in gastric lavage or on laryngeal swabs are considered non-infectious, unless cavitation and/or phthisic progression are evident.

Given this definition of an infectious case, it is likely that the data overestimate the incidence of *smear-positive* disease, and reflects the incidence of *sputum-positive* (smear and/or culture-positive) disease. For reference, Figure 1.4, summarizes the age-specific notification rates of pulmonary and 'infectious forms' of tuberculosis in males in Norway,

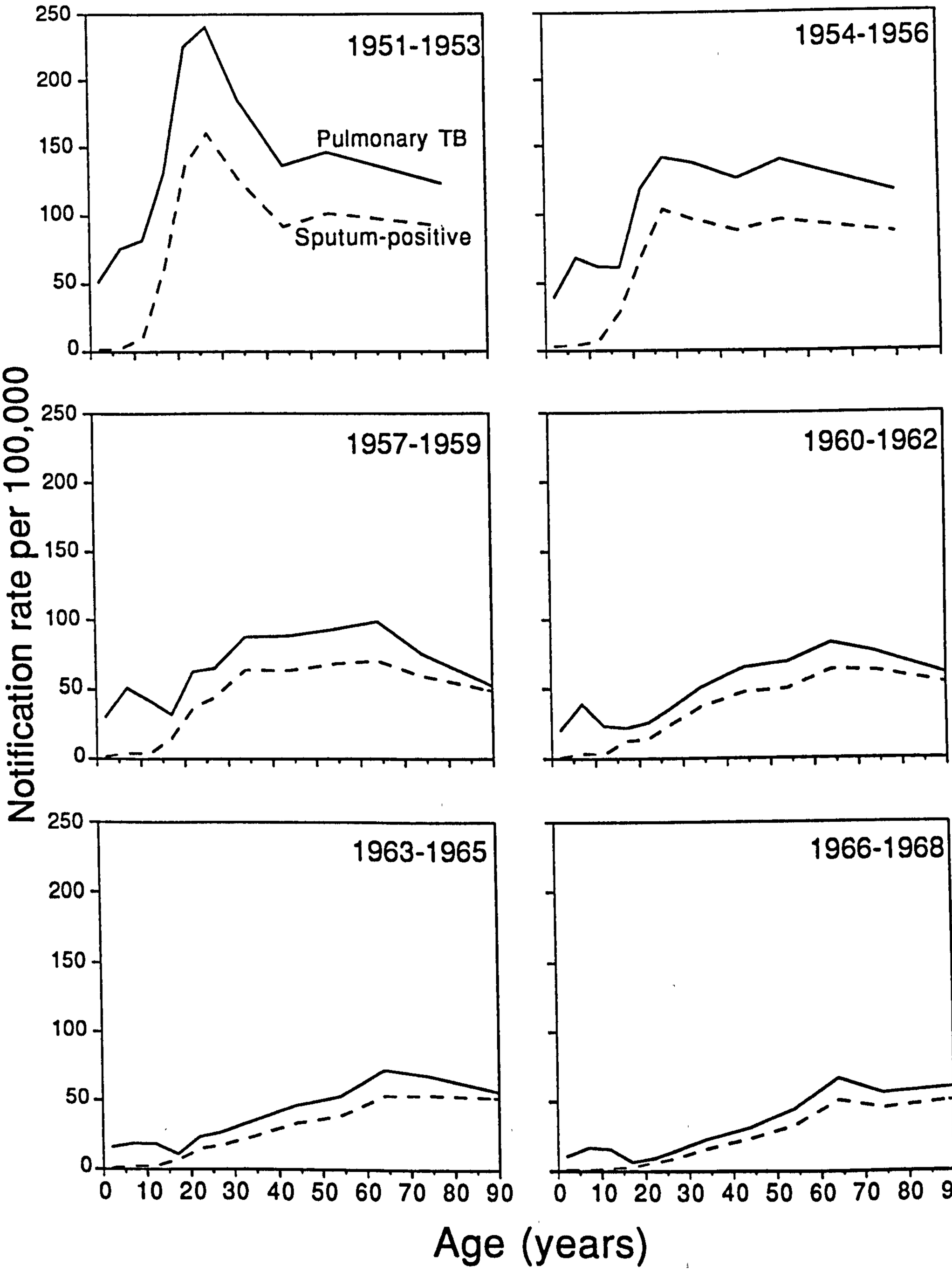


Figure 1.4: Age-specific notifications of pulmonary and 'infectious' (sputum-positive) forms of tuberculosis in males in Norway during the period 1951–1969.
Source: Dr K. Styblo and Dr. K. Bjartveit (Norwegian National Health Screening Service).

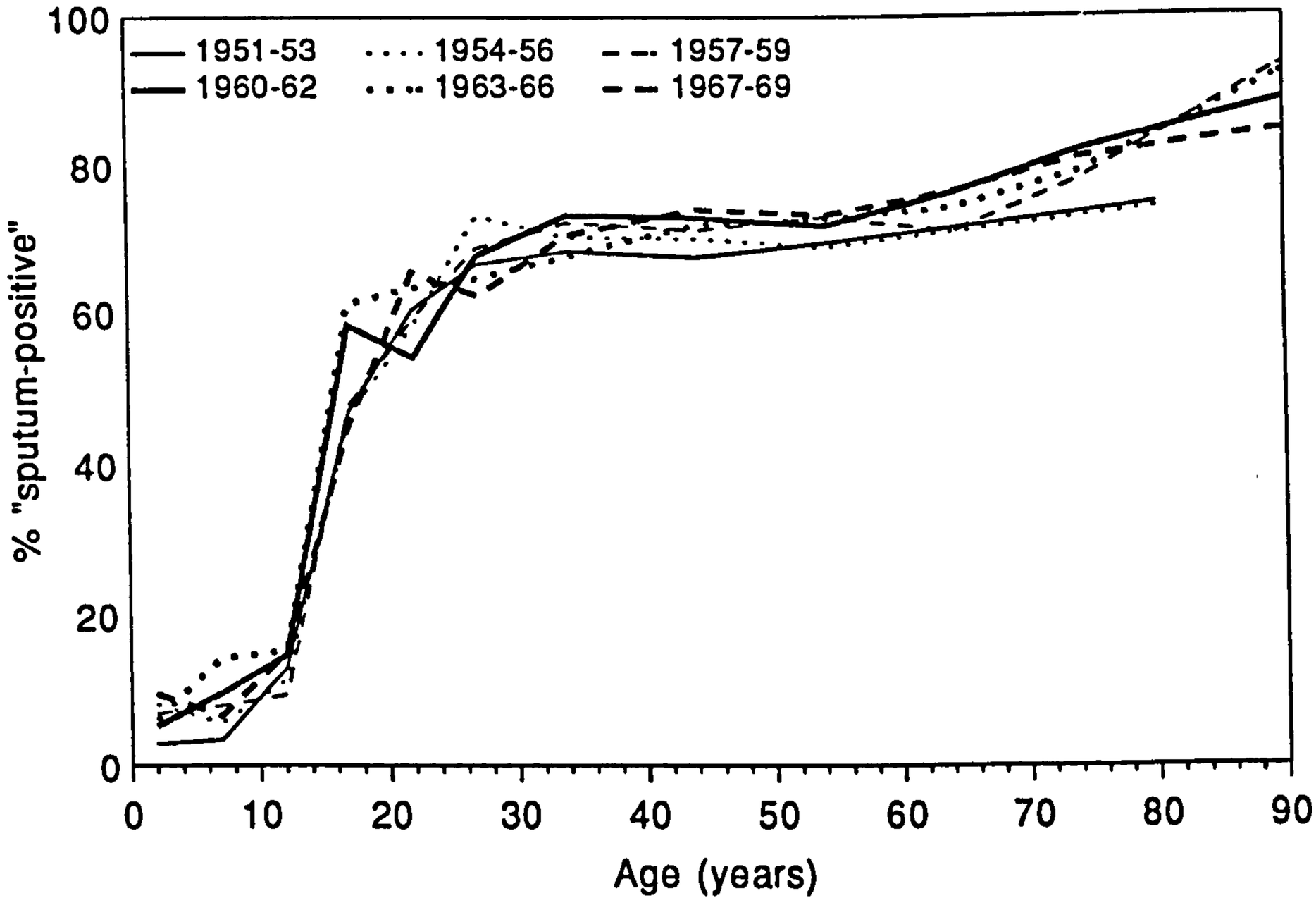


Figure 1.5: Relative contribution of sputum-positive disease to age-specific notifications of pulmonary tuberculosis in males in Norway during the period 1951–1969.

during the period 1951–1969. The relative contribution of 'infectious' forms to the age-specific disease incidence during this period is summarized in Figure 1.5. This shows a consistent age-specific pattern across the time period considered, with less than 20% of pulmonary tuberculosis in children being sputum-positive, as compared with over 60% of that in adults.

Further insight into the relative contribution of smear-positive disease to the age-specific incidence of pulmonary tuberculosis is provided by a study of all cases (1,844 individuals) notified in Saskatchewan during the period 1960–1973 [33]. Isolates were obtained from 98.5% of the cases either directly from sputum or by gastric lavage during three consecutive days. The report of this study does not define the number of positive smear-examinations required before individuals were considered smear-positive. The results are summarized in Table 1.1.

These results are not directly comparable to those found for Norway, where the definition of a 'sputum-positive' case excludes individuals positive on gastric lavage and includes cavitary cases. This is reflected by the high proportion of 0–14 year olds (33%) found to be smear or culture-positive in Saskatchewan, as compared with less than 20% of their counterparts in Norway. Similarly, about 42% of pulmonary cases aged 15–29 years in

Table 1.1: Relative contribution of smear and culture-positive disease to notifications of pulmonary tuberculosis in Saskatchewan during the period 1960–1973 [33].

Age group (years)	Number of cases				% age contribution to pulmonary TB		
	Smear +ive	Culture +ive only	'Sputum' +ive	Total pulmonary	Smear +ive	Culture +ive only	'Sputum' +ive
0–14	7	103	110	315	2	33	35
15–29	113	46	159	272	42	17	59
30–49	134	50	184	330	41	15	56
50^{+}	275	107	382	584	47	18	65

Saskatchewan were smear-positive, and almost 60% were smear or culture-positive, which is slightly lower than the estimates found for Norway. The estimates from Saskatchewan are also based on smaller number of cases, as compared with those from Norway, and are hence subject to greater sampling error.

Without treatment, the case-fatality of tuberculosis varies according to the type of disease. The case-fatality of tuberculous meningitis during the pre-chemotherapy era approached 100%, and, with treatment, varies according to the severity of the case at presentation. Estimates of the case-fatality of bacillary (positive on smear or culture) cases in developing countries in the absence of treatment is provided by a 5-year longitudinal study in a rural district in South India [25]. After one and a half years, 30% of the 126 bacillary cases at the start were dead, 28% were cured, and 42% were still bacillary-positive. Five years later, 49%, 33% and 18% were dead, cured and bacillary-positive respectively. These data are not stratified by either age or by severity of the case, and thus these results cannot be generalized further.

Other follow-up studies during the pre-chemotherapy era, reviewed by Murray and Styblo [6,34], found similar estimates of the case-fatality rate among sputum-positive cases. The study of Berg [35] was the most extensive, following up 6,156 smear-positive cases (defined as individuals with 'tubercle bacilli found by direct examination of sputum') in Gothenburg, Sweden, during the period 1910–1934. Cases aged 15–49 years had a case fatality of 31–33% during the first year after diagnosis, which dropped to 16–18% for the two subsequent years. The overall case-fatality for these individuals during the 20 years of

follow-up was about 60%[1]. The case-fatality among individuals aged over 50 years was over 40% during the first year after diagnosis. This last estimate may well be unreliable, as it was based on a small number of cases.

Drolet attempted to estimate age-specific case-fatality rates by comparing the number of reported cases against the corresponding mortality in London and in several cities in the US during the period 1925–1937 [36]. Overall, he found that among 0–4 year olds, there were twice as many reported cases as deaths, and for individuals aged over 65 years, the number of notified cases only just exceeded the numbers of deaths. Inferences from such data as to the case-fatality rate depend on assumptions about the efficiency of the notification process. There is evidence to suggest that notification has been relatively inefficient in many countries, at least until 1950 (see e.g. section 2.2).

There are three main strategies for the control of tuberculosis, namely BCG vaccination, chemoprophylaxis and case-finding and treatment.

BCG vaccination (Bacillus Calmette Guérin) was first used in humans in 1921 and has been used increasingly since then, with varied success. Evidence suggests that BCG does not protect against infection, but reduces the risk that infected individuals develop disease [37]. In England and Wales, the efficacy of BCG given during adolescence has been estimated at about 77%, but may decline over time [38]. A trial in South India, on the other hand, indicated a zero efficacy for BCG [26]. Various hypotheses have been proposed to explain this varying efficacy of BCG, such as masking by environmental mycobacteria, differences between BCG strains, differences in *M. tuberculosis* or genetic differences between populations [39].

The contribution of BCG vaccination to the decline of tuberculosis in developed countries is unclear. In the Netherlands, for example, where BCG has never been used as a mass control policy, the decline in the crude annual incidence of tuberculosis has paralleled that in other Western European countries [1].

Chemoprophylaxis is sometimes given to individuals who have been infected recently and has been used in many developed countries as a control policy. Its effect on the incidence of tuberculosis in developed countries today, where the annual risk of infection is low, is limited, since much of the disease occurs in older individuals, and is mainly through

[1]Note that these estimates were not corrected for other causes of death. About 97% of the deaths in the entire follow-up population were attributed to tuberculosis.

endogenous reactivation of infections acquired long ago [6].

An alternative strategy for protecting 'susceptibles' is the removal of sources of infection, either through treatment or, in the pre-chemotherapy era, through isolation in sanatoria. Identifying sources of infection is often difficult. Due to the low prevalence of tuberculosis in developed countries, mass screening of the population is no longer recommended as a policy for detecting sources of infection [6].

Chemotherapy was introduced into developed countries during the late 1940s. In England and Wales, streptomycin was first used for treatment in 1947, and, together with PAS (para-aminosalicylic acid), became generally available by 1950. Isoniazid first became available in 1953. The case-fatality improved gradually following the introduction of chemotherapy, as found by a follow-up study of all sputum-positive cases aged over 15 years notified in Birmingham (England) during the period 1947–1959 [40]. Individuals notified in 1947, 1950, 1953, 1956 and in 1959 had overall case-fatality rates of 65%, 38%, 29%, 25% and 27% respectively during 10 years of follow-up. The case-fatality rate for cases notified in 1947 exceeds that found by the follow-up study in Bangalore during the period 1961–1966 [25] (see page 35). This is attributable to deficiencies in the notification process in England and Wales (see section 2.2), which meant that only the most severe cases, and hence those with the highest case-fatality rates were likely to be reported.

These estimates of the case-fatality ratios depend on the severity and the age of the case when treatment was initiated. About 45% (155 out of 333) of the cases followed up from 1959 were aged over 45 years, as compared with 38% (196 out of 521) of those followed up from 1956. This accounts for the slight excess in the estimated case-fatality rate among individuals followed up from 1956, as compared with that for counterparts followed up from 1959.

Multiple drug regimens lasting up to 15 months, and consisting of as many as 5 drugs, are now used to treat clinical tuberculosis. Short course regimens, lasting 9 months have been used in England and Wales since 1976 [41] and have a good cure rate for patients who adhere to the full course. A controlled follow-up study of short-course chemotherapy regimens started in 1972 in England and Wales found that out of 665 patients available for follow-up after 54 months, none of the 52 deaths were attributable to active tuberculosis [41].

In both developing and developed countries, treatment compliance is a problem, and inadequate treatment predisposes to the emergence of drug-resistant bacilli. This was

highlighted as a severe public health problem during the early 1990s, particularly in New York City, where several outbreaks of multi-drug resistant tuberculosis occurred in hospitals [22, 23, 42] and in state prisons [30, 43]. Most of the cases affected in these instances were infected with HIV, and thus had a high risk of developing tuberculosis after infection [44].

1.2 Review of related work

1.2.1 Studies of the risks of developing disease

It is often stated that individuals face a 5–10% lifetime risk of developing tuberculosis after infection [8]. This lifetime risk depends on many factors, such as the individual risks of developing 'primary', 'endogenous' and 'exogenous' diseases, the age at which infection is acquired, the risk of reinfection and the mortality rates experienced subsequent to initial infection. Consequently, it is unlikely that the overall lifetime risk is identical for all age groups or that it has remained constant over time.

Similarly, these factors determine the distribution of the time interval between infection and disease onset (conventionally known as the incubation period), and the serial interval (the time interval between two successive cases in a chain of transmission) for tuberculosis, which are two of the most important parameters underlying the dynamics of any infectious disease.

In the past, no study has attempted to link together these individual factors to obtain age and time-dependent estimates of the overall lifetime risk of developing disease, and the incubation period and serial interval for tuberculosis. Instead, many observational studies of the natural history of tuberculosis have focussed on either the risks of (primary) disease shortly after infection, or on the risks among 'infected' — i.e. tuberculin-positive — individuals. The observed risks among the latter, in turn, are a function of the time since initial infection (often not known) and of the risk of reinfection and subsequent exogenous disease.

Some of the more extensive observational studies are reviewed below. One other study, that of Sutherland *et al* [7] estimated the risks of developing 'primary', 'endogenous' and 'exogenous' disease using a model-based approach. This is described in section 1.3.2.

1.2.1.1 The UK MRC BCG trial

The UK MRC BCG trial during the 1950s provides the most reliable data on the risk of developing disease following 'infection' and its trend with time since infection. These estimates are based on 12,867 tuberculin-negative (0–4mm induration to 100TU) 14–15 year olds chosen at random as controls and left unvaccinated, and followed up with 14 month cycles of tuberculin tests and chest X-rays.

The magnitude of the disease risk estimates depend on the criterion for tuberculin sensitivity used for defining 'infection'. About 4.7% (121 individuals) of those who 'converted' after the start of the trial from tuberculin negativity (0–4mm induration to 100 TU) to positivity (induration of 5mm or more to 3 TU), developed disease over the ten years following 'conversion' [29]. This compares with the 8.1% (108 individuals) of those who had an induration of 8mm or more to 3 TU after 'conversion' who developed disease over the same period [6][2].

A further limitation of the data from the MRC trial is that about 125 initially tuberculin-negative individuals also developed disease but did not have a documented positive tuberculin test and were excluded from the disease risk estimates. Hence the number of individuals who had been infected after the start of the trial and who were at risk of developing disease, but had no documented positive tuberculin reaction, is not known.

The date of 'conversion' for these individuals was assumed to coincide with the first abnormal chest X-ray or clinical symptoms. Given the regularity of the tuberculin testing procedures, this should not have overestimated the 'true' date of conversion by more than 14 months. On the basis of this assumption, the distribution of the time interval between 'conversion' and disease onset was obtained for the unvaccinated individuals who developed disease (see section 1.1 on page 31). This provided the first ever indication that most of those who develop disease do so during the first year following 'infection'[3].

The disease risk estimates from the study, based on the 121 individuals who developed disease following a documented tuberculin 'conversion' (5^{+}mm induration to 3TU) were also broken down by age at 'conversion'. These are summarized in Table 1.2. These results indicate that the risk of developing disease decreased with the age at 'conversion', and was about 10.5% for 14 year olds, as compared with about 2.4% for individuals aged over 19 years. This is inconsistent with age-specific patterns in the mortality rates among cohorts of

[2]This estimate is based on the most recently published review of the risks of developing disease [6], which cites an analysis of the data in 1967 [29]. The cited report however, gives disease risk estimates of 7.5% for individuals with indurations of 8mm or more.

[3]Note that the first estimate of this distribution was based on 238 cases (113 and 125 cases among those with and without a documented tuberculin conversion respectively). All subsequent reviews of the trial base this distribution on *243* cases [6], but do not show the numbers who had had a positive tuberculin test. The final report of the BCG trial states that there were 248 cases in total among initially tuberculin-negative individuals, although the distribution of the time interval between tuberculin 'conversion' and disease onset has not been updated since [38].

Table 1.2: The ten-year incidence of tuberculosis following 'conversion' from tuberculin negativity (0–4mm to 100TU) to positivity (5^{+}mm to 3 TU), according to age at 'conversion', as found by the UK MRC BCG trial during the 1950s [29].

	Age at 'conversion' (years)						
	14	15	16	17	18	19^{+}	All
No. of 'converters'	57	580	761	398	336	418	2,550
No. of cases	6	38	42	17	8	10	121
'Risk' of disease	*10.5*	*6.6*	*5.5*	*4.3*	*2.4*	*2.4*	*4.7*

individuals during the pre-chemotherapy era in developed countries, which suggested that the risks of developing disease increased during young adult life (see page 30 in section 1.1). There are several possible explanations for this apparent inconsistency.

First, individuals who converted when aged over 17 years during the BCG trial had experienced at least 4 tuberculin tests by that time, which could have led to boosting in their tuberculin response (see section 1.1). This could have resulted in an overestimate in the number truly infected, and thus an underestimate in the true risks of developing disease.

Second, the risks of developing disease among those who converted when aged over 19 years were based on small numbers of individuals, and thus may be unreliable.

Third, given the decline in the risk of infection, a greater proportion of individuals who 'converted' when aged 14 years could have been reinfected and experienced exogenous disease, as compared with those who converted when older.

The study also provided estimates of the disease risks according to the degree of tuberculin sensitivity at the start [38]. The annual disease incidence during 20 years of follow-up was greatest (1.04 per 1,000 per year) among individuals with an induration of 15mm or more to 3TU, as compared with 0.45 per 1,000 per year among individuals with a 5–14mm induration to 3TU and 0.98 per 1,000 per year among tuberculin-negative (unvaccinated) individuals. These results are consistent with those from other studies, which we describe below.

1.2.1.2 US Public Health Service chemoprophylaxis studies

Other data on the risks of developing disease given recent and distant infection, and according to the age of individuals, are provided by the US Public Health Service chemo-

Table 1.3: Incidence rate of disease during ten-year follow up of the US Public Health Service chemoprophylaxis trial during the 1950s [45] according to:
(i) initial tuberculin reaction and
(ii) the age at enrolment among individuals who were either initially tuberculin-positive, or who converted to tuberculin-positivity during the first year of the trial.

	Size of population followed up	Total number of cases	Risk (per 1,000 followed up)
(i) Tuberculin reaction			
10^{+}mm	4,992	147	29.4
5–9 mm	1,616	31	19.2
Tuberculin 'converters': (0–5mm induration at entry, 5^{+} mm after 12 months.)	867	32	36.9
(ii) Age at enrolment†			
<15	3,132	52	16.7
15–34	2,344	98	42.3
35–54	1,724	53	30.9
55^{+}	796	12	15.2

† Note that the risk estimates exclude cases among individuals who were non-reactors both at recruitment into the study and 12 months thereafter.

prophylaxis trials carried out during the 1950s among contacts of active tuberculosis cases. These studies were reviewed by Ferebee [45]. Table 1.3 summarizes the incidence rates of disease according to the size of the initial tuberculin reaction (to 5TU PPD-S) and the age at enrolment for the control group.

Considering the results according to the size of the initial tuberculin reaction, this shows that the risk of disease during follow-up was highest (36.9 per 1,000) among 'converters' to tuberculin-positivity, as compared with that among initial tuberculin reactors. This estimate of the risk of developing disease following conversion is slightly lower than that found for the UK MRC BCG trial (see Table 1.2), and could be a consequence of differences between the age composition of the group of 'converters'. Alternatively, the differences could reflect differences in the risk of infection, which may have been higher during the UK study. This could have also led to a greater proportion of individuals in the UK study being reinfected during follow-up and experiencing exogenous disease as compared with that in

the US. Overall, these results agree with the MRC data in that the risks of developing disease immediately after tuberculin 'conversion' exceeded those many years thereafter.

The results in Table 1.3 also indicate that the risks of developing disease may be age-dependent — individuals enrolled when aged 15–34 years experienced the highest risks (42.3 per 1,000 individuals followed up), compared with those for 0–14 year olds and those aged over 55 years. Given that the number of 'converters' to tuberculin-positivity in these age groups is not provided, it is unclear whether the high risk for 15–34 year olds reflects a high risk of developing primary or endogenous disease, or of reinfection followed by exogenous disease.

Data from these studies also suggest that the risks of developing disease decrease with the time since initial 'infection'. The greatest proportion of individuals who developed disease (i.e. 41%) did so during the first year of the trial, corresponding to an annual risk of 12.2 per 1,000. This dropped to 3.1 per 1,000 per year for the following two years and was about 1.6 per 1,000 per year during the sixth and seventh years. These data do not provide reliable insight into the magnitude of the decrease in the risk of developing disease with time since infection, given that the time of conversion to tuberculin-positivity for these individuals could not be established. Hence these results are not directly comparable to those from the UK MRC BCG study.

1.2.1.3 Follow-up studies from the pre-chemotherapy era

Several follow-up studies during the pre-chemotherapy era attempted to quantify the risks of developing disease following 'infection'. The reliability of the results from these studies is restricted, given that most considered only small populations, usually consisting of university or medical students, and generally had short durations of follow-up. These studies are reviewed by Styblo [6] and are not described here. Two of the more extensive studies during the pre-chemotherapy era which followed up individuals for over ten years and attempted to estimate these risks of developing disease according to both age and time since infection are described below.

Oslo Public Health Service, 1929–1944

A study of Meyer in 1949 followed up 889 'converters' to tuberculin-positivity found by the Oslo Public Health Service during the period 1929–1944 [46]. These individuals were

Table 1.4: Summary of the age-specific cumulative risks of developing pulmonary tuberculosis following conversion to tuberculin-positivity, as found by the follow-up study of patients during the period 1929–1944 through the Oslo Public Health Services [46].

Age at 'conversion' (years)	Number of 'converters'	Total who ultimately developed disease	Cumulative risks of developing disease (%)
0–3	58	0	—
4–12	342	4	2.4
13–19	242	26	12.4
20^{+}	247	19	9.3

identified through either routine annual medical examinations of schoolchildren, or through 'environmental and industrial investigations'. The definition of conversion depended on the test used (i.e. 5mm induration for Mantoux test with 1mg tuberculin or "3mm induration and 4mm rubor" for the Pirquet test) and hence was not identical for all individuals.

The cumulative risks of developing pulmonary tuberculosis according to the age at 'conversion' during the 17 years of follow-up found by the survey are summarized in Table 1.4. This shows that individuals who 'converted' when aged 4–12 years faced a cumulative risk of about 2.4%, as compared with about 12.4% for 'converters' aged 13–19 years.

These risks of developing disease following 'conversion' during adolescence exceed those found during the UK MRC BCG trial (see Table 1.2). This could reflect either biases in the way the follow-up population in Oslo was selected or that a greater proportion of individuals developed disease following infection during the pre-chemotherapy era, as compared with counterparts thereafter. The latter could be a consequence of

1. the high risks of infection during the pre-chemotherapy era, which meant that many individuals were reinfected and subsequently experienced exogenous disease, and
2. poor living standards and general health status of the population during the pre-chemotherapy era, which predisposed a large proportion of individuals to develop disease after infection.

Given the small numbers of cases found in the follow-up population, the study provides only limited information on the risks of developing disease according to the time since

'conversion'. For 13–19 year olds, the data suggest the annual risks of developing disease were 2–3% during the first four years after 'conversion'.

Williamson County, Tennessee 1931–1954

Some insight into the age-dependent risks of developing disease is provided by a major study of families of contacts of 828 tuberculosis cases notified to Williamson County (Tennessee) Health Department during the period 1931–1954 [47]. This study was undertaken in Tennessee as a result of its high tuberculosis mortality rates for both white and non-white individuals, as compared with those in the rest of the US. As part of this study, 3,446 white individuals (1,778 males and 1,668 females) and 768 non-white individuals (384 males and 384 females) were followed up for up to 24 years.

Twenty two years after the start of the study, Zeidberg *et al* [48] analysed the time interval between exposure (as measured by the estimated onset of 'cough' of the case first notified to the County Health Department) and disease of 733 children of white, sputum-positive tuberculous cases. Figure 1.6 summarizes the cumulative risks of developing disease according to the age at first exposure, as found by the study. These results suggest that the cumulative risks of developing disease following 'exposure' at age 5–14 years were lower than those for other age groups (e.g. about 12%, as compared with 20% for individuals aged over 15 years). These estimates seem high, given that the degree of exposure varied between individuals and hence only a fraction of the exposed individuals had probably been infected.

The reliability of these results is unclear, given that they were based on small numbers of cases and the size of the follow-up population in each age group was not provided in the original report. A further limitation is that there are no data on when or even whether the 'exposed' individuals had converted to tuberculin sensitivity. The high risk of developing disease observed soon after exposure for individuals aged over 15 years, for example, could reflect a greater prevalence of infection, as compared that in other age groups. It is also impossible to make any inferences about whether the disease individuals developed was primary, endogenous or exogenous in origin, given that the subsequent exposure of individuals to tuberculosis is not provided.

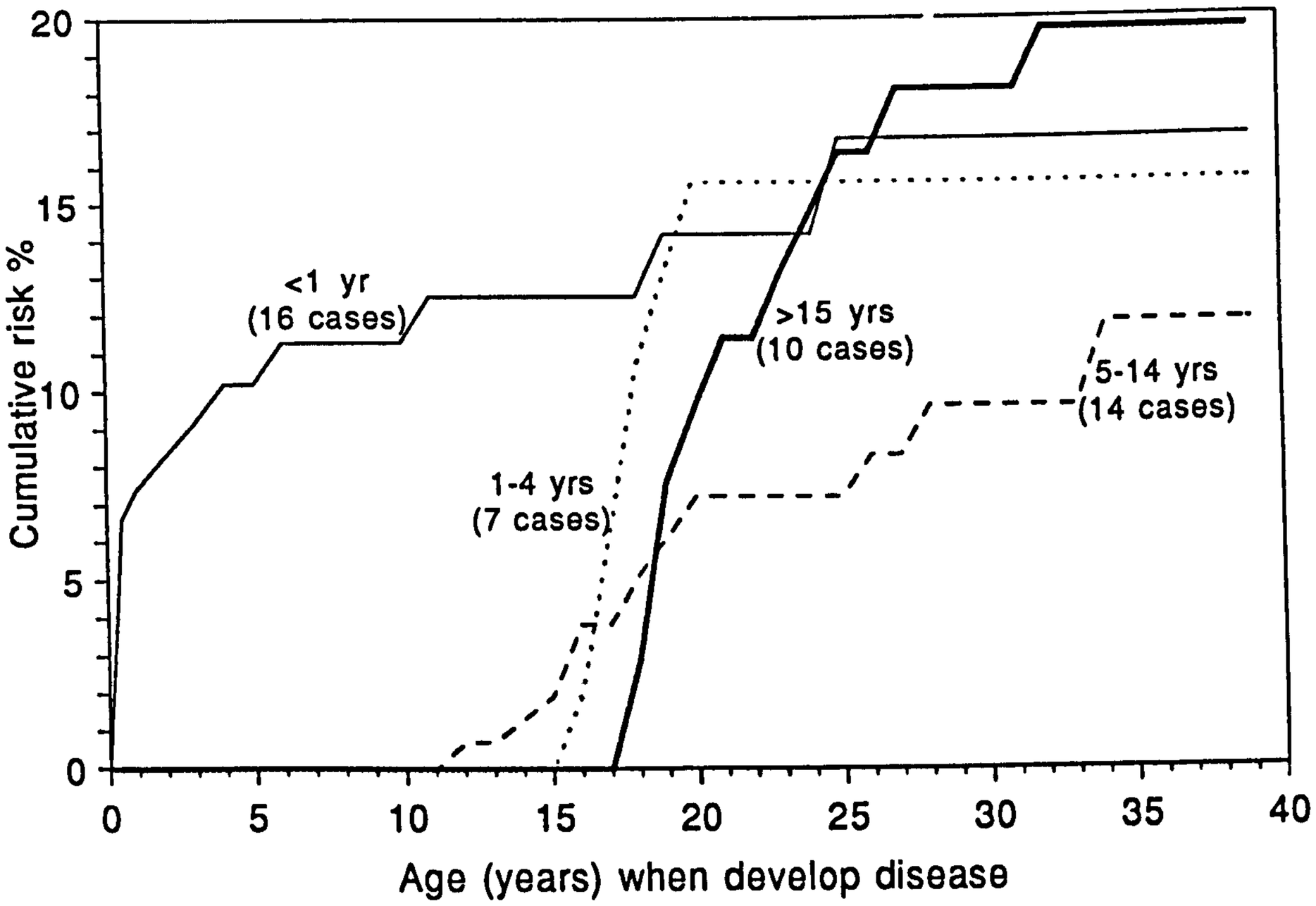


Figure 1.6: Cumulative risks of developing tuberculosis for 733 children of white sputum-positive cases, according to age at first 'exposure', as found by the Williamson County tuberculosis study during the period 1931–1954 [48]. Numbers in brackets denote the total number of cases observed among the individuals followed-up.
Reproduced from Zeidberg et al [48].

1.2.1.4 Estimates of the risks of developing endogenous or exogenous disease

There have been few direct observational studies of the risk of developing endogenous disease and no studies examining the risks of exogenous disease after reinfection.

Estimates of the risk of developing endogenous disease have been inferred in the past by observing the incidence of disease among tuberculin-positive individuals in an area with a low risk of infection. Styblo [6], for example, observed that the incidence of sputum-positive pulmonary disease among 65–74 year olds and those aged over 75 years in the Netherlands was about 10 and 20 per 100,000 respectively during the period 1971–73 (when the risk of infection was low), and concluded that this reflected the risk of developing endogenous disease, given that most individuals in this age range were probably infected and few were reinfected, given the low risks of infection during the period. While this is a realistic assumption, it is complicated by the fact that many older individuals do not mount a response to tuberculin at all. This either implies that they have never been infected or else that they have lost the sensitivity they had earlier in life (see section 1.1).

Similar estimates of the risks of developing endogenous disease were derived by Horwitz *et al* [49], who followed up for 12 years 286,250 tuberculin reactors (6mm or more induration to 10TU) aged over 15 years identified during the period 1950–1952 during BCG vaccination campaigns in Denmark.

In areas with a high risk of infection, the incidence rates among individuals are generally high and do not reflect the risks of endogenous disease, given that some of the disease is attributable to reinfection and exogenous disease.

This is illustrated by the fact the during community chemoprophylaxis trials in Greenland during the 1950s, the incidence of disease among initially tuberculin-positive individuals during the 6 years of follow-up was about 82 per 1000 followed up [45]. This was over two and half times greater than that found among tuberculin-positive individuals during the US Public Health Service trials during *10 years* of follow up (see Table 1.3). Inferences as to the magnitude of the risk of reinfection and of subsequent exogenous disease are complicated by the fact that we have to account for both the age and time since infection in the corresponding study populations.

1.2.1.5 Conclusions

In interpreting the results from these studies, it is important to recognize that they are influenced by a number of factors, such as the type of test and criterion used to determine tuberculin-positivity, the efficiency and duration of follow-up, and the definitions used to describe a tuberculosis case.

Two consistent features nevertheless emerge from these studies, namely that the risk of developing disease depends on both:

1. the age of individuals considered, and appears to be greater among adults than for children. This result is consistent with age-specific patterns in mortality rates among birth cohorts during the pre-chemotherapy era (see section 1.1).

2. The degree of tuberculin sensitivity. Individuals with a small amount of tuberculin sensitivity at the start of follow-up in the above studies faced *lower* risks of developing disease, than did individuals who 'converted' to tuberculin-positivity. Assuming that this weak tuberculin sensitivity is attributable to infection with *M. tuberculosis*, this suggests that infection might confer some protection against disease many years

thereafter[4].

The studies also suggest that the risk of developing disease among tuberculin-positive individuals in populations with a high risk of infection, exceed those for comparable individuals in populations with low risk of infection. Whilst this difference may be attributable to differences in the environmental, genetic and social factors in the populations, which may be associated with high risks of infection, the fact that some of the disease may be attributable to exogenous reinfection should not be ignored.

Given these observations, we conclude that both age and reinfection are important factors underlying the disease incidence, and should be included in a model if we are to provide a realistic representation of past trends in tuberculosis in England and Wales.

[4]Note that weak tuberculin sensitivity may also be attributable to infection with environmental mycobacteria — see section 1.1.

1.2.2 Studies of the transmission of *M. tuberculosis*

The basic and net reproduction numbers are two of the most important parameters underlying the dynamics of any infectious disease, as they describe the number of secondary infectious cases which will result from one infectious case under given circumstances. More formally, the basic reproduction number is defined as:

> 'the average number of secondary infectious cases resulting from the introduction of a *typical* infectious case during its entire infectious period, in a [totally] susceptible population' [50,51].

The net reproduction number can be defined analogously as:

> 'the average number of secondary infectious cases resulting from each infectious case during its entire infectious period, in a given population'.

The definitions and interpretation of both the basic and net reproduction numbers are complicated for tuberculosis, given that individuals face age-dependent risks of developing 'primary', 'endogenous' and 'exogenous', and that even reinfected individuals can be considered 'susceptible' to tuberculosis, given that reinfection can occur. These issues will be discussed in detail in Chapter 5.

A basic or net reproduction number greater than 1 for a given infection implies that one infectious case leads, on average, to more than one subsequent infectious cases and that the infection will persist in the population in the absence of control or intervention. This logic was implicitly recognized by Frost as early as 1937 [2]:

> "It is necessary only that the rate of transmission be held permanently below the level at which a given number of infection-spreading (i.e. open) cases succeed in establishing an equivalent number to carry on the succession. If, in successive periods of time, the number of infectious hosts is continuously reduced, the end result of this diminishing ratio, if continued long enough, must be the extermination of the tubercle bacillus."

The dramatic decline in tuberculosis which has been observed in developed countries since the 1850s, suggests that the net reproduction number has been less than unity over this period. This decline has been attributed to various factors (e.g. reduced crowding or

increased resistance to developing disease), but the effect of these factors on the basic reproduction number of tuberculosis is unclear. Reduced crowding, associated with improved socio-economic conditions, for example, infers a reduction in transmission, and hence a reduction in the basic and net reproduction numbers.

There has been only one published study of the basic reproduction number for tuberculosis [9], using a model in a hypothetical population without age-structure. This is described in greater detail in section 1.3. Until now, no study has attempted even to define the net reproduction number for tuberculosis. Instead, some researchers in the past have attempted to estimate a statistic which is related to the reproduction numbers, namely the average number of *new infections* resulting from a single infectious (smear-positive) case. For diseases in which all infected individuals develop disease, this measure would be identical to the net reproduction number. These estimates are reviewed below.

In practice, such measures represent only average values and are subject to a large degree of variability. Some insight into the extent of this variability and its determining factors is provided by documented reports of outbreaks of tuberculosis. Examples of such outbreaks are described below.

1.2.2.1 Average number of new infections per smear-positive case

The earliest attempt at estimating the average number of new infections resulting from a smear-positive individual was that of Styblo [52], who defined a "contagious parameter", given by the ratio between the annual risk of infection and the prevalence of smear-positive cases. This "contagious parameter" was first obtained for the period 1921–1937 for the Netherlands, which has detailed information on the annual risk of infection since 1911 from tuberculin survey data, but no pre-chemotherapy data on the prevalence of smear-positive cases. Styblo first estimated the latter for the period 1921–1937 by assuming that it was four times the mortality rate of *all forms* of tuberculosis in the general population. This implicitly assumed that

1. The prevalence of smear-positive cases in the general population was twice the corresponding incidence (i.e. that cases are infectious for two years), and

2. The incidence of smear-positive disease was twice the mortality rate from all forms of tuberculosis in the general population.

Using this approach, Styblo estimated that each smear-positive case in the Netherlands infected on average 13–14 individuals in a given year during the period 1921–1937. The main shortcoming of these estimates was that the relationship between the mortality rate of all forms of tuberculosis and the prevalence of smear-positive cases is more complicated than that assumed, given that the risks of developing disease are age-dependent and the fact that the age-distribution of smear-positive cases in the general population has changed over time.

Styblo obtained similar values for the 'contagious parameter' in Uganda and Lesotho using estimates for the annual risk of infection in 1960 and the prevalence of smear-positive cases estimated from surveys in 1957 and 1958. The latter were based on small numbers of cases, which restricts their reliability. By assuming that each infectious case is infectious for 2 years in developing countries and infects 10 others in a given year, Styblo estimated that, on average, each smear-positive case leads to 20 new infections.

These analyses were extended by Styblo [53] and then by Murray *et al* [54], who correlated the annual risk of infection to the *incidence* of smear-positive disease by linear regression. This used estimates of the incidence of smear-positive disease for the Netherlands from the pre-chemotherapy era (derived using mortality data, as above), and WHO tuberculosis prevalence survey data from 13 African countries during the late 1950s. These included Nigeria, Bechuanaland and Somaliland, where the annual risks of infection were estimated to be over 4%, over 5% and about 8% respectively.

The analyses of Murray *et al* indicated that each 1% risk of infection corresponded to an incidence of 49 smear-positive cases per 100,000 general population (95% CI: 39 to 59 per 100,000). As pointed out by Rieder [15], these results may be unreliable, given that the authors could not account for variance in sample size, and, where sufficient data were not available, just assumed that the smear-positive disease incidence was half the prevalence, which was perhaps too simplistic. The estimates of Murray *et al* imply that each smear-positive case infects on average about 20 individuals in a given year. This is roughly twice the estimate derived by Styblo earlier.

1.2.2.2 Overview of outbreak literature

Much of the older literature on factors determining the size of an outbreak, as defined by the number of new infections and/or secondary cases found, was reviewed by Lincoln in

1965 [55], who reviewed 109 outbreaks in 12 countries from both before and during the chemotherapy era. To some extent, documented reports of outbreaks provide an overestimate of the number of new infections and/or secondary infectious cases which can result from a given typical infectious case, as only the remarkable ones are generally reported.

There have been few outbreaks in which it has been possible to estimate reliably the number of new infections resulting from an infectious source, given that the tuberculin sensitivity status of individuals prior to exposure is often not known. This is complicated in many instances by the fact that the outbreak is identified at a late stage, when there are several cases of (infectious) pulmonary tuberculosis, and it is impossible to identify a unique source of infection.

A famous outbreak in a Danish school during 1943 described by Hyge [56] is exceptional, in that several weeks prior to the outbreak (December 1942), 105 pupils out of 368 had been tested and found tuberculin-negative (Mantoux, 100TU). During January–February 1943, 70 of these had converted to tuberculin-positivity, with 41 developing lesions detectable by X-ray. In this instance, there had been only one source case (a physics teacher) and transmission had occured over a 2-month period in 2 blacked-out basement teaching rooms with no ventilation, one of which also served as an air-raid shelter.

Most of the outbreaks reviewed by Lincoln occurred among young adults or schoolchildren. In these instances, children were rarely sources of infection. This reflects the rarity of infectious pulmonary disease among children, as compared with its frequency among adults. Whilst outbreaks must have occurred among older adults, they were less likely to be recognized as such, given the high infection prevalence among adults during the period reviewed. More recently, however, with a lower infection prevalence among adults, there have been several documented outbreaks involving adult exposure to smear-positive cases.

Kline [57] documented one such outbreak during the 1990s, in which the source case was a homeless individual who spent most of his infectious period (approximately 6 months) at a neighbourhood bar in Minneapolis. In total, he was estimated to have infected 41 out of a possible 97 contacts (though this could be an overestimate, as only 8 of these were known to be tuberculin-negative before the outbreak).

Reports of outbreaks provide only limited information on the number of (sputum-positive) cases resulting from each infectious source, as exposed individuals are rarely followed up over a sufficiently long period of time. The outbreak documented by Hyge [56]

involved a very long follow-up of up to 12 years. Before the advent of DNA fingerprinting, even when individuals were followed up for long time-periods, it was typically impossible to determine whether or not a disease episode resulted from an initial infection or a subsequent reinfection.

The main factor in determining the number of secondary (infectious) cases resulting from a given source is the age of those exposed. All the exposed individuals in the outbreak documented by Hyge [56] were aged 14–18 years, and therefore faced high risks of developing disease. As a result, a large number (14) of cases of pulmonary tuberculosis occurred among those presumed to have been infected by the source, with 8 doing so within a year of exposure. It is unclear exactly how many of these cases were sputum-positive, as 8 had a documented positive gastric lavage, all were defined as having 'post-primary pulmonary tuberculosis', but 8 involved cavitation.

Similarly, adults comprised most of those exposed to the infectious source in the Minneapolis bar [57], and in total, 14 active tuberculosis cases developed among the contacts within of 2–3 years. At least eight of these cases had been infected/reinfected by the source case, as determined by identical RFLP patterns. It is not known whether this large number of secondary (sputum-positive) cases was a consequence of an increased susceptibility to infection and subsequent disease among heavy alcohol users in the bar, or the high risks of developing (sputum-positive) disease following initial infection or reinfection in adult life.

The number of secondary cases resulting from a given infectious source also depends on the prior tuberculin sensitivity status of exposed individuals, given the different risks of developing disease after first infection or of reinfection followed by subsequent disease. There have been few documented outbreaks in which exposed individuals developed exogenous disease. One famous exception involved homeless individuals in a shelter in Boston during the 1980s, in which one index case lead to 7 cases of exogenous disease, as determined by the similarity between phage-types and drug-resistance patterns [21].

Since the 1950s, the number of smear-positive cases resulting from an infectious source has also depended on whether exposed individuals received chemoprophylaxis after exposure. In an outbreak in 1974 in Corinth (Mississippi) [58], one adolescent source case led to 4 other pulmonary cases within 2 years of exposure, as determined by the similarity between phage types and drug-sensitivity patterns; 2 of these cases had been prescribed isoniazid at the time of exposure, but uptake had been irregular.

The *total* number of secondary cases resulting from a given infectious source also depends on whether he/she has several (infectious) disease episodes and the number of individuals infected during each of these periods of infectiousness. The index case in the outbreak in the Boston shelter [21] had been treated for culture-positive tuberculosis at least once before the outbreak, but it is not known how many individuals had been infected during the earlier disease episode.

More recently, there have been several outbreaks involving multi-drug resistant organisms. Most of these have been among HIV-positive individuals and involved either hospitalized patients [22, 23, 42, 59] or prison inmates [30, 43], who have different mixing patterns from individuals in the general population. In practice, the number of secondary (infectious) cases resulting from an infectious source in an HIV-positive population depends on many factors, such as the stage in the HIV-infection at which infection with *M. tuberculosis* is acquired. Those who develop tuberculosis early in the course of HIV-infection, for example, are more likely to be smear-positive as compared with those doing so with a low CD4+ T lymphocyte count [60]. Several of these outbreaks have been remarkable in demonstrating the rapid development of disease among HIV-positive individuals. In the outbreak documented by Coronado *et al* [23], for example, three HIV-positive individuals developed multi-drug resistant disease, with identical RFLP patterns, within 5 months of exposure to the index case.

1.2.2.3 Concluding remark

Overall, we see that the number of infections and secondary infectious cases resulting from a given infectious source in the outbreak setting are subject to a large degree of variability, and it is difficult to make generalizations as to the magnitude of the basic or net reproduction numbers for tuberculosis on the basis of such data.

1.3 Models for tuberculosis

Models for tuberculosis fall into several categories, of which transmission models comprise the largest group. These describe the disease process in a population by modelling the transitions of individuals between at least three different epidemiological states, namely:

1. Those not yet infected with the tubercle bacillus,
2. Those infected, but who have not yet developed an infectious form of disease,
3. Infectious cases.

Possible transitions between these groups of individuals are illustrated schematically in Figure 1.7. In this simple model, infectious cases may recover and experience further disease.

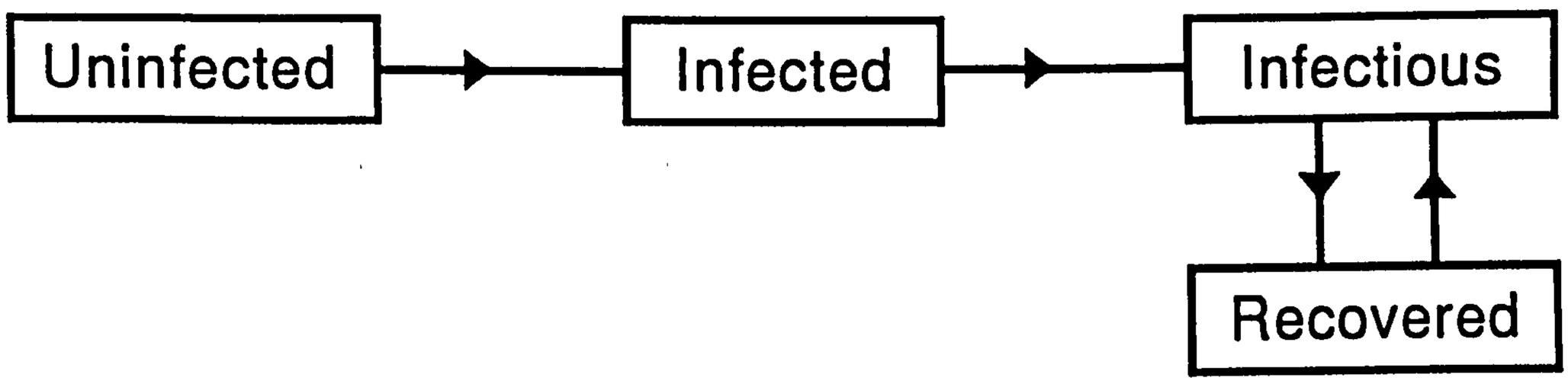


Figure 1.7: Schematic diagram showing the structure of simple models of the transmission dynamics for *M. tuberculosis.*

Models of the transmission dynamics of *M. tuberculosis* have generally used the term 'susceptible' to describe individuals who have not yet been infected with the tubercle bacillus. This term should be used with caution when considering the natural history of tuberculosis, given that even infected individuals can be 'susceptible', as reinfection can occur. As we see below, the models published by other authors ignored the effect of reinfection on the transmission dynamics of *M. tuberculosis*, and hence assume that only uninfected individuals are 'susceptible'.

In most transmission models, the rate at which 'susceptible' individuals become infected is a function of the prevalence of infectious individuals, according to the mass action principle, first postulated by Hamer [61]. Where the endemicity of infectious pulmonary tuberculosis is high, as in some developing countries today or in developed countries at the

beginning of this century, the implicit assumption of homogenous mixing between infectious and 'susceptible' individuals may be realistic. This is not the case in populations with a low prevalence of infectious pulmonary tuberculosis, since transmission is likely to be confined to certain high risk groups, in which the risk of infection greatly exceeds the average.

Transmission models can be categorized according to whether they consider a population with or without any age-structure i.e. whether or not transitions between the epidemiological groups are age-dependent. One of the drawbacks of age-structured models is that more data or assumptions are required to specify the characteristics of the population, as compared with models without age-structure. Published examples of these two categories of models are reviewed below.

A second and smaller group of models for tuberculosis comprises those developed for parameter estimation [7,62]. These follow a statistical approach, and are generally simpler than transmission models. These are discussed in section 1.3.2.

1.3.1 Transmission models for tuberculosis

1.3.1.1 Models considering populations without age-structure

The first and simplest model for the transmission dynamics of *M. tuberculosis* in a population without age-structure was produced by Waaler *et al* in 1962 [63]. This consisted of four difference equations and modelled the transitions between 'susceptible' (defined as uninfected), infected and infectious individuals. The structure of the model was the same as that shown in Figure 1.7, except that recovered individuals returned to the infected category and faced the same risks of developing disease as did other infected individuals.

The initial parameters of the model population were defined using data from a survey carried out during the period 1951–1955 in South India [64]. The model was used to predict the incidence of infectious pulmonary tuberculosis over 20 years, with either no intervention, or with a case-finding and treatment programme or with BCG vaccination. This model was designed only to illustrate the potential contribution of modelling approaches for evaluating different control programmes for tuberculosis.

Models produced by Azuma in 1975 [65], Hock and Loy in 1981 [66] and Trefny and Hejdova in 1982 [67] had the same structure as the model of Waaler [63], and made long-term predictions of the incidence of 'bacteriologically confirmed' cases in Japan from 1953 [65],

in Singapore from 1975 [66] and in the Czech republic from 1949 [67]. All these models were simplistic and their main weakness was that they assumed all infected individuals experienced the same unchanging risk of developing disease over the entire period simulated. This was given by the ratio between the crude incidence of infectious pulmonary tuberculosis and the prevalence of infection in the population at the start of the period, estimated using survey data.

Given the complex natural history of tuberculosis, the average risk of developing disease in a population is a function of the individual risks of developing 'primary' and 'endogenous' disease and of reinfection, followed by 'exogenous' disease, and the number of individuals experiencing each form. These depend most importantly on the age distribution of individuals at risk of developing each of these disease forms and the annual risk of infection. Since neither of these factors remains at a constant level over a long-term period, the predictions from the early models are unlikely to have been reliable. It is also interesting to note the different annual risks of developing disease estimated for these models using survey data, namely 0.85% for South India, 0.05% for Japan, 0.1% for Singapore and 0.14% for the Czech Republic.

Subsequent variants of Waaler's basic model partially remedied this problem regarding unchanging risks of developing disease (e.g. ReVelle *et al* [68] and Chorba and Sanders [69]). The model of ReVelle *et al*, for example, which explored the effect of combinations of control strategies (BCG vaccination, chemoprophylaxis) on the disease incidence in a developing country setting, subdivided infected individuals into those with high and low risks of developing disease, according to whether or not they had received chemoprophylaxis. Additionally, diseased individuals recovered either naturally or after treatment, and faced different risks of developing disease from those who had yet to develop disease.

The model of Chorba and Sanders in 1971 [69] was similar to that of ReVelle *et al*, except that it broke the infected category down still further into those with a recent or a more distant infection. The latter were at a high or a low risk of developing disease according to whether or not they had pulmonary lesions. The structure of the model is illustrated in Figure 1.8. This model explored the effects of vaccination or administering chemoprophylaxis on the incidence of disease in the US during the period 1967–1987 and assessed the costs and benefits of screening for infection in high and low infection prevalence areas.

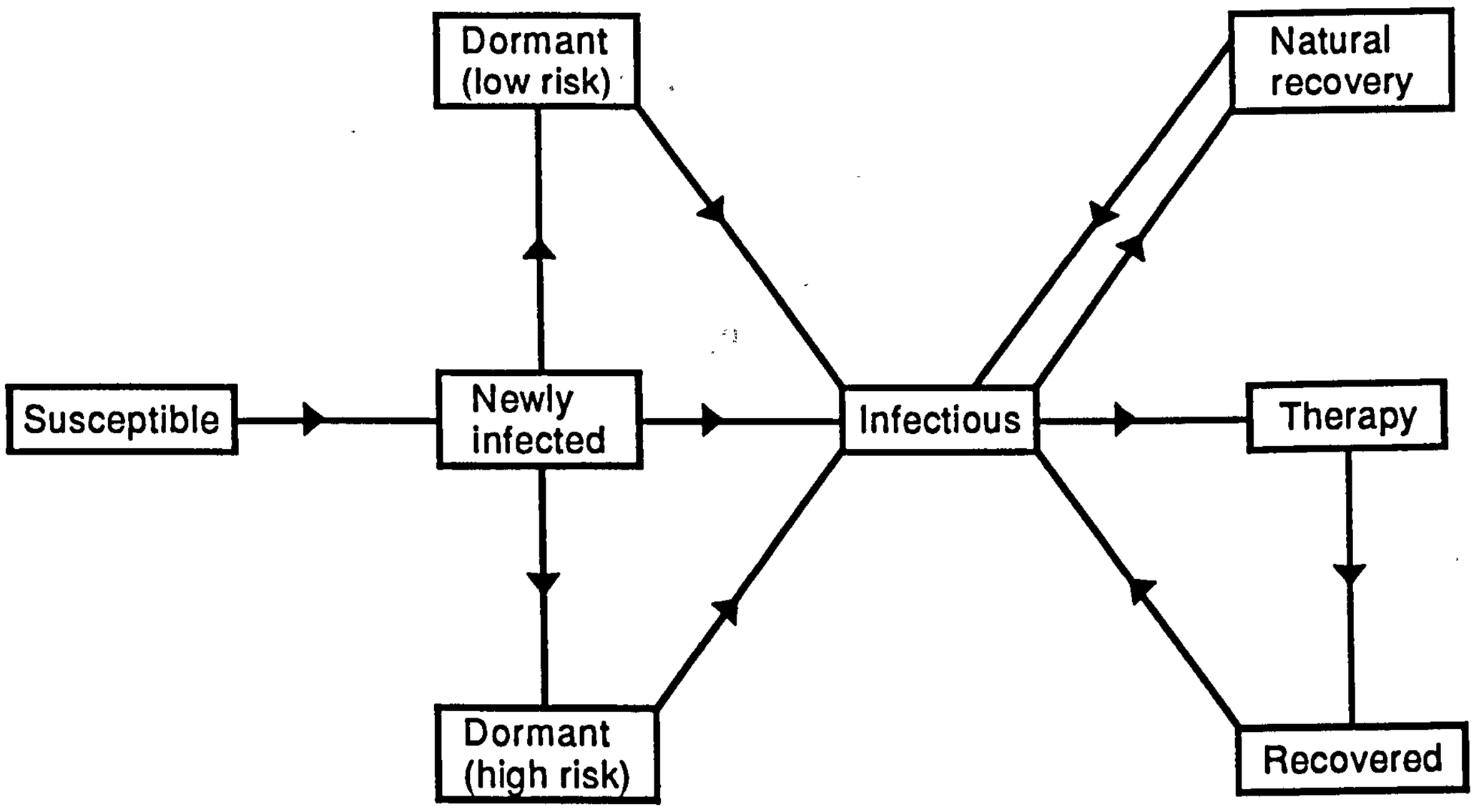


Figure 1.8: Schematic diagram showing the structure of the model produced by Chorba and Sanders in 1971 [69].

The model of Chorba and Sanders was the first to include a separate (disease) state consisting of individuals with pulmonary lesions. The model of Joesoef *et al* in 1989 [70] similarly assumed that infected individuals developed pulmonary lesions before developing bacillary disease. The structure of this model is shown in Figure 1.9. Whilst this is a realistic assumption, it requires reliable data on the prevalence of radiological cases, which is often difficult to acquire. Joesoef *et al* also assumed that vaccinated individuals faced unchanging risks of developing pulmonary lesions (see Figure 1.9), which, in practice, should have been a function of the annual risk of infection. The model of Joesoef *et al* made projections of the incidence of cases with active pulmonary lesions and bacillary disease in Indonesia during the period 1983–2003, and investigated the cost-effectiveness of different treatment strategies.

More recently, models have been developed by Schulzer *et al* [71,72] and Heymann [73] to assess the impact of the HIV epidemic on the tuberculosis situation in Africa. These describe only adult populations (15–49 year olds), and assume that parameters do not vary between different age groups in this age band. Hence these models also fall into the category of non-age-structured models.

In the model of Schulzer *et al*, first published in 1992 [71] and presented in greater detail

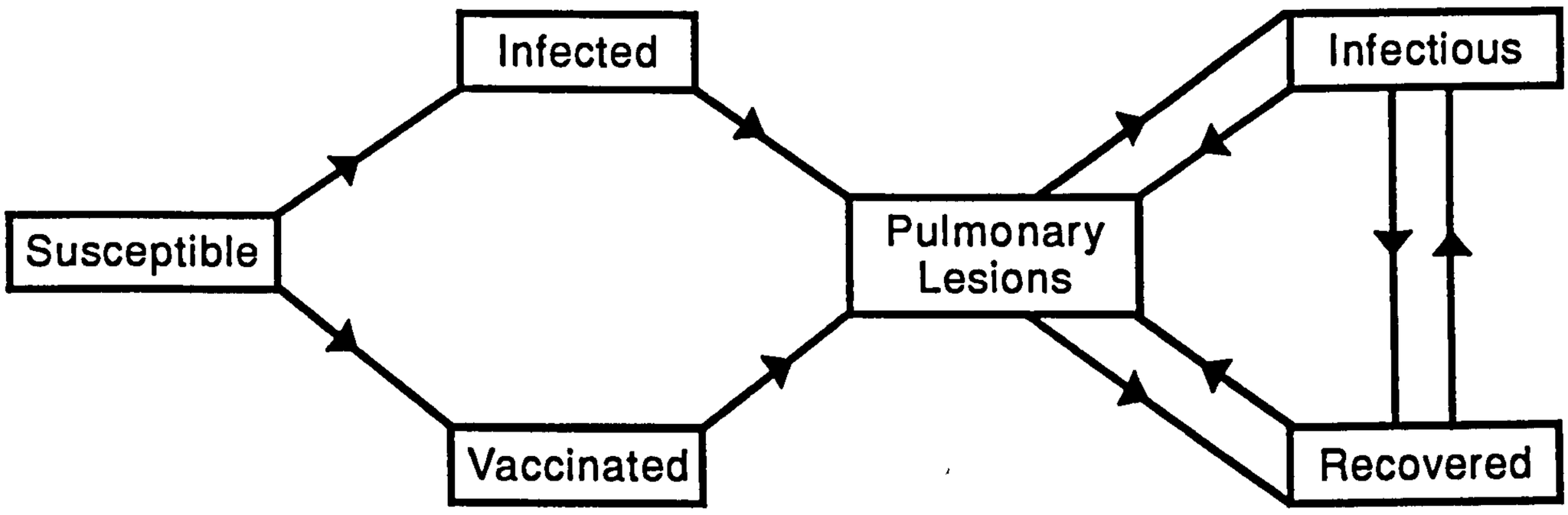


Figure 1.9: Schematic diagram showing the structure of the model produced by Joesoef in 1989 [70].

in 1994 [72], individuals could develop either smear-positive or smear-negative disease, and the risk of doing so depended on the time since first infection. Thus 5% of HIV-negative individuals, for example, developed disease during the first year after infection, 2% during the second year, 0.5% during the third year etc. The risk of developing disease among HIV-positive individuals was assumed to be about 8 times greater than that among HIV-negative individuals during each year after infection. This corresponded to a 42.1% risk among HIV-positive individuals during the first year after infection (which is very high). A further drawback of this model was that they did not consider the role of reinfection, which may well play an important role in the epidemiology of tuberculosis in developing countries.

Schulzer *et al* assumed that any increase in the annual risk of infection occurred as a consequence of new smear-positive cases among HIV-positive individuals, and neglected the contribution from new smear-positive cases among HIV-negative individuals. They also assumed that an annual risk of infection of 1% corresponds to a smear-positive incidence of roughly 50 per 100,000, which is based on a result from linear regression analyses of incidence and infection risk data from both developing and developed countries [34, 53, 54]. The results from this regression analysis are themselves controversial [15] (see section 1.2.2).

The model predicted that with an annual risk of tuberculous infection of 1%, a tuberculous infection prevalence of 45% and a 2% prevalence of HIV infection in 1989, there would be a 68% increase in the number of new cases of smear-positive tuberculosis among adults by the year 2000. The model also considered the effect of HIV infection on the incidence of tuberculosis among drug users in Vancouver. This found that with an annual risk of infection of 0.01%, a 10% prevalence of tuberculous infection, and HIV infection prevalence

of 3.6%, a 21-fold increase in the incidence of smear-positive cases would occur between 1995 and 2000.

The model of Heymann in 1993 [73] used a Markov chain formulation. This categorized HIV-positive and negative individuals into those susceptible to tuberculous infection, those recently infected, those infected for more than 2 years and those with 'active' tuberculosis. HIV-positive individuals had higher risks of developing disease, as compared with HIV-negative individuals. The objective of this model was to assess the impact of chemoprophylaxis and treatment on tuberculosis mortality in Africa. However, it did not incorporate disease through exogenous reinfection and it did not distinguish between smear-positive and smear-negative forms, which restricted the validity of the conclusions.

The most recently published model of the transmission dynamics of *M. tuberculosis*, that of Blower *et al* [9], also considered a population without age-structure. This categorized individuals into those who were 'susceptible' (defined here as uninfected individuals), 'latently infected', infectious or non-infectious cases, and those who were recovered. 0–30% of individuals were assumed to develop disease within a year of infection and reinfection was not incorporated. This model differed from those described thus far as it did not model the tuberculosis dynamics in any specific population and instead explored the long-term dynamics of disease.

The authors simulated an epidemic following the introduction of one infectious individual into an uninfected population, and found that the incidence of infectious cases initially increased, peaked and then stabilized at an equilibrium level within 31 to 7,524 years, assuming that all conditions remained constant.

Blower *et al* concluded that tuberculosis epidemics have slow intrinsic dynamics and suggest that the decline in tuberculosis over the past century in developed countries reflects the natural decline in a long epidemic. This interpretation of the decline in tuberculosis was also postulated by Grigg in 1958, who proposed that the incidence of tuberculosis first increased during the 1600s in developed countries [74].

The main shortcoming of this model was that it oversimplified the natural history of tuberculosis by neglecting to consider an age-structured population and the contribution of disease following reinfection.

1.3.1.2 Models with age-structured populations

The first model to consider an age-structured population was produced by Brogger in 1967 [75]. The population was stratified into single year age bands for those aged under 15 years and into 5 year age bands for 15–89 year olds. The model was similar in structure to that of Joesoef *et al*, in that individuals first entered an intermediary stage containing those with pulmonary lesions, before developing infectious pulmonary disease. Brogger did not distinguish between 'primary', 'endogenous' and 'exogenous' forms of disease and assumed that the risk of developing disease was age dependent but independent from the time since first infection. The structure of the model is shown in Figure 1.10. This model explored the effects of BCG vaccination, case-finding and chemotherapy in Thailand, using data from the national survey carried out during the period 1960–1963.

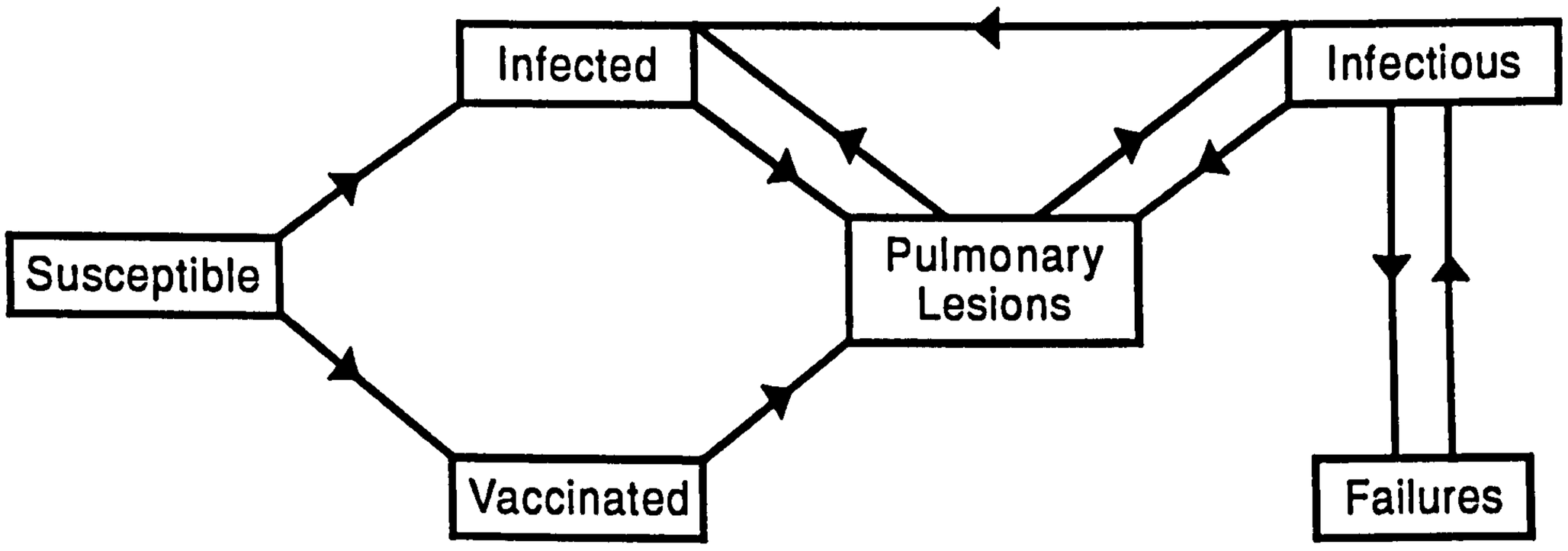


Figure 1.10: Schematic diagram showing the structure of the model produced by Brogger in 1967 [75].

The framework for a more realistic age-structured model was produced by Waaler in 1968 [76]. This extended Brogger's work and Waaler's earlier model [63]. The population was stratified into 5 year age groups. Individuals faced different risks of developing disease according to whether they had been infected for more or less than 5 years, and could develop either smear-positive or smear-negative disease. This was the first model to distinguish between primary and post-primary disease. The overall structure of the model is illustrated in Figure 1.11. In this theoretical framework, Waaler allowed post-primary disease to occur either through endogenous reactivation or exogenous reinfection.

The model was first applied to evaluate the cost-effectiveness of BCG vaccination and its effect on the long-term incidence of disease in a country with a low annual risk of infection,

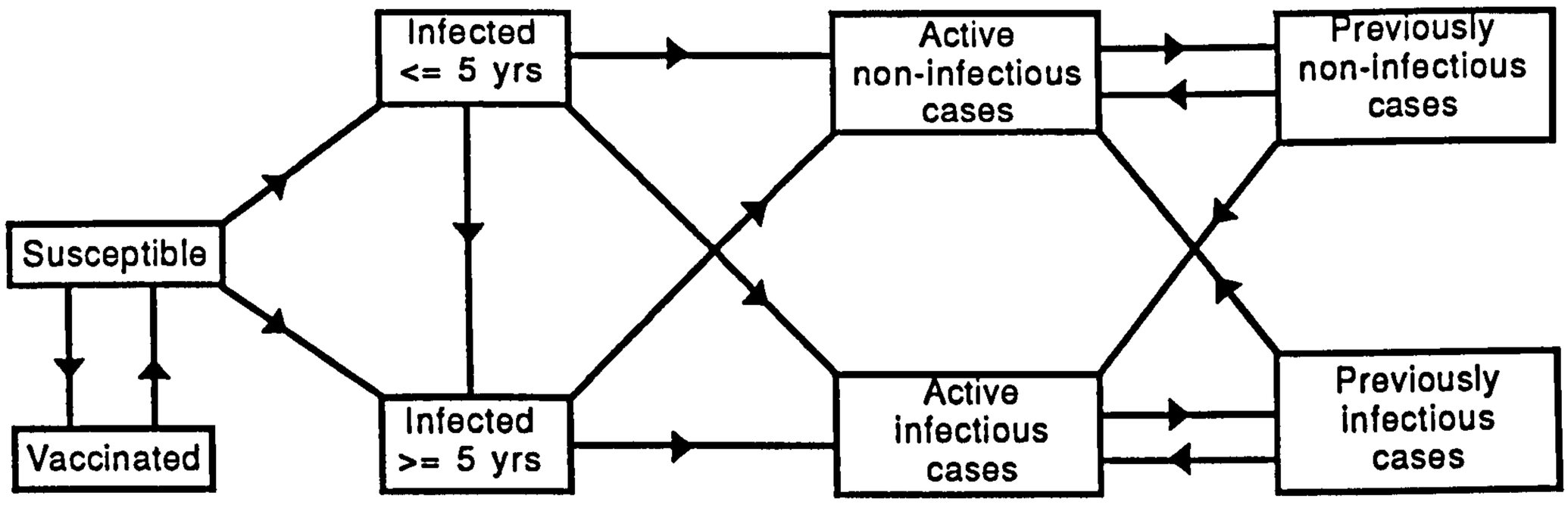


Figure 1.11: Schematic diagram showing the structure of the age-structured model produced by Waaler in 1968 [76].

similar to Norway during the 1970s [77–79]. The effect of BCG vaccination in a region similar to rural South India, was also explored [80]. This analysis was based using data from the longitudinal 5-year survey of tuberculosis carried out in Bangalore during the period 1961–66 [25].

In both applications, Waaler allowed individuals to develop disease only through endogenous reactivation. This was a reasonable assumption to make for Norway, given its low risk of infection during the 1970s. The situation in Bangalore, however, was vastly different, with a high annual risk of infection (1–2%), and with much disease among infected individuals probably occurring through exogenous reinfection. This led to unreliable conclusions from the model, namely that with a vaccine efficacy of 30% and an annual coverage of 66% among 0–20 year olds, the incidence of bacillary cases would decrease to less than one third of its initial level within 40 years. Since then, studies have also shown that the effects of BCG vaccination are unpredictable and may be temporary [81].

Waaler also provided a method for assessing the impact of different control programmes in a population by defining a 'TB problem' as the 'cumulative sum, over all future years of man-years of tuberculosis experienced by a community' [76]. Waaler later refined this definition to assign different weights for cases occurring many years or shortly after the start of a control programme [82]. This measure was used by Azuma [65] and Hock and Loy [66] to assess the relative impact of different combinations of treatment efficiency and BCG vaccination on the tuberculosis situation in Japan and Singapore respectively.

Whilst this is a useful measure of a 'TB problem', it depends on assumptions in the model, i.e. whether all infected individuals experience the same risk of developing disease,

irrespective of their age and the time since infection, and whether individuals can develop disease through both exogenous reinfection and endogenous reactivation.

The most recent age-structured model of the transmission dynamics of *M. tuberculosis*, which also used data from the longitudinal survey in Bangalore, was produced by Balasangameshwara *et al* in 1992 [83]. This divided the population into broader age-bands than used by Waaler [76] (i.e. into 0–4, 5–14, 15–34, 35–54 and >55 year olds). The risks of developing disease were assumed to be constant over time and thus Balasangameshwara *et al* did not distinguish between exogenous and endogenous disease forms. The model explored the effect of alternative coverage levels of active case-finding and treatment on the prevalence of smear and/or culture-positive cases over a 50 year time period.

1.3.2 Models for parameter estimation

Styblo *et al* first used a model-based approach to estimate the annual risk of infection in the Netherlands since 1911 [62]. This relied on two data sets, namely:

1. 4 tuberculin surveys (using the von Pirquet test) of 0–14 year old children who were not home contacts of tuberculosis cases, which were carried out by the Amsterdam Chest Clinic during the periods 1925–1927, 1933–1935, 1938–1940 and 1946–1948, and

2. routine tuberculin sensitivity surveys among Dutch military recruits, aged about $19\frac{1}{2}$ years, during the period 1956–1966. Individuals were considered infected if they responded with an induration of 8mm or more to a Mantoux test of 1TU of RT23.

BCG was not used as a routine general control strategy in the Netherlands, and so the survey data provided a reliable indication of the prevalence of 'infection' in these age groups. The data sets were analyzed separately using the same approach.

Considering the prevalence surveys among Dutch military recruits, Styblo *et al* first estimated the average annual risk of infection experienced by each cohort during its lifetime. These were assumed to be the actual risks experienced in some year during the cohort's lifetime, and were found to decrease at a constant rate. These risks were then extrapolated back and assigned to particular years by fitting the expected prevalence of infection to the prevalence of tuberculin sensitivity obtained at each survey. Combining these results with those from the survey among children in Amsterdam, Styblo *et al* found that the annual risk of infection declined at about 5.5% pa until 1940 and at about 13% thereafter. The

faster decline after 1940 was attributed to the enforced pasteurization of milk during the Second World War. Sutherland *et al* subsequently extended these analyses to estimate the risks of infection until 1979 in the Netherlands [84].

Styblo *et al* discussed some of the problems with this analysis, such as the reliability of the von Pirquet method used for detecting tuberculous infection among children in Amsterdam and possible age-specific fluctuations in the annual risk of infection. A further problem with the analysis is that the data sets were analyzed independently. In estimating the prevalence of infection among recruits tested in 1956, (who were born in 1937) Styblo *et al* did not use the annual risks of infection estimated using the surveys of school-children in Amsterdam, for the period 1937–1940. This would not have substantially altered the risk estimates obtained.

The work of Styblo *et al* constituted the first attempt at quantifying past trends in transmission of tubercle bacilli in any country, and represents an extremely important contribution to tuberculosis epidemiology.

Sutherland *et al* used these estimates of the annual risk of infection to estimate the individual risks of developing 'primary', 'endogenous' and 'exogenous' disease among 15–69 year olds during the period 1951–1970 in the Netherlands [7]. In the first version of the model, presented in 1975 [85], Sutherland *et al* assumed that individuals faced

1. a constant annual risk r_1 of developing ('primary') disease during the first five years after infection,

2. an annual risk r_2 of developing ('endogenous') disease more than 5 years after first infection, and in the absence of reinfection, and

3. a constant annual risk r_3 of developing ('exogenous') disease during the first five years after reinfection, for those who had experienced an initial infection more than 5 years beforehand.

These risks were assumed to be identical for all age groups throughout the time period considered. Sutherland *et al* expressed the expected incidence of disease for each age group and for each year as the sum of the corresponding incidence of 'primary', 'endogenous' and 'exogenous' diseases. These were given by the product of

1. The population in each age group at risk of developing the given disease form, X_1, X_2 and X_3, estimated using the annual risks of infection derived by Styblo *et al*, and

2. The individual risks of developing disease, r_1, r_2 and r_3.

This yielded a system of 220 linear equations in r_1, r_2, and r_3, for the expected incidence of disease in each 5-year age group between 1951 and 1970. Multiple regression techniques were used to estimate the risks r_1, r_2, and r_3, by fitting the expected incidence to the corresponding notification rates of respiratory disease in males and females (separately) in the Netherlands.

These produced the following annual risks of developing respiratory disease in 15–69 year old males:

1. 5.01% per annum during the five years after first infection,

2. 1.91% per annum during the first five years after reinfection,

3. 0.0253% per annum following a distant primary infection with no reinfection.

This corresponds to a risk of 23% and 9% of developing 'primary' and 'exogenous' diseases during the first five years after first infection and reinfection respectively. The risk of developing primary disease appears high, compared with a risk of developing disease of about 10% during a ten year period for those who 'converted' to tuberculin sensitivity during the UK MRC BCG trial (see section 1.2.1.1).

Sutherland *et al* also considered the effect of modifying the assumptions in the basic model on the disease risk estimates, e.g. assuming an initial infection conferred complete, partial or no immunity against subsequent reinfection, or that individuals infected for more than 25 years without subsequent reinfection faced no risk of developing endogenous disease. Further refinements documented in a TSRU (Tuberculosis Surveillance Research Unit) working paper in 1975 [86] show the effects of assuming age-dependent risks of infection and a declining risk of developing disease during the first ten years after first infection. Some of the main findings were summarized by Sutherland in 1982 [7].

Although each of these were realistic extensions of the basic model, they were not incorporated simultaneously, and better fits were obtained only by assuming age-dependent risks of infection.

The chief shortcoming of this approach for estimating risks of developing disease, is that it does not account for the prior disease status of individuals, nor their recovery time, which means that the individuals at risk of developing each form of disease are only crudely defined.

Those infected for less than five years, for example, comprise those at risk of developing 'primary' disease — this definition encompasses individuals who had been infected for 3 years, for example, and had developed disease during the previous year, and who in practice would have been recovering. A further problem relates to the use of the time since infection for defining individuals at risk of developing exogenous and endogenous disease. We discuss this in greater detail section 3.1.1.

Krishnamurthy and Chaudhuri [87] used a similar approach to that of Sutherland *et al* [7] to estimate the risks of developing endogenous and exogenous smear/culture-positive disease among 'infected' individuals in South India. They used data from the 5-year longitudinal tuberculosis survey carried out during the period 1961–1966 in Bangalore [25]. This fitted the expected incidence of smear/culture-positive disease among individuals considered infected (10^{+}mm induration to 1TU PPD RT23) to the observed incidence during two time periods and yielded estimates of 6.55% and 0.21% for the risks of developing exogenous and endogenous disease during one and a half years.

It is difficult to interpret these disease risk estimates, given that they are based only on 2 data points, and that those in the infected category included individuals with signs of active disease, as diagnosed by X-ray.

Schulzer *et al* [88] employed a similar approach to estimate the risk of uninfected individuals developing drug-sensitive or drug-resistant disease during a five-year period in Korea. This used data from tuberculosis surveys carried out at 5-year intervals during the period 1961–1982, which provided information on drug-resistant and drug-sensitive cases which had emerged, or had remained prevalent between surveys in the population.

Schulzer *et al* assumed that new drug-sensitive/resistant cases developed as a consequence of mixing between uninfected individuals and those drug-sensitive/resistant at the preceding survey. The corresponding combined risk of infection and of developing disease of the uninfected individual depended on the duration of time during which the case had been prevalent and his/her drug-sensitivity.

Whilst this is a realistic assumption, Schulzer *et al* neglected the age of uninfected individuals at risk of infection and disease, and did not consider the effect of reinfection on the emergence of new cases.

1.3.3 Conclusions

The objectives of most of the models of the transmission dynamics described in this section were to forecast the likely incidence of tuberculosis under various settings and thereby to assess the potential impact of different control programmes. Their main shortcoming was that they oversimplified the basic natural history of tuberculosis and were overambitious in their aims. Only three transmission models [75,76,83], for example, considered the dynamics in age-structured populations, and only the statistical models of Sutherland *et al* [7] and Krishnamurthy and Chaudhuri [87] considered the effect of reinfection on the disease incidence. As we saw in section 1.2.1, both age and reinfection are important in determining the risks of developing disease experienced.

Intuitively, before we can assess the likely impact of a given intervention on the tuberculosis situation in a given setting, it is necessary to understand how different factors, such as age, the different risks of developing 'primary', 'endogenous' and 'exogenous' disease and trends in the risks of infection determine the dynamics underlying the disease incidence. This has been the objective of the work described in this thesis.

Chapter 2

Epidemiology of tuberculosis in England and Wales — data sources and trends

2.1 Mortality rates

Age and cause-specific mortality rates were first published in England and Wales in 1838. These have been published in the Annual Report of the Registrar General for England and Wales until 1974 and in the OPCS Monitor thereafter. The numbers of deaths by age, sex and cause by single year since 1901 are also available in electronic form from OPCS as 'historic data' files.

From 1847 until 1901, mortality rates among 0–24 and 25–74 year olds were stratified into 5 and 10 year age bands respectively, and those among individuals aged over 75 years were combined into a single age band. Since 1901, the rates have been stratified into smaller age bands.

Published mortality rates distinguished between 4 forms of tuberculosis until 1901, namely 'Pulmonary tuberculosis and phthisis (not otherwise defined)', 'Tuberculous meningitis', 'Tuberculous peritonitis, tabes mesenterica' and 'Other tuberculous diseases'. Tuberculosis mortality rates from 1901 have been classified using the 'International Classification of Disease' codes (ICD codes), which have since undergone nine revisions. During the periods 1901–1910 and 1940–1949, 9 and 15 ICD codes respectively were used to categorize

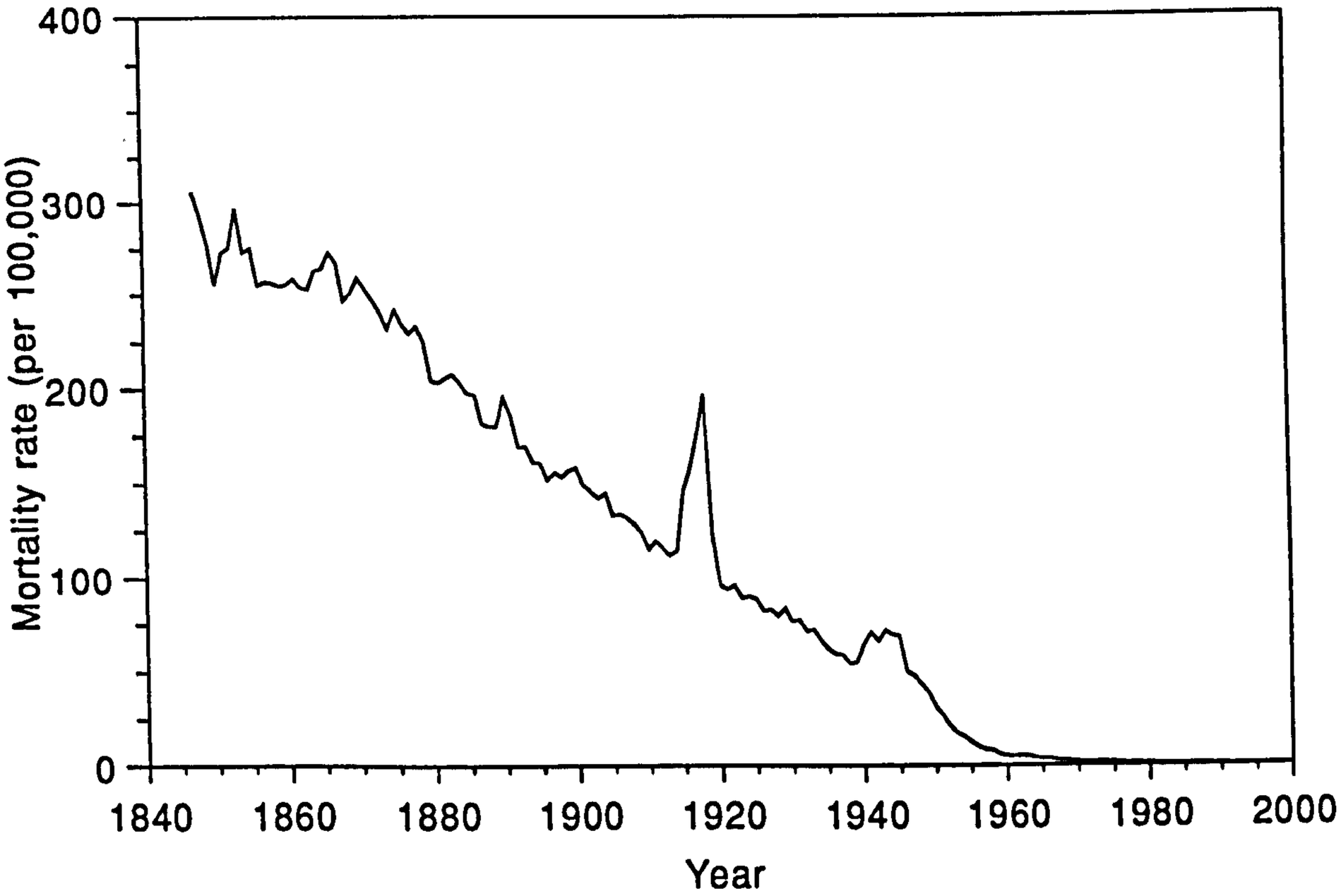


Figure 2.1: Mortality rates from respiratory tuberculosis in males in England and Wales during the period 1847–1990. (Age-standardized to the 1901 male population.)

tuberculosis deaths, as compared with the 47 currently in use.

Figure 2.1 summarizes the annual mortality rates from respiratory tuberculosis in males between 1847 and 1990, which have been age-standardized to the population in 1901. This shows that the mortality rate in England and Wales began to decline at least as early as the mid 1850s, despite the absence of effective treatment or vaccination. The peaks in mortality rates during the periods 1914–1918 and 1940–44 reflect increased mortality during the two World Wars. The dramatic decline in the mortality rate after 1950 occurred as a consequence of the introduction of specific antimycobacterials for treating tuberculosis (see section 1.1).

The corresponding mortality rates within individual age groups during the period 1850–1990 are summarized in Figure 2.2. These declined within each age group in a similar way to the mortality rates in the general population (see Figure 2.1), with sharp increases, especially among 15–24 and 25–34 year olds during the two World Wars. This excess mortality among young adults during the two World Wars is attributable to at least two factors.

First, the rates refer to the civilian population only. This comprised a greater proportion of individuals with high risks of developing disease than during peacetime, since 'healthier'

individuals were conscripted into the army.

Second, some deaths which occurred in or after combat, or as a consequence of desertion could well have been deliberately misdiagnosed as tuberculosis. This would have affected the mortality rates among young adults, given that they composed most of the military force.

Figure 2.2 shows that throughout the time period considered, 0–14 year olds experienced lower mortality rates than did any other age group. The mortality rates also declined more rapidly for these individuals than for older individuals. This is seen more easily in Figure 2.3, which plots the age-specific mortality rates on a log scale. The reasons behind this are complicated and relate to the decline in the risk of infection and how it determined the age at initial infection. We return to this issue in greater detail in section 3.4.2.

The decline in the risk of infection also led to changes in the age-specific pattern in the mortality rates when plotted by year and birth cohort. This is illustrated in Figure 2.4a, which plots the mortality rates by year for the period 1881–1935. For all the years shown, the age-specific mortality rates increased from adolescence to peak at about 40–50 years of age. The pattern changed during the first half of the 20th century, with relatively less decline among young adults than those of middle age. When these rates are plotted by *cohort* (Figure 2.4b), the peak in mortality shifts from individuals aged 30–35 years for the early cohorts (born 1881–1890), to young adult life for those born after 1901.

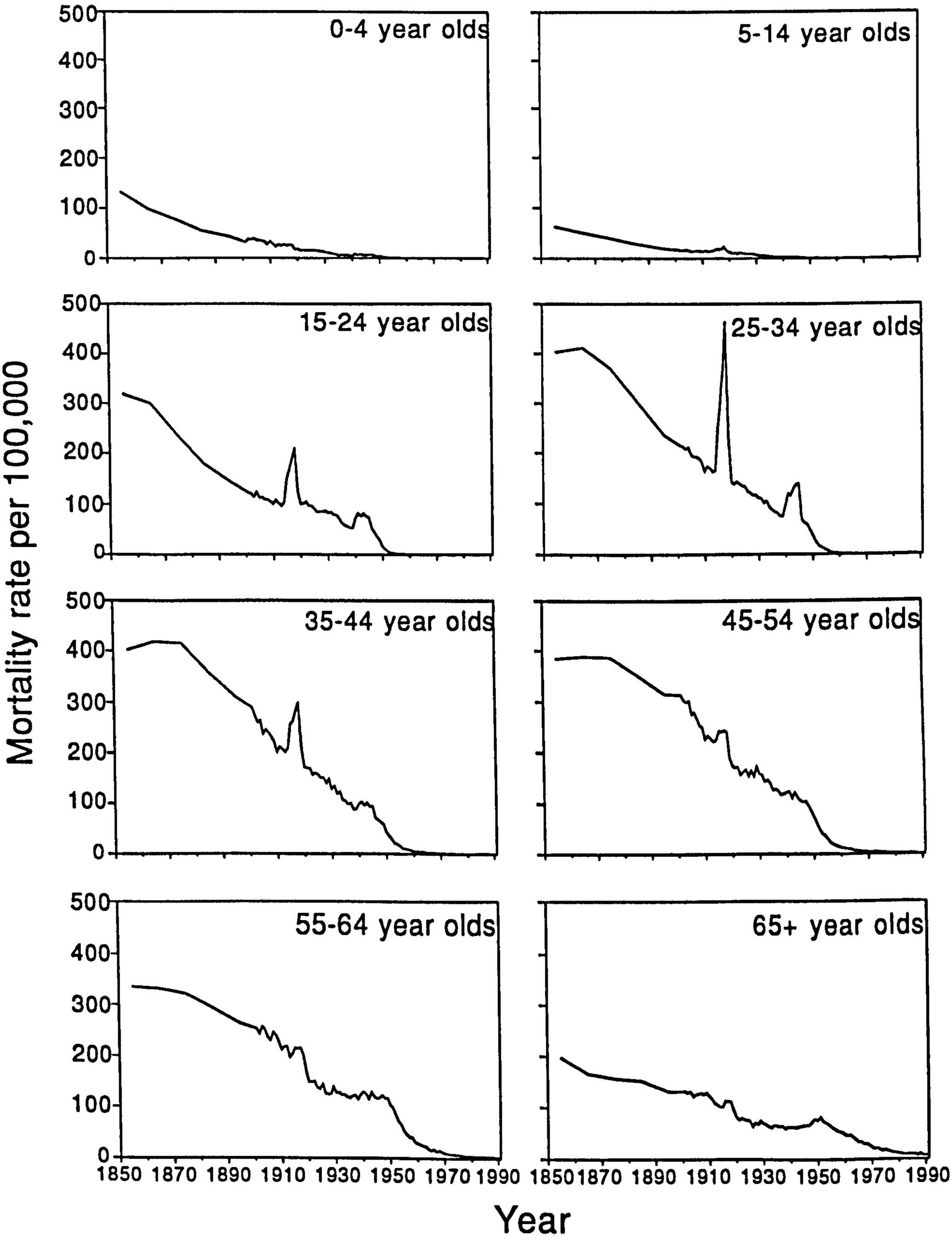


Figure 2.2: Age-specific mortality rates from respiratory tuberculosis in males in England and Wales during the period 1850–1990.

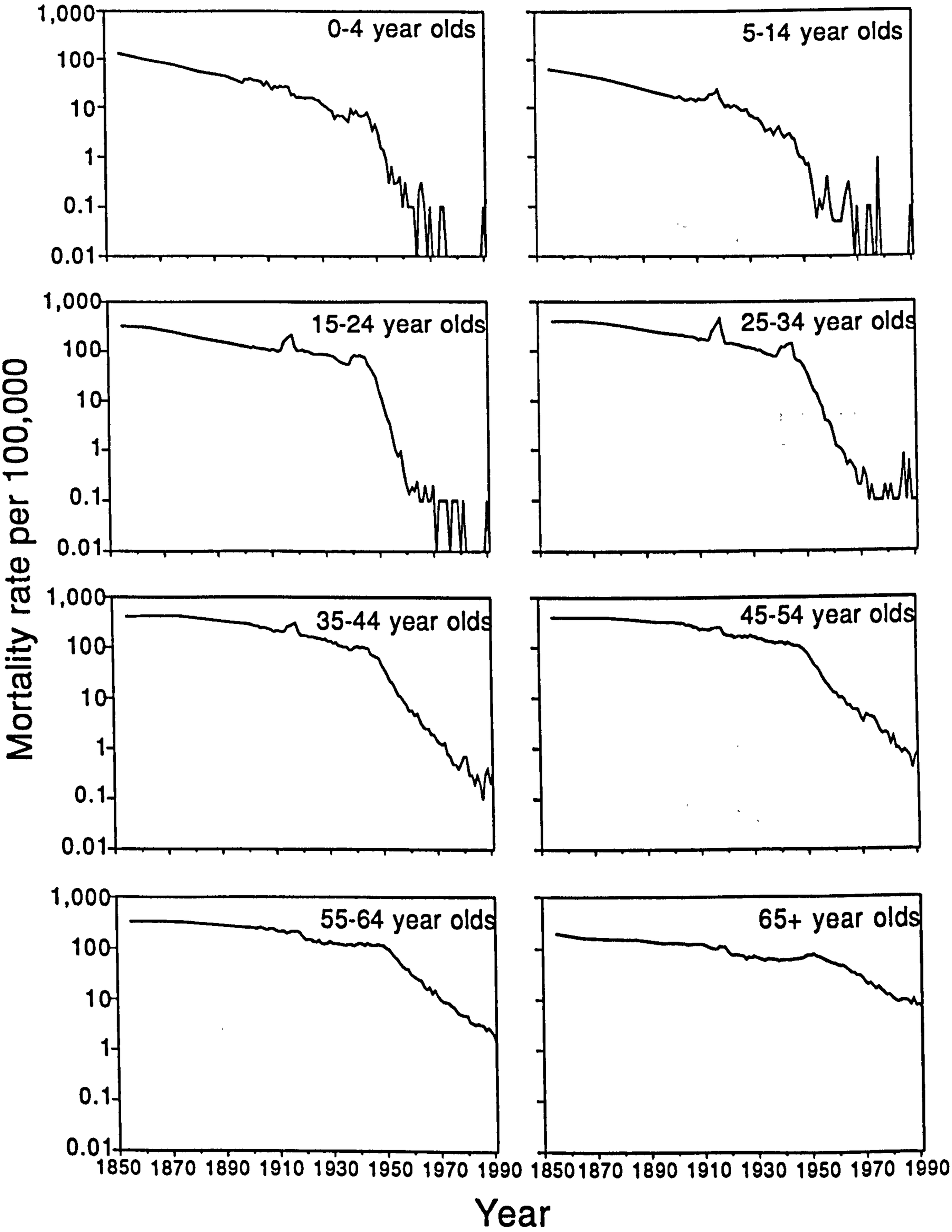
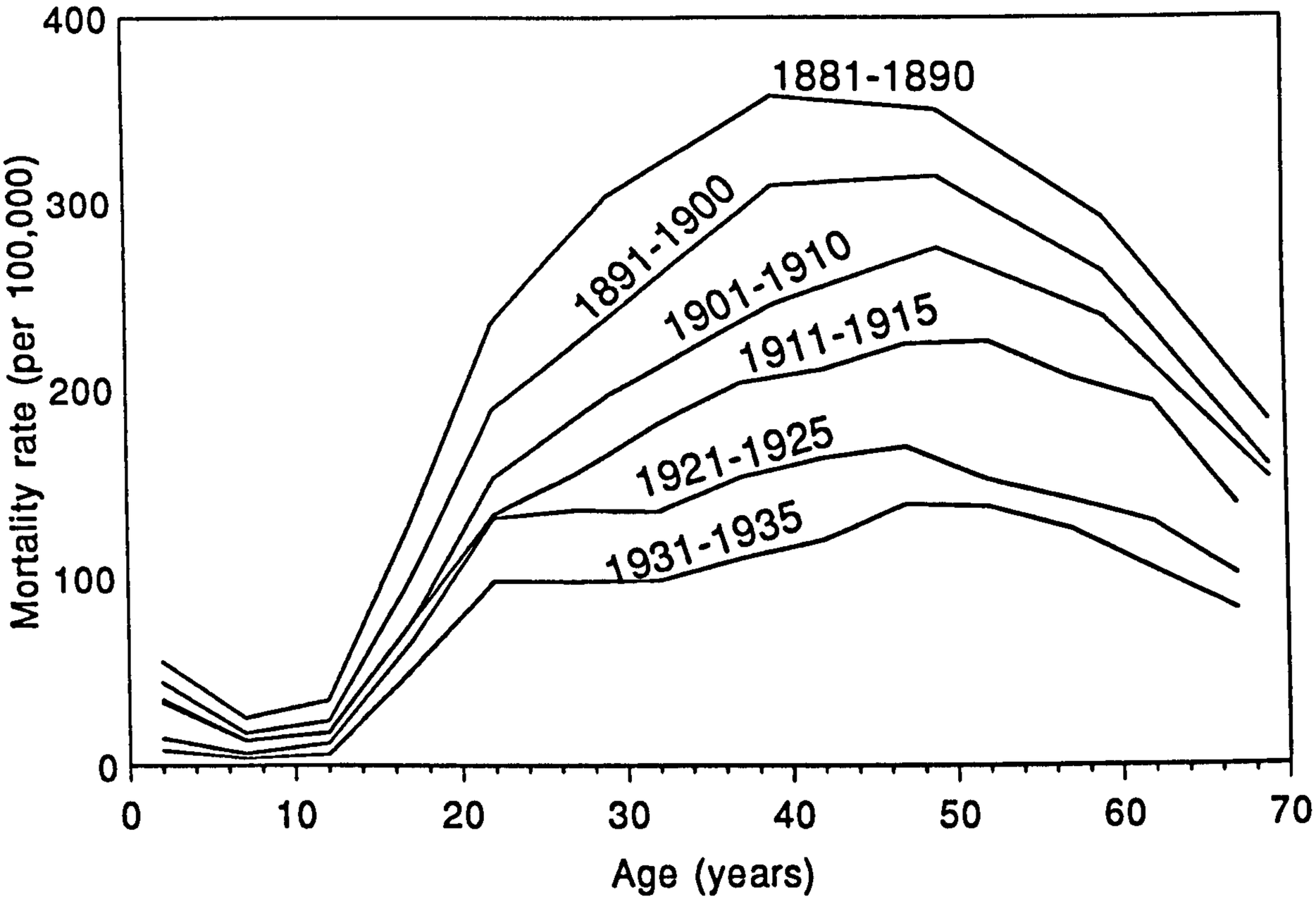
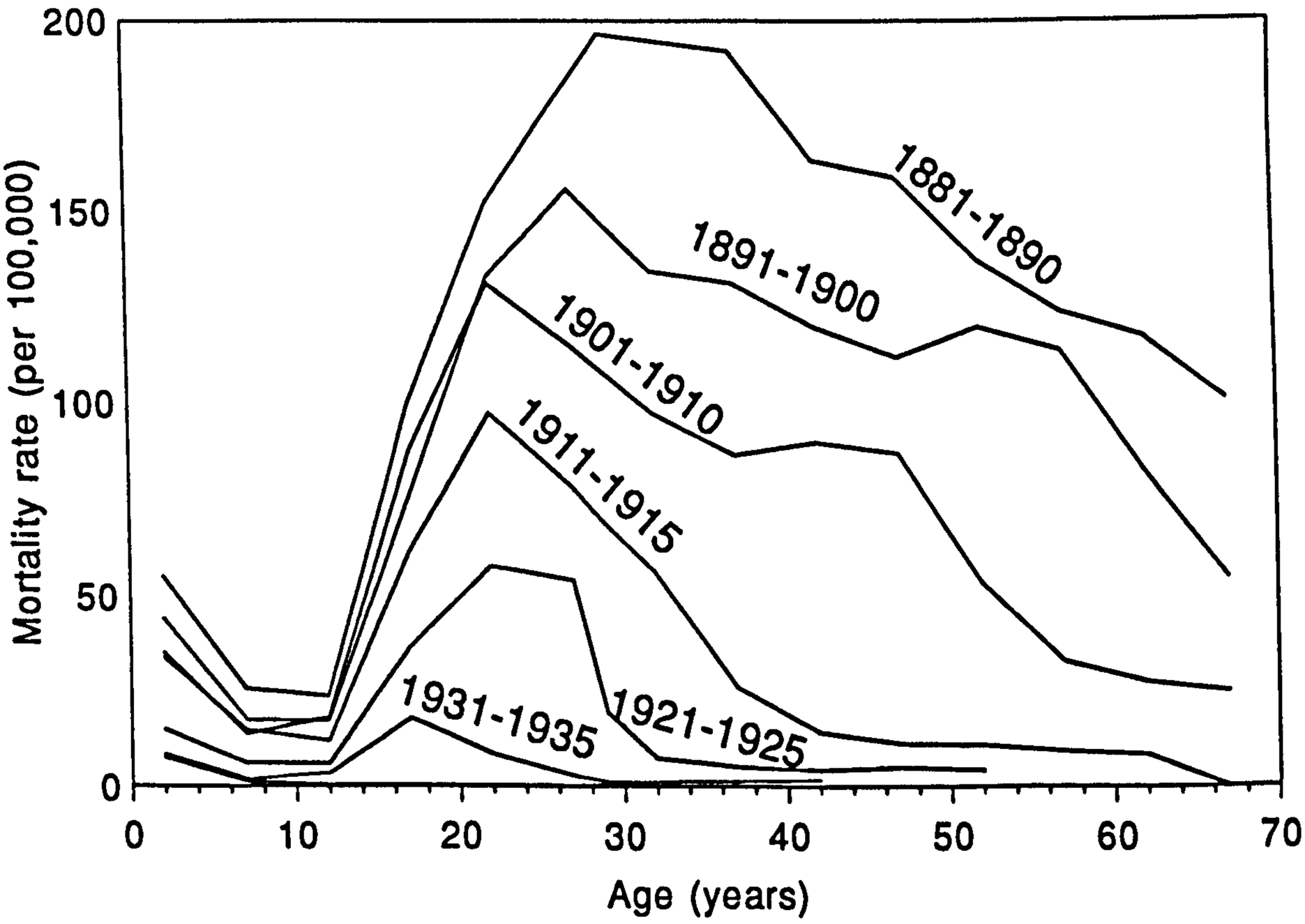


Figure 2.3: Age-specific mortality rates from respiratory tuberculosis in males in England and Wales during the period 1850–1990, plotted on a *log* scale.



(a) Mortality rates by year.



(b) Mortality rates by birth-cohort.

Figure 2.4: Age-specific mortality rates from respiratory tuberculosis in males in England and Wales during the period 1881–1950, plotted by year and birth cohort.
These figures are identical to those on page 30 in section 1.1. They have been reproduced here for convenience.

2.2 Notifications of Tuberculosis

2.2.1 Data sources and case definitions

Respiratory and non-respiratory tuberculosis became notifiable in England and Wales in 1912 and 1913 respectively. Until 1954, notifications were recorded on annual returns submitted by local authorities to the Ministry of Health and the figures were published in the Annual Reports of the Ministry of Health. Since 1954, weekly and quarterly returns of notifications of tuberculosis have been sent by Medical Officers of Health to the General Register Office. The figures from quarterly returns have been included in the Registrar General's Statistical Review of England and Wales; from 1974, these have been published in the OPCS MB2 Monitor Series.

Published notifications were not stratified by age and sex until 1938 and, until 1954, differentiated only between respiratory and non-respiratory forms of tuberculosis. Since 1954, notifications have distinguished between tuberculous meningitis/CNS, respiratory and non-respiratory forms.

Several changes in disease classification were introduced in 1982 as a result of ambiguities found during a national survey of tuberculosis notifications in 1978/9 [89,90]. The main changes are summarized below and are discussed in greater detail in the Report of the Joint Tuberculosis Committee of the British Thoracic Association [91].

1. Since 1982, notified cases with lesions in a respiratory and another site have been included in the totals of both respiratory tuberculosis and of 'other forms'. Until 1982, such cases were classified as having respiratory tuberculosis only, although official guidelines (OPCS 1974) stated that cases involving meningitis and CNS should have priority over "other forms", whilst "other forms" should have priority over "respiratory tuberculosis".

2. Medical officers of health now have to state whether a tuberculous case has:

   (a) pulmonary lesions (with or without mediastinal nodes or pleural effusion or both),

   (b) mediastinal nodes or pleural effusion, or both, without a pulmonary lesion,

   (c) meningitis or CNS or both,

   (d) other forms alone,

(e) more than one of the above.

All sites of involvement have to be specified if possible. Prior to this, there were no guidelines indicating whether individuals with isolated mediastinal lymphadenopathy or pleural effusion or both, should be included in the "respiratory tuberculosis" or "other form of tuberculosis" categories.

3. Since 1982, all individuals with no evidence of disease but notified because they were taking chemoprophylaxis are excluded from published notifications. The 1978/9 survey found that individuals taking chemoprophylaxis comprised about 3% of all notifications.

4. Cases notified after death are now included in the published notifications; the policy was ambiguous until 1982, although the intention was that they should be reported separately from those notified whilst alive.

There have been only two studies estimating the degree of undernotification. The first survey [92] reviewed all pathology reports between 1981 and 1984 at the Pathology department at Edinburgh university, and found that of the 69 firmly diagnosed tuberculosis cases, 26 (40%) had not been notified. About a third of the 62 treated cases also escaped notification.

The second survey reviewed all patients presenting at the London Chest Hospital or the Royal London group of hospitals in the East End of London, between January 1985 and December 1989 [93]. Information was available for 580 patients (95% of all cases), of whom 426 (73%) had been notified, as compared with 177 out of 205 (84%) of smear-positive cases.

In contrast with practice in England and Wales, microbiologists in Scotland have a statutory duty to notify all positive bacteriological results, which should compensate for any failure to notify by clinicians. Hence the lower estimate of the level of undernotification for Scotland is surprising.

2.2.2 Trends in notifications

Figure 2.5 compares the crude notification rates of respiratory tuberculosis in England and Wales until 1990 with the corresponding mortality rates. The latter differ from those

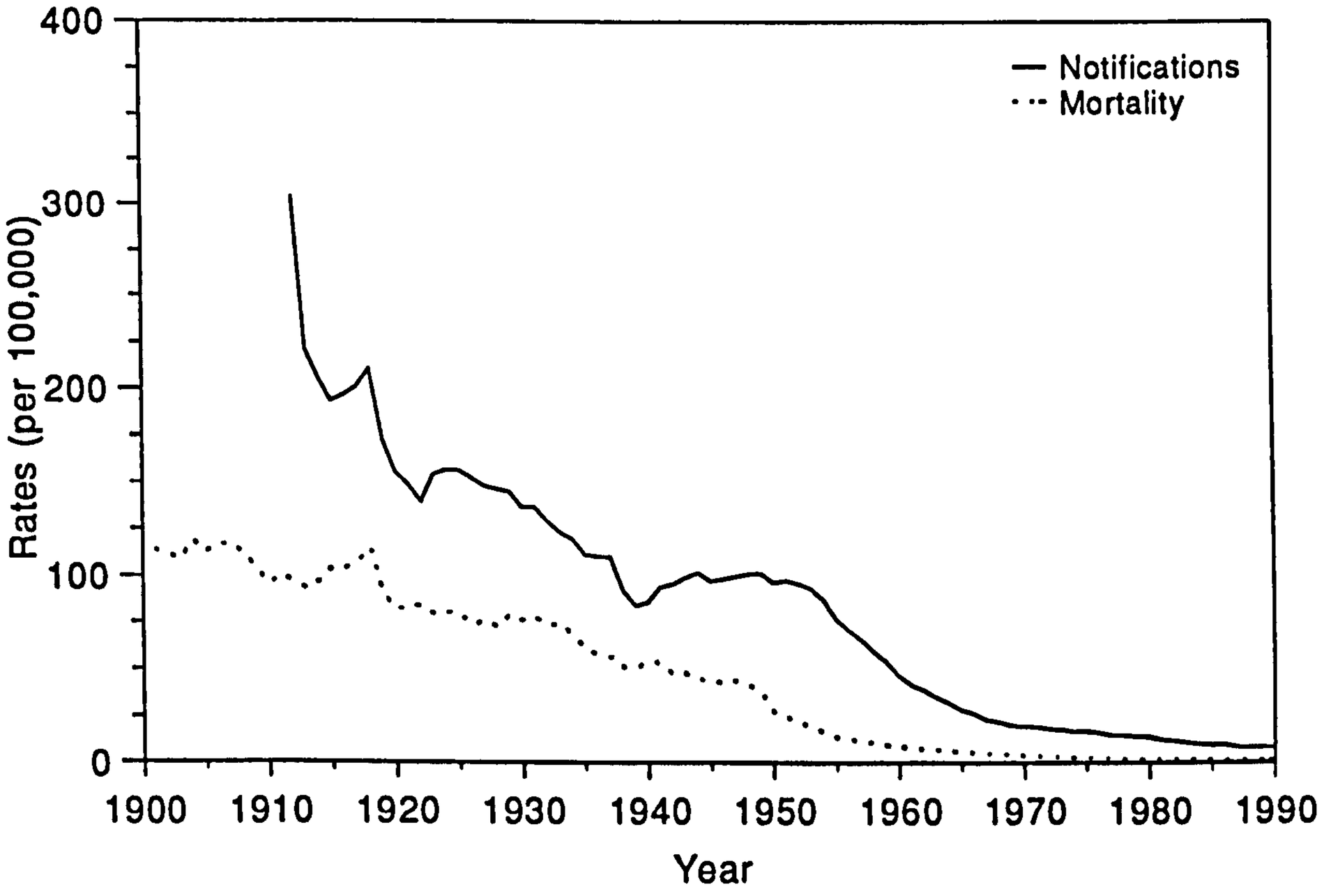


Figure 2.5: Crude annual mortality and notification rates (non-age standardized) of respiratory tuberculosis in England and Wales between 1901 and 1990.

shown in Figure 2.1 since the rates are not age-standardized: published notifications were not published by age and sex until 1938 (see page 74).

In addition to the obvious increase during the First World War, this shows that the notification rates increased in both 1923 and 1938 whereas the mortality rates declined throughout the time period considered. The increase in 1923 reflects a change in the registration of tuberculosis cases: after 1923, the published notifications included cases identified from death certificates and other sources, which had not always been formally notified before this time. The increase in 1938, just before the start of the Second World War, is attributable to improved case detection, as civilians were mobilized into the army and all conscripts were examined for signs of tuberculosis.

Case detection improved further during the 1940s and throughout the 1950s after the introduction of mass miniature radiography into the civilian population in 1943. This is reflected in Figure 2.6, which compares the age-specific notifications of respiratory tuberculosis in males in the general population in England and Wales during the period 1938–1989 against the corresponding mortality rates. This shows that for most age groups, the notification rates first decreased during the late 1940s/early 1950s, whereas the corresponding mortality rates had been declining at least since 1850. Case detection improved relatively

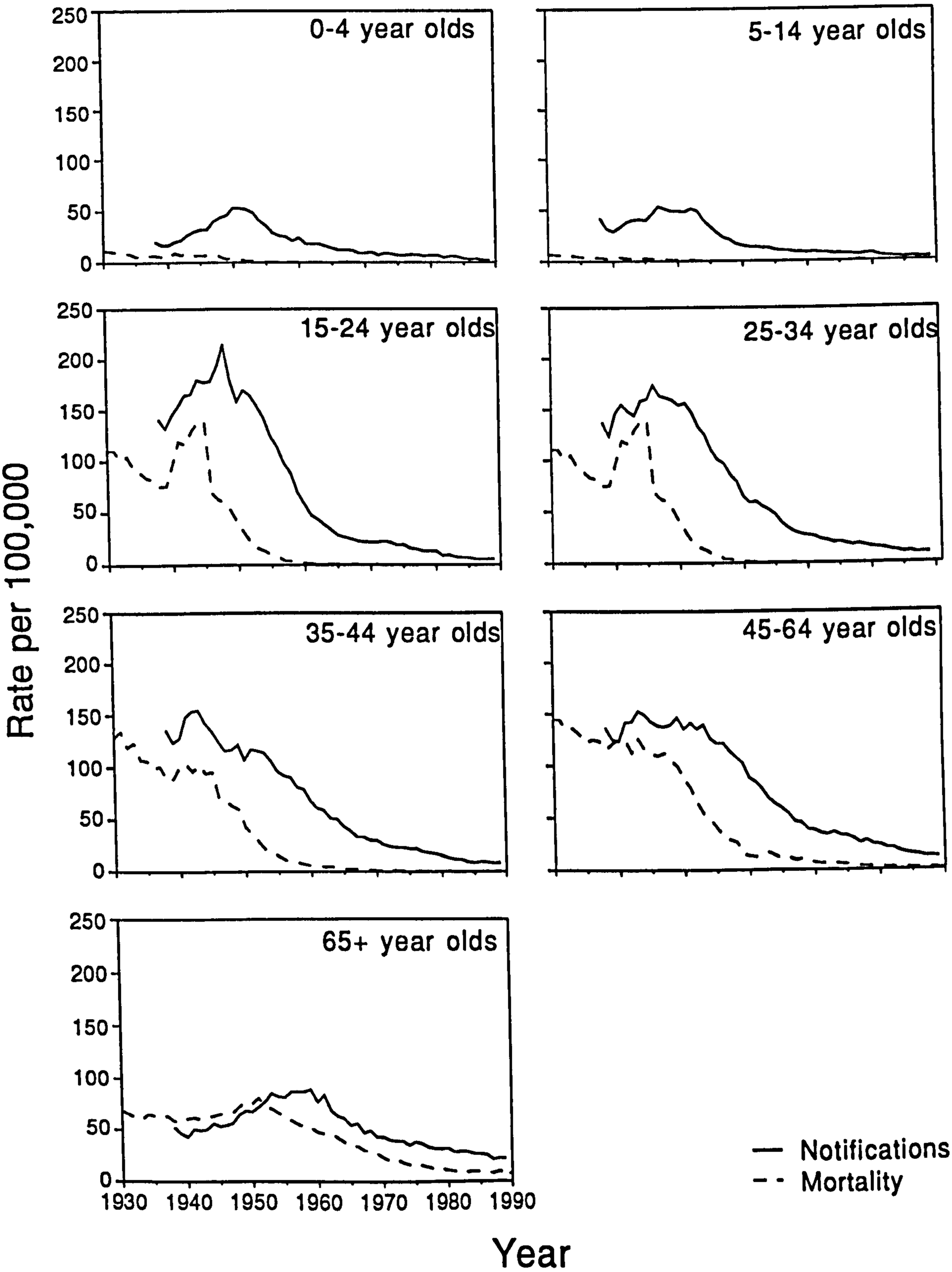


Figure 2.6: Published age-specific notification and mortality rates of respiratory tuberculosis in males in England and Wales during the period 1930–1989.
Note that published notifications for 45–64 year olds were first broken down into those for 45–54 and 55–64 year olds after 1965. Hence the mortality data for 45–64 year olds in this figure are stratified differently from those in Figure 2.4.

slowly for individuals aged over 60 years, as indicated by the fact that the notifications first exceeded the mortality rates after 1950 and started to decline during the 1960s.

The most rapid decline in the notification rates during the 1950s occurred among 15–24 year olds (see Figure 2.6). This is a consequence of two factors. First, BCG vaccination of tuberculin-negative 13 year olds was introduced in 1954 in England and Wales, which protected 84% of adolescents against tuberculosis during the first five years following vaccination in the UK MRC BCG trial [38]. Second, the introduction of chemotherapy in 1950 (see section 2.3) accelerated the decline in the annual risk of infection in England and Wales, and led to a reduction in the incidence of 'primary' disease in all age groups. This particularly affected notifications in adolescents, since they faced high risks of developing disease after infection and a greater proportion of their disease was probably 'primary' in origin, as compared with that in older individuals. This issue is discussed in greater detail in section 3.4.2.

The decline in notifications in most age groups in the general population slowed down during the 1960s at the same time as immigration increased from regions with higher tuberculosis disease prevalence, in particular, the Indian subcontinent (ISC). Routine notifications are not broken down by ethnic group and the extent of the contribution from immigrants was not known until a national survey of tuberculosis notifications was carried out between February and April in 1965 [94]. Further surveys of tuberculosis notifications were carried out during 4 months in 1971 and over 6 months during the period 1978–9, in 1983, 1988 and 1993 [89,95–98]. Two nationwide tuberculosis surveys had been carried out earlier during the period 1955–56 and in 1963, but these focussed on the prevalence of drug-resistance among patients treated for bacteriologically-confirmed disease in chest clinics [99,100].

The relative contribution of the ISC to notifications of respiratory and non-respiratory tuberculosis found by the surveys are summarized in Table 2.1. This shows that 10% of all tuberculosis notifications (380 out of 3,806) in 1965 came from the ISC population. The contribution varied by type of disease, with the ISC constituting 21% and 5% of the notifications of non-respiratory and respiratory forms respectively in 1965. By 1971, 45% and 18% of notifications of non-respiratory and respiratory tuberculosis respectively came from the ISC population, as compared with 57% and 33% by 1988.

The ISC population constituted a small proportion of the general population in these

Table 2.1: Summary of notification rates of all forms, respiratory and non-respiratory forms of tuberculosis in the general population in England and Wales during the period 1965–1988, together with the relative contribution from the ISC population.

Year	All forms		Respiratory†		Non-respiratory†	
	Notif rate (per 100,000)	% from ISC	Notif rate (per 100,000)	% from ISC	Notif rate (per 100,000)	% from ISC
1965	31.87	10	27.6	5	4.3	21
1971	22.4	24	18.2	18	4.3	45
1978/9	16.4	35	12.7	31	4.9	52
1983	12.0	37	9.1	32	4.0	52
1988	8.4	39	6.9	33	3.0	57

Data compiled from OPCS publications (see section 2.2.1) and from reports from the national surveys [89,94–97].

†Include notifications involving both respiratory and non-respiratory sites.

Note also that notifications after 1978/9 are for England only; prior to this, they refer to England and Wales.

years (under 2% and about 3% of that in 1978/9 and 1988 respectively) and, given their large contribution to the reported tuberculosis incidence in England and Wales, had higher notification rates than the white ethnic group. This is illustrated in Table 2.2, which summarizes the crude notification rates of all forms of tuberculosis for the ISC and white ethnic groups found in each survey. This shows that the notification rates were about 178 and 6.9 per 100,000 in 1983 in the Indian and white ethnic populations respectively. This corresponds to a difference by a factor of 38. By 1988 the notification rates in these two populations differed by a factor of about 28.

Springett *et al* performed separate analyses of the notifications of all forms of tuberculosis in the white ethnic group between the period 1953 and 1983 and estimated the rate of decline in individual age groups [101]. 1953 was selected as the starting point since it predated both the large-scale immigration from the Caribbean and the Indian subcontinent to England and Wales, and the introduction of BCG vaccination of tuberculin-negative 13 year olds. This implicitly assumed that the age-specific notifications for the general population in 1953 approximated those in the white ethnic group. Results from a similar comparison between the corresponding notifications between 1978/9 and 1988 are provided in the report of the national survey in 1988 [97]. The average annual declines for each age group and for each

Table 2.2: Crude annual notification rates (per 100,000) of all forms of tuberculosis in the white, Indian and Pakistani ethnic groups found during the national surveys of notifications between 1965 and 1988 in England and Wales.

	Ethnic group†		
Year	White	Indian	Pakistani
1965	27.6	324.0	684.0
1971	15.6	414.0	840.0
1978/9	9.4	354.0	353.0
1983	6.9	178.0	169.0
1988	4.7	134.6	100.5

† determined by an individual's place of birth for surveys in 1965 and 1971; by individual's place of birth and that of his/her parents for subsequent surveys. Population estimates: for 1965 and 1971 — from censuses carried out in 1961 and in 1971 respectively; for 1978/9 — from a National Dwelling and Housing Survey in 1978; for 1983 and 1988 — on the basis of Labour Force Surveys in these years.

period are summarized in Table 2.3.

This shows that the fastest decrease in the notification rate occurred among 15–24 year olds during the period 1953–1965. This is consistent with findings in this age group in the general population (see page 78). The fastest decline among 25–34 year olds occurred during the period 1965–1971. This is attributable to the introduction of BCG vaccination in 1953: most of those aged 25-34 years in 1965 had not been vaccinated, whilst many of their counterparts in 1971 were protected by vaccination.

In any time period, the rate of decline in the notifications was slowest for those aged over 65 years. Most individuals in this age range had been infected much earlier in life and hence were at risk only of developing endogenous disease or of reinfection followed by exogenous disease. Given the low risk of infection at this time, much of the disease among these individuals was probably attributable to endogenous reactivation. Hence the slow decline in the notification rates among those aged over 65 years reflects the gradual change in the corresponding prevalence of infection. This is discussed by Springett [102].

45–54 and 55–64 year olds experienced the most rapid declines in the notification rates during the period 1978/9–1983. As pointed out by Springett *et al* [101], most of these individuals experienced rigorous tuberculosis examinations during the war years, and chemotherapy was available from the 1950s. This suggests that relatively few of these individuals had undetected and untreated lesions and hence were at risk of relapse later on in

Table 2.3: Average annual decline in the notifications of all forms of tuberculosis by age in white males in England and Wales between 1953 and 1988, as estimated by Springett *et al* [101] and MRC [97]

Age group (years)	Average annual decline (% pa)				
	1953–1965	1965–1971	1971–1978/9	1978/9–1983	1978/9–1988
0–14	13.7	8.7	8.4	6.2	7.2
15–24	15.3	10.1	10.4	8.3	13.8
25–34	10.8	13.4	7.3	6.3	9.9
35–44	8.4	9.4	9.6	7.5	7.5
45–54	7.9	9.8	5.9	12.3	8.0
55–64	6.4	9.3	2.9	12.4	10.5
65 and over	3.7	6.0	1.7	3.8	3.7

life. This argument may be justified, given that the rate of decline in the notification rate in this age group remained high between 1983 and 1988.

2.2.3 Discussion

Results from the national surveys suggest that tuberculosis notifications in England and Wales since 1965 had been influenced by contributions from the immigrant population. Given that immigrants experienced different risks of infection from individuals born in the UK, any analysis of the transmission dynamics of tuberculosis has either to take this into account or just consider one ethnic group. For this reason, this thesis analyses tuberculosis only in the white ethnic group. Given gender differences in the epidemiology of tuberculosis (see page 29), we further restrict our analyses to white ethnic males.

The analyses of Springett *et al* were carried out on notifications of *all forms* of tuberculosis. Notifications of *respiratory* tuberculosis are more relevant for studying the transmission dynamics of *M. tuberculosis*, since only a small proportion of extrapulmonary forms are infectious (see section 1.1). The survey in 1988 provides the only data on respiratory tuberculosis for the white ethnic population, and hence the notification rates for earlier years have to be estimated from those of *all forms* of tuberculosis.

Reasonable approximations can be derived by assuming that notifications of respiratory tuberculosis declined from their level in 1953 at the same rate as those of all forms of tuberculosis, summarized in Table 2.3. The estimates derived in this way are compared

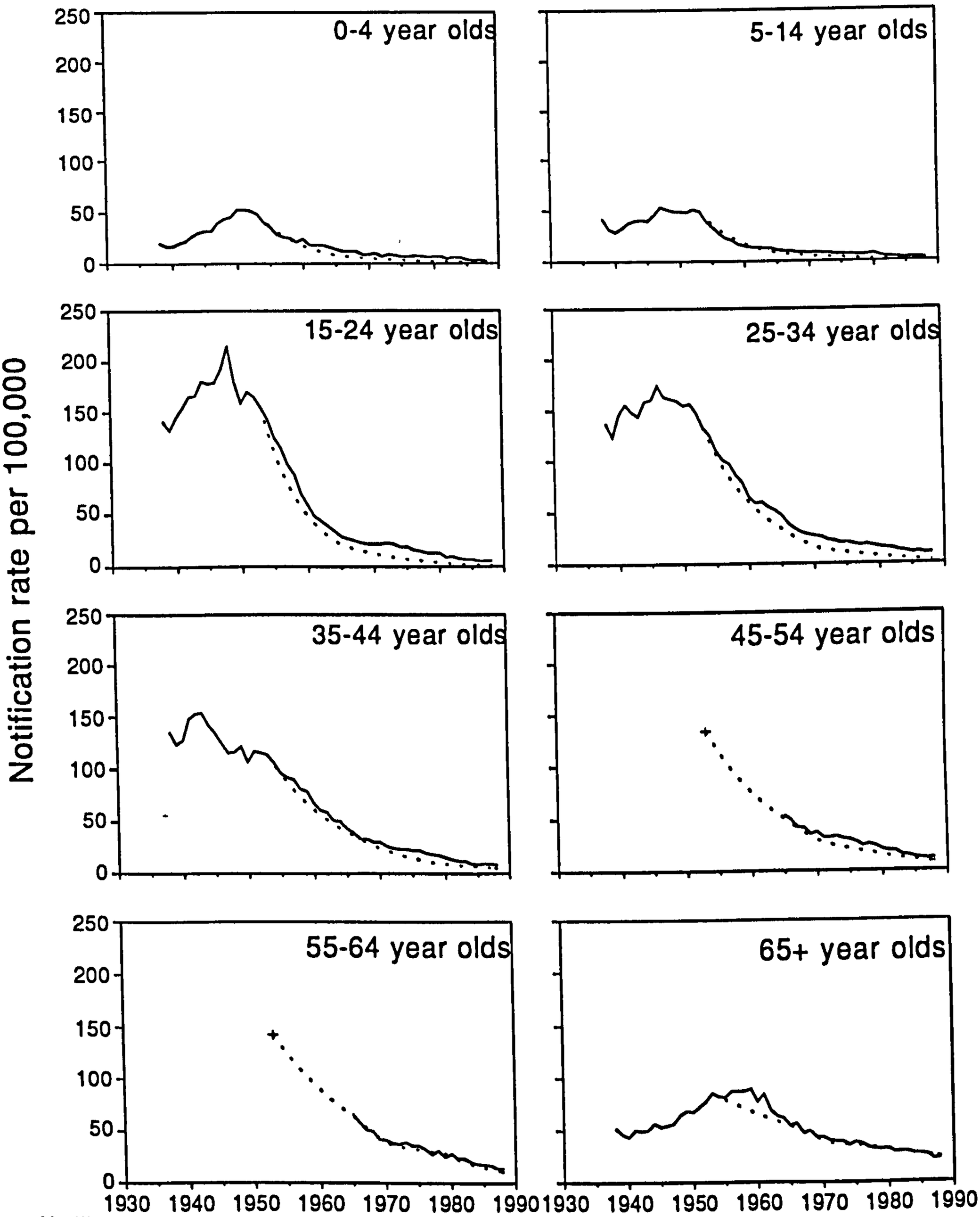


Figure 2.7: Comparison between age-specific notification rates of respiratory tuberculosis in males in the general population in England and Wales and those estimated for the white ethnic group, assuming that they declined from their level in 1953 at the rates given in Table 2.3.

Note that notification rates for 45–54 and 55–64 year olds in 1953 (denoted by +) were assumed to be 95% of those of *all forms* of tuberculosis, given in Springett *et al* [101] (see page 83).

against the corresponding notifications in the general population in Figure 2.7[1]. Note that tuberculosis notification rates for 45–54 and 55–64 year olds in the general population were not officially published until 1965— see also Figure 2.6. The notification rates of respiratory tuberculosis for these age groups in 1953 were derived by assuming that they constituted 95% of those of *all forms* of tuberculosis. The notification rates for the latter are provided by Springett *et al* [101]. This assumption is reasonable, given that respiratory tuberculosis comprised 95% of notifications of all forms of tuberculosis for *45–64* year olds in 1953.

Figure 2.7 shows that the difference between the notification rates among white ethnic males and those in the general population was greatest among 15–24 and 25–34 year olds after the mid 1960s, and was smaller for older individuals. This suggests that notifications in the *general population* among individuals aged over 55 years after 1965 accurately reflect those in the *white ethnic group*.

[1]The method used to generate these data was validated by comparing the notification rates estimated for 1988 against those from the national survey in this year, obtained directly from Dr J Watson and Mr A Charlett at CDSC. The notification rates derived using this method differed by upto 0.7 cases per 100,000 for each age group from those in the national surveys.

2.3 Annual risk of infection

2.3.1 Pre-chemotherapy era

There are few data on the annual risk of infection in England and Wales during the pre-chemotherapy era. The national survey of tuberculin sensitivity carried out during the period 1949–1950 [103] provided the first indication of the nationwide prevalence of tuberculous infection. This covered 22 urban and rural areas and a population of 94,221 5–20 year olds. Individuals were considered tuberculin-positive if they responded with an induration of 5mm or more to tuberculin jelly (equivalent to an intradermal test with 10TU) or to a Mantoux test with 100TU of Old Tuberculin (0.1ml of 1/100 dilution).

Figure 2.8 shows the age-specific percentage of tuberculin-positive males found in the south urban, northern urban and in rural areas in England. About 30% of 11–12 year old males in south urban England were considered tuberculin-positive, as compared with about 44% of those in rural areas. Rural areas also had the highest prevalence of tuberculin-positive males/individuals in other age groups. These differences reflect regional variations in the interpretation of tuberculin reactions, and the greater prevalence of infection with *M. bovis* in rural areas, as compared with urban areas. The latter has been attributed to the greater consumption of raw milk from non-attested cattle in rural areas [103].

The annual age-specific incidence of infection during the survey was estimated by taking the ratio between the prevalence of tuberculin-negative individuals in adjacent 2-year age groups (e.g. 5–6 year olds and 7–8 year olds), who were tested simultaneously [103]. This estimate did not account for the effects of tuberculin boosting and reversion in tested individuals, and implicitly assumed that the prevalence of infection in a given age group did not change over time, which is unrealistic. Hence the estimates of the risks of infection obtained (e.g. 5% and 15% for 17–18 year old males in rural and urban areas respectively and about 4% for 5–6 year olds in both areas) probably overestimated the true values.

Styblo *et al* [62] found that tuberculous meningitis mortality rates among 0–4 year olds provide a reliable indication of the annual risk of infection during the pre-chemotherapy era. Tuberculous meningitis is a severe extrapulmonary form of tuberculous disease, affecting the central nervous system (CNS) [10]. The time interval between infection and disease onset is approximately 3 months [11] and without treatment, its case fatality approaches 100% (see section 1.1). Tuberculous meningitis was responsible for 48–62% of all tuberculosis deaths

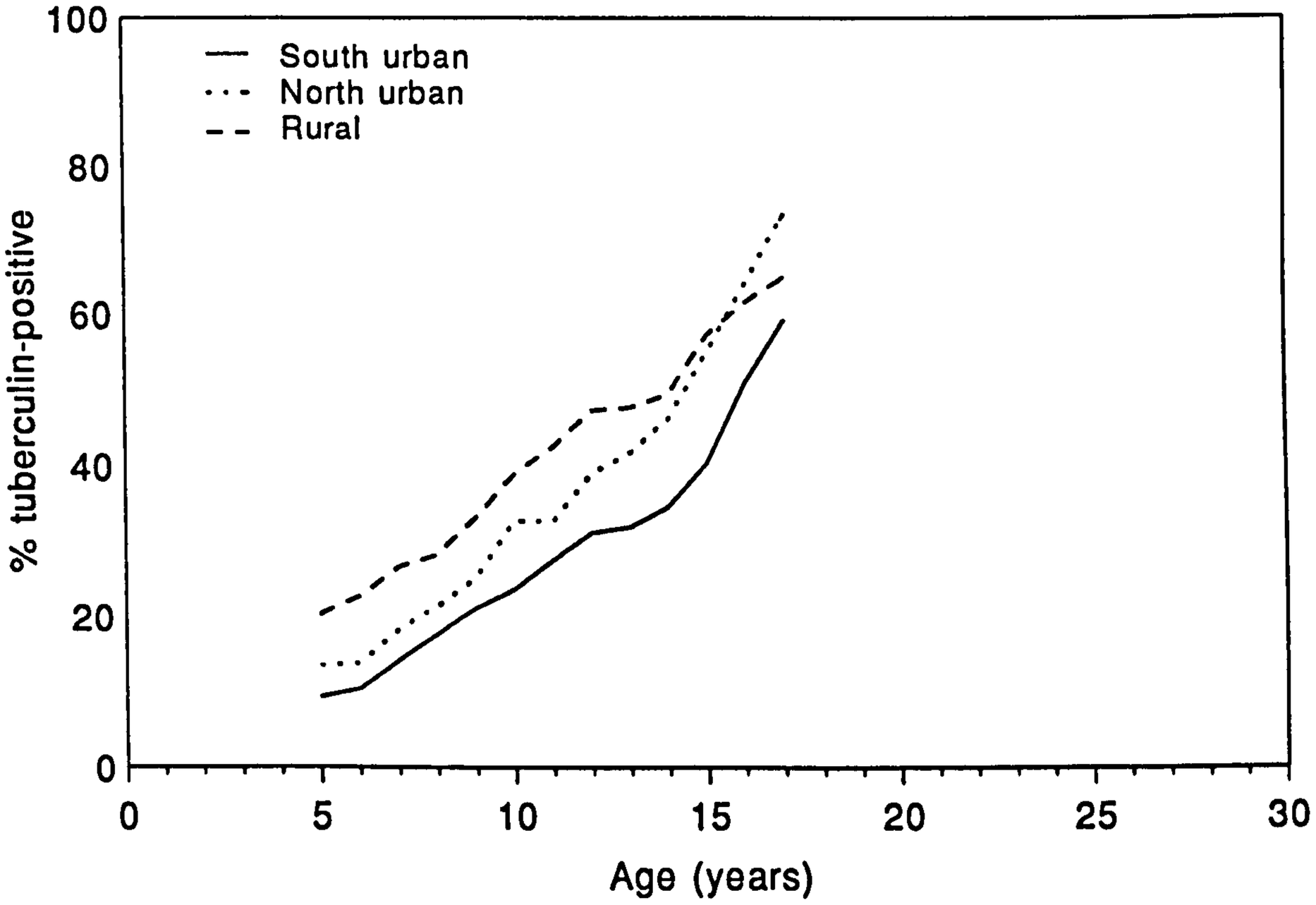


Figure 2.8: Age-specific prevalence of tuberculin-positive males in England and Wales, as found by the national tuberculin survey during the period 1949–50 [103].

among 0–4 year olds in the Netherlands between 1905 and 1940 [104]. Styblo *et al* [62] found that the tuberculous meningitis mortality rate in 0–4 year olds during the period 1920–1949 in the Netherlands was roughly 1% of the corresponding annual risk of infection, and depended on prevailing social conditions. During a harsh famine in 1945, for example, the ratio between the two increased to 2%. This relationship implies that if individuals experience identical risks of infection in a given year, irrespective of age, then 1% of 0–4 year olds who become infected with the tubercle bacillus develop tuberculous meningitis and die if not treated.

A similar correlation between the annual risk of infection and tuberculous meningitis mortality rates existed in Sweden between 1925 and 1935 [105]. The two did not correlate as well thereafter as a result of specific control measures aimed at reducing exposure of infants to infection in the community (e.g. case isolation, BCG vaccination of children in tuberculous homes and the compulsory pasteurization of milk for consumption by 1939). These measures accelerated the decline in tuberculous meningitis mortality rates, but had a lesser effect on the risk of infection.

Figure 2.9 summarizes the tuberculous meningitis mortality rates among 0–4 year old males in England and Wales during the period 1860–1957. Before 1920, deaths from tuber-

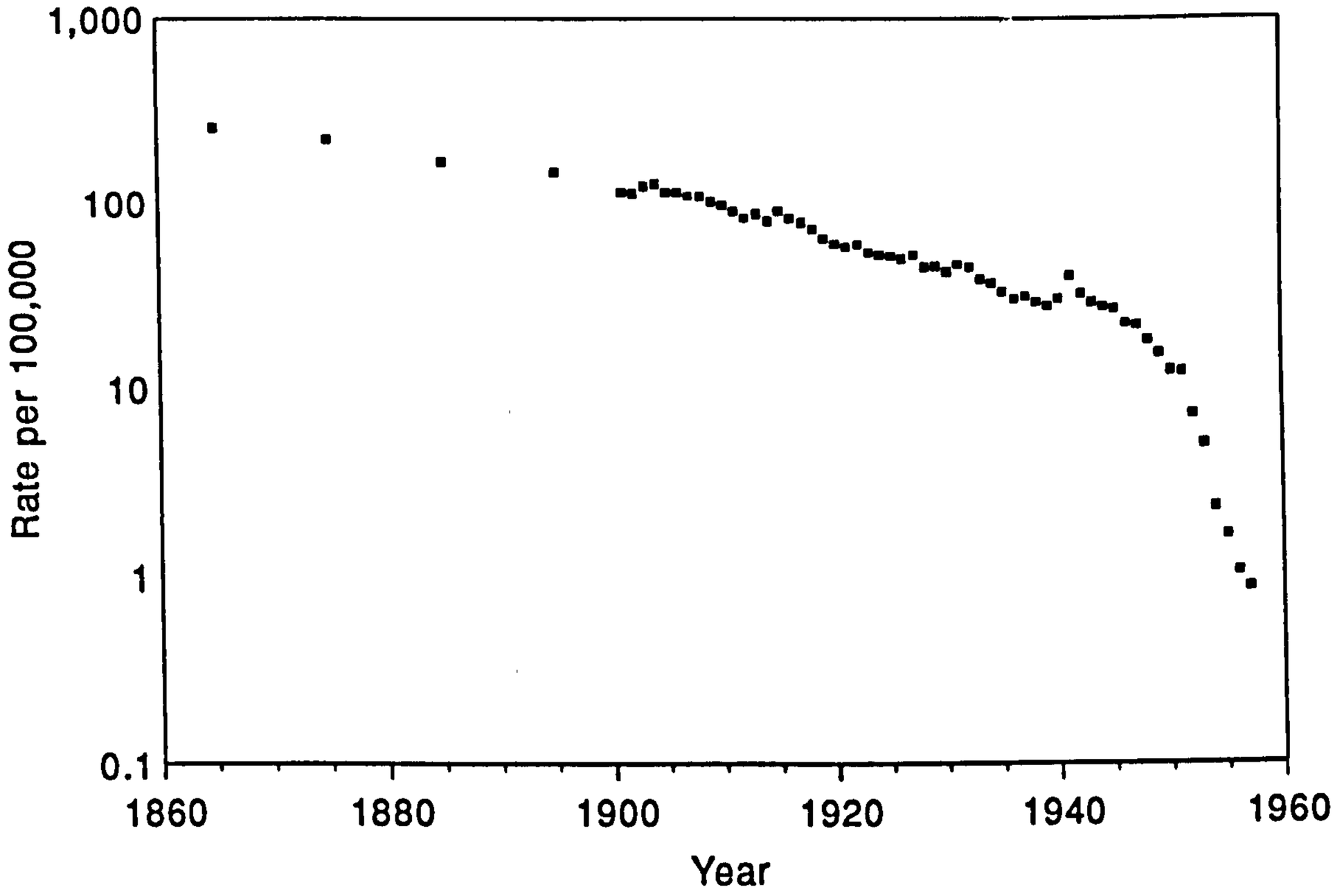


Figure 2.9: Mortality rates from tuberculous meningitis in 0–4 year old males in England and Wales during the period 1860–1957.

culous meningitis were classified separately from other deaths. After 1920, published tuberculous meningitis deaths were grouped with other deaths from tuberculosis of the CNS. Given that tuberculous meningitis is the commonest form of tuberculosis of the CNS [10], the two classifications should be comparable and there is no apparent discontinuity in the trend in mortality before and after 1920.

Figure 2.9 shows that the tuberculous meningitis mortality rate declined steadily until 1950, peaking slightly during the two World Wars. Similar increases in mortality rates from other forms of tuberculosis in other age groups in England and Wales were also observed during these time periods (see section 2.1). Regression analyses of the mortality rate during the period 1901–1950 show that it declined by 4.0% pa (95% CI [3.8, 4.2]) and by about 2% pa before this. Application of the 1% correspondence between the tuberculous meningitis mortality rate and the annual risk of infection implies that the latter was about 14% in 1901, declined to about 6% in 1920 and reached 2% in 1949. This estimate for 1949 is lower than that obtained from the national tuberculin survey during the period 1949–50, and is perhaps more realistic.

This risk of infection for 1901 (i.e. 14%) seems extremely high and exceeds the level found in several developing countries in recent years (see e.g. [25,26]). Similarly high risks

of infection (11%) were estimated for the Netherlands for 1911 on the basis of data on the prevalence of tuberculin sensitivity among schoolchildren and military recruits [62]. Even higher risks of infection (25%) were estimated among Eskimos in Alaska the 1950s (see section 1.1). This suggests that the high risk estimated for England and Wales in 1901 is not unrealistic.

As we have seen, tuberculous meningitis mortality data can only be used to infer the annual risk of infection in the absence of treatment and hence, after 1950, alternative measures are required to estimate the risk of infection.

2.3.2 Chemotherapy era — after 1950

The second national tuberculin survey was carried out during the period 1971–1973, and given wide-spread coverage of BCG vaccination of 13 year olds since 1954 (see section 2.4), covered only 6 and 13 year olds [106]. The latter were tested prior to vaccination. The survey covered 29 local health authorities and tested 24,221 children with two Mantoux tests, using 5TU of human PPD and 5TU of Battey PPD. Individuals reacting with an induration of 8mm or more to human PPD were considered tuberculin-positive[2].

Of 8,699 unvaccinated 6 year olds in the "UK national group" (children who, together with their parents, were born in the UK), 0.31% were tuberculin-positive, as compared with 1.14% of the 11,519 children tested at age 13 years. Christie and Sutherland estimated from these data that the annual risk of infection in the UK national group had been about 0.18% (CI 0.13–0.24) in 1960, and 0.089% (CI 0.06–0.12) and 0.045% (CI 0.03–0.06) in 1965 and 1970 respectively [106]. This corresponded to an decline of about 13% pa, which is similar to that seen in the Netherlands during this time (see section 1.3.2) [84].

Estimates of the risk of infection for other ethnic groups were slightly higher, but more unreliable and difficult to interpret, given that they were based on a small sample size (e.g. a total of 303 children born in the UK but belonging to the Asian and African group).

Given the high tuberculous meningitis case fatality rate during the pre-chemotherapy era, notification rates in an efficient disease notification system should directly reflect the level of the mortality rates, and hence the level of the annual risk of infection, if treatment

[2]The report points out that the definition of tuberculin-positivity also depended on the type of induration resulting from human PPD and the diameter of induration to Battey PPD, but further details have never been published.

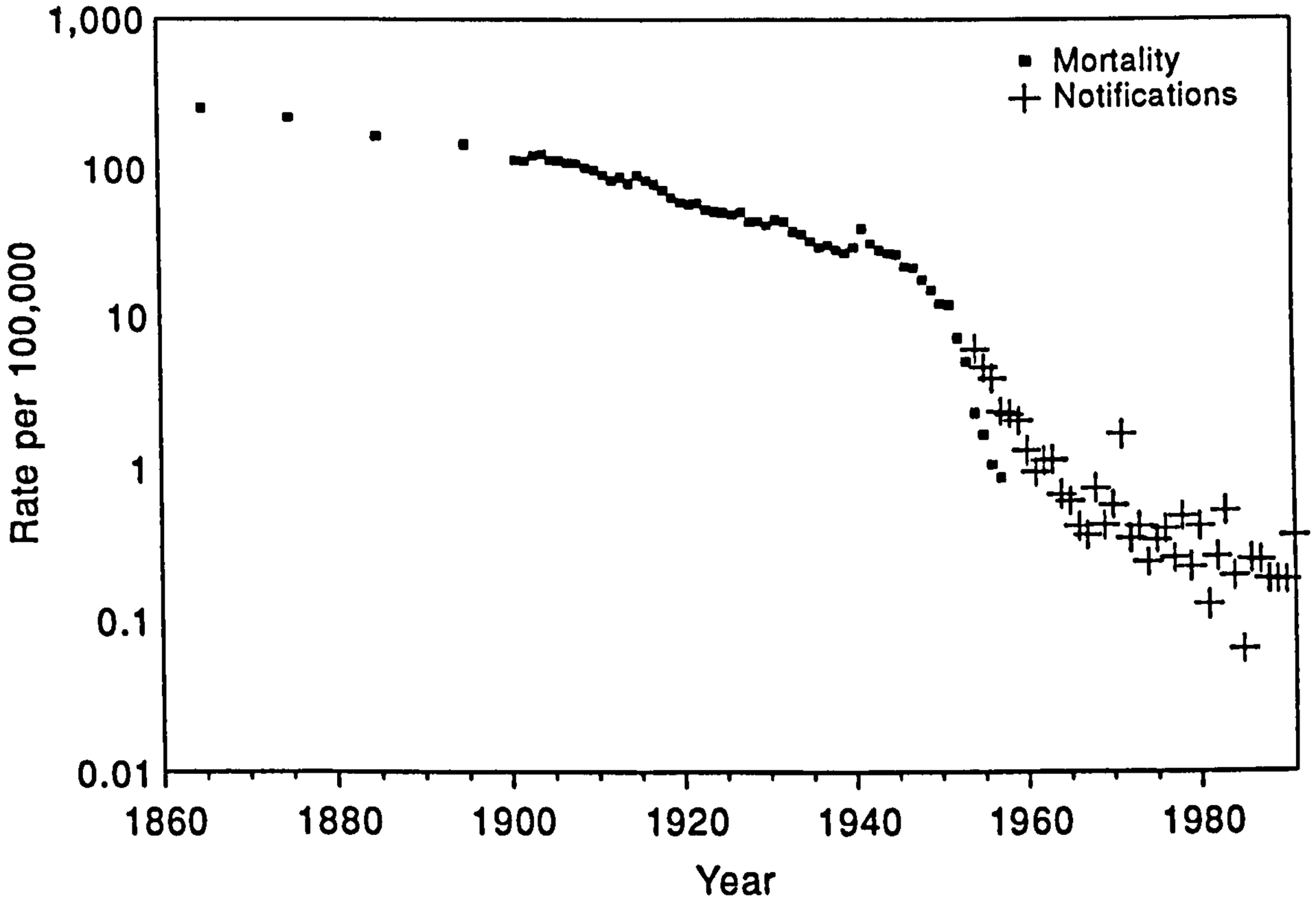


Figure 2.10: Comparison between mortality rates of tuberculous meningitis in 0–4 year old males in England and Wales until 1957 and the corresponding notification rates from 1954.

not been available.

Figure 2.10 summarizes the notification rates of tuberculous meningitis in England and Wales among 0–4 year olds after 1954, together with the mortality rates until 1957. This shows that the notification rate declined more rapidly during the 1950s than the corresponding mortality rate before 1950. This is attributable to three factors:

1. deficiencies in the notification system, which meant that a proportion of tuberculous meningitis cases were never reported,

2. a genuinely faster decline in the annual risk of infection after 1950, as a consequence of the introduction of chemotherapy, which reduced dramatically the prevalence of sources of infection by shortening the duration of infectiousness of infectious pulmonary cases, and

3. BCG vaccination of contacts of infectious cases introduced in 1949. Its protection against tuberculous meningitis among infants is estimated at 65–95% [107], and hence notifications among 0–4 year olds after 1950 exclude those who perhaps would have developed disease before the introduction of BCG vaccination.

This suggests that tuberculous meningitis notification rates are not directly comparable to corresponding mortality rates in 0–4 year olds before 1950. Hence the decline in tuberculous meningitis notification rates during the 1950s probably slightly overestimates the actual decline in the annual risk of infection.

It is difficult to interpret the slow rate of decline in the tuberculous meningitis notification rates after the mid 1960s, as there are few cases both in the white ethnic group and in the immigrant population. This suggests that notification rates of tuberculous meningitis in England and Wales after the mid 1960s cannot provide reliable estimates of the true level of the annual risk of infection. However, tuberculous meningitis notifications until this time provide a useful indication that the annual risk of infection declined at a faster rate after 1950, as compared with that before this time.

2.4 BCG vaccination data

Table 2.4 summarizes the main events relating to BCG vaccination in England and Wales since 1949. These events are discussed in greater detail below.

BCG vaccination was first officially introduced in England and Wales in 1949. Coverage was initially restricted to individuals considered to be at high risk (i.e. medical students, nurses, and contacts of infectious pulmonary cases), and was extended to tuberculin-negative 13 year olds in 1954, with the aim of preventing high morbidity rates during adolescence and early adult life. In 1959, vaccination was extended to cover children aged over 14 years who were still at school, students at further educational establishments, and children below age 13 years. The last amendment enabled entire school classes to be vaccinated simultaneously, rather than in stages.

The vaccination policy varies between regions in England and Wales. A survey of all district health authorities in 1984 found that most (184 out of 201) were vaccinating 11–14 year olds, and 5 had stopped routine vaccination of schoolchildren altogether. Table 2.5 summarizes the age groups covered in these authorities. According to the most recent survey [108] 15 health authorities had stopped vaccinating schoolchildren by 1990, a further 13 were reviewing their policy and the remaining 169 questioned were routinely vaccinating 10–14 year olds.

A summary of the total numbers of individuals vaccinated under the schools and contacts scheme has been published annually in the Chief Medical Officer's Report to the Government until the 1980s. These data have also been published in the Digest of Health Statistics for 1970–1972. The most detailed data are published in the "Health and Personal Social Services Statistics for England and Wales" from 1972. These are broken down according to the number of individuals skin tested, found to be tuberculin-positive, tuberculin-negative, and the number vaccinated under both the schools and contacts schemes. These data are not broken down by age, although most of those vaccinated under the schools scheme were aged 13 years at least until 1959. From 1974, this detailed breakdown is provided only for England.

Since its introduction, vaccination coverage of schoolchildren increased gradually. In 1954, 42,735 children were vaccinated, as compared with 443,236 children in 1959. Estimates suggest that by 1962, about 60% of each cohort had been vaccinated, rising to about

Table 2.4: Summary of the main events in the history of BCG vaccination in England and Wales.

Year	Event
1949	Introduction of vaccination (fluid vaccine from State Serum institute, Copenhagen) of individuals at high risk, i.e. hospital nursing staff, medical students (by special arrangement with the Ministry) and contacts of infectious pulmonary cases. Total vaccinated: 1,000 ($\approx$ half were nursing staff; the rest contacts).
1950	MRC BCG trial of adolescents started.
1953	Ministerial approval for vaccination of children towards year preceding 14th birthday. LHAs to apply to government for permission to introduce vaccination in schools under their jurisdiction. Parental consent also required.
1954	First year of vaccination among schoolchildren — 43,765 children vaccinated. 2 LHAs granted permission to vaccinate diabetic schoolchildren outside the official age range.
1956	Trials of British freeze-dried vaccine begun. Initial results from MRC trial confirms protection for the first 2 and a half years.
1958	2 vaccines in use: freeze-dried and Copenhagen. Advantage of free-dried vaccine — can be stored for longer.
1959	Vaccination extended to children aged over 14 years still at school and students of FE establishments, and those aged under 13 years.
1965	Immigrants of school age and newly arrived adults to be tuberculin tested and vaccinated if necessary.
1972–1983	5 LHAs suspend vaccination of schoolchildren (2 in 1972, one each in 1977, 1980 and 1983).
1983–1900	11 more LHAs suspend vaccination; 1 LHA reverses suspension made during 1972–1983.
1985	Joint Committee on Vaccination and Immunization (JCVI) recommends vaccination to be discontinued in 1990 if decline in notifications continues.
1990	JCVI recommends BCG vaccination to continue until results from national survey in 1993 available.

Table 2.5: Summary of BCG vaccination policy in 201 districts in England and Wales in 1984, as found by Miller *et al* [109].

Number of districts	Policy
	Schools scheme:
184	11–14 years (mainly 13 year olds)
8	10–11 years
1	5–13 years
2	5 years
1	newborn, school entrants and 13 year olds
	Neonates:
98	neonatal contacts and/or immigrants
9	all neonates
88	no routine vaccination of neonates
6	neonatal vaccination under consideration

70% by 1970 and has remained constant at 75% since then [110]. It is likely that the corresponding proportions for *tuberculin-negative* individuals in each cohort are slightly greater; breakdowns for these individuals are not published.

Data on the proportion of individuals found to be tuberculin-positive prior to testing are difficult to interpret and cannot be used to estimate the annual risk of infection for several reasons. First, the definition of a positive reaction varies between regions and according to the test used. Second, the proportion of individuals found to be 'tuberculin-positive' under either the schools or contacts scheme are not broken down by age — an increasing proportion of individuals tested under the former after 1960 were aged 10–24 years. Third, many individuals tested after 1960 were from the immigrant population, who faced different risks of infections earlier in life. For reference, Figure 2.11 summarizes the prevalence of tuberculin sensitivity found among individuals tested under the schools scheme. This shows that the combination of the above factors led to a prevalence of tuberculin sensitivity which was far in excess of that found among 13 year olds during the national tuberculin survey during the period 1971–1973 [106] (see section 2.3.2), indicating that *the data are unreliable.*

Data on the efficacy of BCG vaccination among adolescents in England and Wales is provided by the UK MRC BCG trial during the 1950s. This found that among 13,598 individuals who were vaccinated with BCG, 84% were protected during the first 5 years,

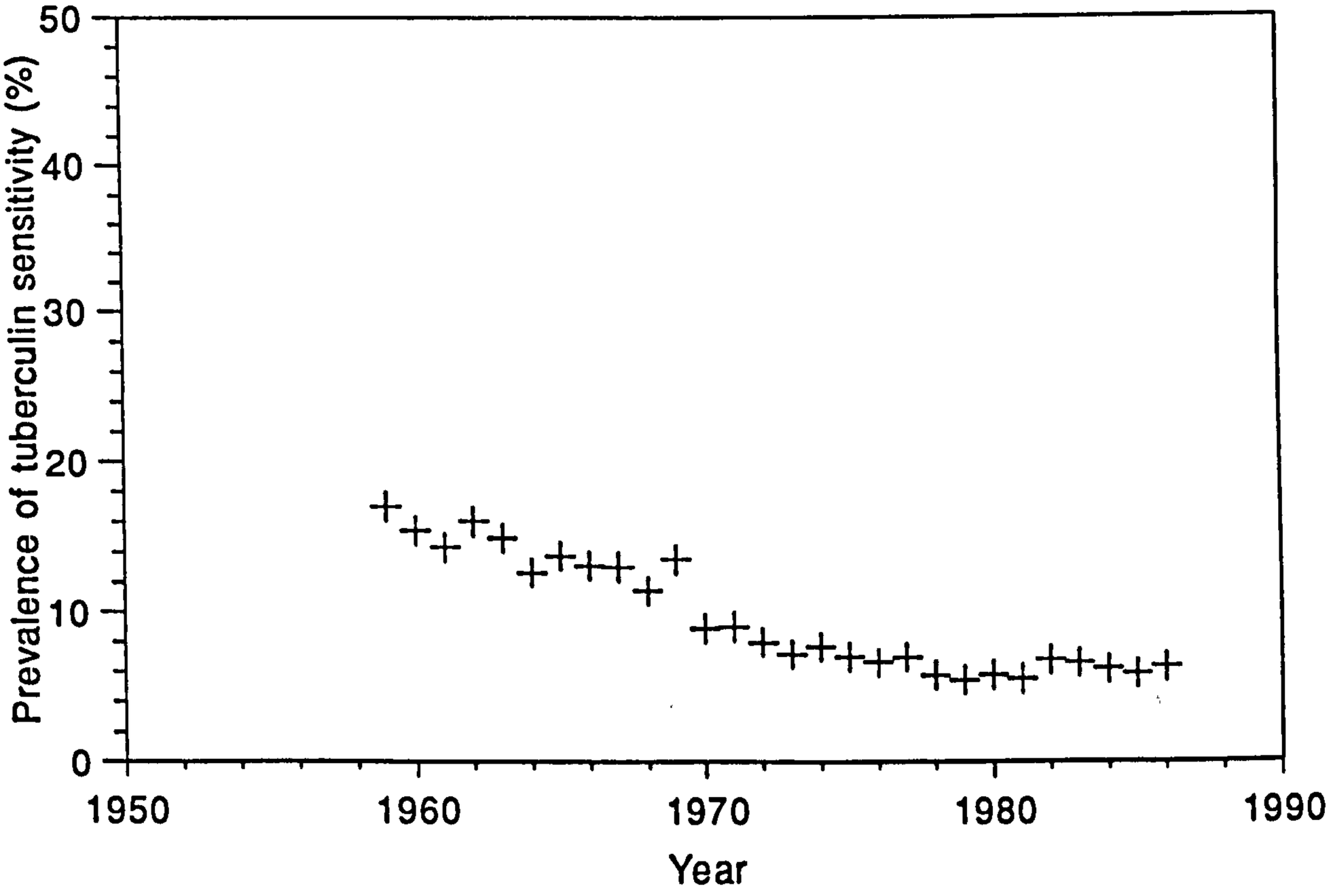


Figure 2.11: Prevalence of tuberculin sensitivity among school children and students tested as part of the national BCG vaccination scheme.

and 69% and 59% during the following 2 five year periods [38]. Estimates of the efficacy thereafter are unreliable, as there were only 5 cases in the unvaccinated group, and 6 among those vaccinated. This corresponds to an overall vaccine efficacy of about 77% during follow-up.

Chapter 3

Development and evaluation of a model for tuberculosis — TBDYN3

3.1 Formulation of TBDYN3

3.1.1 Definitions of 'primary' and 'post-primary' disease

Transmission models typically describe a disease process in a population by modelling the transitions of individuals between different states, i.e. those susceptible to infection, those infected (but not diseased) and those who are diseased (see section 1.3).

The model of the transmission dynamics of *M. tuberculosis* presented here, called TBDYN3[1], considers an age-structured population, and subdivides the states so as to reflect the different risks of developing disease according to time since infection and reinfection. TBDYN3 extends the work of Sutherland *et al* [7], who were the first to estimate the relative contribution of 'primary', 'endogenous' and 'exogenous' forms to the disease incidence in any country (see section 1.3.2). Before describing the structure and assumptions of TBDYN3, we first discuss how the risks of developing disease differ between individuals and how they relate to the definitions of primary, endogenous and exogenous disease.

Tuberculous disease is conventionally defined as either 'primary' (occurring during the first five years after infection) or 'post-primary' (occurring more than five years after infec-

[1]Originally, the name TBDYN3 reflected the model's status as the third stage in an evolution. The name now reflects the fact that it describes the dynamics of tuberculosis, and assumes that there are three mechanisms through which individuals can develop disease.

tion) [16] (see section 1.1). These formal definitions stem from Holm in 1969 [16], although the clinical distinction between the two forms is often unclear in practice. Post-primary tuberculosis is generally associated with disease in the upper part of the lung [111] and cavitation [112], for example, but both can also occur shortly after initial infection. Post-primary tuberculosis is thought to occur either through endogenous reactivation or after exogenous reinfection [16], and Holm defined exogenous disease as disease occurring within five years of reinfection. These definitions, and the analyses of Sutherland *et al* on the risks of developing disease, reflect the views that

1. individuals experience an initial high risk of developing (primary) disease after infection, which then falls to a lower unchanging risk of endogenous reactivation after a few years, and

2. reactivation and reinfection are generally rare during the first few years after infection.

The formal definitions of 'primary', 'endogenous' and 'exogenous' disease lead to several inconsistencies when considering individuals who experience two or more disease episodes within five years of initial infection or reinfection. Both episodes experienced by an individual who develops disease during the first and fifth years after initial infection, for example, would be defined as 'primary' in origin, although the risk of developing the second episode could well have differed from that of someone who was still at risk of developing their first (primary) episode during the fifth year after infection. On this basis, the second disease episode within five years of infection might also be defined as endogenous in origin, if the individual did not experience any reinfection, and exogenous otherwise.

The situation is equally complicated for an individual who develops disease within a year of reinfection, but recovers and experiences disease again two years thereafter (i.e. within five years of reinfection) without having experienced further reinfection. The second disease episode could be defined as 'endogenous reactivation of the reinfection event', but, by the definition of Holm [16], would be defined as exogenous disease.

The conventional definition of post-primary disease implies that reinfection cannot occur within five years of infection. It is likely that reinfection can occur at any time, although it may be less likely than initial infection once an individual has mounted an immune response to *M. tuberculosis* antigens. The effect of reinfection on the disease risk whilst an individual is still at a high risk of developing disease from the *preceding* infection or reinfection event,

is unclear. It is also unclear whether the first (primary) disease episode within 5 years of initial infection among individuals who have also been reinfected can be attributed to bacilli from the initial infection, from the subsequent reinfection or from both infection and reinfection events. This issue becomes important when considering the incubation period, serial interval and basic and net reproduction numbers for tuberculosis.

From these perspectives, it is unrealistic to apply the same 5-year cut-off criterion, used for defining 'primary', 'endogenous' and 'exogenous' forms to distinguish between different risks of developing disease. It is preferable to assume reinfection can also occur less than five years after initial infection and that the disease risks experienced differ according to whether individuals are at risk of developing their first primary episode, endogenous or exogenous disease, where

1. **exogenous disease** is *redefined* as **the first disease episode within five years of reinfection**, and

2. the definition of **endogenous disease** (by convention, disease occurring five or more years after initial infection or most recent reinfection) is *extended* to include **disease occurring within five years of infection or reinfection, if an individual has already experienced disease but no further reinfection in the meantime.**

The disease states modelled in TBDYN3 reflect these assumptions concerning the different risks of developing disease, and are described fully below.

3.1.2 General structure of TBDYN3

Figure 3.1 shows the overall structure of TBDYN3, in terms of the transitions between the individual disease states. Following the discussion in section 2.2, we will restrict our analyses to respiratory forms of tuberculosis in the white ethnic male population in England and Wales.

Diseased individuals are divided into three categories: those experiencing their first primary episode (P), endogenous disease (E_n), or exogenous disease (E_x), where endogenous and exogenous disease are as defined by points 1 and 2 above.

Similarly, TBDYN3 separates infected individuals into three categories, which reflect those *at risk* of developing each form of disease, namely:

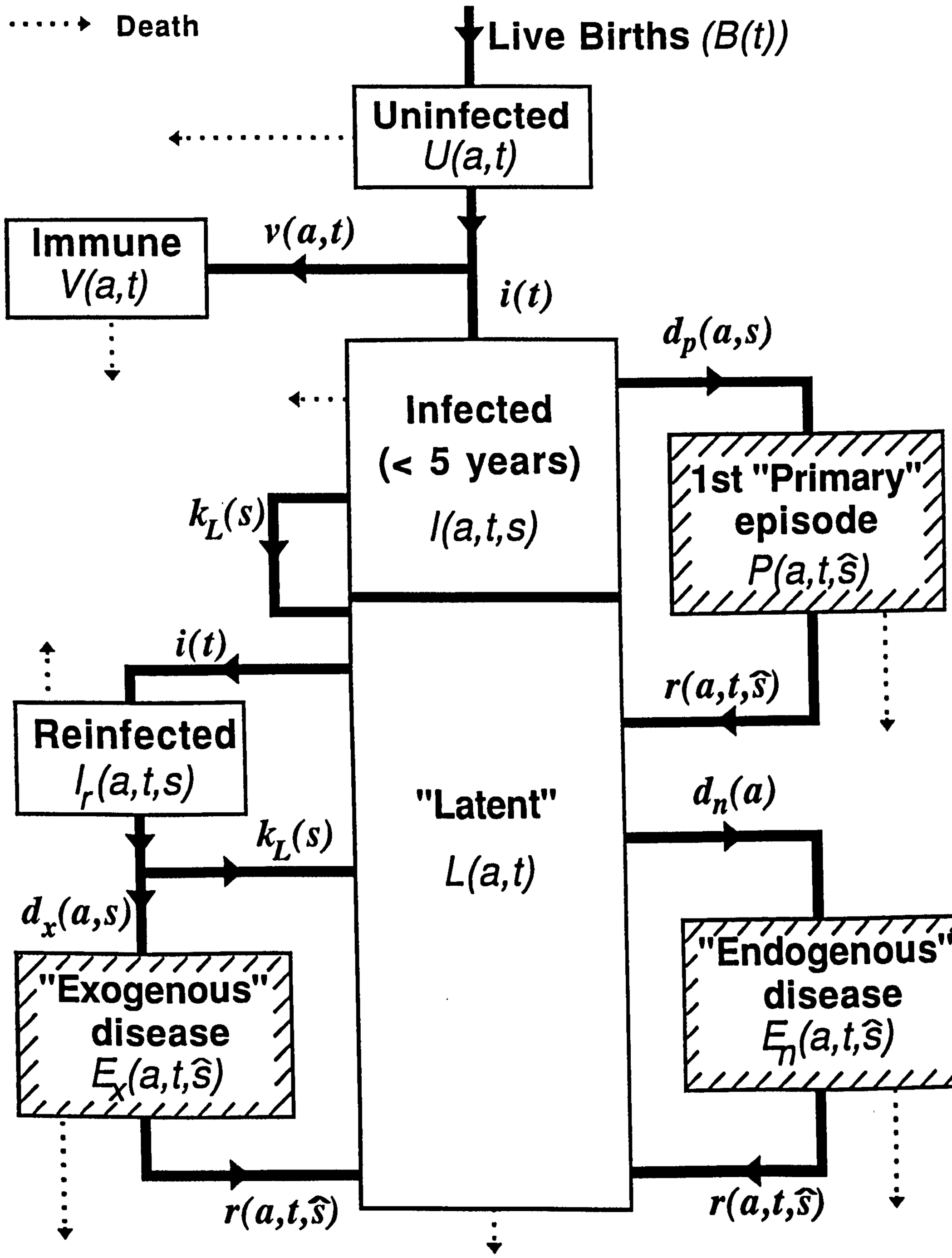


Figure 3.1: Schematic diagram showing the overall structure of the TBDYN3 model.

1. Those in the (recently) infected class (I), defined to be at risk of developing their first primary episode,

2. Those in the 'latent' class (L), defined to be at risk of developing endogenous disease or of being reinfected, and

3. Those in the (recently) reinfected class (I_r), defined to be at risk of exogenous disease.

Individuals first enter the 'latent' class either five years after initial infection in the absence of any disease, or after recovering from their first primary episode. Similarly, they return to the 'latent' class after recovering from endogenous or exogenous disease, or five years after reinfection in the absence of any disease in the meantime.

It is recognized that in reality, individuals can probably be reinfected at any time (see section 3.1.1). For simplicity, we assume that individuals cannot be reinfected whilst they are experiencing disease, or whilst they are at an already high risk of developing their first primary episode or exogenous disease (i.e. whilst in the infected or reinfected classes respectively). Individuals reinfected whilst in the 'latent' class enter the reinfected class R, where the risk of developing exogenous disease behaves analogously to the risk of developing the first primary episode after initial infection.

Two further states are included in TBDYN3: those uninfected (i.e. 'susceptible' to first infection), U, and those protected by BCG vaccination (V). It is assumed that all individuals are born susceptible, and that once infected, they retain infection for life. Individuals effectively vaccinated (determined by the vaccine coverage and the vaccine efficacy) are assumed to be protected for life. The effects of these assumptions on model predictions are discussed in section 3.3.3.

Appendix A.1 summarizes the partial differential equations describing the general formulation of TBDYN3. Table 3.1 summarizes the names and definitions of the individual disease states. Table 3.2 summarizes the transition variables between these states, which are discussed in greater detail below.

3.1.3 Transitions between disease states

The parameter $i(t)$ determines the rate at which uninfected or 'latent' individuals become infected or reinfected respectively. This depends only the time t and is not defined here in terms of the contact between uninfected and 'infectious' individuals. This assumption

Table 3.1: Definitions of class variables used in TBDYN3

Variable name	Definition
$B(t)$	Number of live births at time t.
$U(a,t)$	Number of uninfected individuals of age a at time t.
$V(a,t)$	Number of individuals of age a at time t who are protected by BCG vaccination.
$I(a,t,s)$	Number of individuals of age a at time t who have been infected for time s (≤ 5 years) without having yet developed disease.
$P(a,t,\hat{s})$	Number of individuals of age a experiencing their first primary episode at time t, who have been diseased for time $\hat{s}$.
$L(a,t)$	Number of individuals of age a at time t in the 'latent' class i.e. those who have either just recovered from their first primary episode, or who have been infected for more than five years.
$I_r(a,t,s)$	Number of individuals of age a at time t, who have been reinfected for time s (≤ 5 years) and who have not yet developed exogenous disease.
$E_x(a,t,\hat{s})$	Number of individuals of age a with exogenous disease at time t, who have been diseased for time $\hat{s}$.
$E_n(a,t,\hat{s})$	Number of individuals of age a who have endogenous disease at time t, who have been diseased for time $\hat{s}$.

is modified in Chapter 5, when we consider the net and basic reproduction numbers for tuberculosis. The model assumes homogeneous mixing between individuals, an assumption which is probably more appropriate for the early part of this century in England and Wales, when the risks of infection were high, than for recent years. The risk of infection may in fact vary slightly with the age of individuals, but the extent and nature of this variation is unclear [113].

In reality, the infection rate for uninfected individuals probably exceeds the reinfection rate for those in the 'latent' class (see section 3.1.1). There are no empirical estimates of the magnitude of the difference, and so we assume, as a first approximation, that the infection and reinfection rates are identical.

Individuals in the infected class I develop their first primary disease episode at a rate of $d_p(a,s)$, which depends on the age at infection and declines with time since infection. The rate in each year following first infection is a factor of $d_p(a,0)$ (i.e. the rate *immediately* after first infection). Approximate values for these factors (relative rates) were obtained from the estimated 'relative risks' of developing disease in each year following tuberculin

Table 3.2: Definitions of transition variables used in TBDYN3.

Variable name	Definition
$i(t)$	Rate at which uninfected and 'latent' individuals become infected and reinfected respectively at time t.
$v(a,t)$	Proportion of uninfected individuals of age a who have been immunized at time t.
$d_p(a,s)$	Rate at which individuals infected at age a who have been infected for time s develop their first primary disease episode.
$r(a,t,\hat{s})$	Rate at which individuals of age a at time t recover from their first primary episode, exogenous and endogenous disease, at time $\hat{s}$ after disease onset.
$k_L(s)$	Rate at which individuals who have been infected or reinfected for time s without developing disease move into the 'latent' class. $k_L(s) = 0$ if $0 < s < 5$ and ∞ for $s = 5$ years.
$d_x(a,s)$	Rate at which individuals reinfected at age a, who have been reinfected for time s develop exogenous disease.
$d_n(a)$	Rate at which 'latent' individuals of age a develop endogenous disease.
$d_+(a)$	Relative contribution of sputum-positive forms to the disease incidence among individuals of age a.
$m_+(t,\hat{s})$	Mortality rate of among sputum-positive cases at time t and time $\hat{s}$ since disease onset.
$m_g(a,t)$	Mortality rate of sputum-negative and non-diseased individuals in the general population of age a at time t.

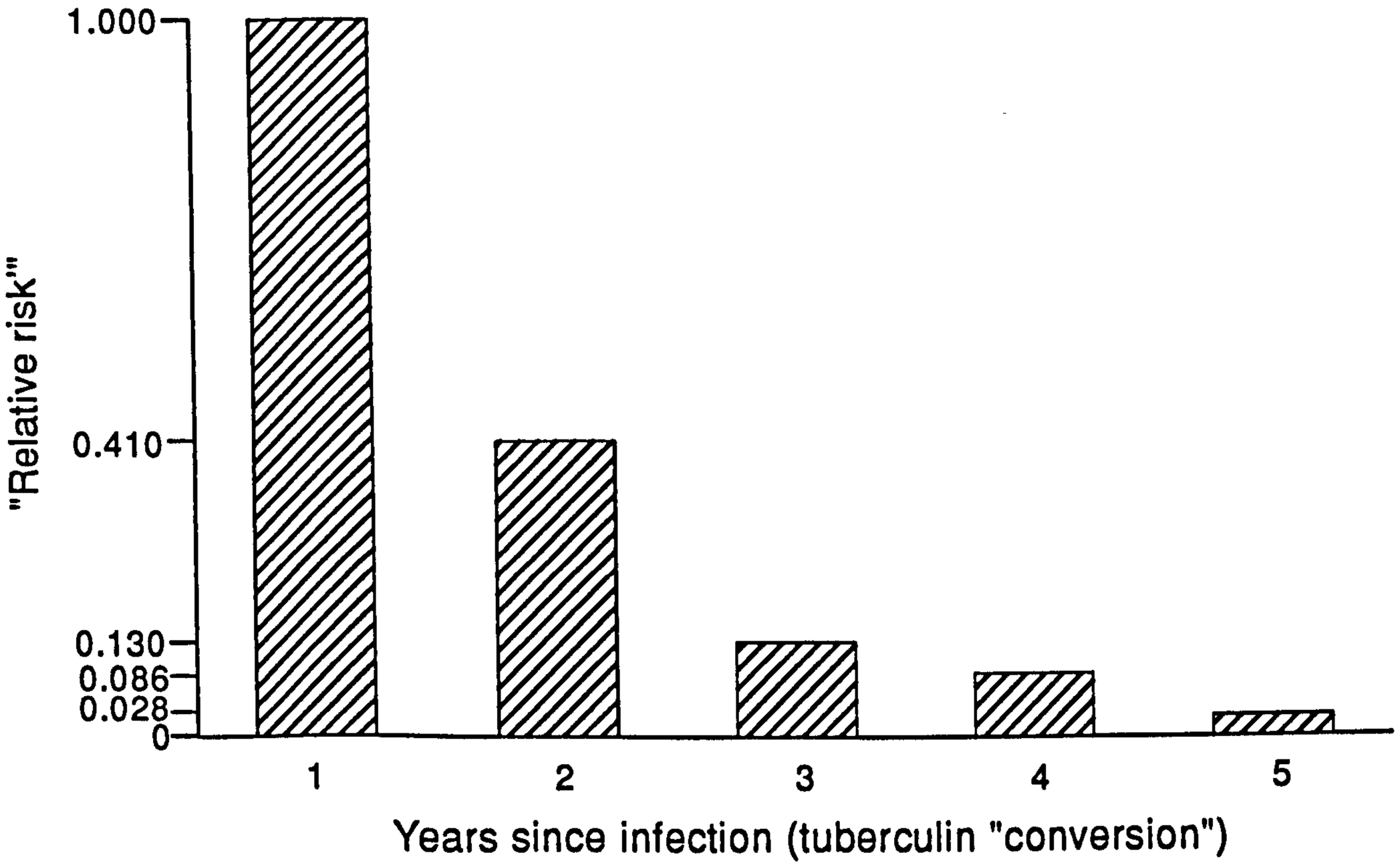


Figure 3.2: Approximate risks of developing the first primary disease episode, relative to those experienced during the first year after infection, as estimated from the distribution of the time interval between 'tuberculin conversion' and disease onset of those who were tuberculin-negative at the start of the UK MRC BCG trial [29].

conversion among those who were tuberculin-negative at the start of the UK MRC BCG trial [29], as shown in Figure 3.2. The 'relative risk' for a given year after 'conversion' is assumed to be the ratio between

1. the proportion of the total disease incidence among initially tuberculin-negative individuals which occurred in that year following 'conversion', and

2. the corresponding proportion which occurred during the first year after 'conversion'.

The estimates provide only a rough indication of the magnitude of the actual 'relative risks'; these could not be estimated more accurately given that the date of tuberculin 'conversion' was not established for almost half of the initially tuberculin-negative individuals who developed disease in the UK MRC trial (see section 1.2.1.1). The data from the trial were not stratified according to the type of disease individuals developed in each year following 'conversion', and hence we assume that similar estimates for the 'relative risks' would be obtained for respiratory forms alone.

According to the approach employed here, the rate at which infected individuals develop

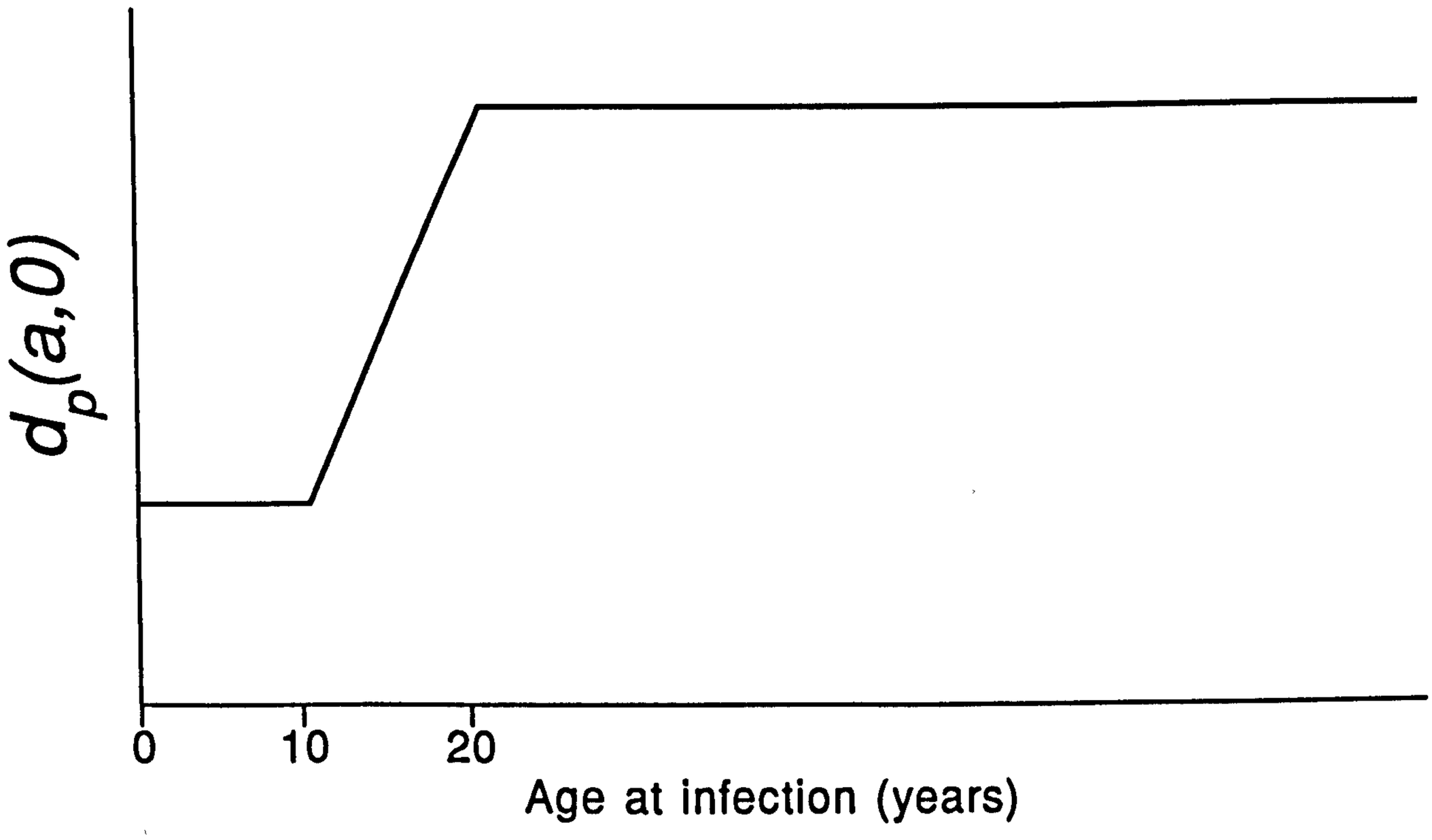


Figure 3.3: Relationship between the risk of developing the first primary disease episode and the age at infection assumed in TBDYN3.
The relationships (a) between the risk of developing exogenous disease and the age at reinfection, and (b) between the risk of developing endogenous disease and the current age of an individual are assumed to be analogous.

disease immediately after infection ($d_p(a,0)$) defines the rate during the following five years. $d_p(a,0)$ is assumed to depend on the age at infection in the manner shown in Figure 3.3 (i.e. constant for those infected when aged 0–10 years, increasing linearly to age 20 years, and constant thereafter). This formulation means that the rate at which individuals infected between the ages 10 and 20 years develop disease immediately after infection can be expressed in terms of that for 0–10 year olds ($d_p(10,0)$) and individuals aged over 20 years ($d_p(20,0)$)[2].

The assumed increase between the ages 10 and 20 years in the rate at which individuals develop disease is consistent with observed age-specific patterns in mortality rates among cohorts during the pre-chemotherapy era (see section 1.1). Reasons for the decreasing risks of developing disease according to time since 'conversion' found during the UK MRC BCG trial were discussed in section 1.2.1.1.

The number of individuals experiencing their first primary episode when aged a at time t

[2] $d_p(a,0) = \frac{(d_p(20,0) - d_p(10,0))}{10} a + 2d_p(10,0) - d_p(20,0)$ for $10 < a < 20$.

and at time s following initial infection is given by the product of the number of individuals in the infected class of age a at time t who have been infected for time s ($I(a,t,s)$), and the corresponding rate at which they develop disease, $d_p(a-s,s)$. Summing over the possible values for s for individuals in the infected class I ($0 \leq s < 5$), the total number of new cases experiencing their first primary episode at age a at time t is given by the integral term in equation A.4 in Appendix A.1.

Individuals in the reinfected class R experience exogenous disease at a rate $d_x(a,s)$. This is assumed to depend on the age at reinfection and on the time since reinfection in the same way as that at which $d_p(a,s)$ depends on the age at and the time since *first infection* (see Figures 3.2 and 3.3). This is intuitively reasonable, although there are in fact no empirical data indicating whether the risk of exogenous disease is a function of age or time since reinfection. Equation A.7 in Appendix A.1 describes the rate of change in the incidence of exogenous disease with age and time.

Individuals in the 'latent' class develop endogenous disease at a rate $d_n(a)$, which is assumed to depend only on their *current age*, as shown in Figure 3.3. In reality, this probably increases even after age 20 years, but the extent and nature of this increase is not clear from the available data (see section 1.2.1.4).

A proportion $d_+(a)$ of all diseased individuals is assumed to be sputum-positive, which depends only on their age and not on whether they are experiencing their first primary episode, endogenous or exogenous disease. We use data from Norway from the time period 1951–1968 (see section 1.1) to estimate $d_+(a)$. These values are illustrated in Figure 3.4. On this basis, it was assumed that 10%, 65% and 85% of respiratory disease among 0–10, 20 and 90 year olds respectively was sputum-positive, and that the proportions increased linearly between these age groups.

Individuals are assumed to be diseased for two years, unless they die in the meantime. The mortality rate experienced depends on their sputum status. Individuals who are *sputum-negative* face a mortality rate $m_g(a,t)$, which is identical to that of other individuals of the same age in the general population. *Sputum-positive* cases are assumed to face a mortality rate $m_+(t,\hat{s})$, which depends only on time t and the time since they developed disease ($\hat{s}$), and not on their age. Until 1950, 35% and 23% of sputum-positive cases are assumed to die during the first and second years after they develop disease, which is equivalent to an overall case-fatality rate of 50%. This is consistent with estimates of the *overall* case-fatality

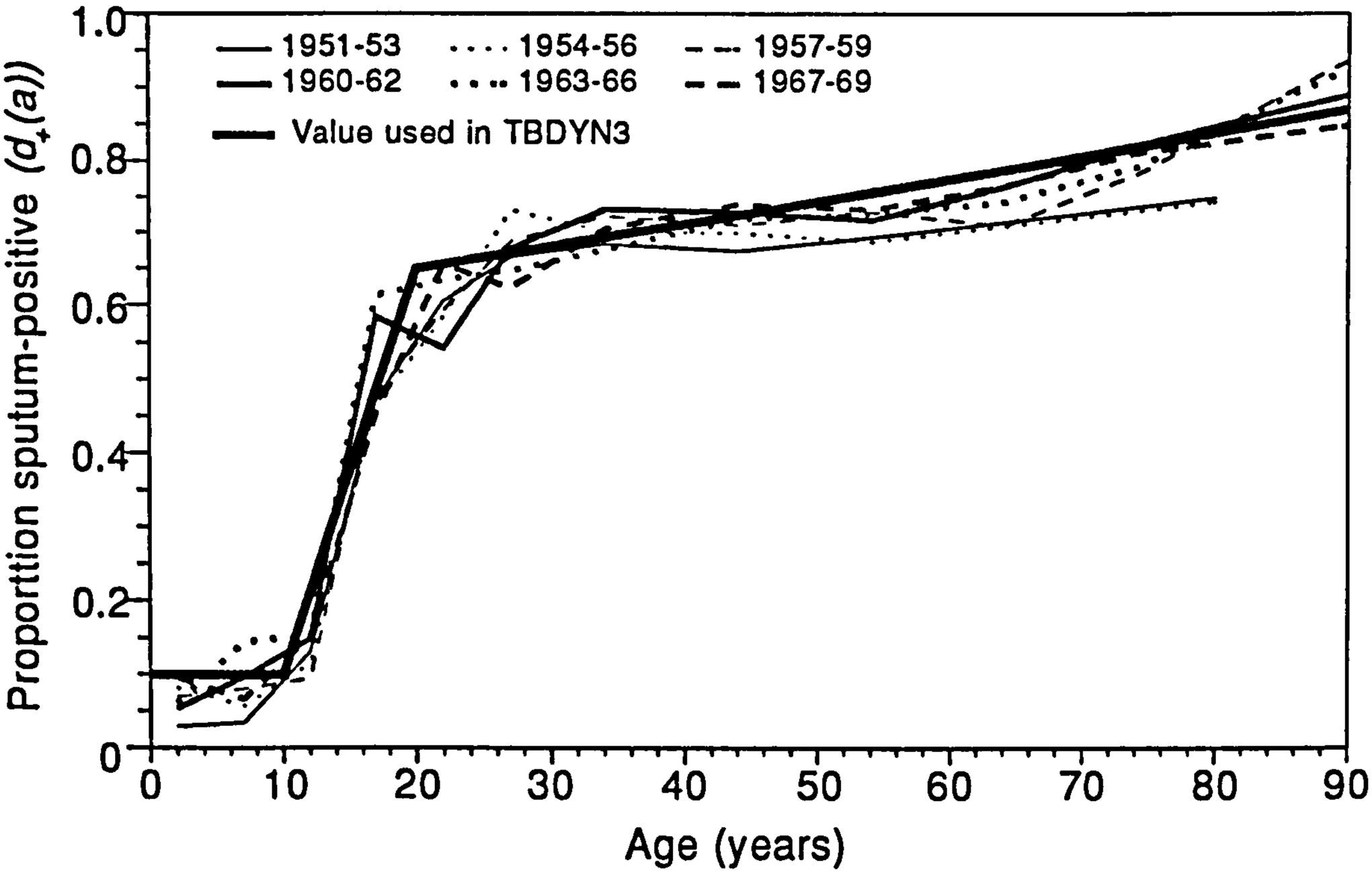


Figure 3.4: Summary of the relative contribution of sputum-positive disease to age-specific notifications of pulmonary tuberculosis in males in Norway during the period 1951–1969, together with the corresponding values assumed by TBDYN3 ($d_+(a)$).

rates from follow-up studies during the pre-chemotherapy era (see section 1.1), and from a longitudinal study of the natural history of tuberculosis in South India during the period 1961–1966 [25]. The ratio between the case-fatality rate for sputum-positive cases in the first and second years after disease onset is assumed to be constant throughout the time period considered (i.e. 35/23 $\approx$1.4).

The *overall* case-fatality rate for sputum-positive cases is assumed to decline steadily after 1950, reaching 30% and 25% by 1953 and 1956 respectively. This is held constant until 1976, and is assumed to be identical to the mortality rate in the general population thereafter. These assumptions are consistent with findings from follow-up studies of treated patients after 1950 in England and Wales [40,41] (see section 1.1), although it is recognized that the assumed case-fatality for sputum-positive cases after 1976 may be slightly low.

3.1.4 Incidence of disease

The total number of individuals of age a at time t ($N(a,t)$) in the population is given by the sum of individuals in each of the disease states. The total number of incident cases of age a at time t ($C(a,t)$) is given by the sum of new cases experiencing their first primary episode

($P(a,t,0)$) and endogenous and exogenous disease ($E_n(a,t,0)$ and $E_x(a,t,0)$ respectively):

$$C(a,t) = P(a,t,0) + E_n(a,t,0) + E_x(a,t,0) \tag{3.1}$$

Each of these depend on the past disease history of individuals born at time $t-a$, and are functions of the age-specific rates at which individuals develop the first primary episode, endogenous and exogenous disease. These, in turn are functions of the rates among 10 and 20 year olds ($\{d_p(10,0), d_p(20,0)\}$, $\{d_n(10), d_n(20)\}$ and $\{d_x(10,0), d_x(20,0)\}$ (see footnote on page 102), denoted by $\underline{\mathbf{d}}_\mathbf{p}$, $\underline{\mathbf{d}}_\mathbf{n}$ and $\underline{\mathbf{d}}_\mathbf{x}$ respectively.

The incidence rate of disease in individuals of age a at time t, $Y(a,t;\underline{\mathbf{d}}_\mathbf{p},\underline{\mathbf{d}}_\mathbf{n},\underline{\mathbf{d}}_\mathbf{x})$ is given by:

$$Y(a,t;\underline{\mathbf{d}}_\mathbf{p},\underline{\mathbf{d}}_\mathbf{n},\underline{\mathbf{d}}_\mathbf{x}) = \frac{C(a,t;\underline{\mathbf{d}}_\mathbf{p},\underline{\mathbf{d}}_\mathbf{n},\underline{\mathbf{d}}_\mathbf{x})}{N(a,t)} \tag{3.2}$$

and the corresponding incidence in specific age groups (e.g. 5–14 year olds) at time t is therefore given by:

$$Y(a_{5-14},t;\underline{\mathbf{d}}_\mathbf{p},\underline{\mathbf{d}}_\mathbf{n},\underline{\mathbf{d}}_\mathbf{x}) = \frac{C(a_{5-14},t;\underline{\mathbf{d}}_\mathbf{p},\underline{\mathbf{d}}_\mathbf{n},\underline{\mathbf{d}}_\mathbf{x})}{N(a_{5-14},t)} \tag{3.3}$$

$$= \frac{\sum_{a=5}^{14} C(a,t;\underline{\mathbf{d}}_\mathbf{p},\underline{\mathbf{d}}_\mathbf{n},\underline{\mathbf{d}}_\mathbf{x})}{\sum_{5}^{14} N(a,t)} \tag{3.4}$$

Hence we see that for given input data on the incidence of infection, the age and time-specific mortality rate, and the numbers of live births and vaccinations, only the parameters $\underline{\mathbf{d}}_\mathbf{p}$, $\underline{\mathbf{d}}_\mathbf{n}$ and $\underline{\mathbf{d}}_\mathbf{x}$ are required to describe the disease incidence for a given age group and for a given year. By their definition, these parameters also determine the relative contribution of endogenous and exogenous forms to the age and time-specific disease incidence. In this chapter, we first derive estimates of these parameters, and explore how they determined age and time-specific patterns in the overall disease incidence since 1900 in England and Wales.

3.2 Methods and input data

3.2.1 Input data

To describe fully the disease incidence among individuals aged 0–100 years since 1900 in England and Wales, TBDYN3 uses the following input data:

1. **Annual age and time-specific mortality rates for sputum-negative and non-diseased individuals, $m_g(a,t)$.** Age and time-specific mortality rates for *all individuals in the general population* are provided on an annual basis since 1841 by the UK Government Actuary's Department. The corresponding mortality rates for sputum-negative and non-diseased individuals were derived from these data by taking away the deaths among sputum-positive individuals, which were estimated using TBDYN3.

2. **Annual number of live births, $B(t)$.** These are available from OPCS publications also since 1841.

3. **The annual risk of infection, $i(t)$.** Following the discussion in section 2.3, the annual risk of infection was assumed to be 14% in 1901, declining at 4% pa until 1950, and at 13% pa thereafter. We assume the annual risk of infection declined at 2% pa between 1880 and 1900, and was constant before this. The rate of decline from 1880 is consistent with that found for tuberculous meningitis mortality rates among 0–4 year olds (see section 2.3), although it implies that the annual risk of infection was extremely high at 22% in 1880.

4. **BCG vaccination data.** It is assumed that most of the individuals vaccinated in England and Wales were aged 13 years (see section 2.4). The actual numbers of individuals vaccinated under the schools scheme from 1954 until 1960 were used in TBDYN3, as published in the Chief Medical Officer's Report [110]. These data are not stratified by sex and hence males are assumed to have comprised 50% of all those vaccinated. From 1960, it is assumed that 80% of uninfected 13 year olds males were vaccinated each year. The vaccine efficacy is assumed to be 77% which is consistent with corresponding estimates from the UK MRC BCG trial during the 1950s [38] (see section 2.4).

The mortality and births estimates were extrapolated back from 1841 to 1800, so that the transmission dynamics could be considered among all individuals from 1900.

3.2.2 Estimating the age-specific risks of developing disease

The system of partial differential equations in Appendix A.1 was first reduced to a set of ordinary differential equations following individual cohorts from birth throughout their lifetime. These equations were programmed in 'C', using time-steps of 1 year, using the Euler method — this meant that the continuous time transition rates could be replaced by annual transition risks, and hence yearly data on the number of births, age and time-specific mortality, vaccination and the risk of infection, provided in their original format, could be used[3].

The risks of developing the first primary, endogenous and exogenous disease for 0–10 year olds and individuals aged over 20 years ($\underline{d}_p$, $\underline{d}_n$ and $\underline{d}_x$) were estimated by fitting model predictions of the disease incidence to age-specific notification data for white ethnic males, available for the period 1953–1988 (see section 2.2). The annual risk of infection can be estimated reliably on the basis of tuberculous meningitis mortality data only since 1901 (see section 2.3). Hence the fitting was restricted to notification data only among individuals born since then, illustrated in Figure 3.5. The notifications for 55–64 and 65–74 year olds are those in the general population, since they closely correspond to those in the white ethnic population (see section 2.2). In total, the fitting was based on 252 notification data points.

Best-fitting estimates for the 6 parameters, $\{d_p(10,0), d_p(20,0)\}$, $\{d_n(10), d_n(20)\}$ and $\{d_x(10,0), d_x(20,0)\}$ were obtained by finding a minimum sum of squares of the differences between the incidence predicted by the model and the notification rates shown in Figure 3.5, using the Levenburg-Marquardt method [114]. This uses an iterative approach, and converges to a minimum sum of squares using

1. a set of starting parameter estimates, and
2. expressions for the partial derivative of the expected age and time-specific disease incidence with respect to each of the six parameters.

The latter determines only the speed at which the procedure converges to a minimum

[3]It is generally more conventional to use Runge-Kutta or Bulirsch-Stoer methods to solve ordinary differential equations [114]. These methods are less convenient than the Euler method for implementing the equations in TBDYN3, given that they require input data (i.e. mortality rates and numbers of births) representing continuous-time, rather than discrete-time, transitions.

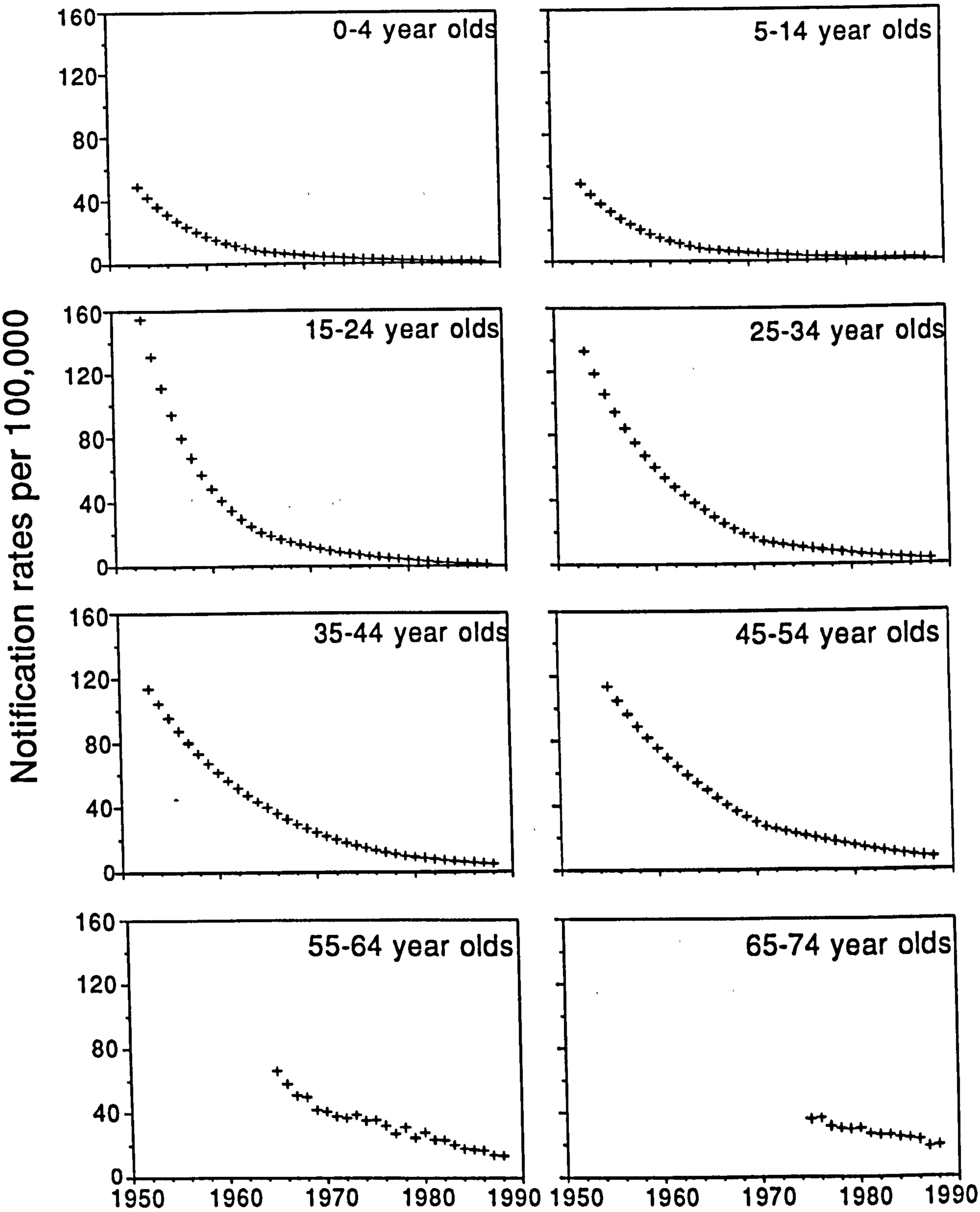


Figure 3.5: Estimated age-specific notification rates of respiratory tuberculosis in white ethnic males used to derive the risks of developing disease by TBDYN3. These notification rates were estimated from data on all forms of tuberculosis in white ethnic males during the national surveys. See section 2.2 for further details. We restrict the fitting to data points corresponding to individuals born since 1901.

sum of squares, and does not affect the best-fitting parameter values themselves. Hence, given that these partial derivatives cannot be expressed analytically in terms of the 6 parameters, realistic approximations were used (see Appendix A.2).

The function of the sum of squares of the difference between the expected incidence and the actual notifications may have several (local) minima, which correspond to different sets of parameter values. Iterative procedures, such as the Levenburg-Marquardt method, sometimes converge to such local minima, depending on the set of starting parameter values used. To compensate for this problem, the fitting process was repeated several times starting from different sets of parameter values. The best-fitting parameter values were considered to be those yielding the *overall* minimum sum of squares.

3.2.3 Sensitivity analyses of parameter estimates

Formal standard errors for the parameter values derived using TBDYN3 cannot be obtained easily. For general (non-)linear models, standard errors on fitted parameter values can be obtained from the covariance matrix of the converged estimates if

1. the (notification) data points used in the fitting procedure are independent and identically distributed and
2. the differences between the estimated incidence and the notifications used are normally distributed [115].

The notification rates of respiratory tuberculosis for white ethnic males used in the fitting procedure were estimated by assuming that they declined from their level in 1953 at the same rate as those of all forms of tuberculosis (see section 2.2), as estimated from data from the national tuberculosis surveys [97, 101]. Hence the notification rate of *respiratory* tuberculosis used for a given year is determined by that in both the preceding year and in 1953, and also by the magnitude of the notification rates of *all forms* of tuberculosis found during the survey years. The estimated notification rates used in the fitting procedure are therefore not independent and identically distributed and hence the first condition above is not satisfied for the parameters estimated using TBDYN3. On this basis, formal standard errors for the parameter estimates cannot be derived using the corresponding covariance matrix.

It is likely that the notification rates of *all forms* of tuberculosis during the national surveys were themselves subject to some random variability, which would have influenced the magnitude of the estimated notification rates of *respiratory* tuberculosis for white ethnic males. This could have led to slightly different estimated risks of developing tuberculosis from those obtained using the original data. This only applies to the notifications used for 0–54 year olds, as those used for older individuals are assumed to approximate notifications in the general population (see section 2.2.3).

To explore the sensitivity of the disease risk estimates to the magnitude of the notification rates, the fitting procedure was repeated on 50 other simulated data sets of notifications of *respiratory* tuberculosis among white ethnic males, which could have been derived from the notifications of all forms of tuberculosis from the national surveys, if the numbers of cases during the survey years had been subject to some random variability. This variability in the number of cases of *all forms* of tuberculosis during the survey years was assumed to follow a Poisson distribution[4] with a mean given by the 'observed' numbers of cases. Hence the notification rates of respiratory tuberculosis obtained by assuming that the numbers of cases of all forms of tuberculosis followed a Poisson distribution during the survey years represent the plausible range in which the true notification rates would be expected to lie.

We provide further details of the method used to generate the data sets in Appendix A.3, and Figure A.1 summarizes the 95% range of the notification rates of respiratory tuberculosis covered by the simulated notification data sets.

3.2.4 Application to Dutch data

Estimates of the age-specific risks of developing disease were also obtained using data from the Netherlands for comparison against those obtained by Sutherland *et al* [7].

Sutherland *et al* based their estimates of the disease risks on notifications of respiratory tuberculosis among 15–69 year olds during the period 1951–1970. The analyses using TB-DYN3 were based on the same notifications, but covering the age groups 0–69 years, among individuals born since 1892, as birth and mortality data (obtained from the Dutch Central Statistics bureau) are available only since this year. The notifications were generously provided by Dr. K. Styblo at the International Union Against Tuberculosis. Following the assumptions of Sutherland *et al* [7], the annual risk of infection was taken to be 18% in

[4]This distribution is generally used to model random variability in 'count' data [116]

1892, declining by 2.5% pa until 1911, at 4% pa until 1940 and at 13% pa thereafter.

3.3 Estimates of the risks of developing disease

3.3.1 Results

Table 3.3 summarizes the best-fitting estimates of the risks of developing disease obtained, together with the corresponding values for 15 year olds. Figure 3.6 compares the resulting age-specific incidence of disease against the notification data used in the fitting procedure.

Table 3.3: Best-estimates of the risks of developing disease (to 3 significant figures (sf)), as derived using TBDYN3 by fitting to notifications in white ethnic males in England and Wales. Values for the first primary and exogenous disease episodes refer to the risk during the first year after infection and reinfection respectively and the cumulative risks experienced during the first five years after infection/reinfection.

	Risks (% pa)				
Age (years)	First primary		Endogenous	Exogenous	
	1st year	*Cumulative (1st 5 yrs)*		1st year	*Cumulative (1st 5 yrs)*
0–10	2.42	*3.96*	1.39×10^{-7}	3.69	*6.00*
15	5.59	*9.01*	0.0144	4.38	*7.10*
> 20	8.76	*14.0*	0.0288	5.07	*8.19*

Table 3.3 shows that about 4% of 0–10 year olds were estimated to develop disease within 5 years of infection, as compared with about 14% of those infected when aged over 20 years. This higher risk of developing disease following infection during young adult life, as compared with that in childhood, is consistent with patterns in age-specific mortality rates seen among birth cohorts during the pre-chemotherapy era (see section 1.1).

The risks of developing the first primary episode among individuals aged over 15 years slightly exceeds those estimated from the UK MRC trial, which found that 6.9% and 5.5% of individuals developed disease within 10 years of 'conversion' to tuberculin-positivity at age 14–15 and 16 years respectively (see section 1.2.1.1). This suggests that TBDYN3 may slightly *underestimate* the population considered to be at risk of developing the first primary episode, and hence the estimated risk of developing the first primary episode has to exceed that found during the MRC trial to attain the notification rate. Such an overestimate could be a consequence of the simplifying assumptions incorporated into TBDYN3 (e.g. that vaccination confers full protection against infection for 77% of individuals, and that

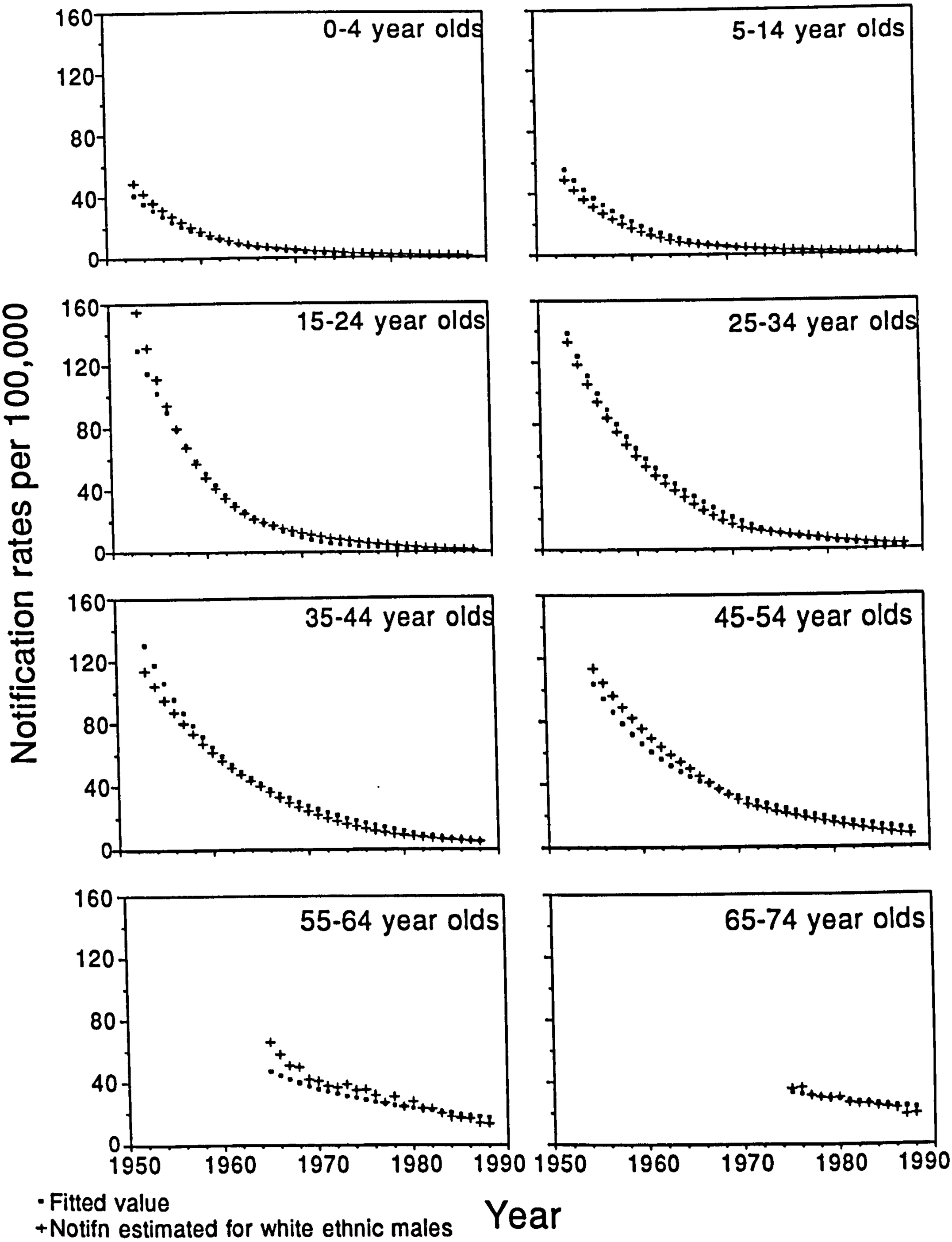


Figure 3.6: Comparison between the incidence of respiratory tuberculosis in 0–74 year old males estimated using TBDYN3 for England and Wales, and the notifications used in the fitting process.

the risk of infection depends only on the calendar year and not on the age of an individual). The effect of these simplifications on the magnitude of the parameter estimates obtained is discussed in detail in section 3.3.3.

Figure 3.6 suggests that the overall fit of model predictions to the disease incidence was good, although, for most age groups, TBDYN3 underestimated the given notification rate during the early 1950s. This could again be a consequence of simplifying assumptions incorporated in TBDYN3. Sutherland *et al* [7] also found that the fit of their model to the Dutch data was worse for the early 1950s.

The estimated risk of endogenous disease among individuals aged over 20 years is of a similar order of magnitude to that derived by Styblo [6] and Horwitz [49] for the Netherlands and Denmark respectively (see section 1.2.1.4). For individuals aged over 20 years, the risk of developing the first primary episode exceed both the risks of developing endogenous and exogenous disease. Sutherland *et al* [7] obtained similar findings for the Netherlands. This is consistent with results from observational studies in developed countries, which found a lower disease incidence among individuals with weakly tuberculin-positive reactions than among tuberculin-negative individuals (see section 1.2.1).

The estimates for the risks of developing endogenous and exogenous disease among 0–10 year olds are based on fewer data points than the corresponding risk of developing the first primary episode, and hence, are less reliable. This stems from the assumption in TBDYN3 that individuals can first develop endogenous and exogenous disease either five years after infection or after recovering from the first primary episode. This constrains 3 year olds to be the youngest at risk of developing these forms of disease, whereas individuals can develop a first primary episode at any age. This may explain the higher estimated risk of developing exogenous disease, as compared with that of developing the first primary episode, for 0–10 year olds.

3.3.2 Sensitivity analyses

Table 3.4 summarizes the range of parameter estimates derived by repeating the fitting procedure on the 50 simulated data sets described in section 3.2.3.

Considering the risk of developing the first primary episode during the first year after infection at age 0–10 years, this shows that 95% of the parameters obtained were within the range 2.18–2.58%, whereas the risk estimated using the original data set was 2.42%.

Table 3.4: Summary of the disease risk estimates obtained by repeating the fitting procedure on 50 simulated notification data sets, derived by sampling the numbers of notifications from the national tuberculosis surveys from the Poisson distribution. Values for the risks of developing the first primary and exogenous disease refer to those experienced during the first year after infection/reinfection respectively.

Parameter	Age group (years)	Original data	Simulated data			
			Mean	Lower 95%	Median	Upper 95%
First Primary Episode	0–10	2.42	2.40	2.17	2.41	2.58
	>20	8.76	8.55	7.96	8.64	9.05
Endogenous	0–10	1.39×10^{-7}	9.55×10^{-7}	2.36×10^{-8}	1.23×10^{-5}	9.06×10^{-4}
	>20	0.0288	0.0295	0.0287	0.0296	0.0304
Exogenous	0–10	3.69	5.15	3.33	4.68	8.52
	>20	5.07	5.07	4.79	5.06	5.28

Similarly, the estimated risks of developing the first primary episode, endogenous and exogenous disease for individuals aged over 20 years using the original data set compare well against the 95% range of those derived using the simulated data.

This 95% range is not equivalent to a 95% confidence interval — on the basis of these results, for example, it is not possible to say with 95% certainty that the true risk of developing disease would lie in the given interval for the given set of assumptions. At best, the 95% range can be interpreted as

> "the range in which 95% of parameter estimates would lie, given that the 'true' notification rates of respiratory tuberculosis were in a given (realistic) interval".

Table 3.4 shows that the risks of developing endogenous and exogenous disease for individuals aged 0–10 years are relatively insensitive to the level of the notification rates, and hence are less reliable than the other disease risk estimates. 95% of the estimates for the risks of developing exogenous disease ranged between 3.33% and 8.52%. This range overlaps the range of values obtained for the risks of developing exogenous disease among individuals aged 20 years.

The risk of developing endogenous disease for 0–10 year olds also lies in a broad range, with the upper and lower 95th percentiles differing by a factor of 10^4, although the absolute magnitude of the risk is small, relative to the other disease risk estimates.

These results are consistent with the fact that estimates of the risks of developing endogenous and exogenous disease among individuals aged 0–10 years are based on fewer data, as compared with those for other age groups (see section 3.3.1).

Overall, this suggests that

1. the estimated risks of developing the first primary episode for all age groups and those of developing endogenous and exogenous disease for individuals aged over 20 years are relatively insensitive to slight variation in the notification rates of respiratory tuberculosis and

2. of all the parameter estimates, the risks of developing endogenous and exogenous disease among 0–10 year olds are the least reliable.

At best, TBDYN3 represents a simplification of a complex disease process, and thus the estimates for the risks of developing disease are valid only for the set of assumptions incorporated. The extent to which they reflect the 'true' risks of developing disease experienced in England and Wales since 1953 depends on how closely the simplifications approximate the true processes involved. These simplifications and their possible effects on the best-fitting parameter estimates are discussed in the next section.

Table 3.5: Comparison between the best-fitting values for the risks (to 3sf) of developing the first primary episode and endogenous disease derived assuming individuals cannot be reinfected, and the values obtained by including reinfection (original TBDYN3 model). Risks of the first primary episode and exogenous disease refer to those experienced during the first year after infection/reinfection respectively.

Assumption	Age (years)	Risks (% pa) 1st primary	Endogenous	Exogenous	Sum of squares
No exogenous disease	0–10	2.58	5.72×10^{-7}	—	
	15	8.09	0.0206	—	
	> 20	13.6	0.0411	—	25517
Original TBDYN3	0–10	2.42	1.39×10^{-7}	3.69	
	15	5.59	0.0144	4.38	
	> 20	8.76	0.0288	5.07	5232

3.3.3 Effect of simplifications on the disease risk estimates obtained

3.3.3.1 Relevance of reinfection disease

The relative frequency of endogenous and exogenous disease is controversial (see section 1.1), and most of the disease in developed countries in recent years among older individuals is attributed to endogenous reactivation [6].

The importance of reinfection in determining the observed disease incidence was explored by repeating the fitting procedure assuming that no reinfection, and hence no exogenous disease could occur. Table 3.5 summarizes resulting disease risk estimates, together with the goodness of fit, as given by the sum of squares of the differences between the predicted incidence and the notification rate used. The estimates derived using the original assumptions are also shown.

This shows that the estimated risks of developing the first primary episode and endogenous disease for all age groups are appreciably greater when it is assumed that individuals cannot experience exogenous disease, as compared with those derived assuming the converse, and the fit to the data is considerably worse. This is illustrated in Figure 3.7, which compares model predictions of the disease incidence against the notifications used. The fit is especially poor for 45–54 year olds, for whom the model fails to match the decline in the disease incidence between 1955 and 1970. For younger individuals, the effect of excluding

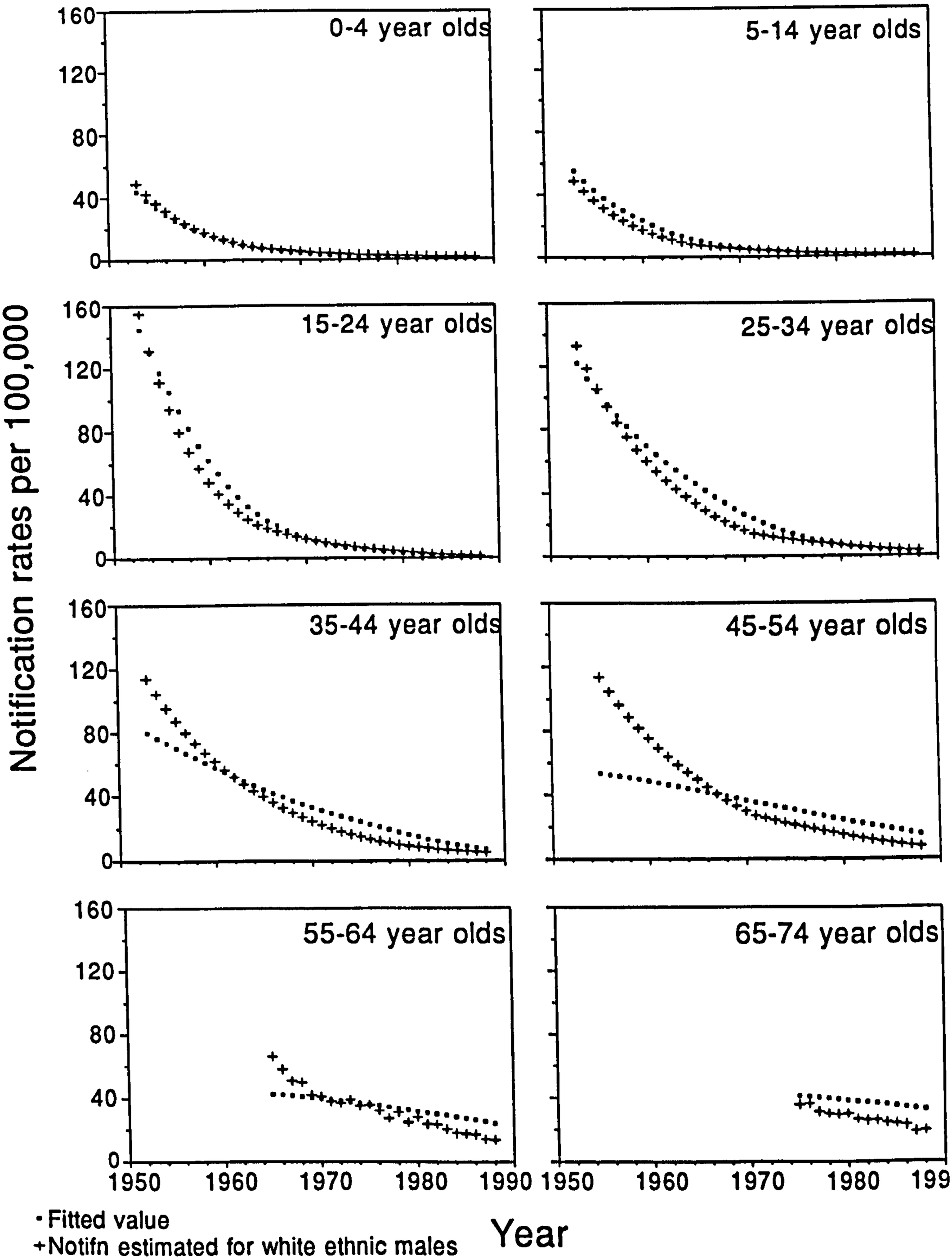


Figure 3.7: Comparison between the incidence of respiratory tuberculosis in 0–74 year old males, derived using the best-fitting parameters assuming individuals could not be reinfected and experienced *constant* risks of developing the first primary episode and endogenous disease after age 20 years, and the notifications used in the fitting process.

reinfection from TBDYN3 on the fit to notification rates is smaller.

Given the high risks of infection before 1950, most 45–54 year olds during the period 1955–1970 must have already been infected. Hence it is unlikely that the reduction in the notification rates among these individuals is attributable to a decline in the incidence of first primary episodes. If it is assumed that 45–54 year olds could not have been reinfected, then the rapid decline in the notification rate could only have occurred as a consequence of a dramatic reduction in the risk of developing endogenous disease over time, which seems unrealistic. This suggests that to attain a realistic fit to the notifications rates especially for 45–54 year olds for the given set of assumptions, reinfection has to be included into TBDYN3.

3.3.3.2 Effect of age-independent risks of infection

Some studies have suggested that the risk of infection increases with age, although the magnitude of the increase is not clear [113]. TBDYN3 assumes that the risk of infection depends only on the calendar year and not on the age of an individual, and hence may misrepresent the prevalence of infection and the size of the population at risk of developing disease in given age groups and years. To explore this question, the prevalence of infection predicted by TBDYN3 was compared against data from the national tuberculin surveys in 1949 and during the period 1971-1973 [103, 106] (see section 2.3).

Figure 3.8 compares the expected prevalence of infection among 0–50 year olds in 1949 used by TBDYN3, against the prevalence of tuberculin sensitivity found during the national tuberculin survey during the period 1949–1950 [103]. The two agree well for 5–16 year olds from south urban areas, although TBDYN3 may overestimate slightly the infection prevalence among 5–8 year olds. The prevalence of tuberculin sensitivity in southern urban areas probably more accurately reflected actual trends in the annual risk of infection with *M. tuberculosis* than did that in rural and north urban areas, as an appreciable proportion of tuberculin sensitivity was attributable to infection with *M. bovis* in the latter areas (see section 2.3).

The prevalence of infection predicted by TBDYN3 for individuals aged over 18 years, underestimates the corresponding prevalence of tuberculin sensitivity found in south urban areas during the survey. This is consistent with results from other studies [113], which suggests that the risk of tuberculous or other mycobacterial infection increased slightly

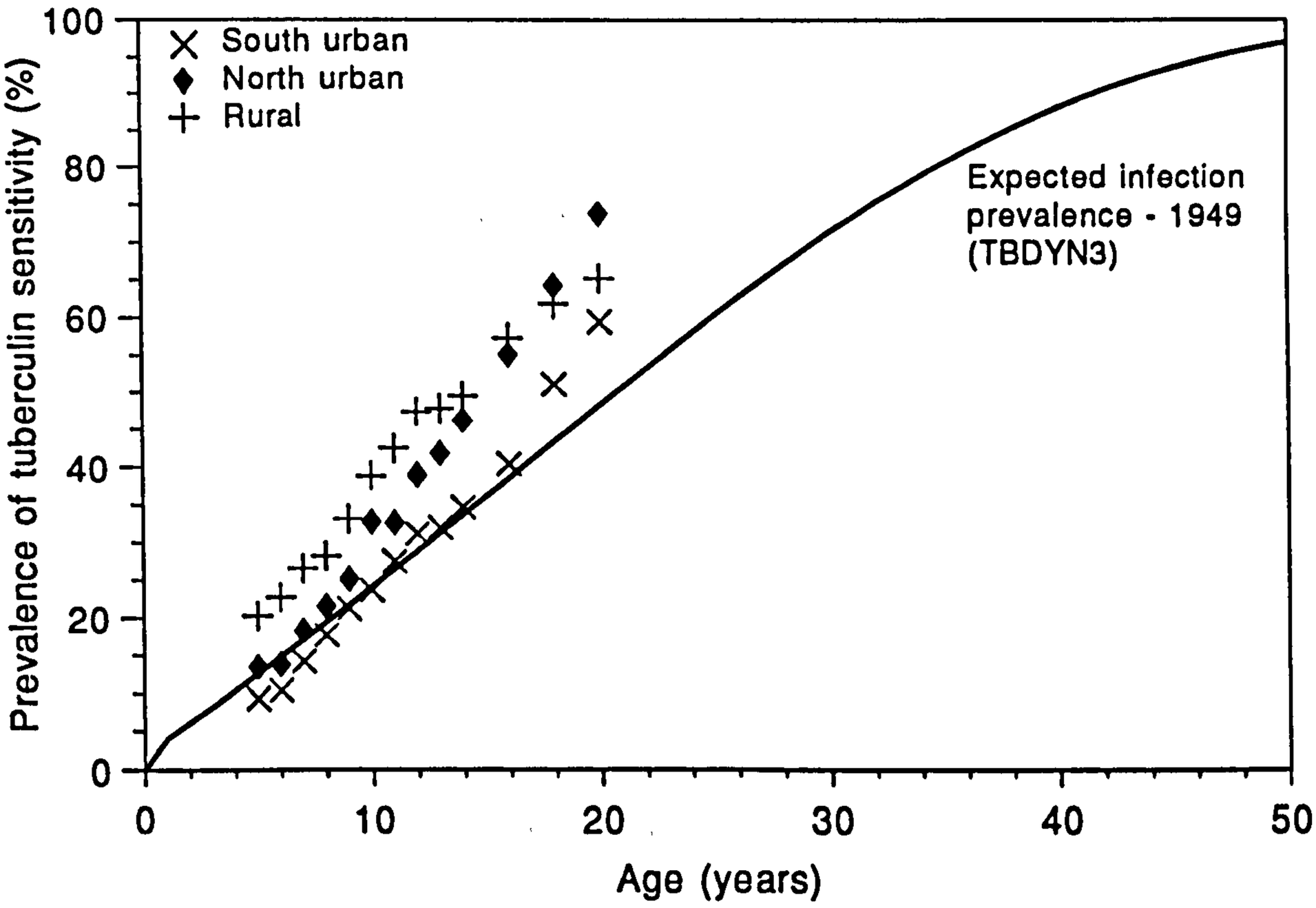


Figure 3.8: Comparison between the expected age-specific prevalence of infection in England and Wales in 1949, as used in TBDYN3 and the prevalence of tuberculin sensitivity in south urban, north urban and rural England, as found during the national tuberculin survey during the period 1949–1950 [103].

during adolescence.

The high risk of infection during the early 1900s (e.g. 14% in 1901) leads to an estimated prevalence of infection approaching 100% among 50 year olds in 1949 (see Figure 3.8). It is difficult to determine whether such high estimates for the risk and prevalence of infection are realistic. Tuberculin surveys in South India during the period 1961–1966 [25], for example, found a lower prevalence of tuberculin sensitivity among those aged over 55 years, than for 45–54 year olds. It is unclear whether this reflected genuine differences in the prevalence of infection or a greater tuberculin reversion rate among those aged over 55 years, as compared with that among 45–54 year olds. Surveys carried out among Eskimos during the late 1950s, on the other hand, found that all adults were sensitive to tuberculin [117] (induration of 8mm or more to 1 TU PPD-S), corresponding to an annual risk of infection of about 25%. This suggests that in extreme circumstances, a prevalence of infection approaching 100% is not impossible.

The national tuberculin survey during the period 1971–1973 found that the prevalence of tuberculin sensitivity among 6 and 13 year olds in the "UK national" group was about

0.31% and 1.14% respectively [106] (see section 2.3). These estimates are slightly lower than the expected prevalence of infection used by TBDYN3 for 1972, namely 0.89% and 3.2%. However, the differences between the two are difficult to interpret given the absence of tuberculin sensitivity data among other age groups.

Overall, this suggests that TBDYN3 slightly overestimates the risk of infection among young children, and hence overestimates the proportion of these individuals who are at risk of disease, although the size of this overestimate is probably too small to affect substantially the corresponding disease risk estimates.

TBDYN3 also slightly underestimates the prevalence of infection among young adults during the early 1950s, and probably overestimates that among individuals aged over 40 years. Given that the risks of developing disease are assumed to be identical for all individuals aged over 20 years, it is likely that the net effect of the under- and the overestimates in the risk of infection among young adults and those aged over 40 years respectively on the estimated risks of developing disease is small.

3.3.3.3 Vaccine efficacy

TBDYN3 probably underestimates the proportion of the population at risk of first infection and of subsequent disease by assuming that vaccination confers solid immunity against infection for 77% of individuals. This may be unrealistic, given that the UK MRC BCG trial found that the vaccine efficacy declined over time (see section 2.4). The assumption could have led to an overestimate in the risk of developing the first primary episode, and would have affected the estimated risks of developing endogenous and exogenous disease, given that the parameter values are correlated (e.g. a high risk of developing the first primary episode may require a lower risk of developing endogenous or exogenous disease to attain the required notification rate).

The effect of this assumption on the parameter estimates was explored by repeating the fitting procedure by deliberately overestimating the population at risk of developing the first primary episode by assuming a zero vaccine efficacy. Whilst this is unrealistic, it yields a *lower* bound on the risk of developing the first primary episode, had a waning vaccine efficacy been incorporated into TBDYN3. Table 3.6 summarizes the corresponding parameter values obtained and Figure 3.9 compares the resulting age-specific disease incidence against the corresponding notification data.

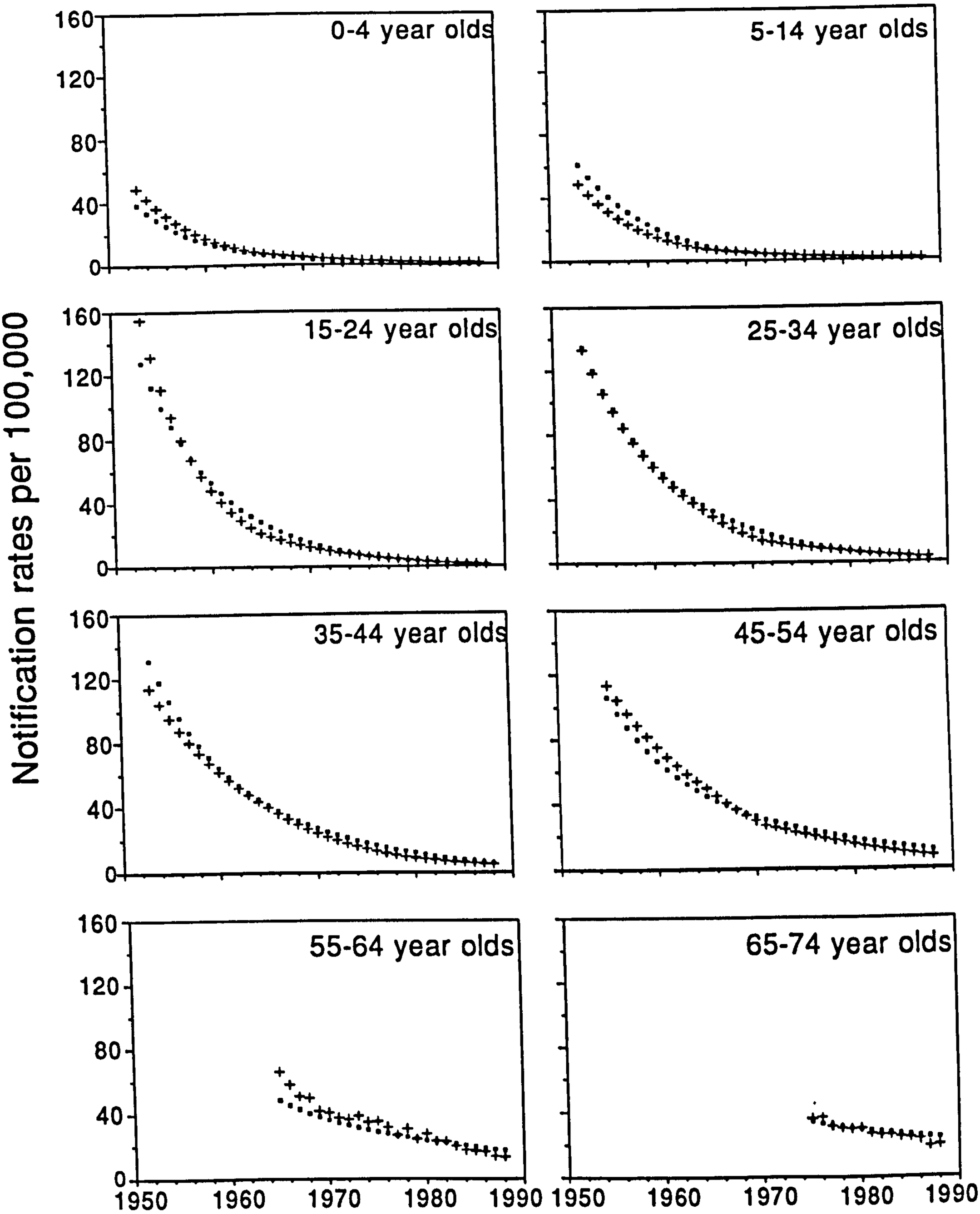


Figure 3.9: Comparison between the incidence of respiratory tuberculosis in 0–74 year old males estimated using TBDYN3 for England and Wales, assuming vaccination conferred no protection against infection, and the notifications used in the fitting process.

Table 3.6: Comparison between the best-fitting values for the risks (to 3sf) of developing the first primary episode and endogenous disease derived assuming vaccination confers no protection against infection, and the values obtained by assuming 77% of individuals are protected from infection for life (original TBDYN3 model). Risks of the first primary episode and exogenous disease refer to those experienced during the first year after infection/reinfection respectively.

Assumption	Age (years)	Risks (% pa) 1st primary	Endogenous	Exogenous	Sum of Squares
0% vaccine efficacy	0–10	2.25	1.81×10^{-9}	9.36	
	15	5.00	0.0146	7.33	
	> 20	7.75	0.0293	5.30	5823
Original TBDYN3	0–10	2.42	1.39×10^{-7}	3.69	
	15	5.59	0.0144	4.38	
	> 20	8.76	0.0288	5.07	5232

As expected, the fit to the notification data is worsened by assuming a zero vaccine efficacy. The assumption leads to only slightly higher estimates for the risks of developing endogenous and exogenous disease among individuals aged over 20 years, and to slightly lower risks of developing the first primary episode for all age groups, as compared with those derived using the original assumption. This suggests that these disease risk estimates are not greatly sensitive to the assumptions regarding vaccine efficacy.

Note also that even though we have assumed that only 13 year olds have been vaccinated since 1954, the assumption of a zero vaccine efficacy leads to different disease risks even for 0–10 year olds. This follows from the facts that

1. we carried out the fitting procedure with this assumption considering all age groups, and
2. the risks of developing disease among 10–20 year olds (and hence among vaccinated individuals) are a function of those for 0–10 year olds and individuals aged over 20 years (see footnote on page 102).

Table 3.6 also shows that the risk of developing exogenous disease for 0–10 year olds estimated by assuming a zero vaccine efficacy exceeds that derived for individuals aged over 20 years. This follows from the fact that this parameter is correspondingly less reliable (see

sections 3.3.1 and 3.3.2).

3.3.3.4 Mixing patterns

TBDYN3 also assumes that both the risks of first infection and reinfection are independent from mixing patterns in the population. High risks of infection or reinfection among individuals in a given age group may in fact be associated with high risks of developing disease.

As mixing patterns changed over time, the relative contribution to the disease incidence from individuals at high infection and disease risk must also have increased, and hence the overall 'observed' risks of developing disease must have changed. The magnitude of this change is unclear. Given that the risks of developing disease obtained lead to a good fit to the notification data throughout the time period considered, this suggests that the assumption of random mixing is not excessively simplistic.

3.3.3.5 Concluding remarks

Overall, we see that the simplifications incorporated into TBDYN3 have not greatly affected the estimates of the risks of developing disease obtained. This suggests that the assumptions incorporated into TBDYN3 are sufficient to provide a realistic representation of the basic tuberculosis situation in England and Wales. Intuitively, identical risks of developing the first primary episode, endogenous and exogenous disease should be obtained if TBDYN3 were to be applied to data from other developed countries. We explore this question in the next section, in which we present results obtained from applying TBDYN3 to data from the Netherlands. This also provides the opportunity to compare the results from the model of Sutherland *et al* [7] with those of TBDYN3.

3.3.4 Disease risk estimates for the Netherlands

Table 3.7 compares the parameter values obtained for the Netherlands using TBDYN3 against those obtained by Sutherland *et al* [7]. Sutherland *et al* assumed that the risks of developing 'primary' and 'exogenous' disease were constant during the first five years after infection. Hence the *cumulative* risks of developing these disease forms within five years of infection and reinfection are best compared against those from TBDYN3. For reference,

Table 3.7: Comparison between the risks (to 3sf) of developing the first primary episode, endogenous and exogenous disease derived with TBDYN3 for the Netherlands, and the corresponding risks derived by Sutherland *et al* [7]. Values for the risks of developing the first primary episode and exogenous disease refer to the cumulative risks experienced during the first five years after infection/reinfection.

Model & data used	Age (years)	Risks (% pa) 'Primary'	Endogenous	Exogenous
TBDYN3 — Netherlands	0–10	12.3	4.17×10^{-7}	19.8
	15	22.0	0.0106	15.2
	20	31.1	0.0213	10.5
Sutherland *et al* [7]	15–69	22.9	0.025	9.19
TBDYN3 — E & W	0–10	3.96	1.39×10^{-7}	6.00
	15	9.01	0.0144	7.10
	20	14.0	0.0288	8.19

Figures A.2 and A.2 summarize compare model predictions of the disease incidence for the Netherlands, against the notifications used.

This shows that the slightly more flexible assumptions regarding the natural history of tuberculosis incorporated into TBDYN3, as compared with those used by Sutherland *et al* led to relatively small differences between the risks of developing endogenous and exogenous disease among adults. The corresponding risk of developing the first primary episode derived using TBDYN3, on the other hand, is almost one and a half times greater than that derived by Sutherland *et al.* The magnitude of this risk is unrealistically high and exceeds even the disease risk estimates found from observational studies during the pre-chemotherapy era (see section 1.2.1).

Table 3.7 shows that the risks of developing endogenous and exogenous disease among Dutch adults derived using TBDYN3 compare well against those for England and Wales. The age-specific risks of developing the first primary episode are over two times greater than those for England and Wales, and the estimated risk of developing exogenous disease *decreases* between the ages 10 and 20 years in the Netherlands. There are no obvious *biological* reasons for these differences between the two countries. Insight into why such differences occurred can be obtained by comparing the annual risk of infection and the notification rates in the Netherlands with those for England and Wales.

As shown in Figure 3.10, the annual risk of infection during the period 1951–1970 was

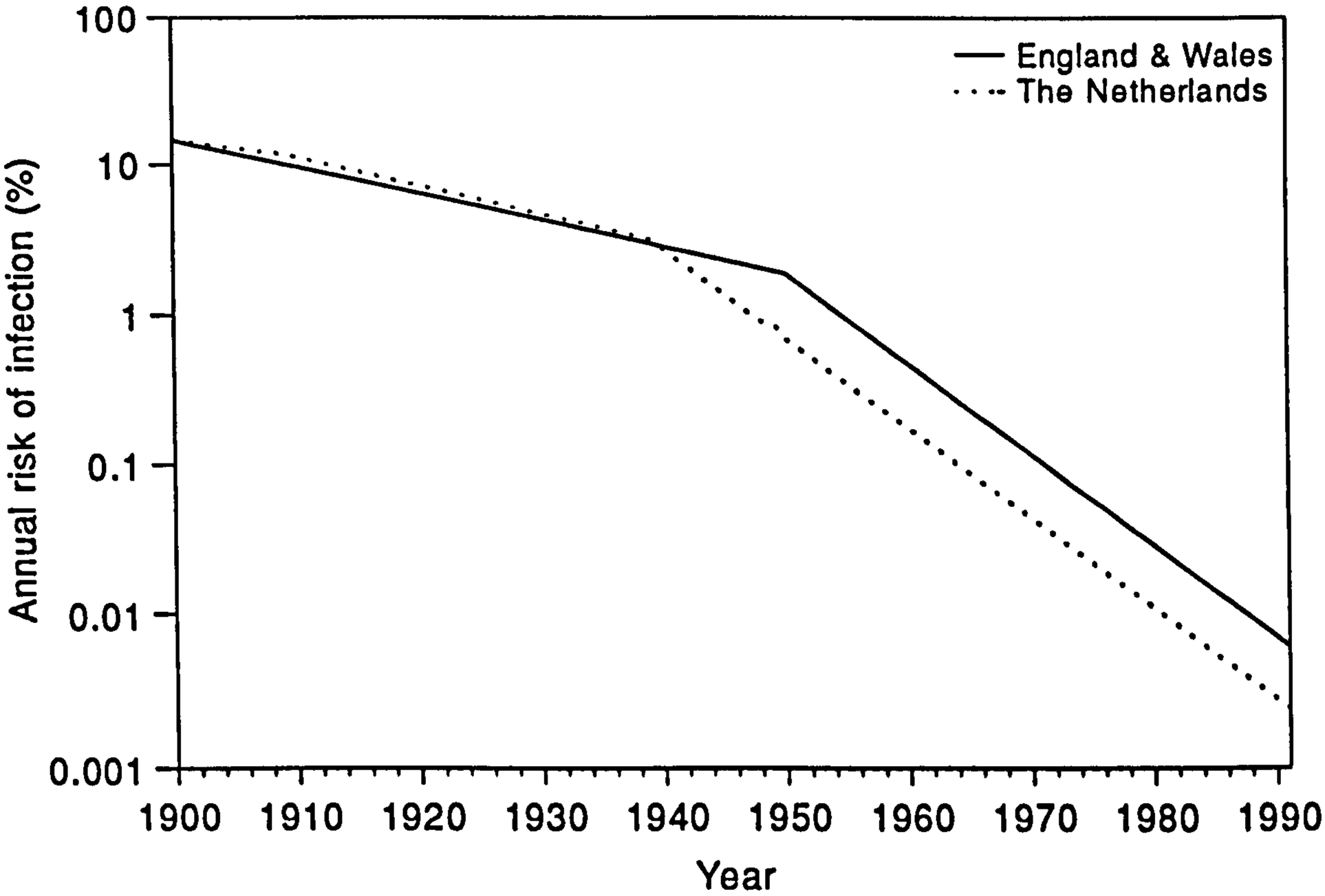


Figure 3.10: Comparison between the estimated annual risk of infection in the Netherlands and that in England and Wales during the period 1901–1990.

lower for the Netherlands than for England and Wales. A smaller proportion of the Dutch population had therefore been newly infected or reinfected and was *eligible* to develop the first primary episode or exogenous disease, as compared with that in England and Wales. This implies that if the risks of developing the first primary episode and exogenous disease were identical for the two countries, the age-specific disease *incidence* in the Netherlands should have been lower than that for England and Wales, after correcting for individuals protected by BCG vaccination in England and Wales.

As shown in Figure 3.11, the notification rates in the Netherlands were lower than those for England and Wales only among individuals aged over 25 years. The notifications among 0–4 and 15–24 year olds were similar in the two countries, whilst those for 5–14 year olds in the Netherlands exceeded those in England and Wales. Given that BCG vaccination was not routinely used in the Netherlands, a greater proportion of 5–14 and 15–24 years olds in England and Wales were protected from developing disease, as compared with those in the Netherlands. This accounts for the 'unexpected' similarity in the notifications in these age groups between the two countries. However, even though BCG vaccination was incorporated into TBDYN3 for England and Wales, the fitting procedure resulted in different estimates

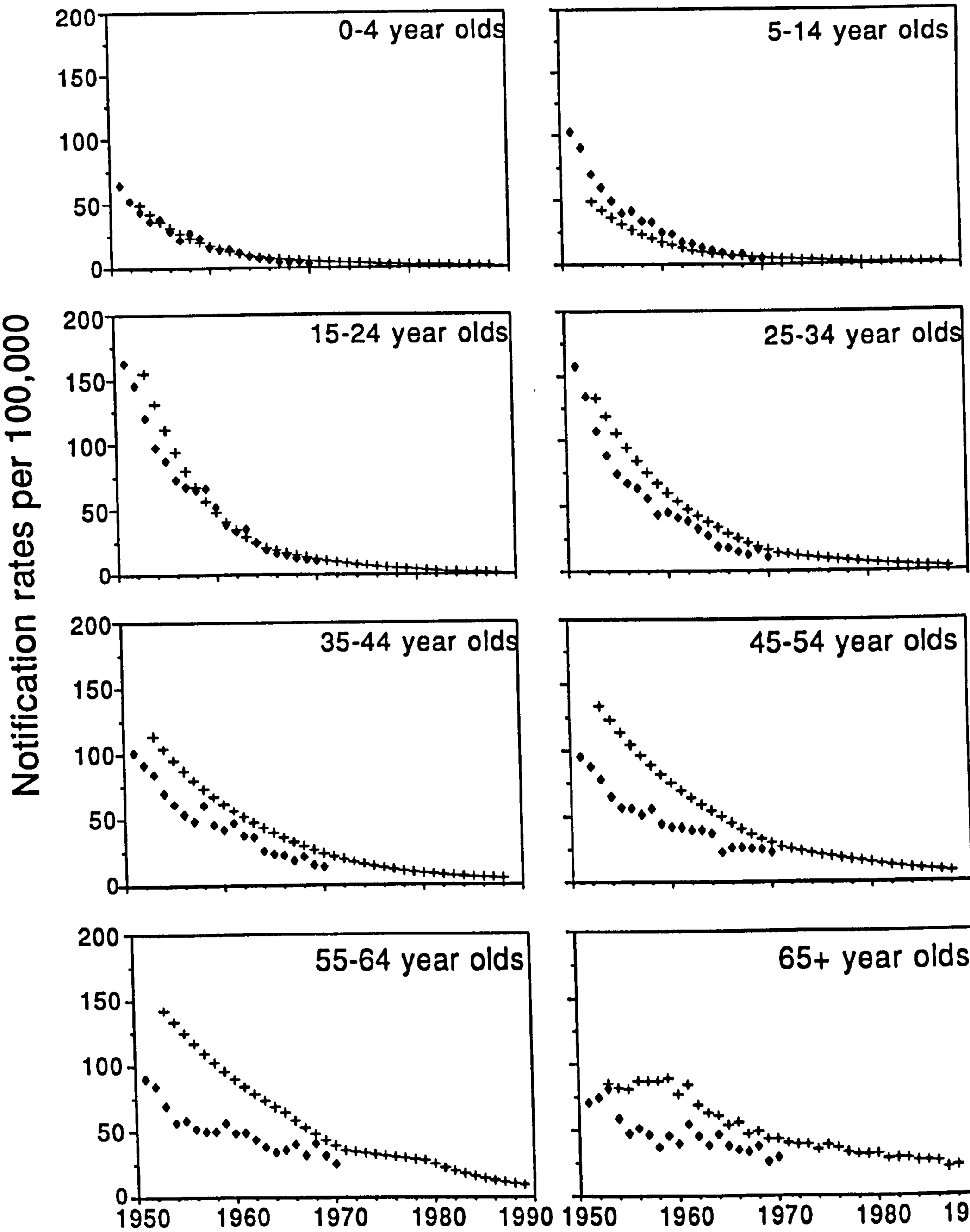


Figure 3.11: Comparison between notification rates of respiratory tuberculosis in males in the Netherlands during the period 1951–1970 and those estimated from survey data (see section 2.2) among white ethnic males in England and Wales.
Notification data supplied by Dr. K. Styblo at the International Union against Tuberculosis
Note: Notifications shown in the 65^{+} year olds category for the Netherlands correspond to those of 65–69 year olds. Notifications for England and Wales in this age groups correspond to those in the general population.

for the risks of developing the first primary episode for the two countries.

Other factors may also account for the unexpected similarity between the notification rates among 0–4 and 15–24 year olds in the Netherlands and England and Wales, and the high rates for 5–14 year olds, such as differences in the notification efficiency, degree of overnotification, or case definitions. These factors would have affected the notification rates in older age groups in a similar way, and this would have resulted in widely different values for the risks of developing endogenous and exogenous disease obtained for individuals aged over 20 years in the Netherlands and England and Wales.

These differences could also be attributable to other biological differences between the two countries, such as the extent of crowding, infection with *M. bovis* or other mycobacteria. Alternatively, this suggests that the annual risk of infection in the Netherlands was actually higher than that estimated, or the increase in the risk of infection with age was greater than that for England and Wales. The most likely explanation is that the parameter values obtained for the Netherlands are based on a smaller data set as compared with those in England and Wales, and as a consequence, are less robust.

3.4 Discussion and implications for past trends in the age-specific incidence of disease

3.4.1 Discussion

In spite of simplifying assumptions incorporated into TBDYN3, the parameter estimates obtained are of a realistic order of magnitude. The risk of developing the first primary episode among adolescents compares well against the risk estimates from the UK MRC BCG trial, and the risk of developing endogenous disease among adults is similar to those derived by Styblo [6] and Horwitz [49] for the Netherlands and Denmark respectively (see section 1.2.1.4).

Both the risks of developing endogenous and exogenous disease exceed the risk of developing the first primary episode for individuals aged over 20 years. This suggests that initial infection may confer some protection against subsequent disease or reinfection. This finding is consistent with results from observational studies in developed countries, which found a lower disease incidence among individuals with weakly tuberculin-positive reactions than among tuberculin-negative individuals (see section 1.2.1).

The risk of developing exogenous disease can be interpreted as a combined risk of reinfection and of developing subsequent disease, since it is assumed that the risks of reinfection and infection are identical. It is likely that the risk of reinfection is in fact lower than that of first infection (see section 3.1.1), which suggests that the risk of developing exogenous disease could be greater than that obtained, and perhaps exceeds that of developing the first primary episode. This is not unrealistic, given that the immune response must already be low, by definition, for reinfection to occur, which could increase the likelihood of subsequent disease.

It is likely that the risks of developing disease also depend on the type of disease considered (e.g. pulmonary, as opposed to extrapulmonary). The relationship between tuberculous meningitis mortality rates and the annual risk of infection during the prechemotherapy era suggests that about 1% of 0–4 year olds develop tuberculous meningitis after infection (see section 2.3). This risk estimate is about 4 times smaller than that estimated for respiratory tuberculosis for 0–4 year olds (see table 3.3), even though the corresponding mortality rates from tuberculous meningitis exceeded those of respiratory tuberculosis until 1957 (see Figure 3.12). A similar comparison between the notification rates after 1953 of the two

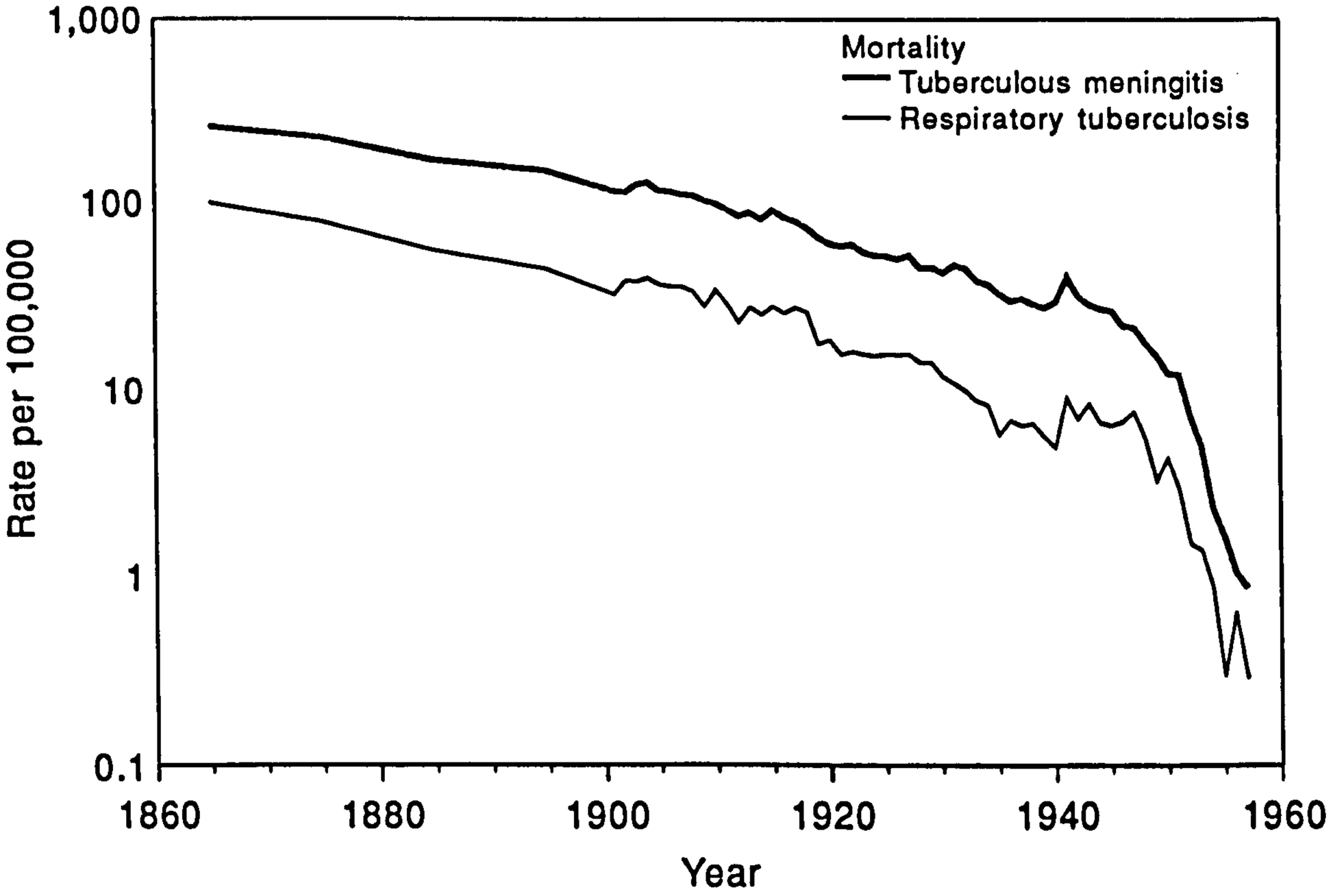


Figure 3.12: Comparison between tuberculous meningitis and respiratory tuberculosis mortality rates among 0–4 year old males in England and Wales during the period 1860–1957.

Note that the tuberculous meningitis and respiratory tuberculosis mortality rates declined at the same rate. This results from the fact that disease among 0–4 year olds is attributable to recent infection, and hence the mortality rates reflect the trend in the annual risk of infection.

forms of disease suggests that the morbidity from tuberculous meningitis was lower than that for respiratory tuberculosis for 0–4 year olds (see Figure 3.13). This shows that the case-fatality rates must have been considerably lower for respiratory tuberculosis than for tuberculous meningitis.

It is unclear whether similar disease risks estimates would be obtained if TBDYN3 were to be applied to data from developing countries. During the BCG trial in South India during the 1970s [26], for example, where the risk of infection was in the range 2–6%, the average annual risk of disease was about 21, 55 and 192 per 100,000 during follow-up for individuals with tuberculin reactions measuring 0–7mm, 8–11mm and 12–15mm respectively. This suggests that

1. if it is *assumed* that individuals in South India face *identical risks of infection as of re-infection*, then the risk of developing endogenous and exogenous disease in South India must have been greater than that estimated for England and Wales, or alternatively

2. if it is *assumed* that the risks of *developing endogenous and exogenous disease* are

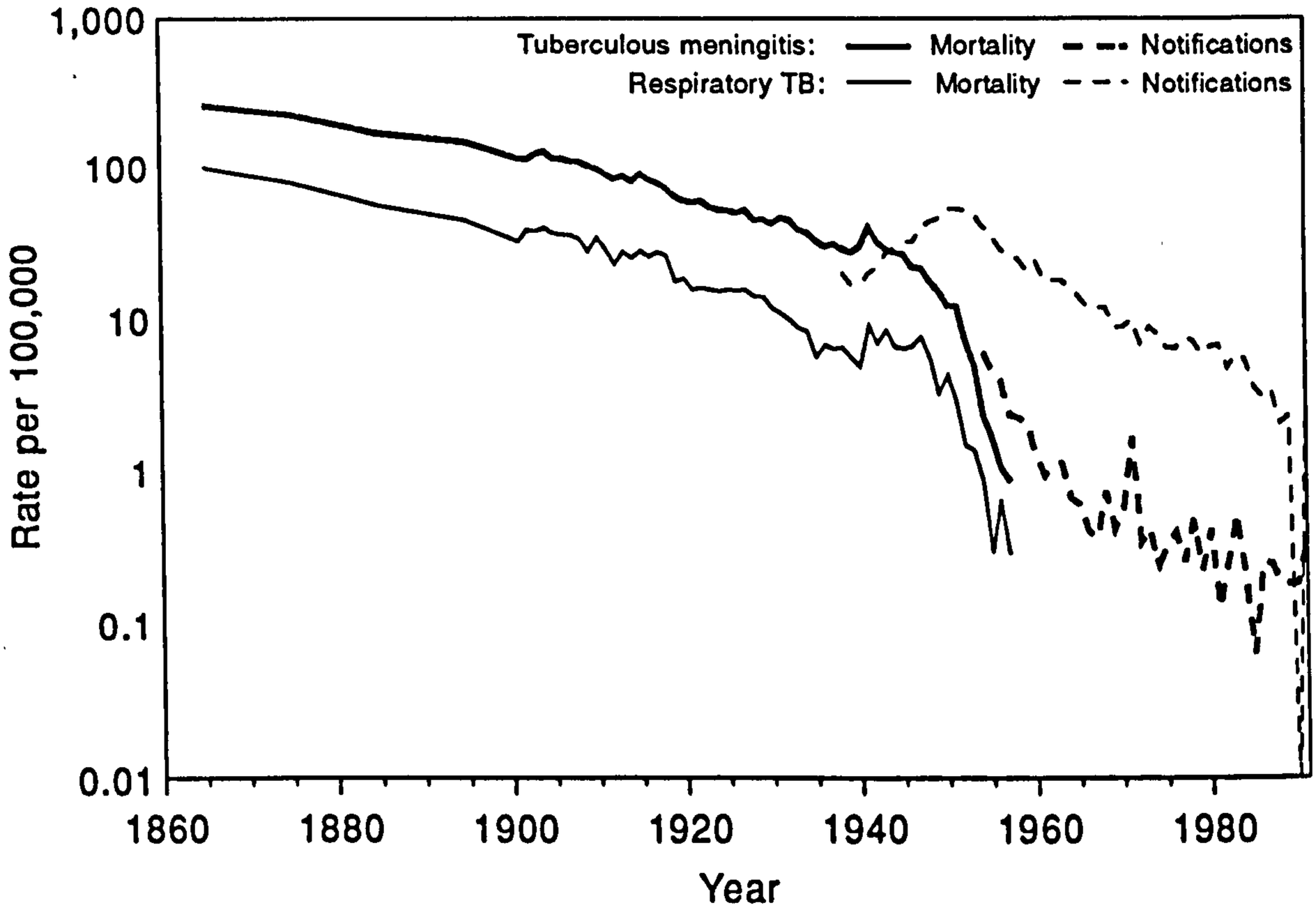


Figure 3.13: Comparison between tuberculous meningitis and respiratory tuberculosis mortality rates among 0–4 year old males in England and Wales until 1957, and the corresponding notification rates in the general population from 1953.

identical to those in England and Wales, then the risk of reinfection in South India must exceed that of first infection,

The situation is further complicated in South India, given that there is high prevalence of infection with atypical mycobacteria, as compared with that in England and Wales, which may impart some protection against disease [26]. Overall, this suggests that the results obtained for developed countries cannot be easily extrapolated to developing countries, and that the transmission dynamics in the two should be analyzed separately.

As it stands, the model formulated in this chapter provides a sound framework with which to explore the transmission dynamics of *M. tuberculosis*. We explore the implications for past trends in the age and time-dependent disease incidence in the next section.

3.4.2 Implications for age-specific patterns in the disease incidence predicted by TBDYN3

Figure 3.14 summarizes the age-specific incidence of disease among 2–70 year olds since 1900 predicted with TBDYN3 using the best-fitting parameter values. This also shows the relative contribution of disease attributable to first primary episodes, exogenous and endogenous disease. The corresponding incidence of disease after 1950 is shown in greater detail in Figure 3.15.

These indicate that much of the disease experienced by individuals aged over 10 years at least until 1940 was attributable to exogenous disease. This is intuitively reasonable, given the high annual risk of infection, and hence the high prevalence of infection predicted for these years.

Figure 3.14 shows that the predicted incidence of first primary episodes particularly for 20 and 30 year olds *increased* between 1900 and 1950. This is attributable to an increase in the incidence of first infections occurring during this time in these age groups, which, paradoxically, was a consequence of the decline in the annual risk of infection. During the early 1900s, for example, only a small proportion of individuals aged over 20 years experienced initial infection, since most had been infected early in life. By the 1920s, however, a greater proportion of 20 years olds had escaped infection, as compared with their counterparts during the early 1900s, since the annual risk of infection during their childhood was considerably lower. Styblo predicted similar increases in the incidence of infection among adults for the Netherlands after 1911 [6].

As a result of this increase in the incidence of first primary episodes, the rate of decline in the *overall* disease incidence among adolescents and young adults was slower than that for younger individuals. This also resulted in a slower decline in the corresponding observed mortality rates for adolescents, as compared with those among infants and young children (see section 2.1).

Figure 3.14 shows that the decline in the predicted incidence of disease in all age groups accelerated in 1950. This is attributable to the rapid decline in the annual risk of infection after 1950 (associated with improvements in case-finding and treatment — see section 2.3), which led to a dramatic reduction in the incidence of first primary episodes and exogenous disease. This translated into an equally dramatic decline in the overall notification rates of

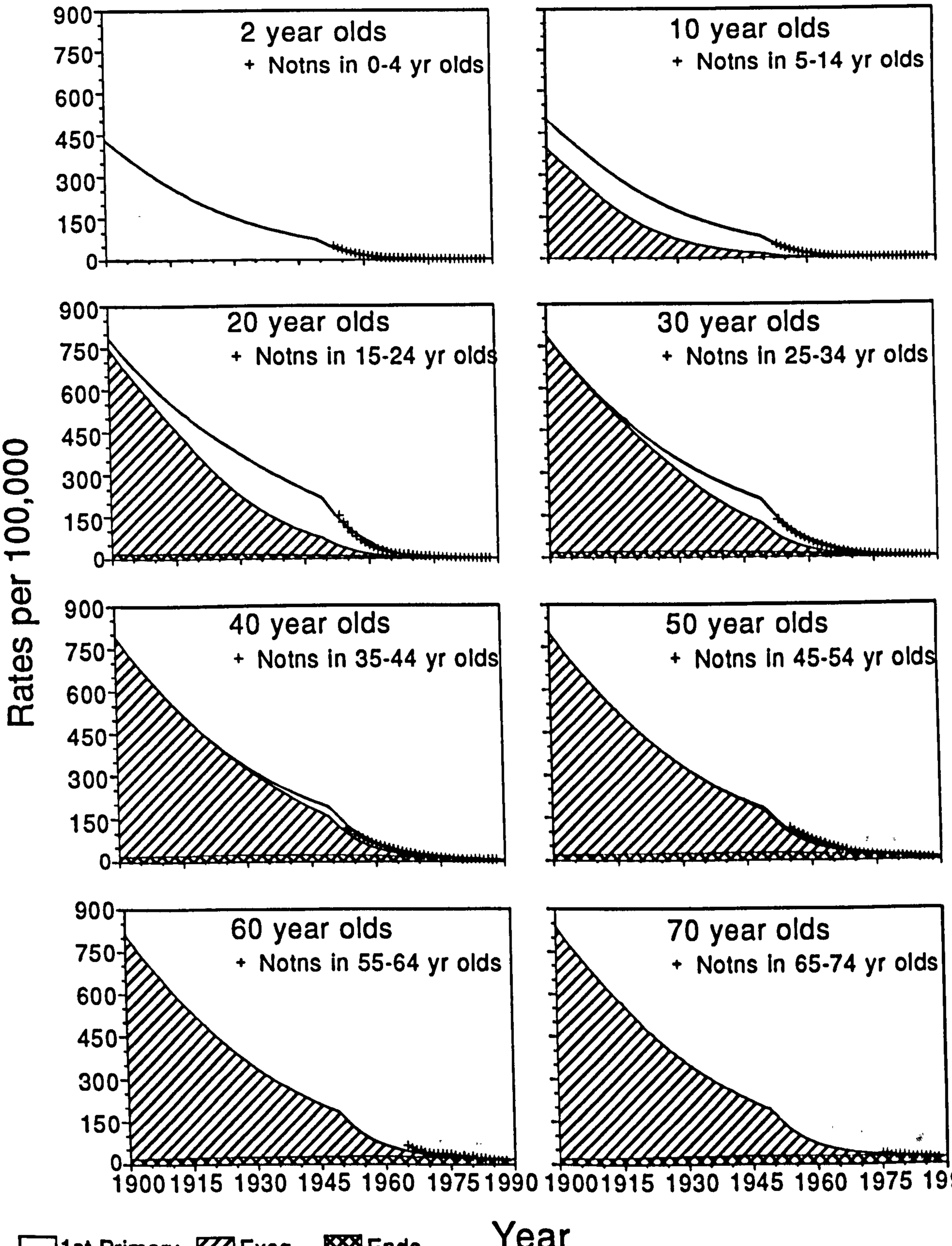


Figure 3.14: Estimated age-specific incidence of respiratory tuberculosis since 1900 in England and Wales attributable to first primary episodes, exogenous and endogenous disease, as predicted using TBDYN3 using the best-fitting parameter values.

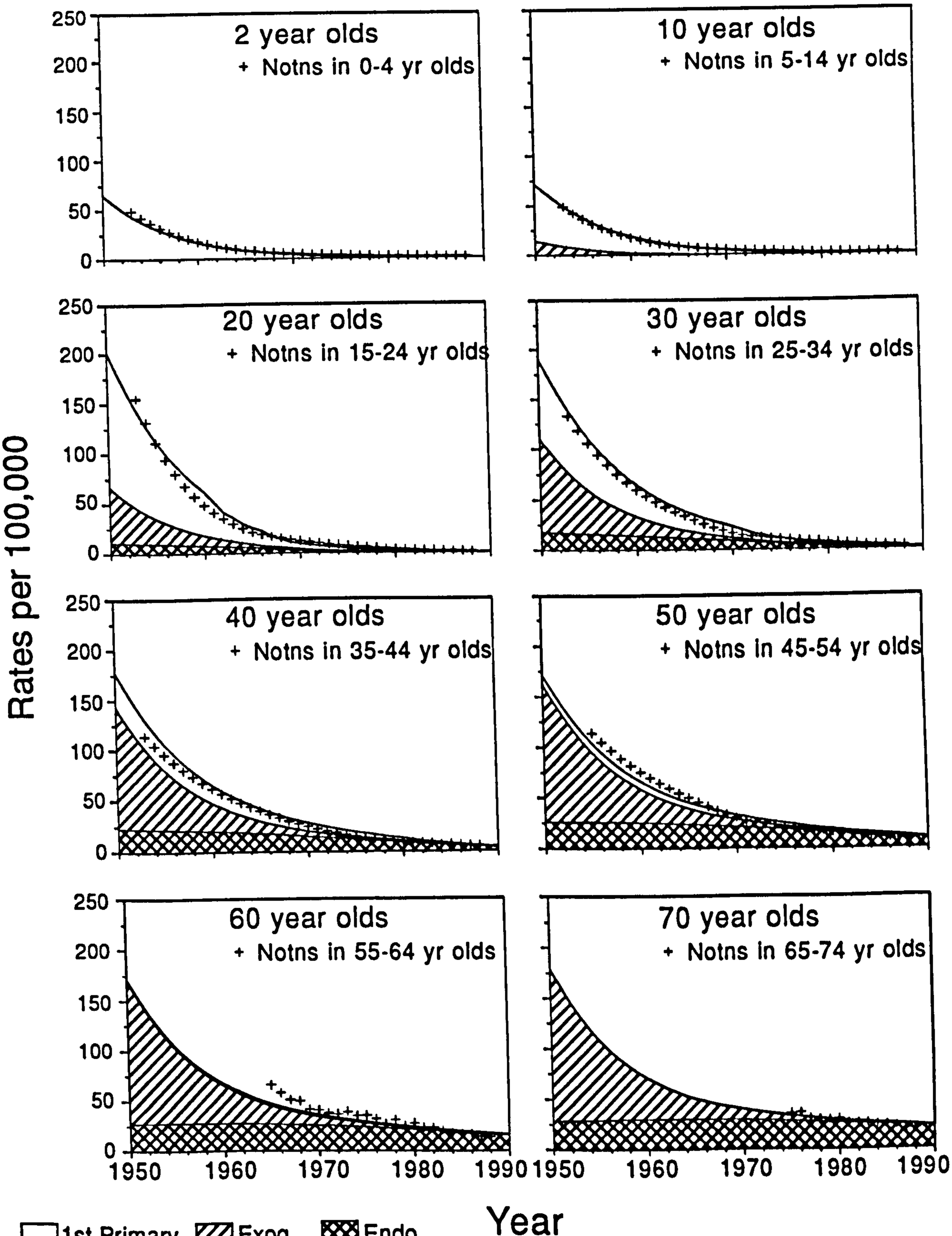


Figure 3.15: Estimated age-specific incidence of respiratory tuberculosis since 1950 in England and Wales attributable to first primary episodes, exogenous and endogenous disease, as predicted using TBDYN3 using the best-fitting parameter values.

respiratory tuberculosis for all age groups (see Figure 3.15).

Figure 3.16 compares the disease incidence predicted using TBDYN3 for given years during the period 1901–1950, against the corresponding mortality rates. This shows that the overall pattern in the age-specific disease incidence predicted using TBDYN3 is consistent with that seen in the mortality rates (i.e. low for 0–10 year olds, increasing during adolescence and remaining consistently high thereafter). Figure 3.17 compares the overall disease incidence rates of respiratory tuberculosis predicted by TBDYN3 against the corresponding published and predicted mortality rates. This shows that the predicted and the observed mortality rates compare well for all years, and for most age groups, except the elderly. For these individuals, model predictions of the mortality rates exceeded those published for most years. This is attributable to at least two factors:

1. problems with case-recognition among elderly individuals, which meant that many deaths attributable to tuberculosis were not recognized as such, and

2. a large proportion of individuals with tuberculosis in older age groups could have also died from causes other than tuberculosis (i.e. the issue of 'competing mortality'), and hence would not have contributed to the overall tuberculosis mortality rates.

The relative contribution of endogenous, exogenous and first primary episodes to the age-specific incidence of disease *after 1950* predicted using TBDYN3 is shown in Figure 3.18. The corresponding disease incidence assuming no individuals had been vaccinated is given by the dotted line. This shows that overall, the pattern in the age specific incidence predicted using TBDYN3 agrees well with that in the notification rates. The peak in the disease incidence for young adults in 1955, for example, is consistent with a similar peak in the notification rates. Similarly, the absence of a peak in young adult life in later years predicted using TBDYN3 is also confirmed by the notification data.

Another perspective on the validity of the results from TBDYN3 may be obtained from Figure 3.19, which compares the incidence and mortality rates of disease among cohorts born during the period 1901–1935 predicted using TBDYN3 against the corresponding mortality rates from respiratory tuberculosis in England and Wales. The similarity between the published and the predicted mortality rates for young adults is particularly striking, especially for cohorts born during the periods 1901–1910 and 1911–1915.

Figure 3.19 shows that the peak in mortality rate shifted from 30–35 year olds for

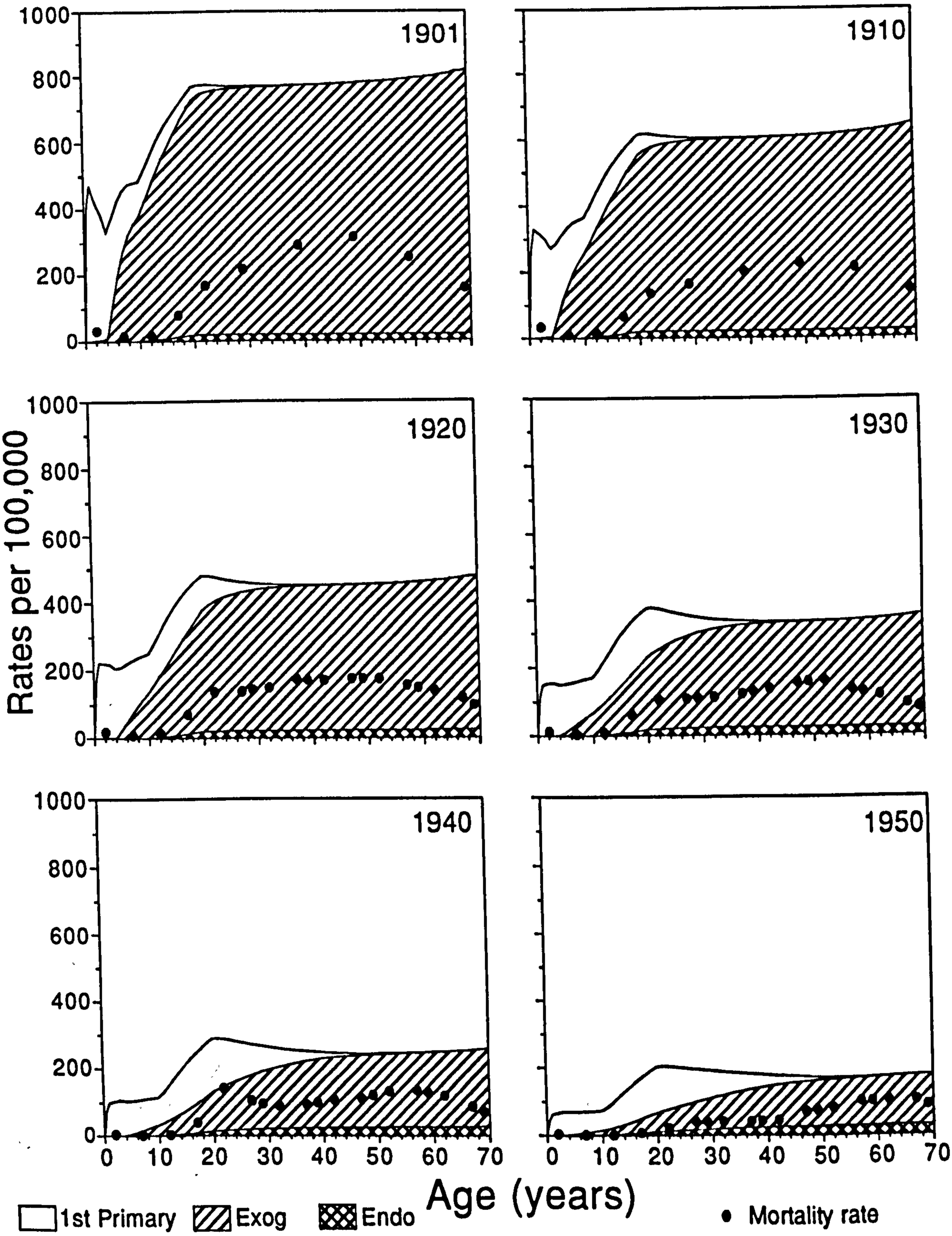


Figure 3.16: Comparison between the incidence of respiratory tuberculosis attributable to first primary episodes, exogenous and endogenous disease for males during the period 1901–1950 in England and Wales, as predicted using TBDYN3 and the corresponding male mortality rates.

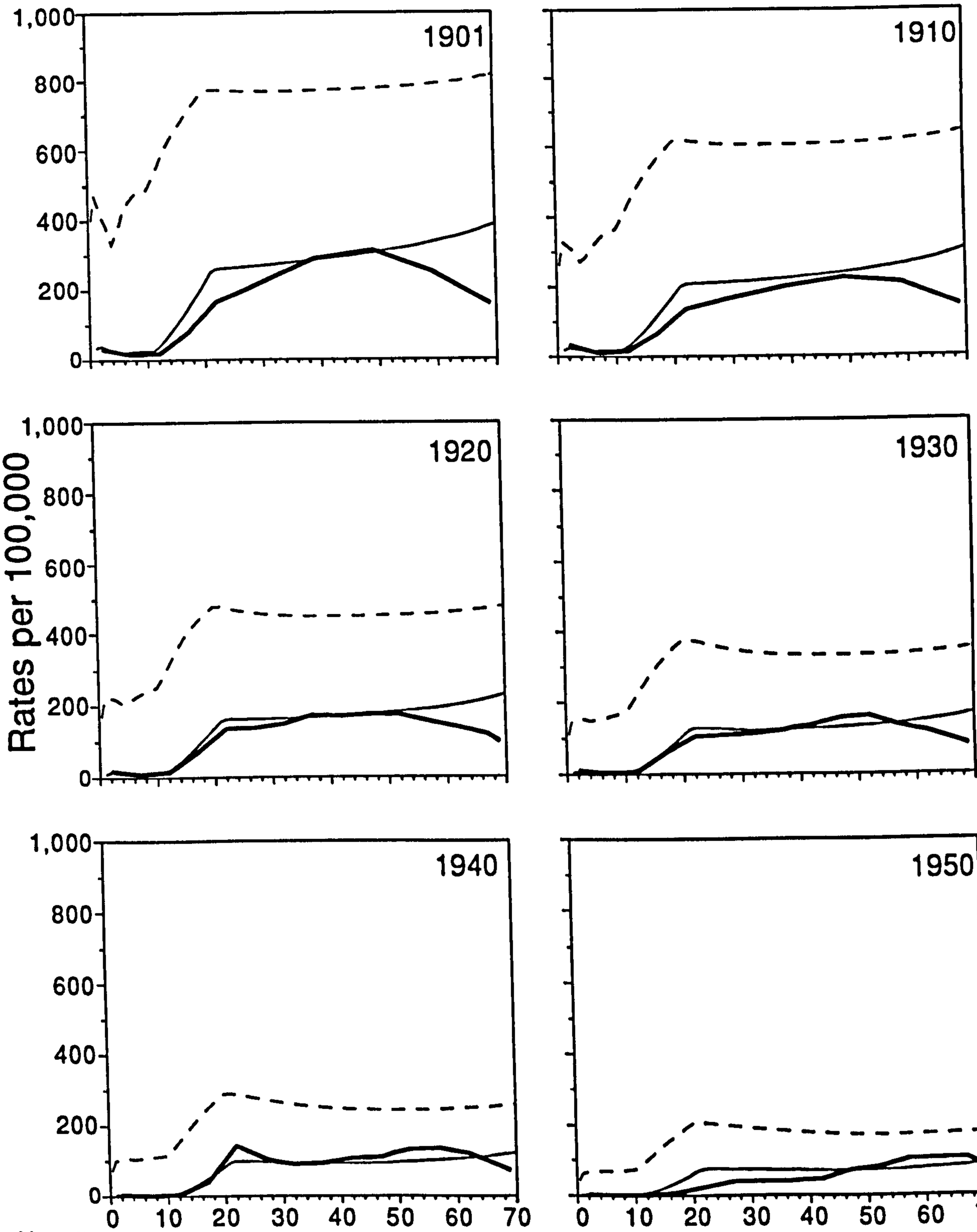


Figure 3.17: Comparison between the mortality rates and incidence rates of respiratory and sputum-positive tuberculosis for given years during the period 1901–1950, as predicted using TBDYN3, and the actual mortality rates from respiratory tuberculosis in the general population.
This assumes a case-fatality of 50% for *sputum-positive* cases and a mortality rate identical to that of the general population for *sputum-negative* cases (see section 3.1.3).

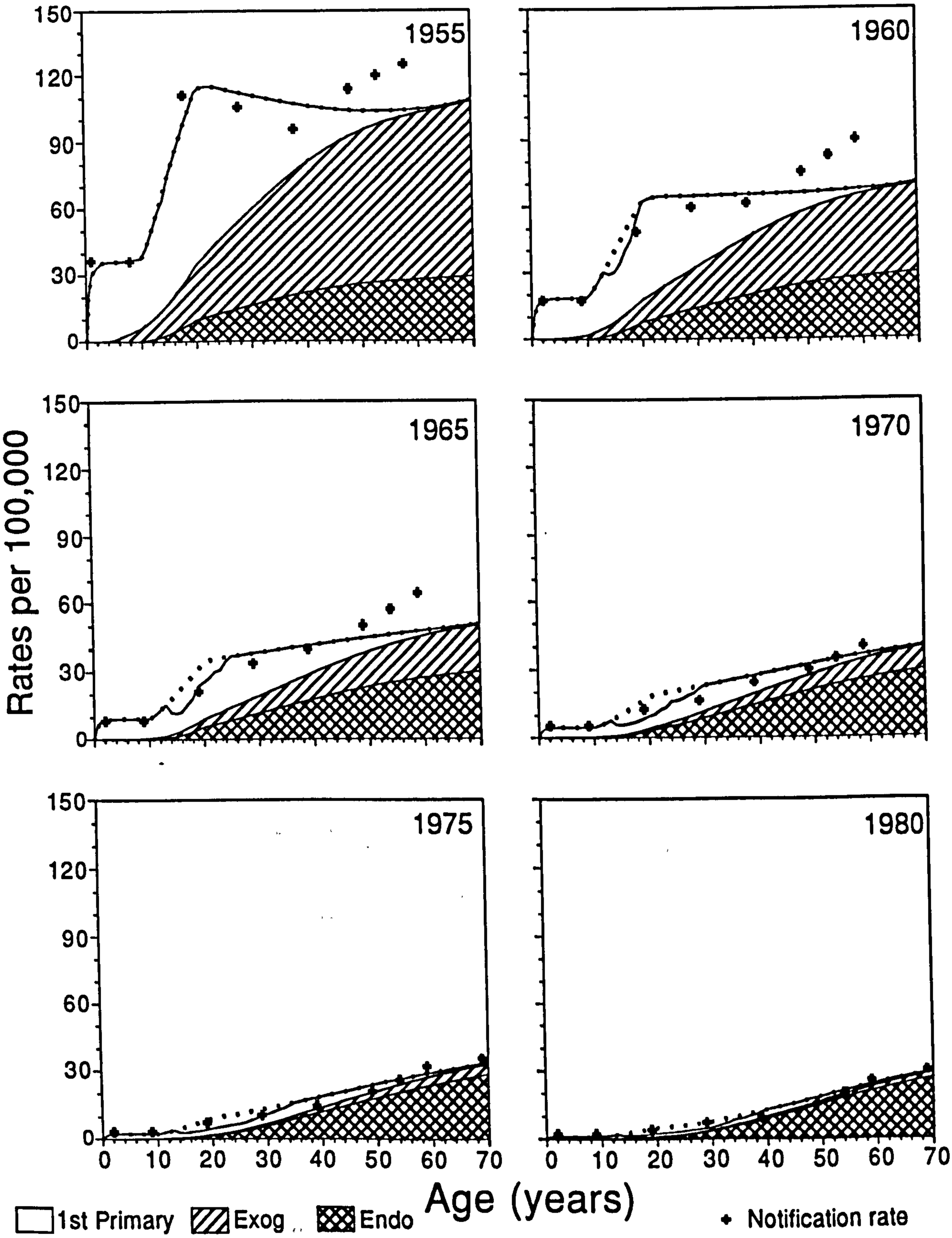


Figure 3.18: Comparison between the incidence of respiratory tuberculosis attributable to first primary episodes, exogenous and endogenous disease for males during the period 1955–1980 in England and Wales, as predicted using TBDYN3, and the corresponding notification rates for white ethnic males.

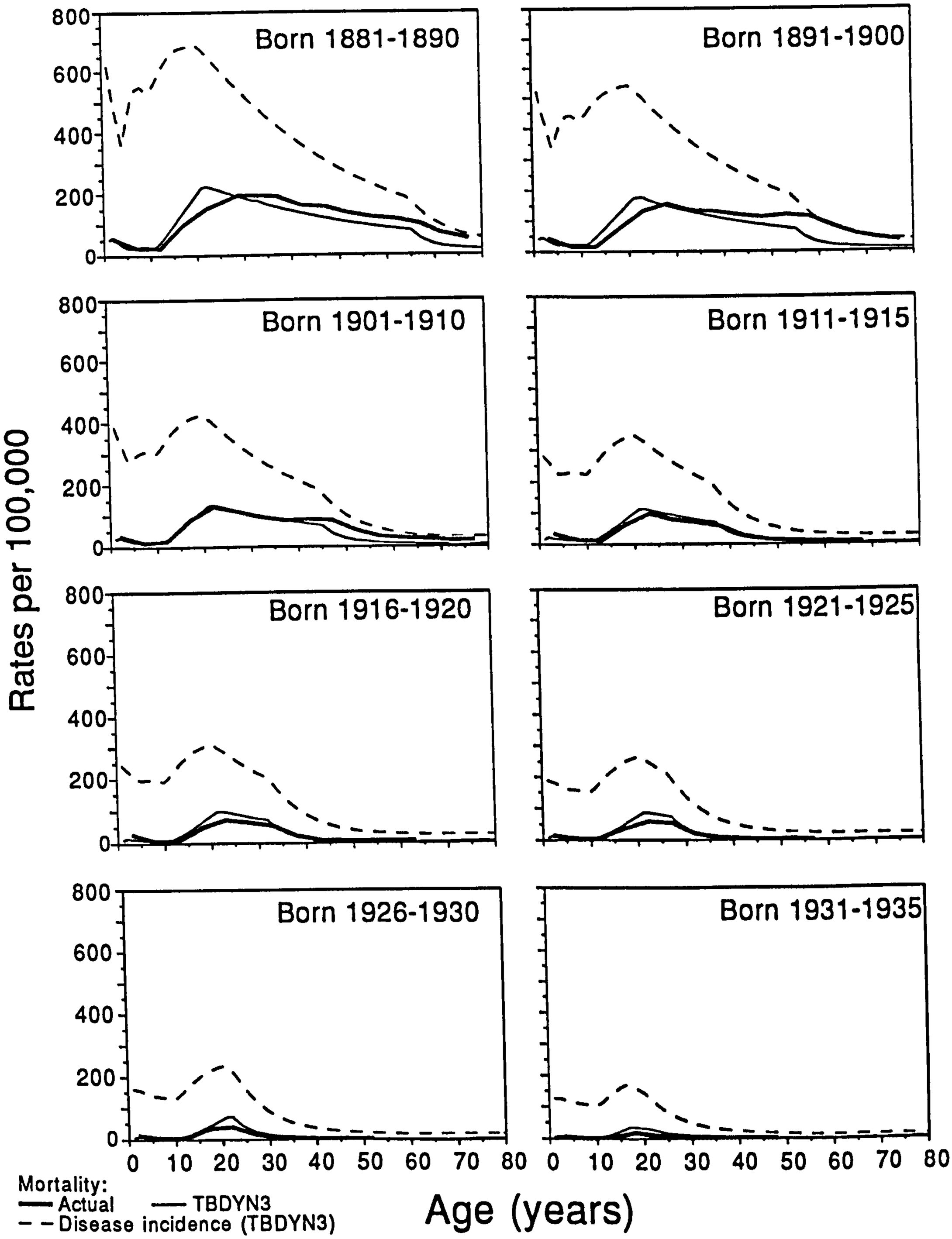


Figure 3.19: Comparison between the mortality rates among male cohorts born since 1880 in England and Wales and the corresponding incidence and mortality rates from respiratory tuberculosis among cohorts, as predicted using TBDYN3 using the best-fitting parameter values.

the 1881–1890 birth cohort to young adult life for individuals born after 1901, whereas TBDYN3 predicts that the peak occurred at age 20 years for each birth cohort. This is largely attributable to the fact that individuals born during the period 1881–1890 were aged 30–40 years during the First World War, when they faced high case-fatality rates: TBDYN3 assumes that, until 1950, the case-fatality rate is independent from the calendar year. A similar argument can be applied to explain analogous differences between the published and predicted mortality rates for the cohort born during the period 1891–1900.

Overall, we see that the assumptions incorporated into TBDYN3 lead to realistic patterns in the age and time-specific disease incidence and mortality rates for both the pre-chemotherapy and the chemotherapy eras. We explore the implications of these assumptions for the definitions and interpretation of the incubation period, serial interval and basic and net reproduction numbers for tuberculosis in the following chapters.

Chapter 4

The incubation period, serial interval and lifetime risks of developing tuberculosis

4.1 Definitions and relevance

The incubation period of an infectious disease is defined as the time interval between infection and disease onset. For many common infectious diseases, such as measles, mumps and rubella, most infected individuals subsequently develop disease and the incubation period is short (measured in days). Moreover, such infections induce a permanent solid immunity and hence the single disease episode experienced is attributable the initial infection event.

The definition of an incubation period for tuberculosis is complicated, given that individuals can be reinfected and may then experience exogenous disease within a few years of reinfection, or may reactivate many years thereafter as a consequence of the reinfection event. Because of this, the time interval since *initial* infection might be used as the basis to a definition of the incubation period for tuberculosis, but would not be ideal, as the beginning (initial infection) and terminal (disease) events might not even be causally linked.

For tuberculosis, and for other infections in which individuals can develop disease following reinfection, such as herpes varicella-zoster [118], the **incubation period** may better be defined as:

The time interval between infection or reinfection and the disease

episode which occurs *as a result* of that infection/reinfection event.

The *serial interval*, also known as the "generation time" [119], was first defined by Hope-Simpson in 1948 [120] as the time interval between successive cases in a chain of transmission. It is equivalent to the incubation period if the onset of infectiousness and disease coincide and if the duration of infectiousness is relatively short.

For tuberculosis, since not all disease is infectious, the **serial interval** may best be defined as:

> **The time interval between the clinical onsets of infectious pulmonary disease in successive cases in the chain of transmission.**

By the above logic, it is equivalent to the incubation period of *infectious pulmonary disease* if the duration of infectiousness is short. This is complicated by the fact that disease may arise in different ways, as either the first primary episode, reactivation or reinfection disease. The mechanisms underlying all these processes must be considered in evaluating both the serial interval, and incubation period for *M. tuberculosis*.

The incubation period of tuberculosis can be derived either *prospectively* as:

> the distribution of the time interval *until* the first disease episode as a result of an infection/reinfection event among individuals of a given age in a given year,

or *retrospectively* as:

> the distribution of the time interval *since* the infection or reinfection event causing disease among individuals with disease onset at a given age in a given year.

Serial intervals can be derived analogously.

For many communicable diseases, such prospective and retrospective analyses would yield identical distributions, irrespective of the age groups or the time period considered. This is not the case for tuberculosis, because of the three different mechanisms underlying disease, and their dependence on the age at infection and on the annual risk of infection. A few general comments will serve to point out the complexity of the incubation period and serial interval concepts as they apply to tuberculosis.

First, the incubation period (and serial interval) of tuberculosis may be very long, as long as a lifetime. Because of this, its frequency distribution will be a function not only

of the age at which infection/reinfection occurs, but also the time period, both of which determine the subsequent mortality experience of infected individuals.

Second, the distributions of both the incubation period and serial interval of tuberculosis will also depend on the magnitude and trend in the annual risk of infection, irrespective of the age group considered. Some individuals who succumb to exogenous disease would not have developed disease at all if they had escaped reinfection, and if they had been at risk only of endogenous reactivation. A higher annual risk of infection therefore implies a shorter time from infection to disease than in conditions characterized by a low risk of infection. This implies an inverse relationship between the annual risk of infection and the average duration of both the incubation period and serial interval.

Third, another complication arises because of the age-dependence of tuberculosis pathogenesis. As indicated in section 1.1, infectious pulmonary disease is rare, and non-infectious forms are comparatively common, in children under the age of 15 years. Many years may elapse before an individual infected in childhood develops infectious pulmonary disease through either exogenous reinfection or endogenous reactivation. For this reason, a high annual risk of infection may be associated with much infection in early childhood but with long average serial intervals.

A different situation arises if the annual infection risks are low, and most individuals escape infection until after childhood. Adolescents and young adults experience high risks of developing disease, given infection, and much of their disease is sputum-positive. Therefore, if an individual becomes infected after childhood, and infects another adolescent or young adult, the serial interval can be relatively short, measured in months or a few years. In this instance, the serial interval will approximate the incubation period. This last point raises an even more difficult issue, that of the pattern of infection transmission within and between different age groups.

We thus see that the definitions of the incubation period and serial interval are complicated for tuberculosis. A greater understanding of the relationship between the incubation period and the serial interval should help to clarify age-specific patterns and trends in tuberculosis in different populations.

4.2 Methods

4.2.1 Definition of a 'causal infection/reinfection' event

TBDYN3 has been adapted to analyze the distributions of the incubation period both 'retrospectively' and 'prospectively', as defined below. *Retrospective* analyses considered the frequency distributions of time intervals from the

1. **initial infection event to onset** of the current disease episode for (diseased) individuals of a given age in a given year, and

2. **"causal" infection or reinfection event to onset** of the current disease episode for (diseased) individuals of a given age in a given year,

Prospective analyses of the incubation period considered the distribution of the time interval between first infection at a given age a and time t and

1. The first episode of respiratory disease, and

2. The first episode of sputum-positive disease.

These two distributions were broken down according to whether the first disease episode was attributable to the initial infection event, or to subsequent reinfection.

Before describing the methods used to derive these distributions, we must first clarify how TBDYN3 defines a causal infection or reinfection event for individuals experiencing a first primary episode, endogenous or exogenous disease.

TBDYN3 assumes that individuals cannot experience reinfection if they are currently experiencing disease, or if they are already at (high) risk of developing their first primary episode or exogenous disease (see section 3.1.2). For simplicity, it is assumed that a given disease episode is attributable to the most recent infection or reinfection event. It is recognized that in reality, disease among reinfected individuals could be attributable to bacilli from

1. both the infection and reinfection events together,

2. only from the reinfection event, or

3. only from the initial infection or a more distant reinfection event

(see section 3.1.1).

The situation is less complicated for individuals experiencing endogenous disease — TBDYN3 attributes such disease to the most recent infection or reinfection event.

4.2.2 Retrospective analyses of the incubation period/serial interval

4.2.2.1 Distribution of time intervals since initial infection

TBDYN3 assumes that a proportion $d_+(a)$ of the disease incidence among individuals of age a is sputum-positive irrespective of the time t, irrespective of whether they are experiencing their first primary, endogenous or exogenous disease, or whether it constitutes their first or a subsequent episode (see section 3.1.3). As a result of this assumption, the distributions of the time interval since initial infection for individuals diseased at a given age at a given time do not depend on whether we stratify the individuals according to their sputum status. This follows from the following argument:

The distribution of the time interval since first infection of individuals experiencing onset of respiratory tuberculosis at age a at time t, $\mathbf{D_i}(a,t,s)$, is given by

$$\mathbf{D_i}(a,t,s) = \frac{D_i(a,t,s)}{\sum_{s=0}^{a} D_i(a,t,s)} \qquad \text{for } 0 \leq s \leq a \qquad (4.1)$$

where $D_i(a,t,s)$ denotes the number of individuals with *all forms of respiratory* disease of age a at time t who had been infected s years previously (i.e. when aged $a-s$ years at time $t-s$). The denominator in this expression is the total number of individuals of age a with respiratory disease at time t.

The number of individuals experiencing *sputum-positive* disease of age a at time t, who had been infected s years previously is given by $d_+(a)D_i(a,t,s)$. Similarly, the *total* number of individuals experiencing *sputum-positive* disease of age a at time t is given by $\sum_{s=0}^{a} d_+(a)D_i(a,t,s)$. Hence the distribution of the time interval since initial infection for individuals experiencing sputum-positive disease at age a at time t is given by:

$$\mathbf{D_i^+}(a,t,s) = \frac{d_+(a)D_i(a,t,s)}{\sum_{s=0}^{a} d_+(a)D_i(a,t,s)} \qquad (4.2)$$

$$= \frac{d_+(a)D_i(a,t,s)}{d_+(a)\sum_{s=0}^{a} D_i(a,t,s)} \qquad \text{for } 0 \leq s \leq a \qquad (4.3)$$

Cancelling out $d_+(a)$ from both the numerator and denominator, we see that this expression is identical to that for the distribution of the time interval since initial infection

for individuals experiencing all forms of respiratory tuberculosis at age a at time t (see equation 4.1), as follows:

$$\mathbf{D_i^+}(a,t,s) = \frac{D_i(a,t,s)}{\sum_{s=0}^{a} D_i(a,t,s)} \tag{4.4}$$

$$= \mathbf{D_i}(a,t,s) \qquad \text{for } 0 \leq s \leq a \tag{4.5}$$

The distribution $\mathbf{D_i}(a,t,s)$ is therefore interpretable as both an 'observed' incubation period and serial interval for individuals diseased at a given age and in a given year. It does not correspond to a 'true' incubation period or serial interval experienced by such individuals (see definition in section 4.1), given that a current disease episode is not always attributable to the initial infection event.

We estimated the distributions $\mathbf{D_i}(a,t,s)$ for individuals diseased at a given age a between 1900 and 1990 by first deriving the corresponding $D_i(a,t,s)$ for all possible values of s, namely between 0 and a years. Hitherto, the disease incidence has not been broken down according to the year in which the initial infection event, or the infection/reinfection event causing the disease episode occurred, and so TBDYN3 had to be modified slightly to derive these estimates.

TBDYN3 derives the disease incidence in a given age group and year by following cohorts since birth and applying the corresponding infection, reinfection and disease risks and mortality rates. At any time, individuals in a cohort belong to either the uninfected, 'immune', infected, 'latent' or reinfected classes, or are diseased with either their first primary episode, endogenous or exogenous disease (see section 3.1.2).

The approach used to derive $D_i(a,t,s)$ for all values of s between 0 and a years was identical, except that TBDYN3 was restricted to track the disease dynamics *only in the group of individuals infected at age $a-s$ at time $t-s$ for duration s.* As above, individuals in this group belong to either the infected, 'latent' or reinfected classes, or are experiencing either their first primary episode, endogenous or exogenous disease. $D_i(a,t,s)$ is thus given by the total number of individuals in the first primary, endogenous and exogenous disease classes in the group after time s. The system of differential equations representing the disease dynamics in these individuals is analogous to that used to derive the incidence of disease in given cohorts of individuals tracked from birth (see section 3.1.2 and Appendix A.1).

4.2.2.2 Distributions of time intervals since the infection/reinfection event causing a disease episode

The distribution of the time interval since the infection/reinfection event causing disease among individuals of a given age a at time t, $\mathbf{D_c}(a,t,s)$ is given by:

$$\mathbf{D_c}(a,t,s) = \frac{D_{i_c}(a,t,s) + D_{r_c}(a,t,s)}{\sum_{s=0}^{a}\{D_{i_c}(a,t,s) + D_{r_c}(a,t,s)\}} \qquad \text{for } 0 \leq s \leq a \qquad (4.6)$$

where $D_{i_c}(a,t,s)$ denotes the number of individuals of age a at time t experiencing disease as a consequence of an infection event s years previously, and $D_{r_c}(a,t,s)$ denotes the number of individuals of age a at time t experiencing disease as a consequence of a reinfection event s years previously. The denominator in this expression is the total number of diseased individuals of age a at time t, and is therefore identical to that in expression 4.1.

By an analogous argument to that used in section 4.2.2.1, this distribution does not depend on whether we stratify diseased individuals according to their sputum status. The distribution $\mathbf{D_c}(a,t,s)$ is therefore interpretable as both an 'actual' incubation period and serial interval for individuals diseased at a given age and in a given year (see definition in section 4.1).

TBDYN3 assumes that the first primary episode and exogenous disease are attributable to tubercle bacilli from the initial infection and most recent reinfection events respectively (see section 4.2.1). Endogenous disease is attributed to the most recent infection or reinfection event experienced by the individual.

$D_{i_c}(a,t,s)$ therefore corresponds to the number of individuals of age a at time t, who were infected s years previously, experiencing either a first primary episode or endogenous reactivation of the initial infection, and who therefore have never been reinfected. Similarly, $D_{r_c}(a,t,s)$ corresponds to the number of diseased individuals of age a at time t, who were *reinfected* s years previously, experiencing either exogenous disease or endogenous reactivation of this reinfection event, and who therefore have not been reinfected further.

The distribution $\mathbf{D_c}(a,t,s)$ was obtained for individuals diseased between 1900 and 1990 using the same approach as that used in section 4.2.2.1, namely by estimating the corresponding $D_{i_c}(a,t,s)$ and $D_{r_c}(a,t,s)$ for all possible values of s (i.e. between 0 and a years). These were estimated by first deriving the numbers of individuals infected and reinfected respectively at age $a-s$ at time $t-s$, and simulating the disease dynamics for duration s in the subset of individuals in these groups who *did not experience further*

reinfection before this time. Individuals in this group belong to either the infected or 'latent' classes or are experiencing their first primary episode or endogenous disease as a consequence of the initial infection. The total number of individuals in the last two disease classes after time s gives $D_{i_c}(a,t,s)$.

Individuals reinfected at age $a-s$ at time $t-s$, on the other hand, belong to either the reinfected and 'latent' classes, or are experiencing exogenous or endogenous disease as a consequence of the reinfection event at time $t-s$. The total number of individuals in the last two classes in this group after time s gives $D_{r_c}(a,t,s)$. The corresponding differential equations representing the disease dynamics to derive $D_{i_c}(a,t,s)$ (and by extension $D_{r_c}(a,t,s)$) are summarized in Appendix B.1.

4.2.3 Prospective analyses of the incubation period and serial interval

4.2.3.1 Extensions to TBDYN3

In order to derive the distributions of the time interval until the first episode of respiratory or sputum-positive disease, individuals infected at a given age and in a particular year must in theory be followed up over their entire lives, given that they can develop endogenous and exogenous disease many years after an initial infection. The distributions ultimately obtained therefore depend on the "definition" of a lifetime. In the context of this work, the *maximum* duration of a lifetime is defined to be 100 years.

This potentially complicates the derivations of distributions for individuals infected when aged 20 years in 1950, for example, since follow-up would have to last until the year 2030, when they reach 100 years of age, whereas TBDYN3 has hitherto traced the disease dynamics among individuals only until the year 1990. Truncated distributions of the incubation period and serial interval could be obtained, considering the disease dynamics among infected individuals only until the year 1990, but this would complicate the comparison between distributions among individuals infected at the same age but in different years. A truncated distribution for individuals infected when aged 20 years in 1920, for example, would be based on 70 years of follow-up, as compared with 40 years of follow-up for similar individuals infected in 1950.

In order to derive distributions of the time interval until the first respiratory or sputum-positive disease episode based on the full maximum 100 years of life for all individuals

infected before 1990, TBDYN3 was extended to consider the disease dynamics in England and Wales beyond 1990. This required projections for the age and time-specific mortality rates and the annual risk of infection into the future. The officially predicted mortality rates were obtained from the Government's Actuary's department. In the first instance, it was assumed that the decline of 13% pa in the annual risk of infection after 1950 continued after 1990, and thus the distributions obtained reflect what would be observed if current trends in the risk of infection are maintained.

4.2.3.2 Lifetime risks and the distribution of the time interval from initial infection until the first respiratory or sputum-positive episode

The distribution of the time interval until the first episode of respiratory tuberculosis for individuals first infected at age A at time t denoted by $\mathbf{D_f}(A,t,s)$, as a function of the time since initial infection s is given by

$$\mathbf{D_f}(A,t,s) = \frac{D_1(A+s,t+s,s)}{\sum_{s=0}^{100-A} D_1(A+s,t+s,s)} \qquad \text{for } 0 \leq s \leq 100-A \qquad (4.7)$$

where $D_1(A+s,t+s,s)$ denotes the number of individuals infected at age A at time t experiencing *their first disease episode* after time s (i.e. when aged $A+s$ years at time $t+s$). Individuals have a maximum of $100-A$ years of life available in which to develop disease, given infection at age A. In expression 4.7, $D_1(A+s,t+s,s)$ does not stratify first respiratory episodes according to whether they are attributable to initial infection or subsequent reinfection. An analogous expression holds for the distribution of the time interval between first infection at age A at time t and the first subsequent sputum-positive episode.

The denominator in expression 4.7 gives the total number of individuals infected at age A at time t who develop at least one respiratory disease episode during their lifetime. Hence the ratio between the denominator and the total number of individuals infected at age A at time t, $I(A,t)$, gives the overall lifetime risk of developing at least one respiratory disease episode given initial infection at age A.

The distribution of the time intervals until the first respiratory disease episode given infection at a particular age, is not directly related to the corresponding distribution considering sputum-positive disease. Individuals experiencing their first sputum-positive episode, for example, may have already experienced a sputum-negative episode sometime before and hence do not contribute to the two distributions simultaneously.

These two distributions were derived by estimating the number of individuals experiencing their first respiratory or sputum-positive episode in each year following infection at age A and time t using an analogous approach to that used in section 4.2.2.1. The distributions of time intervals until the first respiratory or sputum-positive episode attributable to the initial infection (i.e. without further reinfection) were derived similarly. Hence the number of individuals infected at age A at time t was first obtained using TBDYN3, and individuals were followed up *until they developed their first respiratory or sputum-positive disease episode* with or without further reinfection. This involved subdividing the corresponding follow-up population into the appropriate disease classes and tracking the transitions between them. These disease classes, together with the corresponding follow-up population used are summarized in Table 4.1.

To derive the distribution of time intervals between infection and the first respiratory disease episode, as a consequence or otherwise of the initial infection, for example, *individuals infected simultaneously who had not yet experienced disease* were followed up. These individuals belonged to either the infected, 'latent' and reinfected classes (see row 1(a) in Table 4.1). $D_1(A+s, t+s, s)$ is therefore given by the total number of individuals entering the first primary, endogenous and exogenous disease categories from the infected, 'latent' and reinfected classes respectively after time s.

Similarly, to derive the distribution of the time interval between infection and the first sputum-positive episode as a consequence of the initial infection event, *individuals infected simultaneously who had experienced neither sputum-positive disease nor reinfection* were followed up. These individuals belong to either the infected or 'latent' classes or are experiencing sputum-negative first primary or endogenous disease (see row 2(b) in Table 4.1). In a given year, the total number of individuals experiencing sputum-positive disease for the first time as a consequence of the initial infection is given by the total number of individuals entering the sputum-positive first primary and endogenous disease categories from the infected and 'latent' classes respectively.

The lifetime risks of developing respiratory and sputum-positive disease were also obtained by following up infected individuals until age 100 years, and summing the numbers who developed their first respiratory or sputum-positive episode in each year of life.

Table 4.1: Follow-up population and disease categories used to analyse the incubation period and serial interval for tuberculosis prospectively.

Distribution of the time interval between initial infection and:	**Follow-up population**	**Subdivisions within follow-up population**		**Outcomes which contribute to distribution**
		Non-diseased	Diseased	
1(a)1st respiratory episode	Individuals infected simultaneously†, who have not yet experienced respiratory disease	Infected Latent Reinfected	None	1st Primary Endogenous Exogenous
1(b)1st respiratory episode attributable to the initial infection	as 1(a) above, but who have not been reinfected.	Infected Latent	None	1st Primary Endogenous
2(a) 1st sputum-positive episode	Individuals infected simultaneously†, who have not yet experienced sputum-positive disease	Infected Latent Reinfected	1st Primary Endogenous Exogenous } sputum -ive	1st Primary Endogenous Exogenous } sputum +ive
2(b) 1st sputum-positive episode attributable to the initial infection	as 2(a) above, but who have not been reinfected	Infected Latent	1st Primary Endogenous } sputum -ive	1st Primary Endogenous } sputum +ive

† By simultaneously, we mean individuals of age a infected at the same time.

4.3 Results

4.3.1 Retrospective analyses of the incubation period/serial interval

4.3.1.1 Distributions of time intervals since initial infection

Figure 4.1 summarizes the distribution of the time interval since first infection of individuals who developed disease when aged 20 years during the period 1900–1990, and also shows the relative proportion developing their first primary episode, endogenous and exogenous disease.

This shows that only a small proportion ($\approx$ 3%) of diseased 20 year olds in 1900 had first been infected during the preceding year, and the greatest proportion ($\approx$ 25%) had first been infected in their first year of life (i.e. 20 years beforehand). Hence only a small proportion of disease among 20 year olds in 1900 constituted first primary episodes and most were exogenous in origin. This occurred despite the higher risks of developing disease following initial infection in adolescence, as compared with those following reinfection.

The very small contribution to the disease incidence from 'recently' infected individuals follows from the fact that over 95% of 20 year olds in 1900 had already been infected by age 19 years, as they were born in 1880, when the annual risk of infection was estimated to be high ($\approx$ 20%), and declining by about 2% pa (see sections 2.3 and 3.2.1). Hence only a small proportion of the cohort was still at risk of infection and of subsequently developing their first primary episode in 1900.

The reasons underlying the large contribution to the disease incidence among 20 year olds in 1900 from individuals infected in their first year of life are complicated, and relate to the secular decline in the annual risk of infection, and how it determined the age-specific prevalence of infection in birth cohorts. By definition, individuals infected for more than five years comprise most of the 'latent' class, and face risks of endogenous disease, or of reinfection followed by exogenous disease (see section 3.1.2), which are independent of the time since initial infection. The proportion of all diseased 20 year olds in any year, who had been infected for 20, 19, 18 etc. years therefore reflects the proportion of the 'latent' class who had first been infected 20, 19, 18 etc. years beforehand, or whilst in their 1st, 2nd, 3rd etc. years of life. Given the secular decline in the risk of infection, the greatest proportion of any birth cohort was infected in its first year of life, when the risk of infection exceeded that in subsequent years. Hence individuals infected in their first year of life constituted a

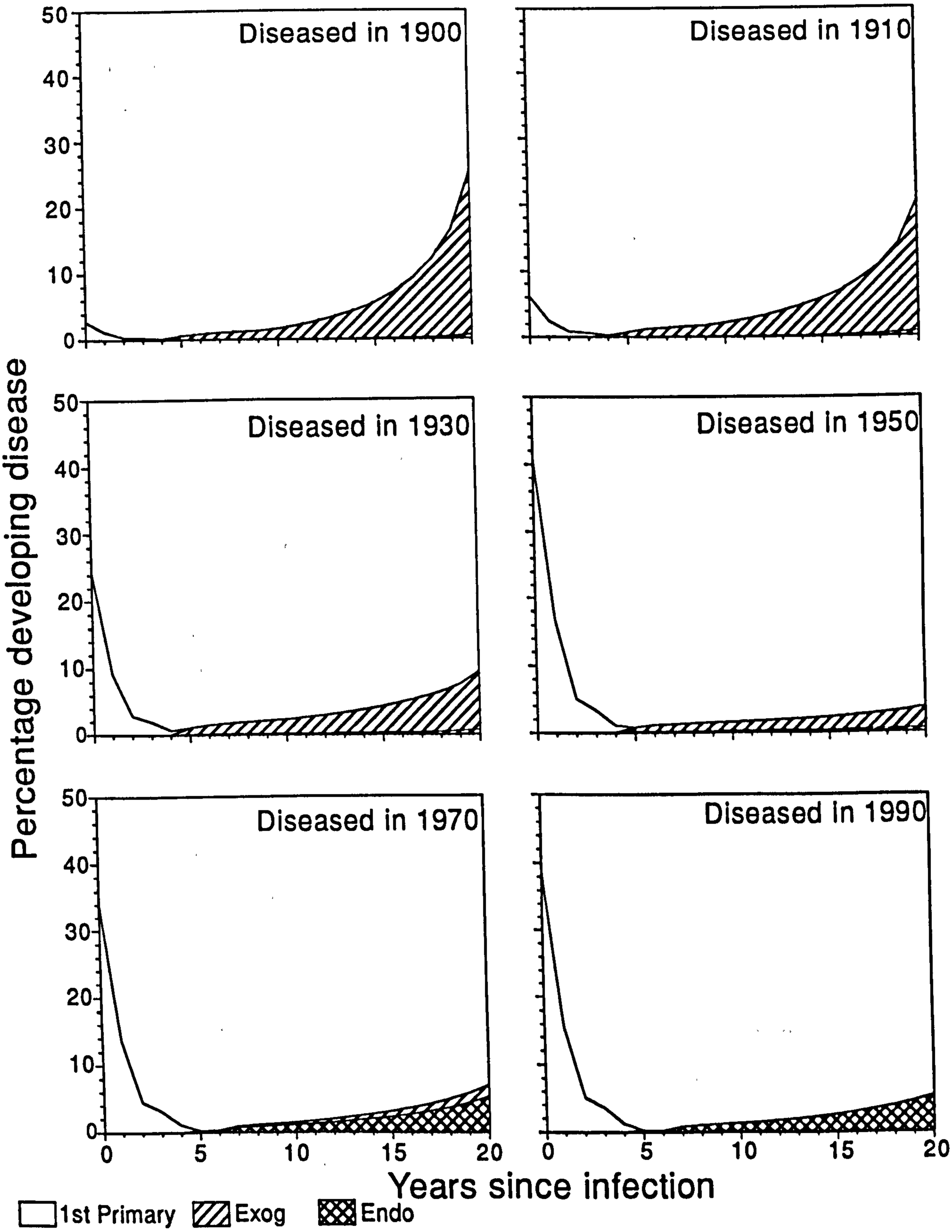


Figure 4.1: Distribution of the time interval *since first infection* of individuals who developed respiratory tuberculosis when aged 20 years during the period 1900–1990, as estimated using TBDYN3.

larger proportion of diseased 20 year olds in 1900, as compared with those infected when older.

Between 1900 and 1990, the shape of the distribution of the time interval since initial infection changed dramatically. The proportion of disease occurring among individuals infected during the preceding five years increased, as the contribution of exogenous disease to the disease incidence among individuals infected for more than five years decreased. Hence by 1990, the disease incidence among individuals infected for more than 5 years was attributable largely to endogenous reactivation. Similar changes are seen in the corresponding distributions considering older individuals diseased in these years (see Figures B.1 and B.2 in Appendix B.2).

These changes are attributable also to the secular decline in the risk of infection, which meant that increasing proportions of later cohorts reached adult life without having been infected (when they faced high risks of developing the first primary episode), as compared with earlier cohorts. The decline in the risk of infection also meant that smaller proportions of individuals in the 'latent' class in later cohorts were reinfected and subsequently developed exogenous disease, as compared with their counterparts in earlier cohorts. *Given the shapes of the distributions obtained*, we see that *it is inappropriate to make general conclusions about the **average** time interval since infection* of individuals developing disease at particular ages and in particular years.

These distributions were also broken down to reflect the proportions of individuals experiencing disease for the first, second or subsequent time. Figure 4.2 summarizes the results obtained for 20 year olds experiencing disease during the period 1900–1990.

This suggests that most of the cases in any year who had been infected for more than 5 years, were experiencing disease for the first time, and only a tiny fraction (not visible on the plots) had already experienced two or more episodes. Of the diseased 20 year olds in 1900 who had been infected for 20 years (i.e. in their first year of life), about 10% were experiencing their second episode, and less than 1% were experiencing their third or subsequent episode. For diseased individuals in 1990, the corresponding proportions were 4% and less than 0.01%. This shows that, overall, diseased individuals in a given year were more likely to have experienced a prior episode if they had experienced *initial infection* early in life and early in the century, when the risk of infection was still high. This is intuitively reasonable, given that:

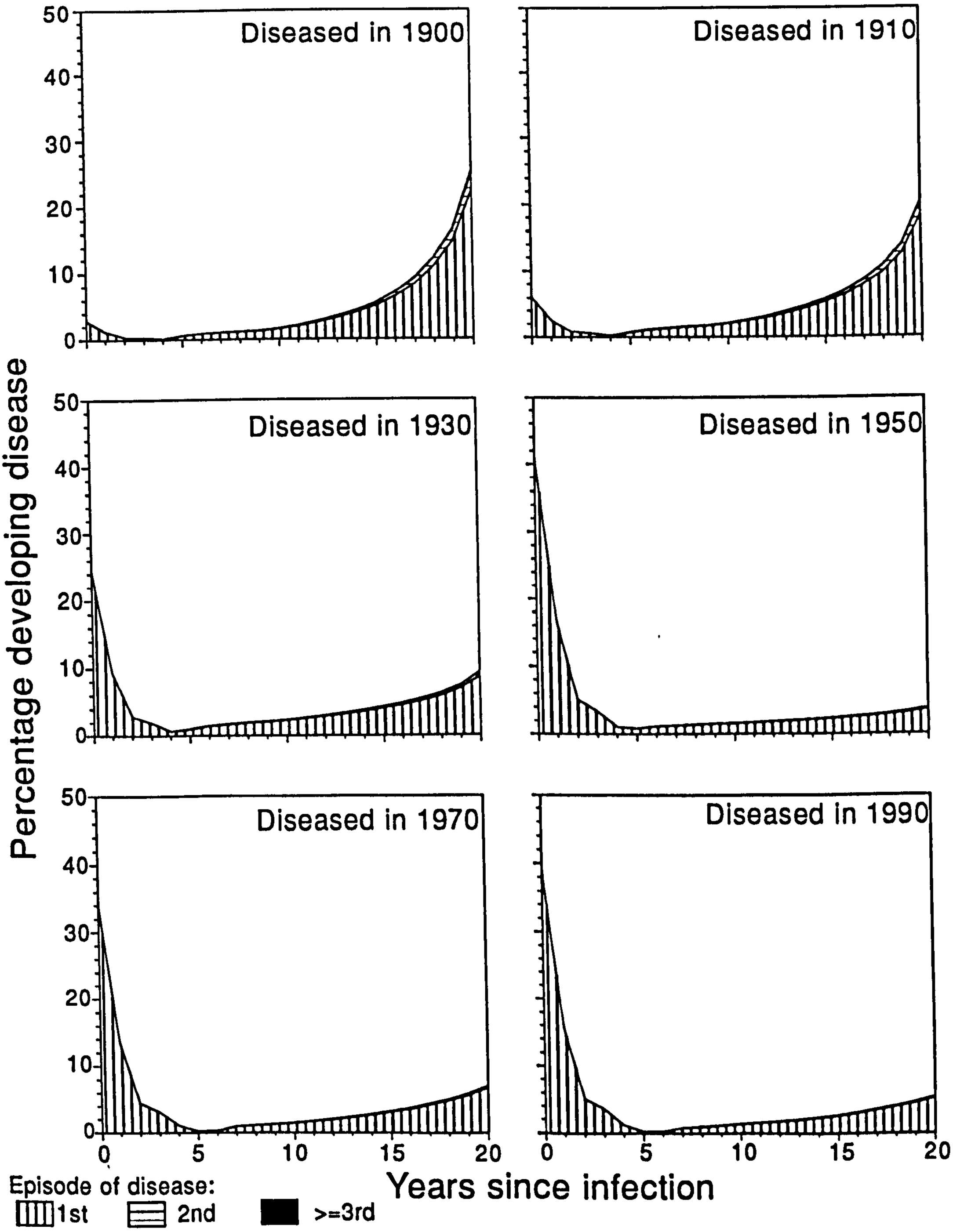


Figure 4.2: Distribution of the time interval since first infection of *individuals who developed respiratory tuberculosis when aged 20 years* during the period 1900–1990, as estimated using TBDYN3. Shaded areas denote the proportion of individuals experiencing their first, second or subsequent episode.
Note that the shapes of these distributions are identical to those in Figure 4.1.

1. individuals infected in infancy have more years of life available in which to develop several disease episodes before age 20 years, as compared with counterparts infected when slightly older, and

2. individuals first infected during the early 1900s experienced a higher risk of infection, and hence a greater proportion were reinfected and experienced exogenous disease, as compared with counterparts infected during the 1980s, for example.

Analogous patterns are seen in the corresponding distributions for 50 and 70 year olds diseased during the period 1900–1990 (Figures B.3 and 4.3 respectively). Of the 70 year olds diseased in 1900, who had been infected in their first year of life, for example, only about one third were experiencing their second episode, as compared with about 10% of their counterparts diseased in 1990.

4.3.1.2 Distribution of the time interval since the infection/reinfection event causing the disease episode

The distribution of the serial interval, or the time interval since the infection/reinfection event causing the current (sputum-positive) disease episode among 20 year olds during the period 1900–1990 is shown in Figure 4.4. This complements Figure 4.1, which shows the distributions of the time interval since *initial infection* for these individuals.

This shows that of the 20 year olds diseased in 1900, almost all had been reinfected during the preceding five years, and were therefore experiencing exogenous disease, whereas a negligible proportion was experiencing endogenous reactivation of an earlier infection or reinfection. Of those diseased in 1990, on the other hand, about 38% had first been infected during the preceding year, and about 8% of disease episodes were attributable to infection in the first year of life.

The change in the overall shape of the distribution between 1900 and 1990 is attributable to the decline in the risk of infection. This meant that later birth cohorts were more likely to reach adult life without having been infected (when they faced high risks of developing first primary episode if infected), and those who had been infected for many years were less likely to be reinfected as compared with earlier cohorts, and mainly developed endogenous disease as a consequence of the initial infection. Hence by 1990, diseased 20 year olds were either experiencing endogenous reactivation as a consequence of the initial infection event

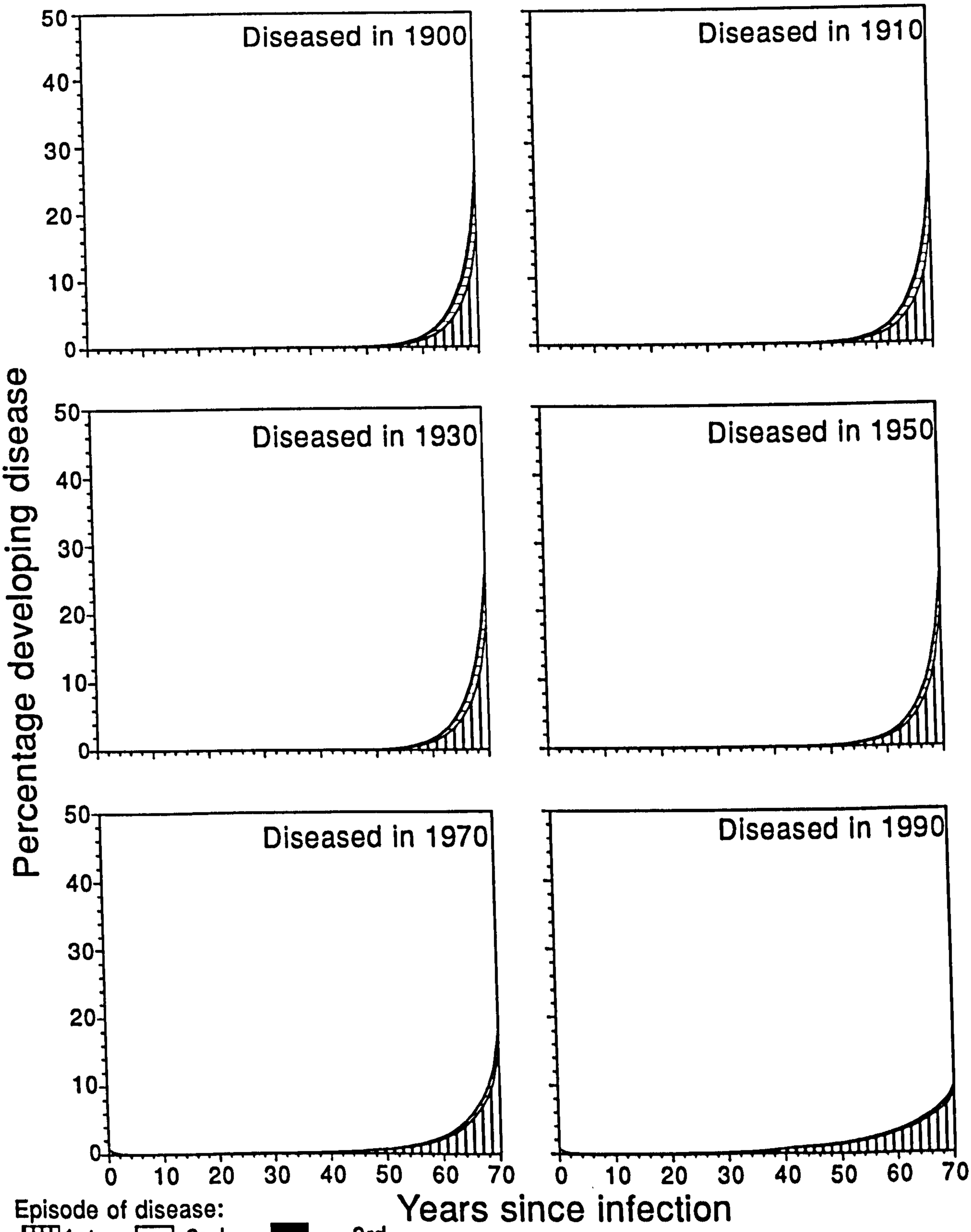


Figure 4.3: Distribution of the time interval since first infection of *individuals who developed respiratory tuberculosis when aged 70 years* during the period 1900–1990, as estimated using TBDYN3. Shaded areas denote the proportion of individuals experiencing their first, second or subsequent episode.

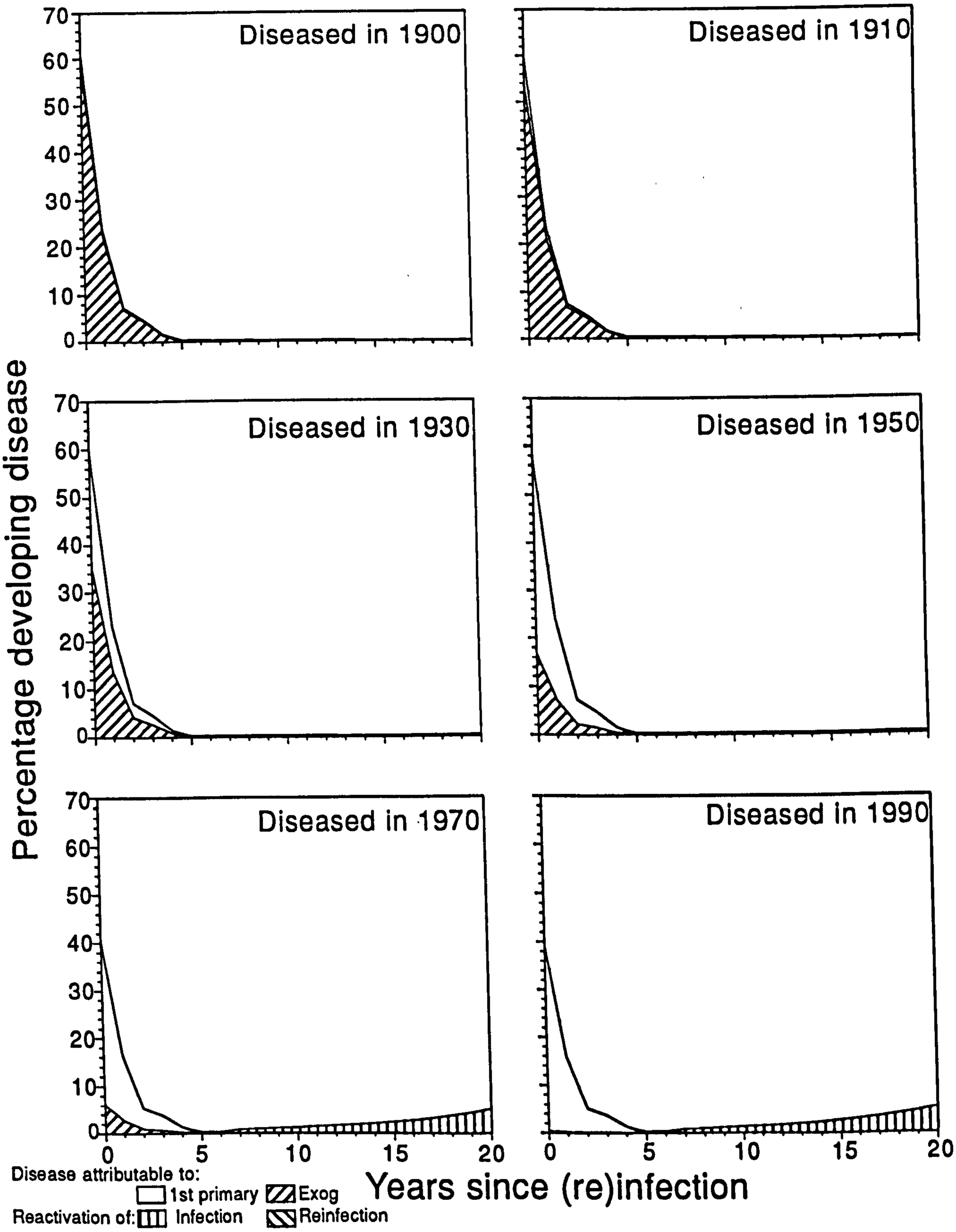


Figure 4.4: Distribution of the time interval *since the infection or reinfection event* causing the current (sputum-positive) episode among *individuals developing disease when aged 20 years* during the period 1900–1990, as estimated using TBDYN3. Shaded areas denote the proportion of individuals experiencing their first primary episode, exogenous disease, or reactivation of an earlier infection/reinfection.

or their first primary episodes after a recent infection. The distribution of the time interval since the infection/reinfection event among diseased 20 year olds in 1990, moreover, closely resembles that of the time interval since initial infection (see Figure 4.1).

Similar patterns are also seen in the distributions of the time interval since the infection/reinfection event causing the disease episode among 50 and 70 year olds during the time period 1900–1990 (see Figures 4.5 and 4.6). For these individuals, a greater proportion of disease in 1970 and 1990 was also attributable to reactivation of a reinfection event more than 5 years previously, as compared with that among diseased 20 year olds in these years. This is attributable to the high risks of infection when the corresponding 50 and 70 year olds were younger. These high infection risks meant that over 60% of 70 year olds in 1990 infected in their first year of life (in 1920), for example, had also experienced further reinfection at least once.

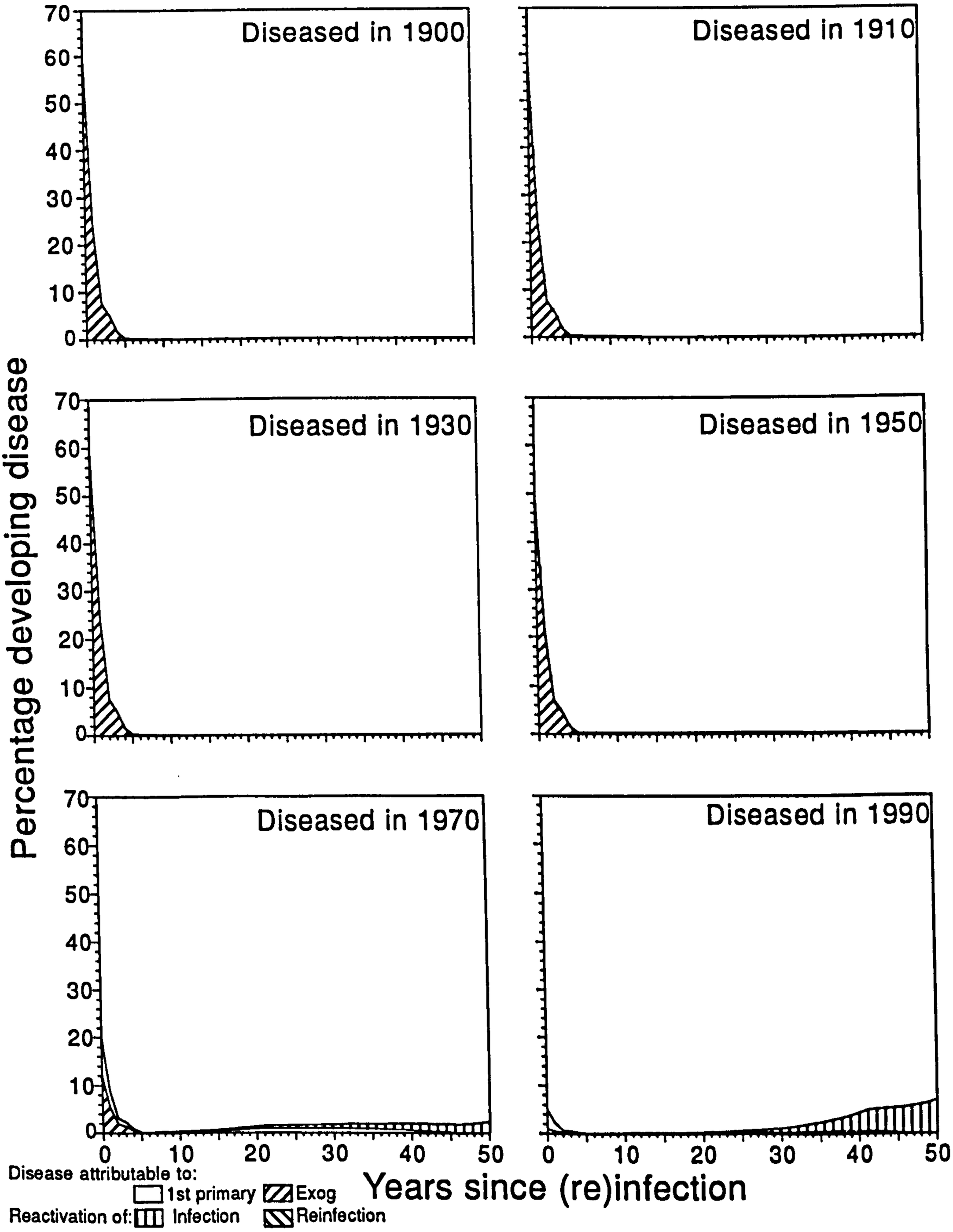


Figure 4.5: Distribution of the time interval *since the infection or reinfection event* causing the current (sputum-positive) episode among individuals developing disease when aged 50 years during the period 1900–1990, as estimated using TBDYN3. Shaded areas denote the proportion of individuals experiencing their first primary episode, exogenous disease, or reactivation of an earlier infection/reinfection.

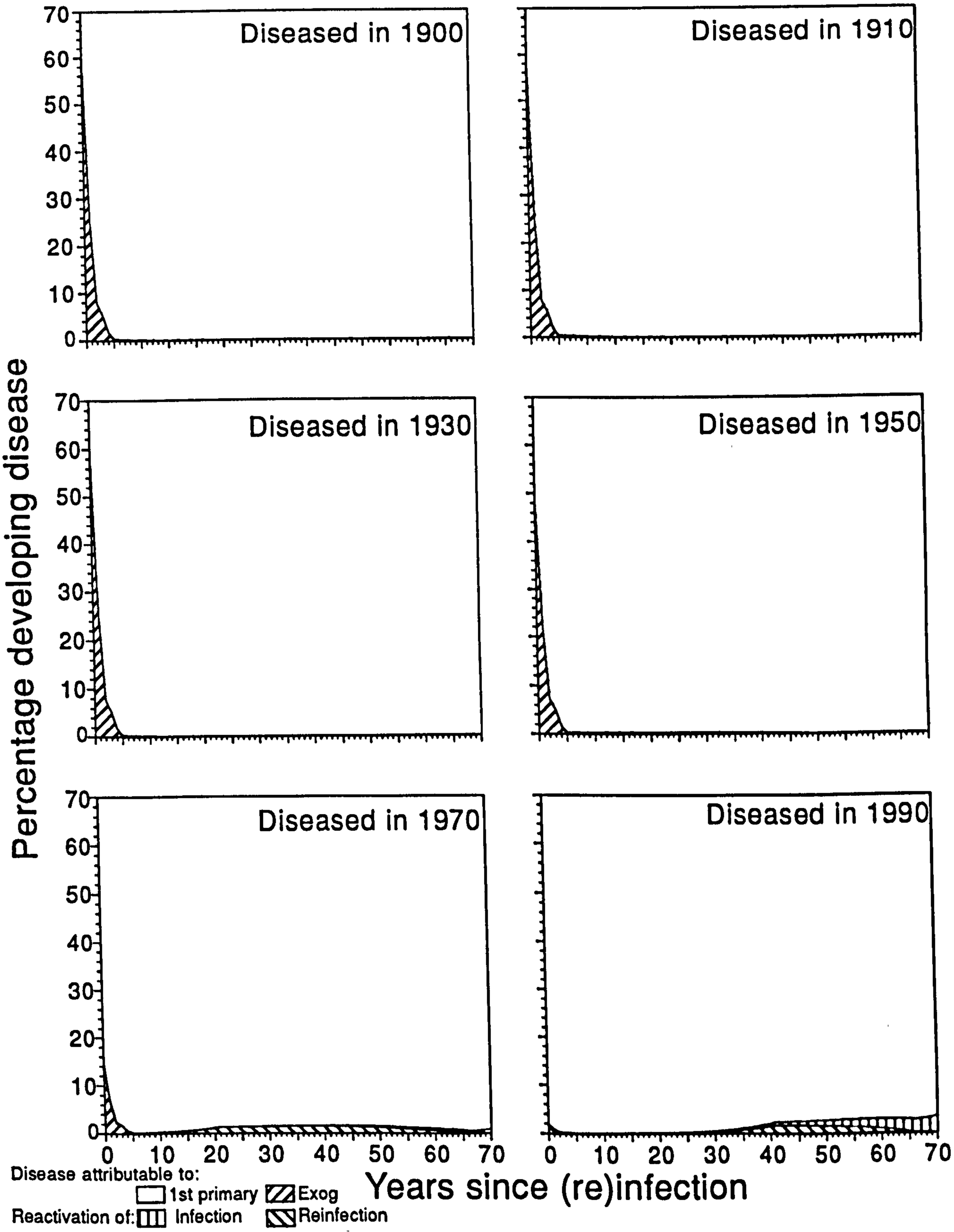


Figure 4.6: Distribution of the time interval *since the infection/reinfection event* causing the current (sputum-positive) episode among *individuals developing disease when aged 70 years* during the period 1900–1990, as estimated using TBDYN3. Shaded areas denote the proportion of individuals experiencing their first primary episode, exogenous disease, or reactivation of an earlier infection/reinfection.

4.3.2 Prospective analyses of the incubation period and serial interval

4.3.2.1 Overall lifetime risks of developing respiratory disease and the 'observed' incubation period

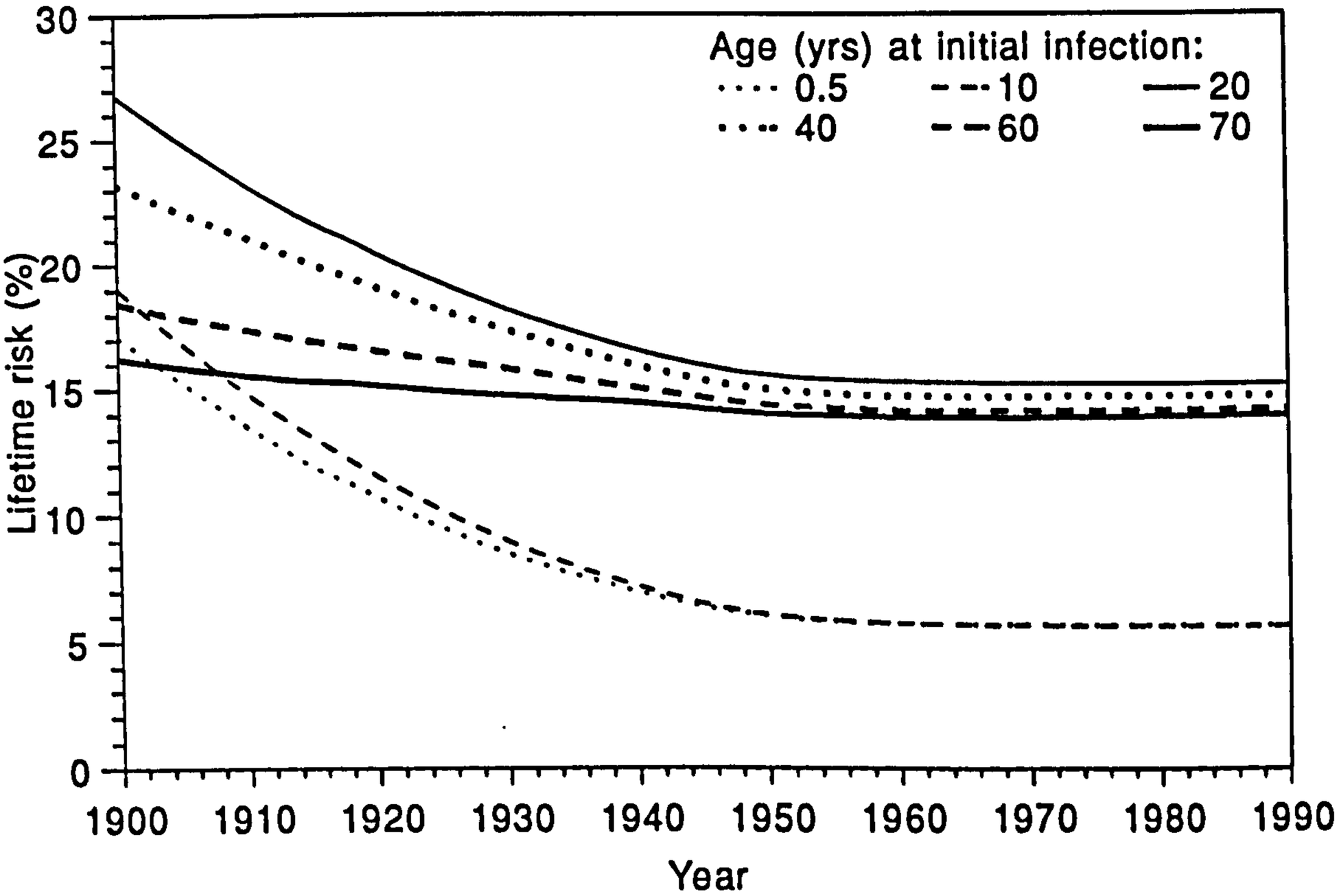


Figure 4.7: Lifetime risk of developing *respiratory* tuberculosis for individuals infected at different ages during 1900–1990 in England and Wales, as derived using TBDYN3.

Figure 4.7 summarizes the lifetime risks of developing respiratory tuberculosis for individuals initially infected at different ages during the period 1900–1990, as estimated using TBDYN3. According to this prediction, approximately 27% of individuals first infected at age 20 years in 1900 developed respiratory tuberculosis at least once in their lifetime, as compared with about 16% of their counterparts infected after 1950. For individuals infected in their first year of life in 1900 and after 1950, the corresponding lifetime risks were about 17% and 6% respectively.

The overall age and time-specific patterns in these estimated lifetime risks depend on several factors, in particular:

1. The age-specific risks of developing the first primary episode, endogenous and exogenous disease,

2. The level and trend in the risk of (re)infection experienced, and

3. The number of years of life available in which to develop disease following infection or reinfection.

Figure 4.7 shows that, regardless of the age at initial infection, the lifetime risk of developing respiratory disease declined between 1900 and 1950 and remained almost unchanged thereafter. This decline between 1900 and 1950 was a consequence of the decline in the risk of infection, which meant that successive cohorts of individuals infected simultaneously were less likely to be reinfected and subsequently experience exogenous disease than their predecessors.

The annual risk of infection was comparatively low after 1950 (e.g. ≈2% and 0.4% in 1950 and 1960 respectively), and hence few individuals first infected during this time also experienced reinfection. Most individuals therefore experienced either a first primary episode or endogenous disease, if they developed disease at all, as the probability of experiencing these forms of disease is insensitive to the level and trend in the risk of (re)infection. This accounts for the small change in the overall age-specific lifetime risks for individuals infected after 1950, despite the fact that the decline in the risk of infection accelerated in 1950 from 4% pa to 13% pa.

All individuals are assumed to face identical risks of developing their first primary episode, endogenous and exogenous disease following infection after age 20 years (see section 3.1.3). Differences between the lifetime risks of developing respiratory tuberculosis for 20 and 70 year olds first infected during the period 1900–1990 can therefore result only from differences in the years of life available in which to develop disease between these two age groups. Seventy year olds first infected in 1900, for example, had a lower lifetime risk as compared with their twenty year old counterparts, as they had fewer years of life available in which to develop disease.

In contrast, the lifetime risk of developing disease among 20 year olds first infected in 1900 was greater than that among corresponding 0–10 year olds. This difference is a consequence of the high risk of developing the first primary episode following infection at age 20 years (14% — see section 3.3.1), which exceeded even the lifetime risk of developing respiratory disease following infection in childhood.

Figure 4.7 shows that the lifetime risk among 10 year olds infected in 1900 was greater than that among younger counterparts, even though they had fewer years of life available in which to develop disease. This follows from the fact that TBDYN3 predicts that a greater

proportion of 10 year olds developed *exogenous* disease following infection during the early 1900s, as compared with younger counterparts, as described below.

TBDYN3 assumes that the risks of developing the first primary episode, endogenous and exogenous disease are identical for 0–10 year olds and first increase after age 10 years. Individuals infected in their *first year of life* in 1900 therefore faced the higher risks of developing exogenous disease after 1910 (i.e. when aged over 10 years) when the risk of infection was less than 6%. In contrast, *10 year olds* first infected in 1900 first experienced the high risks of developing exogenous disease when the risks of infection were slightly higher. Hence a greater proportion of 10 year olds infected in 1900 were reinfected and developed (exogenous) disease, as compared with their counterparts infected in their first year of life.

Given the high risks of infection estimated until 1950, it is likely that most adults during the early part of this century were first infected in childhood, and relatively few actually experienced the maximum lifetime risks predicted by TBDYN3. This is suggested in Figure 4.8, which compares the age-specific lifetime risks of developing respiratory tuberculosis shown in Figure 4.7 against the overall lifetime risk weighted by the number of individuals infected at each age.

This shows that the weighted lifetime risk also declined between 1900 and 1950, and, until 1920, was of a similar level to the risk among 0–10 year olds. This follows from the fact that most new infections during this time period occurred among infants and young children. The decline in the risk of infection meant that an increasing proportion of new infections occurred among adolescents and adults, who have high lifetime risks of developing disease, and hence we see that the decline in the weighted lifetime risk slowed until about 1950. The increase in the weighted average after 1950 is an artefact of the BCG vaccination coverage after 1954 incorporated into TBDYN3. BCG vaccination is assumed to confer full protection against infection for 77% of individuals, and hence large proportions of vaccinated individuals after 1954 in TBDYN3 do not contribute to the overall weighted average.

Given the dependence of the lifetime risks of developing disease on both the age at infection and the annual risk of infection, these results suggest that the distributions of the time interval between initial infection and the first episode of respiratory disease must be age and time-dependent.

This is illustrated in Figure 4.9, which summarizes the distributions of time intervals

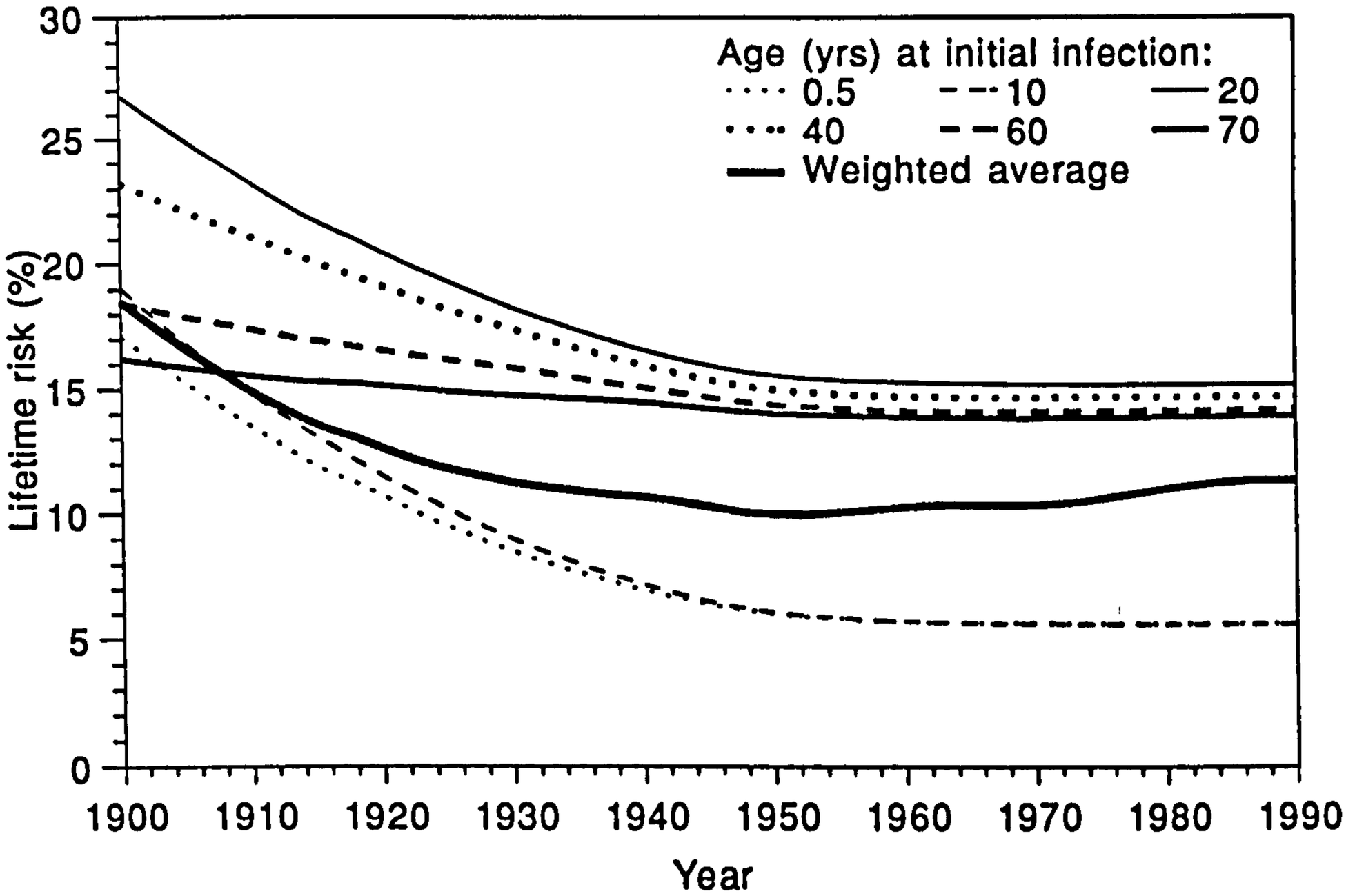


Figure 4.8: Comparison between the age-specific and overall lifetime risks (weighted by the number of individuals experiencing initial infection at each age) of developing *respiratory* tuberculosis for individuals infected during the period 1900–1990 in England and Wales, as derived using TBDYN3.

between initial infection for individuals infected in their first year of life during the period 1900–1990, and the first episode of respiratory disease. These show the relative contribution of first primary, endogenous and exogenous disease to the disease incidence following infection. These distributions correspond to the incubation period which would have been 'observed' among these individuals in practice. They do not correspond to the 'true' incubation period (defined by the time interval between initial infection/reinfection and the first disease episode *attributable* to this infection/reinfection — see section 4.1), since endogenous disease among these individuals can be attributed either to a subsequent reinfection or to the initial infection event.

This shows that, of those infected in 1900 who developed disease sometime thereafter, about 23% did so within 5 years, as compared with about 65% and 70% of their counterparts infected in 1950 and 1990 respectively. The shapes of the distributions were similar for individuals infected in 1970 and in 1990, and most of the disease experienced constituted either first primary episodes or endogenous disease. This is consistent with the fact that the estimated lifetime risk of developing respiratory disease differed little between individuals

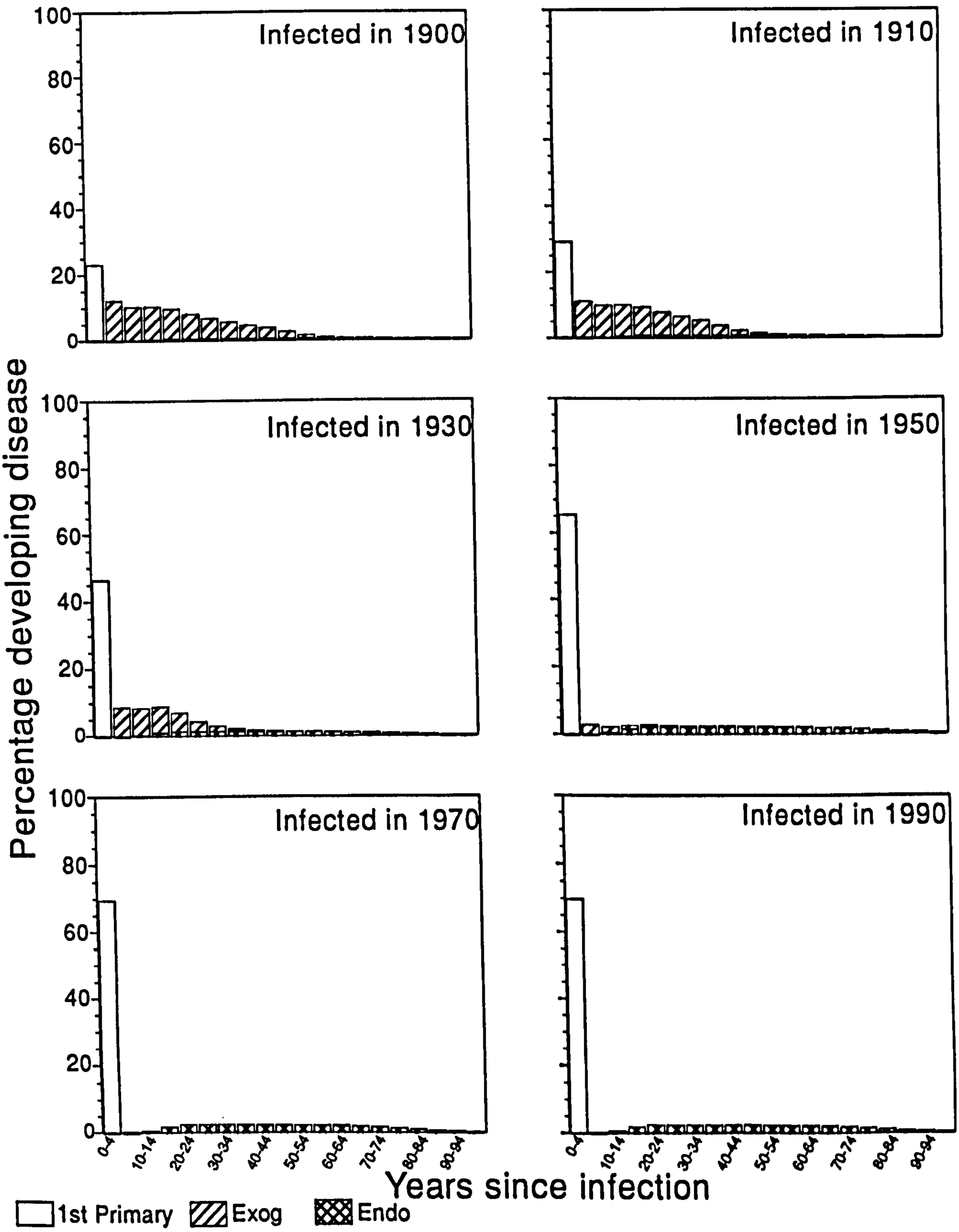


Figure 4.9: Distribution of the time interval between first infection and the first episode of respiratory disease for individuals *infected in their first year of life* during the period 1900–1990, as estimated using TBDYN3. This corresponds to an 'observed incubation period'.

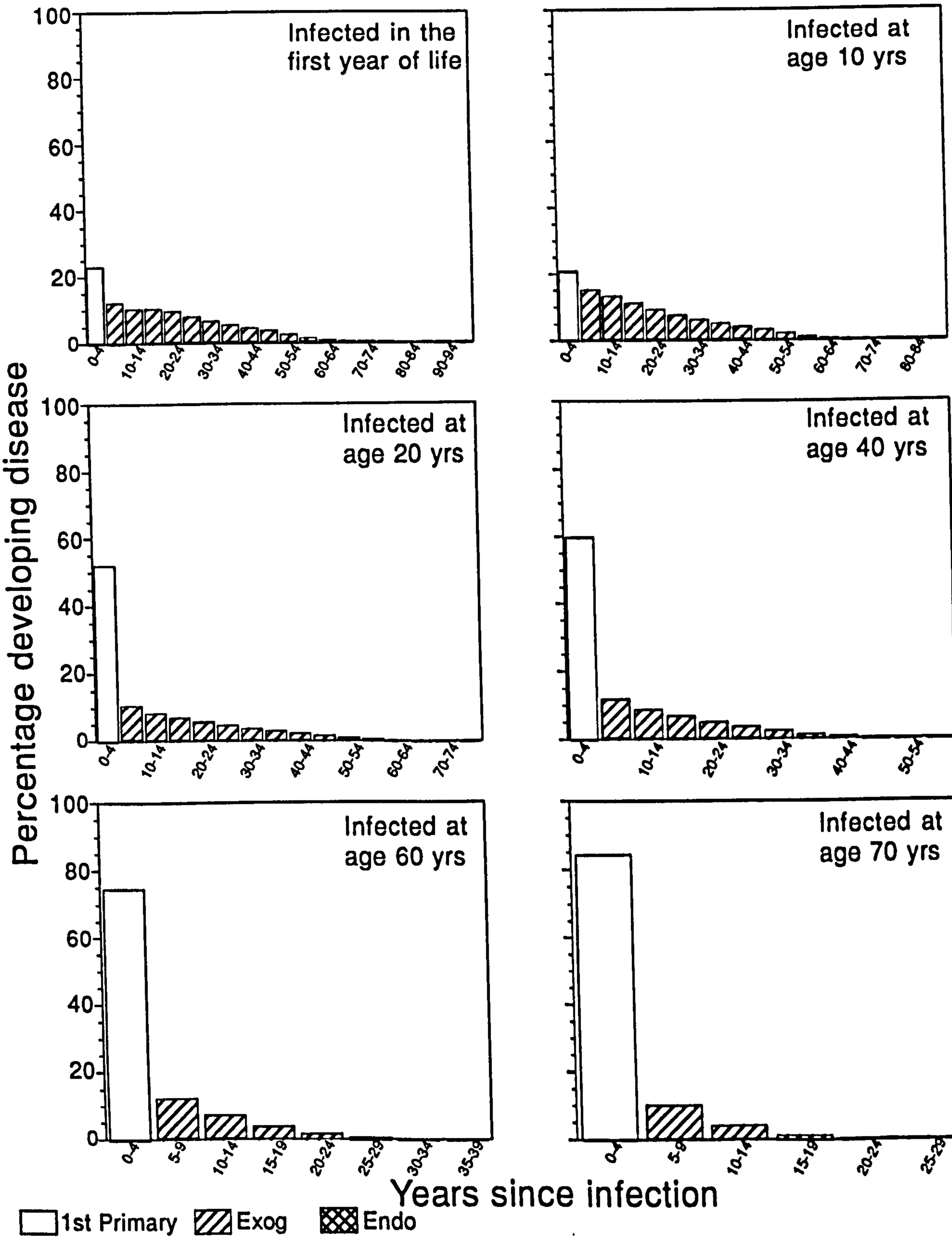


Figure 4.10: Distribution of the time interval between first infection and the first episode of respiratory disease for individuals *infected at different ages in 1900*, as estimated using TBDYN3. This corresponds to an 'observed incubation period'.

infected in 1960 and 1990. This follows from the low risk of infection during this time period, which meant that few individuals were reinfected and subsequently experienced exogenous disease.

The overall shape of the distribution of the time interval until the first episode of respiratory disease following infection in infancy after 1950 reflects the high risk of developing the first primary episode and the change in the risks of developing endogenous disease experienced as the individuals aged (i.e. low for 0–10 year olds, increasing until age 20 years and constant thereafter — see section 3.1.3). The proportion developing their first episode decreased 60 years after initial infection as a consequence of the high mortality rates, which reduced the number of surviving individuals from the initial cohort.

Most of the disease among individuals infected before 1950 is attributable either to a first primary episode or to exogenous disease, although the relative contribution of the latter decreased with successive generations of infected individuals whereas that of first primary episodes and endogenous disease increased. This is consistent with the decline in the estimated lifetime risk of developing respiratory tuberculosis between 1900 and 1950.

The shapes of the distributions of the time interval until the first respiratory episode for individuals infected until 1950 are therefore mainly a function of the size of the risk of developing the first primary episode and the secular decline and magnitude of the risk of infection. The decline in the risk of infection meant that the proportion of individuals who were reinfected and experienced exogenous disease also declined as the time since initial infection increased, even when the risks of developing exogenous disease increased, such as between the ages 10 and 20 years. This translated into a decline in the *overall* proportion developing respiratory disease as the time since initial infection increased *when the risk of infection experienced was still high*, e.g. for individuals infected in 1900 and 1910.

This decline in the proportion developing disease as the time since initial infection increased is also seen in the corresponding distributions for individuals infected at other ages during this time. This is illustrated in Figures 4.10 and B.4 (Appendix B.2.2), which show these distributions for individuals infected when aged 0–70 years in 1900 and 1930 respectively.

Of the individuals who developed disease following initial infection in the first year of life in 1900, Figure 4.10 shows that first primary episodes comprised about 23% of the subsequent disease incidence, as compared with about 53% and 85% of that among counterparts

infected at age 20 and 70 years respectively. The low contribution for individuals infected in their first year of life, as compared with that among 20 and 70 year old counterparts follows from the lower risk of developing the first primary episode following infection in infancy, as compared with that following infection after age 20 years. The higher contribution of first primary episodes to the disease incidence among individuals infected when aged 70 years, as compared with that among 20 year old counterparts follows from the facts that

1. 70 year olds had fewer years of life available in which to develop disease, as compared with their 20 year old counterparts, and

2. newly infected 70 year olds are less likely to survive 5 years and to experience endogenous or exogenous disease, as they faced rapidly increasing mortality rates, as compared with their 20 year old counterparts.

Similar patterns are seen in the corresponding distributions for individuals infected in 1960 and 1990 (see Figures B.5 and B.6 in Appendix B.2.2).

4.3.2.2 Overall lifetime risks of developing sputum-positive disease and the 'observed' serial interval

Figure 4.11 summarizes the age-specific lifetime risks of developing *sputum-positive* tuberculosis, for individuals infected during the period 1900–1990 in England and Wales. For each age at infection considered, these risks are lower than the corresponding lifetime risks of developing respiratory tuberculosis (see Figure 4.7), since only an age-dependent fraction of all respiratory disease is assumed to be sputum-positive (see Figure 3.4 in section 3.1.3).

Overall, the temporal and age-specific patterns in the lifetime risks of developing sputum-positive disease resemble those found for all respiratory disease. For each age at infection considered, for example, the estimated lifetime risk of developing sputum-positive disease also declined over time and the risks among 20 year olds infected in 1900 exceeded those among their 70 year old counterparts. This is consistent with the greater number of years of life available for 20 year olds infected in 1900 to become reinfected and subsequently develop exogenous disease or develop endogenous disease, as compared with those infected when aged 70 years in 1900.

Whilst individuals aged over twenty years are assumed to face identical risks of developing the first primary episode, endogenous and exogenous disease, the probability that this

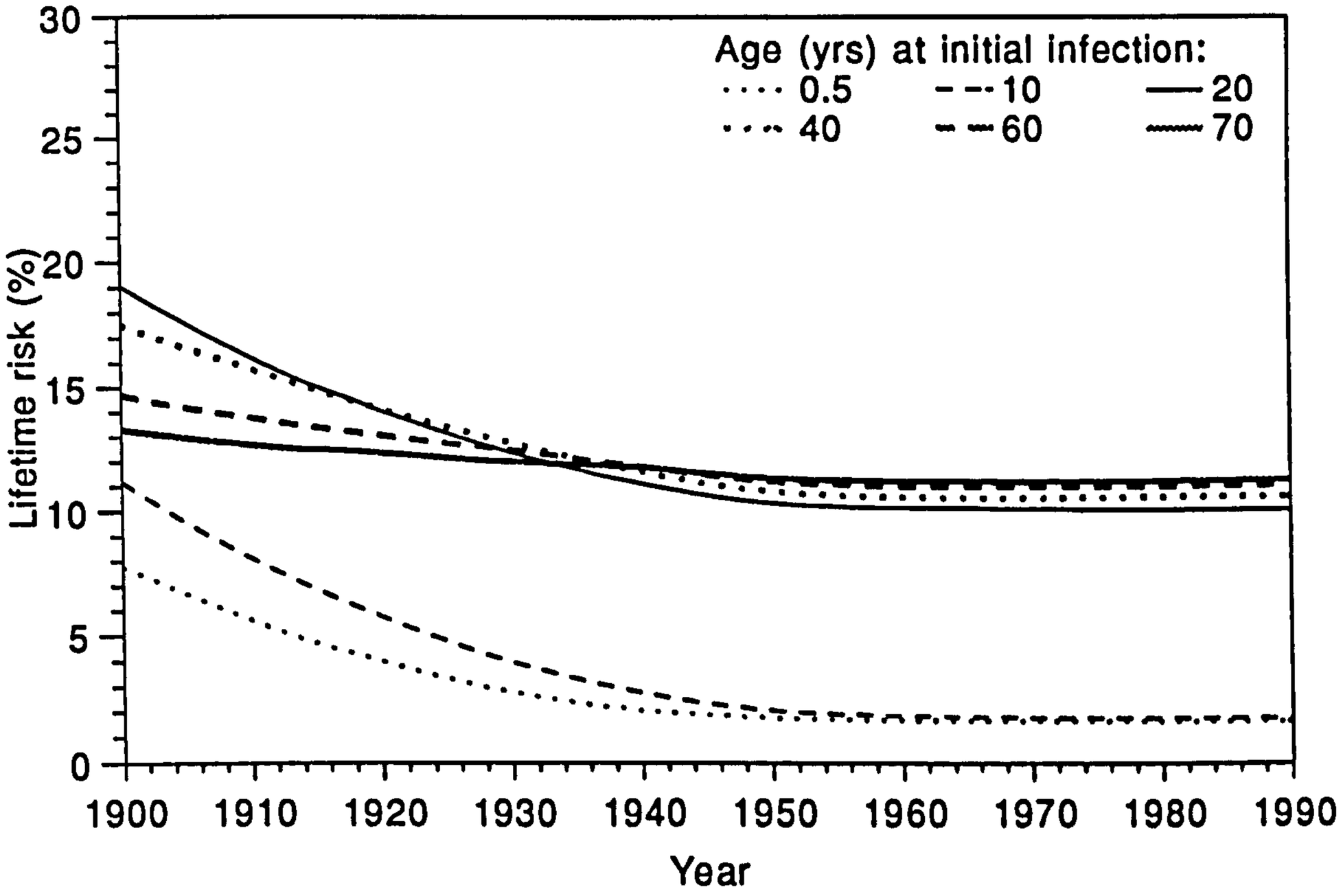


Figure 4.11: Lifetime risk of developing *sputum-positive* tuberculosis for individuals infected at different ages during the period 1900–1990 in England and Wales, as derived using TBDYN3.

disease episode is sputum-positive is assumed to increase with age (see Figure 3.4 in section 3.1.3). When the risk of infection is low (e.g. after 1950), this initial high risk following infection among 70 year olds exceeds the overall lifetime risk of developing sputum-positive disease among younger counterparts. This explains the lower lifetime risks following infection among 20 year olds after 1940, as compared with that among 70 year olds. In contrast, the lifetime risks of developing *respiratory disease* were found to be consistently lower following infection for 70 year olds, as compared with those among 20 year olds (see Figure 4.7).

For reference, Figure 4.12 compares these lifetime risks against the corresponding overall lifetime risk, weighted by the number of individuals experiencing initial infection at each age. This exhibits very similar temporal patterns to the corresponding weighted lifetime risk of developing respiratory tuberculosis, declining until about 1920, and then increasing from 1950.

The differences between the age and time-specific lifetime risks of developing sputum-positive tuberculosis and the corresponding risks of respiratory disease, suggest that the distributions of the time interval between initial infection and the first sputum-positive

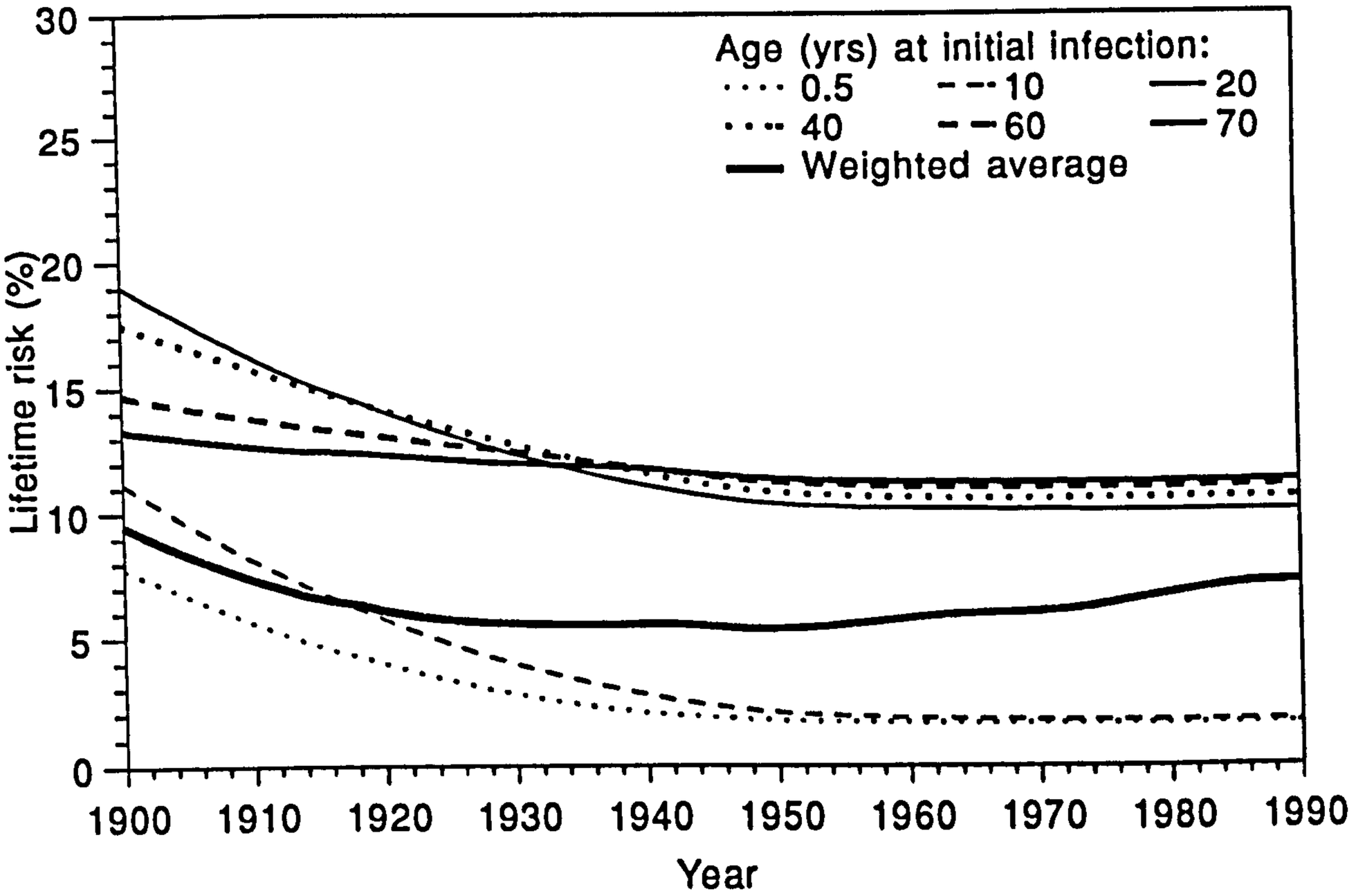


Figure 4.12: Comparison between the age-specific and overall lifetime risks (weighted by the number of individuals experiencing initial infection at each age) of developing *sputum-positive* tuberculosis for individuals infected during the period 1900–1990 in England and Wales, as derived using TBDYN3.

episode must also differ from those for respiratory disease.

This is reflected in Figure 4.13, which summarizes the distribution of the time interval until the first sputum-positive episode for individuals infected in their first year of life during the period 1900–1990. These distributions correspond to the serial interval which would have been 'observed' among these individuals in practice. They do not correspond to the 'true' serial interval (see definition in section 4.1) since endogenous disease among these individuals can be attributed either to the initial infection event or a subsequent reinfection.

The increase in the risk of developing sputum-positive exogenous disease between the ages 10 and 20 years, for example, affects the shape of these distributions for individuals who experience high risks of infection. This increase leads to a peak in the proportion of individuals experiencing disease during the 20–24th year after infection for those infected in infancy before 1950, whereas the greatest proportion of individuals who developed *respiratory* disease did so during the first five years after infection (see Figure 4.9).

When the risk of infection is low (e.g. after 1950), the shapes of these distributions reflect the relatively high risk of experiencing the first primary *sputum-positive* episode and

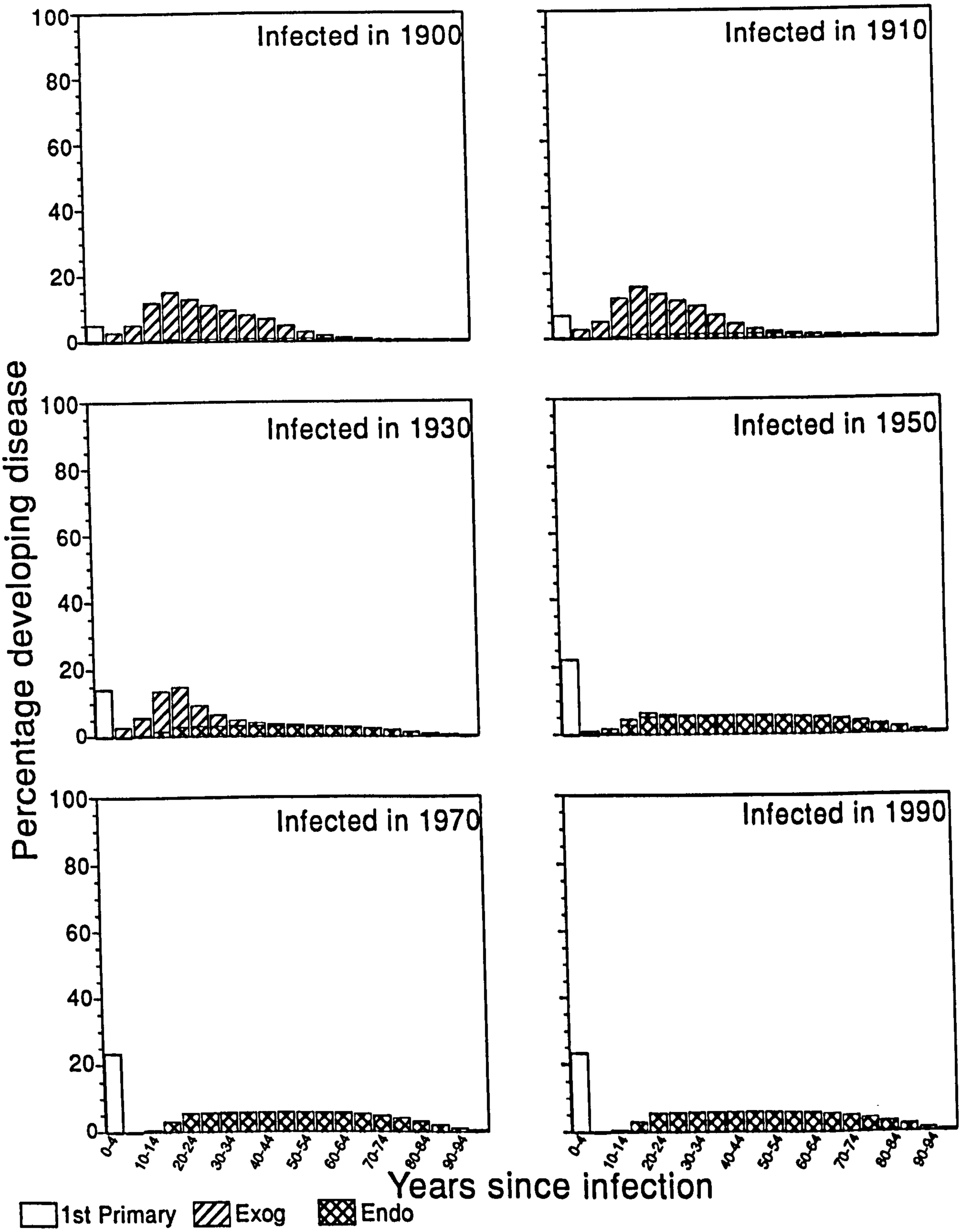


Figure 4.13: Distribution of the time interval between first infection and the first episode of sputum-positive disease for individuals *infected in their first year of life* during the period 1900–1990, as estimated using TBDYN3. This corresponds to an 'observed serial interval'.

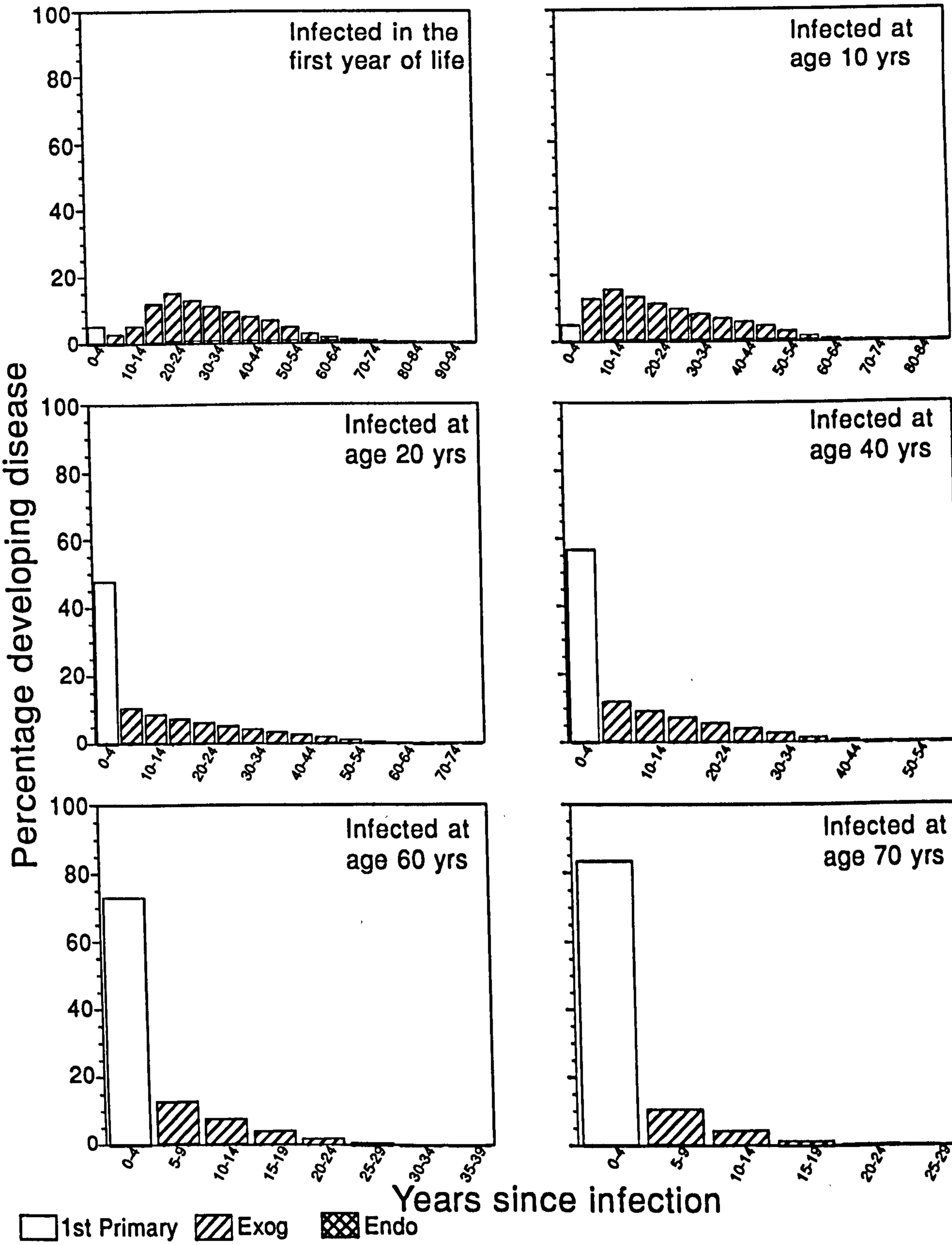


Figure 4.14: Distribution of the time interval between first infection and the first episode of sputum-positive disease for individuals *infected at different ages in 1900*, as estimated using TBDYN3. This corresponds to an 'observed serial interval'.

the change in the risks of experiencing *sputum-positive* endogenous disease experienced as the individuals age. Additionally, the shapes of the distributions are similar to those of the distributions considering respiratory tuberculosis (see Figure 4.9), except that the overall relative contribution of first primary episodes to the sputum-positive disease incidence is slightly lower. For those infected in 1990, for example, first primary episodes comprised 23% of the subsequent sputum-positive incidence, and about 70% of the respiratory incidence (see Figure 4.9). This follows from the assumption that the risk of a given disease episode being sputum-positive increases as individuals age, i.e. as the time since initial infection increases.

Figure 4.14 summarizes the distributions of the time interval since initial infection and the first sputum-positive episode for individuals infected when aged 0–70 years in 1900. Considering those infected when aged over 20 years, the overall shape of the distributions differ only slightly from those considering respiratory tuberculosis (see Figure 4.10), e.g. first primary episodes comprise a slightly *smaller* proportion of the sputum-positive incidence, as compared with that of the respiratory incidence. This is consistent with the findings for individuals infected in their first year of life discussed above. Similar findings are seen in the corresponding distributions considering counterparts infected in 1930, 1960 and 1990 (see Figures B.7, B.7 and B.7 in Appendix B.2).

4.3.2.3 Lifetime risks of developing disease *attributable* to initial infection, and the 'true' incubation period and serial interval

The general conclusions from the analyses of the lifetime risks of developing disease *attributable* to the initial infection are similar to those in the previous sections. They are included here for documentary thoroughness.

Figure 4.15 summarizes the lifetime risks of developing respiratory disease *attributable* to the initial infection event for individuals infected when aged 0–70 years during the period 1900–1990. This refers to the lifetime risk of developing disease without experiencing any reinfection in the meantime. This shows that of the individuals infected in their first year of life in 1900, about 4% ultimately developed respiratory disease without having been reinfected in the meantime, as compared with about 14% of those infected when aged 20 years. Figure 4.16 summarizes the corresponding risks of developing sputum-positive disease. Both also show the overall lifetime risks in each year weighted according to the

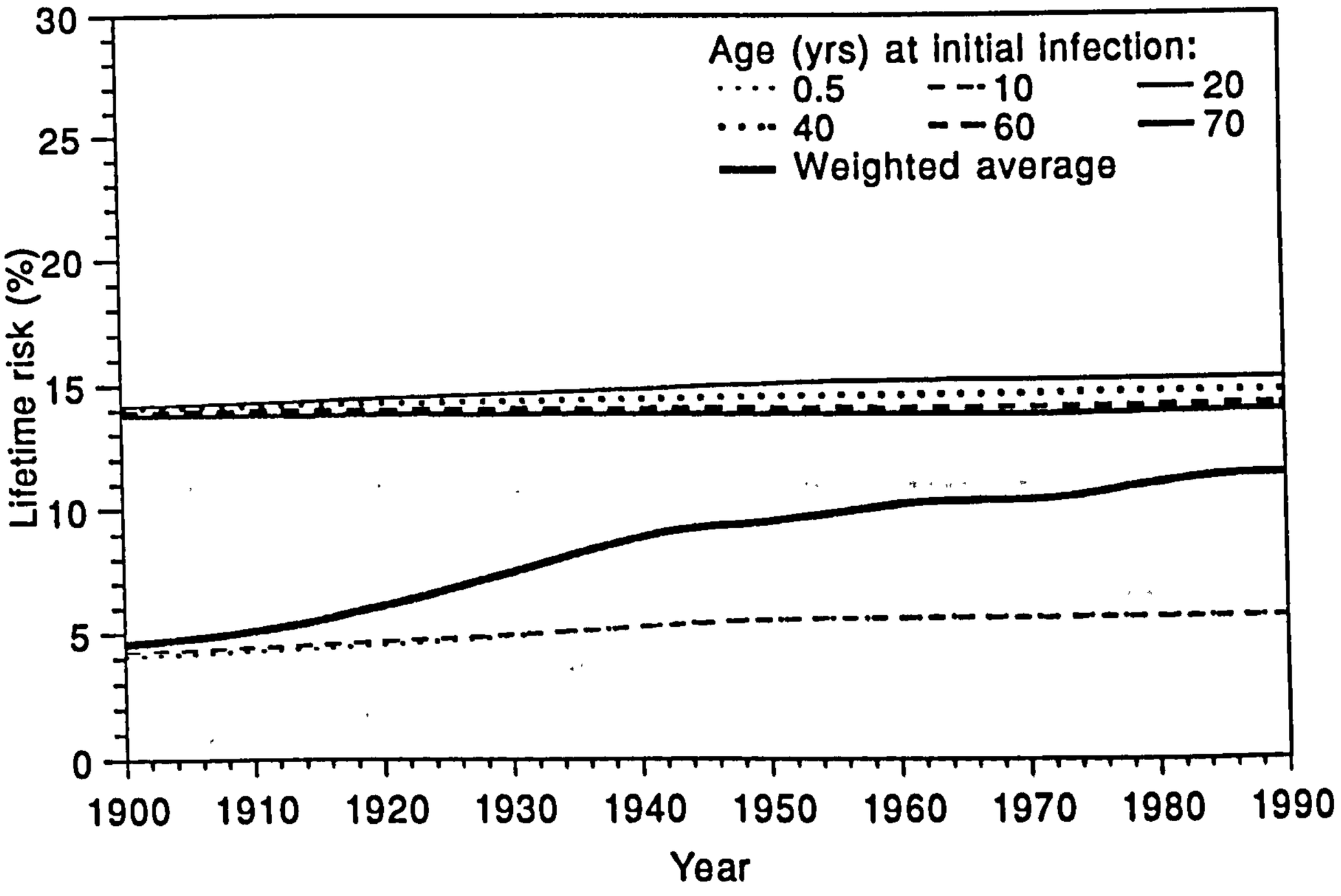


Figure 4.15: Lifetime risk of developing respiratory tuberculosis *attributable to initial infection* for individuals infected at different ages during 1900–1990 in England and Wales, as derived using TBDYN3.

number of individuals experiencing initial infection at each age.

These figures show that for all ages at infection considered, the lifetime risks of developing sputum-positive or respiratory disease *attributable to initial infection* increased slightly over time. Considering individuals infected in their first year of life in 1900 and after 1950, the lifetime risk of developing respiratory disease was about 4% and 6% respectively. This paradoxical increase, albeit slight, resulted from the decline in the annual risk of infection, which meant that individuals infected after 1950 for example, were more likely to experience disease without subsequent reinfection and thus apparently as a consequence of the initial infection, as compared with their counterparts infected in 1900. We also see that the decline in the risk of infection, which led to changes in the age distribution of individuals experiencing initial infection, led to *increases* in the overall *weighted* lifetime risks of developing respiratory and sputum-positive disease over time.

The overall age-specific patterns in the lifetime risks of developing respiratory tuberculosis *attributable to initial infection* are identical to those of the *overall* lifetime risks (see Figure 4.7). Given any year of infection, these risks are lowest for individuals infected in their first year of life, and are higher for 70 year olds, as compared with those among 20

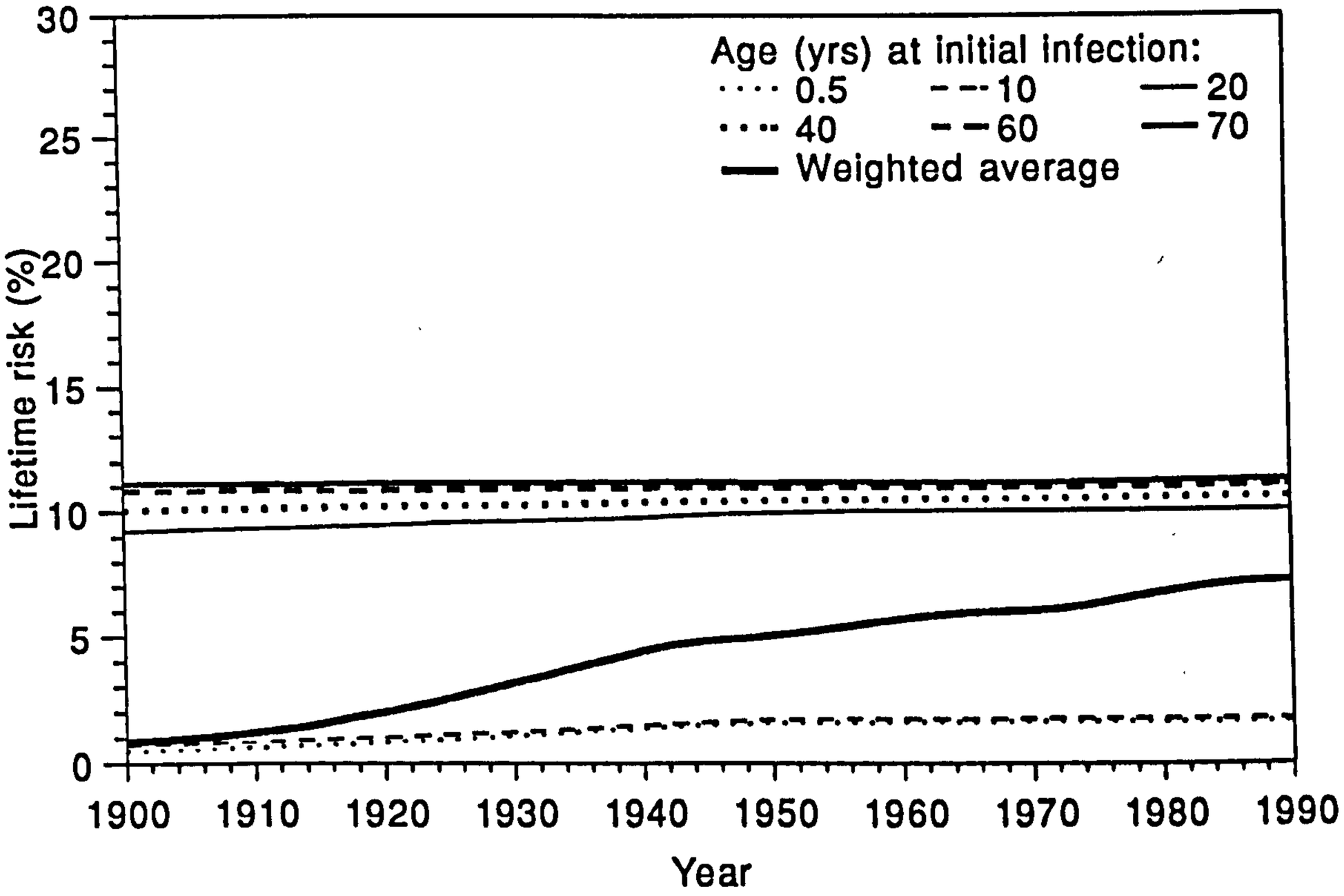


Figure 4.16: Lifetime risk of developing sputum-positive tuberculosis *attributable to initial infection* for individuals infected when aged 0–70 years during 1900–1990 in England and Wales, as derived using TBDYN3.

year old counterparts. The reasons behind these age-specific patterns are identical to those explaining the analogous differences in the overall age-specific lifetime risks of developing respiratory tuberculosis (see page 163).

In contrast, the lifetime risks of developing sputum-positive disease attributable to the initial infection is estimated to increase with the age at infection, irrespective of the year of infection considered. This is consistent with the age-specific patterns in the *overall* lifetime risks of developing sputum-positive disease (see Figure 4.11). These patterns reflect the increasing risk (with age) of developing sputum-positive disease immediately after infection, which outweighs the overall lifetime risk among younger individuals (see page 170).

Further insight into these age and time-specific patterns is provided by the distributions of the time interval between infection and the first onset of respiratory and sputum-positive disease for individuals who developed disease as a consequence of initial infection. These are described below.

Figure 4.17 shows the corresponding distributions considering respiratory tuberculosis for individuals infected in their first year of life during the period 1900–1990. These distributions reflect the 'true incubation period' for individuals infected at these ages in these

years (see definition in section 4.1).

This shows that of those infected in their first year of life in 1900, who developed respiratory disease as a consequence sometime thereafter, about 95% did so during the first 5 years after infection as compared with about 70% of counterparts infected in 1990. This again reflects differences between the risk of infection experienced by these individuals. The high risk of infection before 1950, for example, meant that individuals were more likely to be reinfected and hence less likely to experience disease attributable to the initial infection event *more than five years after infection*, as compared with counterparts infected after 1950. Similar patterns are also seen in the corresponding distributions for individuals infected at other ages. For reference, Figures B.10, B.11 and B.12 show these distributions for 0–70 year old individuals infected in 1900, 1960 and in 1990 respectively.

These figures also show that the distribution of the time interval until the first episode of respiratory disease attributable to the initial infection event closely resembles that of the 'observed incubation period' (see Figure 4.9) for individuals infected after 1950. This is consistent with the low level of the risk of infection during this time, which meant that few individuals were reinfected and developed disease as a consequence.

The same general patterns are seen in the corresponding distributions considering *sputum-positive* tuberculosis (i.e. the 'true serial interval' — see definition in section 4.1). This is shown in Figure 4.18, which summarizes these distributions for individuals infected in their first year of life during the period 1900–1990. We see, for example, that the proportion who developed sputum-positive disease during the first five years after infection was lower for those infected in 1900 (80%), as compared with that for individuals infected in 1990 (23%).

Figure 4.18 also shows that a greater proportion of those who developed sputum-positive disease as a consequence did so more than five years after infection, in any year of infection as compared with the corresponding individuals who developed respiratory tuberculosis. This follows from differences between the risks of developing sputum-positive disease and respiratory disease, which were discussed in relation to the distribution of the 'observed' serial interval (see section 4.3.2.2). This also holds for the corresponding distributions for individuals infected at other ages in other years. These distribution considering 0–70 year olds infected in 1900, 1930, and 1990 are shown in Figures B.13, B.14 and B.15 in Appendix B.2.

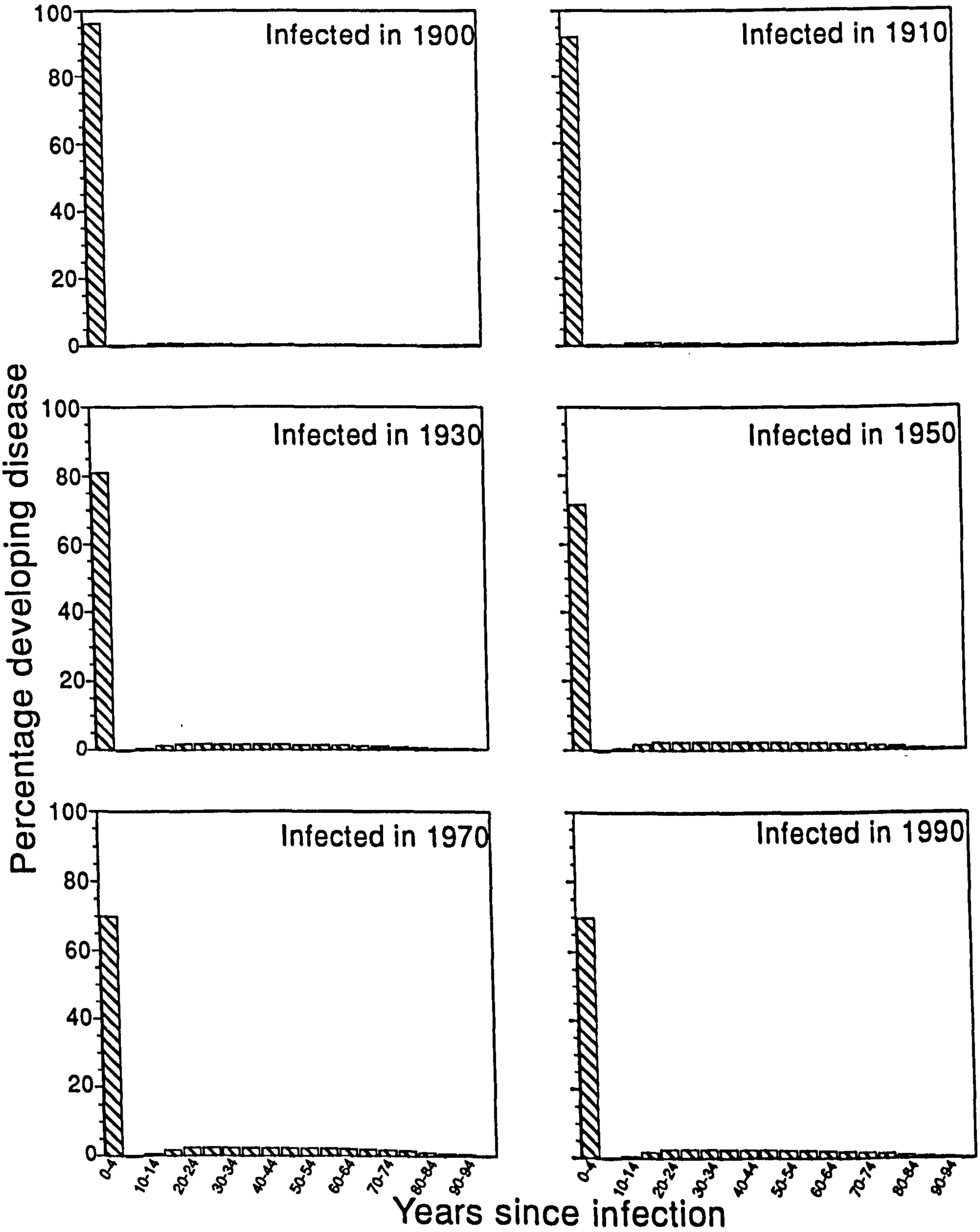


Figure 4.17: Distribution of the time interval between first infection and the first episode of respiratory disease *attributable to the initial infection* for individuals infected in their first year of life during the period 1900–1990, as estimated using TBDYN3. This corresponds to an 'actual incubation period'.

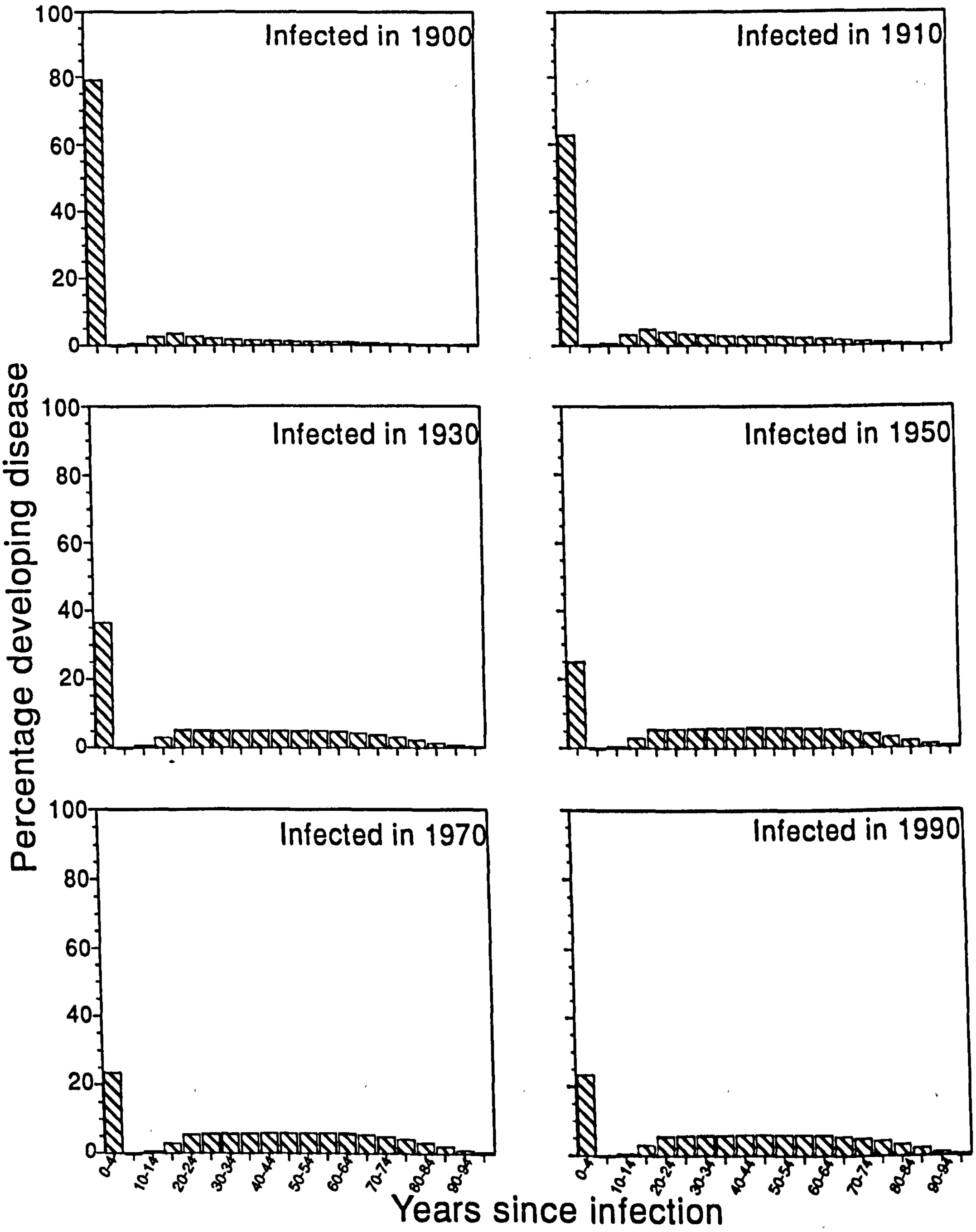


Figure 4.18: Distribution of the time interval between first infection and the first episode of *sputum-positive disease attributable to the initial infection* for individuals infected in their first year of life during the period 1900–1990, as estimated using TBDYN3. This corresponds to a 'true serial interval'.

4.3.3 Effects of simplifications on the distributions obtained

In considering the results, it is important to recognize that TBDYN3 oversimplifies the natural history of tuberculosis in several ways. The effect of each of the main simplifications is discussed below.

4.3.3.1 Effect of age-independent risks of infection

We noted in section 3.3.3 that TBDYN3 oversimplifies the natural history of tuberculosis by assuming that the risk of infection is not age-dependent, whereas in reality, it may increase slightly during adolescence [113]. In practice, it is likely that the magnitude of the increase is too slight to affect substantially the results obtained from the retrospective or prospective analyses of the incubation period and serial interval, or the lifetime risks of developing disease.

In any case, the lifetime risks of developing disease and the prospective analyses of the incubation period and serial interval are based on follow-up of individuals who have already been infected. Hence an age-dependent risk of infection would have affected only the estimated numbers of individuals infected in a given year and at a given age, but not the subsequent disease history of the individuals.

It is also likely that the risk of reinfection increases with age, although the magnitude of the increase has never been determined. The effect of this assumption on the results described in this chapter is likely to be small, given that TBDYN3 probably overestimates the risk of reinfection by assuming that it is identical to that of first infection (see sections 3.1.1 and 3.4.1).

4.3.3.2 Effect of reinfection

TBDYN3 assumes that a disease episode is attributable to the most recent infection or reinfection event, and that individuals cannot be reinfected whilst they are already at risk of developing their first primary episode or exogenous disease (see section 3.1.2).

It is likely that reinfection can occur within five years of a previous infection or reinfection event, although it is probably rare (see section 3.1.1). The effect of such a reinfection event on the risks of developing disease is not known. If it is assumed that the risk of developing the first primary episode increases if individuals are also reinfected, for example, then the

overall lifetime risks of developing disease for those infected before 1950 would have exceeded those derived in this chapter. The extent to which this is realistic is unclear.

4.3.3.3 Sputum-positivity assumptions

TBDYN3 assumes that a fixed, age-dependent proportion of all respiratory disease is sputum-positive, irrespective of whether individuals are experiencing disease for the first or a subsequent time (see section 3.1.3). This is probably unrealistic, particularly for the pre-chemotherapy era. Individuals who have already experienced a sputum-positive episode, for example, may be more likely to experience further sputum-positive disease than those who have never experienced disease at all. This simplifying assumption has probably not substantially affected the results obtained from these analyses, given the relatively small proportion of individuals found to be experiencing their second or subsequent episode (see section 4.3.1.1).

4.3.3.4 Risks of developing endogenous disease

TBDYN3 assumes that the annual risk of developing endogenous disease remains constant for individuals aged over 20 years. In reality, it is likely that this risk increases with age, although the magnitude of this increase is not known. The question is complicated, given that 20 year olds alive during the pre-chemotherapy era, for example, would in general have experienced better general living standards and nutrition when older, which could have led to a *reduction* in the risk of developing endogenous disease. Consequently, it is likely that the results for the retrospective and prospective analyses of the incubation period and serial interval presented here do not differ substantially from those which would have been observed in reality.

4.4 Discussion

The work described in this chapter constitutes the first ever attempts to analyze the lifetime risks of developing respiratory disease. We began by noting that estimates of the incubation period and serial interval could be based from either a 'prospective' or 'retrospective' perspective.

The distributions of the time interval since the infection or reinfection event causing disease among individuals of a given age and in a given year (Figures 4.4, 4.5, and 4.6), or the 'true retrospective' incubation periods and serial intervals, suggest that most of the disease occurring during the early part of this century was attributable to recent transmission. In contrast, most of the disease among the elderly in recent years is attributable to infections or reinfections acquired many years ago. Recently, there has been a growing interest in applying DNA fingerprinting techniques to understanding transmission patterns in populations (see section 1.1). The results from retrospective analyses of the incubation period and serial interval potentially provide some insight into how results from DNA fingerprinting of sputum isolates would have changed over time. A greater proportion of isolates taken from elderly individuals from before 1950, for example, should have been 'clustered', as compared with those taken in recent years.

The fact that the lifetime risk of developing tuberculosis given infection is age-dependent is not widely appreciated. The only published discussion of this question is that of Comstock [121] in 1975, who suggested that the lifetime risk of developing disease following infection was probably *higher* for infants than adolescents, as infants have more life years *available* in which to develop disease. We draw the opposite conclusion using results from TBDYN3, namely that the high risk of developing the first primary episode during adolescence exceed even the cumulative risk of developing disease during a lifetime following infection in infancy (see Figures 4.7 and 4.11).

The lifetime risks of developing respiratory tuberculosis during the pre-chemotherapy era (e.g. 26% following infection at age 20 years) seem high in comparison with the 10% lifetime risk which is often quoted (see e.g. [6]). This 10% lifetime risk is estimated from the disease incidence during 10 years following 'conversion' to tuberculin-positivity among 2,550 adolescents participating in the UK MRC BCG trial during the 1950s [29] (see sections 1.1 and 1.2.1). However, this estimated 'lifetime' risk from the UK MRC BCG trial does

not represent that experienced by individuals alive when the risk of infection was high, e.g. during the early 1900s, given that the risk of infection during the 1950s was already relatively low in England and Wales (see section 2.3).

Whilst the lifetime risks of developing disease estimated for individuals infected until 1950 are high, they are not unrealistic. A 22-year follow-up study of children of sputum-positive cases in Williamson County (Tennessee) Health Department during the period 1931–1954, found risks of developing disease of 19.7% following 'exposure' (as measured by the estimated onset of 'cough' of the case first notified to the County Health Department) for individuals aged over 15 years [48] (see section 1.2.1). These results may well be unreliable, given that they were based on small numbers of cases (e.g. 10 among individuals aged over 15 years followed up after exposure).

Intuitively, the secular decline in the overall lifetime risks of developing either respiratory or sputum-positive disease seems reasonable. The decline estimated using TBDYN3 is attributable to the decline in the risk of infection, which meant that successive generations of individuals were less likely to experience reinfection and subsequent exogenous disease than their predecessors. These lifetime risks could have also declined as a consequence of improvements in general living standards and the nutritional and health status of individuals. This would have meant that once infected, individuals were less likely to develop disease than counterparts in preceding generations. However, the amount by which this actually diminished the lifetime risks is unclear, given that other factors, such as smoking could have also contributed to an increase in the risk of developing disease [122].

The slightly higher lifetime risk of developing sputum-positive disease following infection at age 70 years after 1950, as compared with that among 20 year olds follows from the assumption that the probability that a disease episode is sputum-positive increases with age. This probability was estimated from the relative contribution of sputum-positive disease to age-specific notifications of respiratory tuberculosis in Norway. The reliability of this estimate depends on both the efficiency of the notification system and the rigour with which sputum examinations were carried out. Potentially, only the most severe tuberculosis cases are investigated thoroughly, notified and found to be sputum-positive, and this could have led to an overestimate in the 'true' relative contribution of proportion of sputum-positive disease to the age-specific disease incidence. The extent to which this is true must also depend on the age group and the time period considered.

Given that consistent age-specific patterns in the relative contribution of sputum-positive disease to the respiratory disease incidence were obtained, irrespective of the time period considered, this suggests that the assumption of an age-increasing risk of developing sputum-positive disease is realistic. More reliable data would perhaps have led to slightly different magnitudes in the relative contribution of sputum-positive disease to the respiratory incidence but the overall age-specific patterns would have been similar. This suggests that the age-specific patterns in the lifetime risks of developing sputum-positive disease would not have changed substantially from those presented here.

The results in this chapter illustrate that the distribution of the time interval until the first respiratory or sputum-positive episode, and that of the time interval since initial infection among diseased individuals (or the observed incubation periods and serial intervals among infected or diseased individuals respectively) depend on the trend and level of the risk of infection and on age. This means the distributions derived here differ from those applicable in developing countries, where the risk of infection is high and has changed little during the past few decades. The issue is even more complicated, given that the magnitude of the age-specific risks of developing the first primary, endogenous and exogenous disease may differ from those derived for England and Wales (see section 3.4.1).

In reality, the serial interval of any infection depends on many factors, including the time interval between onset of infection and infectiousness, the duration of infectiousness, the degree of infectiousness and amount of contact between infectious and uninfected individuals. As we have seen in this chapter, the definitions of the serial interval, and the incubation period are particularly complicated for tuberculosis, given that individuals can develop disease following reinfection, and the risks of developing disease are age dependent. These complications must also arise for other infectious diseases, for example, herpes varicella-zoster and leprosy, in which individuals may be able to develop disease following reinfection.

The results for the distributions of the time interval between initial infection and the first sputum-positive episode have implications for the net reproduction numbers. The high risk of infection until 1950 in England and Wales, for example, implied that most of the new infections in these years were among infants. As shown in Figure 4.18, about 20% of individuals infected in their first year of life in 1900, who developed disease as a consequence, did not manifest disease until about 20 years thereafter. Thus overall, these

individuals comprised a long link in the chain of transmission, which must have influenced the value for the net reproduction number. This is discussed in detail in the next chapter.

Retrospective analyses provide more insight into the dynamics underlying the decline in tuberculosis than do prospective analyses, as they reflect the 'true' and 'observed' incubation periods and serial intervals that diseased individuals actually experienced. Prospective analyses of the incubation period provide some insight into the incubation that would have been experienced given infection at specific ages in given years.

Overall, the results obtained seem intuitively reasonable, in spite of the simplifying assumptions incorporated into TBDYN3, and it is likely that the actual lifetime risks of developing sputum-positive and respiratory tuberculosis and the incubation period and serial intervals behaved in a similar way to that described here.

4.5 Brief review of key findings

For reference, we briefly summarize the key findings in this chapter below.

1. The incubation period and serial interval for tuberculosis can be analyzed 'retrospectively' (as the time interval since initial infection for individuals diseased at a given age and in a given year) and 'prospectively' (as the time interval until the first disease episode for individuals infected simultaneously at a given age and in a given year).

2. *Retrospective* analyses suggest that about 3% of diseased 20 year olds *in 1900* had first been infected during the preceding year, and the greatest proportion ($\approx$ 25%) had first been infected in their first year of life. The proportions for diseased 20 year olds *in 1990* are 38% and 6% respectively.

3. *Prospective* analyses of the *incubation period* suggest that about 25% of infants infected in 1900 who developed respiratory disease, did so during the first five years, and the proportion doing so thereafter declined with time since initial infection.

4. *Prospective* analyses of the *serial interval* (by 5 year periods) indicate that the proportion of individuals infected in the first year of life *in 1900* who developed sputum-positive disease sometime thereafter, was maximum between the 20th and 24th years after initial infection ($\approx$ 15%).

5. Twenty year olds faced higher lifetime risks of *respiratory* disease following infection than did any other age group (e.g. 26% and 17% in 1900 and after 1950). The estimated lifetime risks of developing *sputum-positive* disease increased consistently with the age at infection and were about 19% and 12% for 20 year olds infected in 1900 and after 1950 respectively.

Chapter 5

The basic and net reproduction numbers for tuberculosis

5.1 Introduction

5.1.1 Relevance and definitions of the net and basic reproduction numbers for infectious diseases

Much of the thinking behind the basic and net reproduction numbers for infectious diseases stems from the work of Ross on malaria in 1910 [123]. Using a simple model, Ross illustrated that, by maintaining the density of mosquitoes below a critical value for a given human population, malaria would disappear from that area. Macdonald extended this thinking during the 1950s [124, 125], showing that if, on average, each infectious case (of malaria) leads to less than one other case in a given population, then transmission will eventually cease, and the infection will disappear from the area. Macdonald referred to the number of secondary cases resulting from an infectious individual as the '*reproduction rate*', or the '*net effective reproduction rate*', which depended, among other things, on the proportion of 'susceptible' individuals in the population. Macdonald formally defined the '*basic reproduction rate*' as the limit of the 'net effective reproduction rate' as the proportion of 'susceptible' individuals in the population approached 100%, and described it as

> 'the potential reproductive capacity of a totally non-immune and untreated case surrounded by a non-immune population' [125].

For complete control of the infection in an area, the 'net effective reproduction rate' had to remain below one.

For many infectious diseases, transmission of infection and hence the 'reproductive capacity' of an individual case depends on age and mixing patterns in the population. For such diseases, the basic reproduction rate (now often referred to as the 'basic reproduction number') is defined as:

> 'the average number of secondary [infectious] cases resulting from the introduction of a *typical* infectious individual during its entire infectious period, in a [totally] susceptible population' [50,51][1].

The 'typical' infectious case in this definition refers to a hypothetical individual who is *representative* of the different age groups and subgroups in the population. This implies that if the basic reproduction number is greater than one, there is a high likelihood of continued transmission following the introduction of *any* infectious individual with *some*[2] degree of contact with other individuals into a 'totally susceptible' population, and the infection does not persist otherwise. This, of course, assumes that there is no additional control or intervention subsequent to the introduction of the infectious case into the population. An intervention such as chemotherapy, or one which changes the behaviour of individuals, changes important epidemiological parameters, such as the duration of infectiousness, or the contact between individuals, and must therefore change the magnitude of the basic reproduction number, and hence the likelihood of continued transmission [126].

The net effective reproduction rate (now often referred to as the 'net reproduction number') can be defined in an analogous way to the basic reproduction number, as:

> 'the average number of secondary infectious cases resulting from a *typical* infectious case during its entire infectious period, in a given population under any given conditions.'

For any immunizing infectious disease, the net reproduction number is initially approximately equal to the basic reproduction number following the introduction of a *typical*

[1]Note that the definitions given in these references use the word 'infected' rather than 'infectious'. The word 'infectious' is sometimes more appropriate in defining the basic reproduction number, given that for some diseases (e.g. tuberculosis), not all infected individuals develop infectious forms of disease.

[2] Note that this excludes cases who have *no* contact with other individuals.

infectious case into a 'totally susceptible' population. By definition, the net reproduction number has to change as the prevalence of 'immune' individuals changes in a population.

For diseases such as measles, mumps and rubella, the relationship between the basic and net reproduction numbers in terms of the prevalence of immune individuals provides a useful means of estimating the minimum effective vaccine coverage required in a given population for eradication. If a proportion p of individuals in a population are immune, for example, leaving a proportion $1 - p$ 'susceptible' to infection, then the net and basic reproduction numbers (R_n and R_0 respectively) are related as follows [119]:

$$R_n = R_0(1 - p). \tag{5.1}$$

The infection does not persist when the net reproduction number is less than one, and thus when:

$$R_n = R_0(1 - p) < 1 \tag{5.2}$$

Rearranging this expression, the effective vaccine coverage (p) has to satisfy:

$$p > 1 - \frac{1}{R_0} \tag{5.3}$$

for eradication of the infection.

The derivation of the basic and net reproduction numbers for diseases in which age and mixing patterns determine transmission of infection is often not straightforward. In theory, *individual* age or subgroup-specific 'basic reproduction numbers' can be obtained for all possible age groups or population subgroups. These are defined as the expected number of cases produced by an individual of a given age-group or population subgroup following his/her introduction into a 'totally susceptible' population. Analogously, individual age or subgroup-specific 'net reproduction numbers' can also be defined. The overall basic reproduction number can be interpreted as an overall average of these individual age or subgroup-specific basic reproduction numbers.

These individual age or subgroup-specific basic reproduction numbers are rarely calculated in practice, and more commonly, the basic reproduction number for a given disease is estimated by considering the number of cases resulting from each *generation* of infectious cases, shortly after the introduction of an infectious individual into a 'totally susceptible' population [50,51][3]. The derivation of the basic reproduction number by this method relies

[3]More formally, the basic reproduction number is often defined as the 'dominant eigenvalue of the next-generation matrix' [50, 51].

on assumptions relating to the density of infectious cases in the population shortly after introduction of the infectious case.

5.1.2 Definitions of the basic and net reproduction numbers for tuberculosis

We begin our discussion of the basic and net reproduction numbers for tuberculosis by clarifying the definition of a **'totally susceptible population'**.

For many infectious diseases, individuals who have already experienced prior infection with or without disease are generally considered to be fully protected against subsequent infection. For these diseases, a 'totally susceptible population' is therefore one in which no individual has hitherto experienced infection or effective vaccination. For tuberculosis, both uninfected and infected individuals can be considered susceptible — uninfected individuals may be infected and develop disease attributable to this initial infection and infected individuals may be reinfected and develop disease attributable to the reinfection event. Given that the risk of developing disease subsequent to *reinfection* differs from that following *initial infection*[4], a **'totally susceptible population'**, as used in the definition of the basic reproduction number is best defined for tuberculosis as one in which no individual has hitherto experienced infection or disease.

On this basis, and given that pulmonary forms comprise the vast majority of infectious tuberculosis cases, the definition of the basic reproduction number given on page 188 might best be rephrased in the following way to be applicable to tuberculosis:

> **The average number of secondary infectious (pulmonary) cases resulting from a *typical* infectious (pulmonary) case during its entire infectious period, following its introduction into a population in which no individual has hitherto been infected with the tubercle bacillus.**

Similarly, we propose the following definition for the net reproduction number for tuberculosis:

> **The average number of secondary infectious (pulmonary) cases resulting from infection or reinfection by a *typical* infectious (pulmonary)**

[4]It is also likely that the risk of reinfection differs from that of initial infection — see section 3.4.1.

case during its entire infectious period in a given population under any given conditions.'

In the following discussion, we shall refer to the infectious pulmonary tuberculosis case introduced into an uninfected population as the **founder case.** Similarly, we shall refer to the disease episode experienced by the founder case when introduced into the population as the **founding disease episode.**

The definition of a **'typical' infectious case** for tuberculosis has to take into account the fact that individuals can be reinfected or reactivate several times and thus have several episodes of infectiousness. These disease episodes can be viewed as 'continuations' of preceding disease episodes if *they are attributable to the same infection/reinfection event* i.e. if the case experiences no (further) reinfection in the meantime. The number of infectious pulmonary disease episodes experienced by an individual which are attributable to a given infection or reinfection depends mainly on two factors:

1. the risk of subsequent (re)infection, and

2. the age at (re)infection. This determines both

   (a) the number of years of life available in which to develop disease (and/or experience reinfection), and

   (b) the likelihood of the disease episode being sputum-positive. Less than 10% of respiratory disease among infants and children is sputum-positive, for example, as compared with over 50% of that in adults (see section 1.1).

For these reasons, the definition of a 'typical infectious pulmonary tuberculosis case', as used in the definitions of the basic and net reproduction numbers above, should account for all possible ages of founder cases, and the risk of (re)infection subsequent to his/her introduction into an uninfected population.

These factors also determine the **number of secondary infectious cases** resulting from a given case. If a *founder case* is introduced into an uninfected population, for example, is *reinfected and subsequently experiences disease* several years after recovering from the founding episode, then the number of secondary cases arising from the second disease episode should *not* contribute to estimates of the basic reproduction number. The converse should hold if the founder case reactivates several years after recovering from the founding episode

without having been reinfected in the meantime. Similarly, if a *secondary case* is *reinfected and subsequently experiences disease* after recovering from disease attributable to infection by the founder case, then the subsequent episodes should not contribute to estimates of the basic reproduction number. The converse holds if a secondary case experiences further disease without having been reinfected.

For these reasons, the **number of secondary cases** resulting from a given founder case must depend on the age of the individuals infected, and therefore on the age structure of the population, and on the risk of (re)infection experienced.

The fact that cases who have been reinfected by the founder case can be defined as 'secondary cases' also leads to the paradoxical result that, if a given founder case reactivates many years after being introduced into an uninfected population, then the *basic reproduction number* would also have to depend on the *net reproduction number* in the year in which the reactivation occurs.

In practice, basic reproduction numbers for infectious diseases are usually estimated by assuming that the population is in a stable demographic state and that all disease parameters (e.g. those defining the contact between individuals) are constant over the time period considered. For diseases with short incubation periods and serial intervals (i.e. measured in days), occurring in developed countries, these are reasonable assumptions. For tuberculosis, the potentially long incubation period and serial intervals, depending on both the age of individuals and the calendar year, mean that demographic and epidemiologic changes in the population will affect the number of cases attributable to a founder case, and hence the basic reproduction number. Because of this, we are led to define two variants of the basic reproduction number concept to take these temporal changes into account. This argument is presented in the following section.

5.2 Method

5.2.1 Derivation of an effective contact number

Following the discussion in the last section, TBDYN3 was used first to estimate the average number of individuals 'effectively contacted' by each sputum-positive case in a particular year, $c(t)$, and to monitor its change over time. We refer to this measure as the **'effective contact number'**. An 'effective contact' between a sputum-positive and any other individual is here defined, as by Frost [127], as one sufficient to lead to infection if the contacted individual had never been infected in the past.

Considering the population in England and Wales modelled by TBDYN3, individuals contacted by a sputum-positive case at a given time t can belong to one of three categories:

1. Those who have not hitherto been infected ($U(t)$). By definition, an effective contact between a sputum-positive case and such an individual must lead to infection.

2. Those in the 'latent' class ($L(t)$). TBDYN3 assumes that the risk of reinfection among these individuals is identical to that of first infection among uninfected individuals.

3. All other individuals ($O(t)$), who are either experiencing disease, or who have been recently infected/reinfected and who are at an already high risk of developing disease attributable to this infection/reinfection. We have assumed that these individuals cannot be reinfected (see section 3.1.2), and hence that an effective contact between a sputum-positive case and these individuals does not lead to transmission of tubercle bacilli.

We denote the average number of individuals in each of the above classes effectively contacted by a sputum-positive case at time t by $c_i(t), c_r(t)$, and $c_o(t)$ respectively.

Assuming that individuals mix randomly in the population, then the *total* number of individuals effectively contacted at a given time t by each sputum-positive case is independent of the composition of uninfected, 'latent' and all other individuals in the population. Hence the effective contact number for a given time t, $c(t)$, is given by the following expression:

$$c(t) = c_i(t) + c_r(t) + c_o(t). \tag{5.4}$$

The annual risk of infection ($i(t)$) is defined as the proportion of uninfected individuals who become infected in a given year t. Given $U(t)$ uninfected individuals present in the

population in a given year t, then the number of new infections is given by $i(t)U(t)$. If there are $D_+(t)$ sputum-positive cases present, then, on average, the total number of individuals infected (i.e. effectively contacted) by each sputum-positive case ($c_i(t)$) is given by:

$$c_i(t) = \frac{U(t)i(t)}{D_+(t)}. \tag{5.5}$$

Analogous expressions hold for the number of individuals in the 'latent' and 'other' classes contacted by each sputum-positive case.

The assumption of homogeneous mixing means that the probability that a sputum-positive case has effective contact with individuals in the uninfected, or any other class, depends on their relative size in the population. Given $U(t)$ uninfected individuals in a population of size $N(t)$ at time t, a proportion $\frac{U(t)}{N(t)}$ of all those effectively contacted by each sputum-positive case will be uninfected. Hence $c_i(t)$ is given by

$$c_i(t) = c(t)\frac{U(t)}{N(t)} \tag{5.6}$$

Equating this expression for $c_i(t)$ with that given above (equation 5.5) yields:

$$c_i(t) = c(t)\frac{U(t)}{N(t)} = \frac{U(t)i(t)}{D_+(t)} \tag{5.7}$$

Cancelling out $U(t)$ gives:

$$\frac{c(t)}{N(t)} = \frac{i(t)}{D_+(t)} \tag{5.8}$$

and then rearranging this expression implies:

$$\begin{aligned} c(t) &= \frac{i(t)}{D_+(t)}N(t) \\ &= \frac{i(t)}{\mathbf{D_+(t)}} \end{aligned} \tag{5.9}$$

where $\mathbf{D_+(t)}$ denotes the prevalence *rate* of sputum-positive cases at time t. Hence the effective contact number in a given year t is given by the ratio between the annual risk of infection and the prevalence of sputum-positive cases. This is analogous to the 'contagious parameter' defined by Styblo [52], which we described in section 1.2.2.

Figure 5.1 summarizes the crude incidence and prevalence of sputum-positive cases during the period 1900–1990, as estimated using TBDYN3. This shows that at least until 1950, the estimated prevalence of sputum-positive cases exceeded the incidence by a factor of a little over 1.5 (i.e. 1.65). This results from the assumption in TBDYN3 that individuals

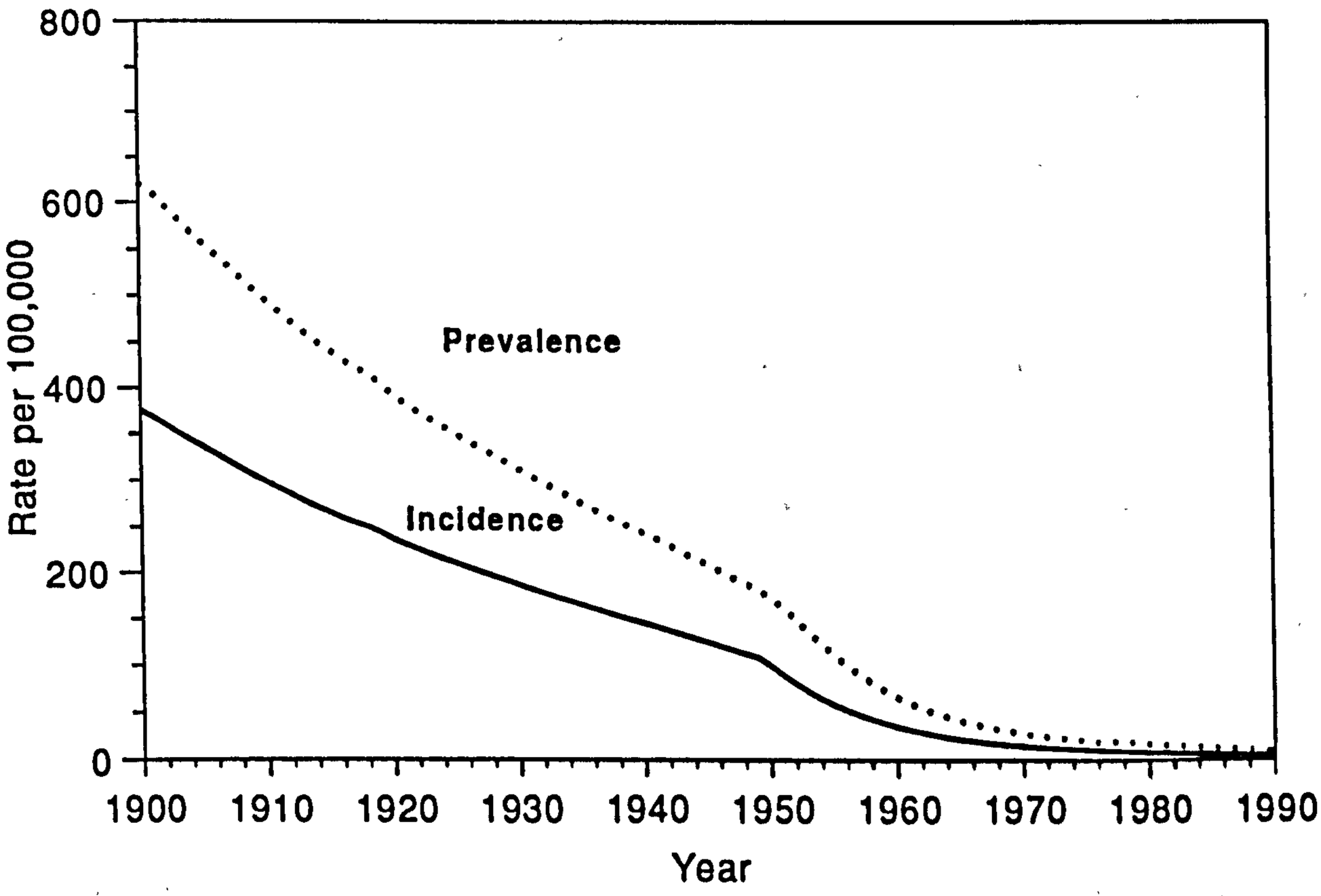


Figure 5.1: Crude incidence and prevalence of sputum-positive tuberculosis per 100,000 males aged 0–100 years in England and Wales during the period 1901–1990, as estimated using TBDYN3.

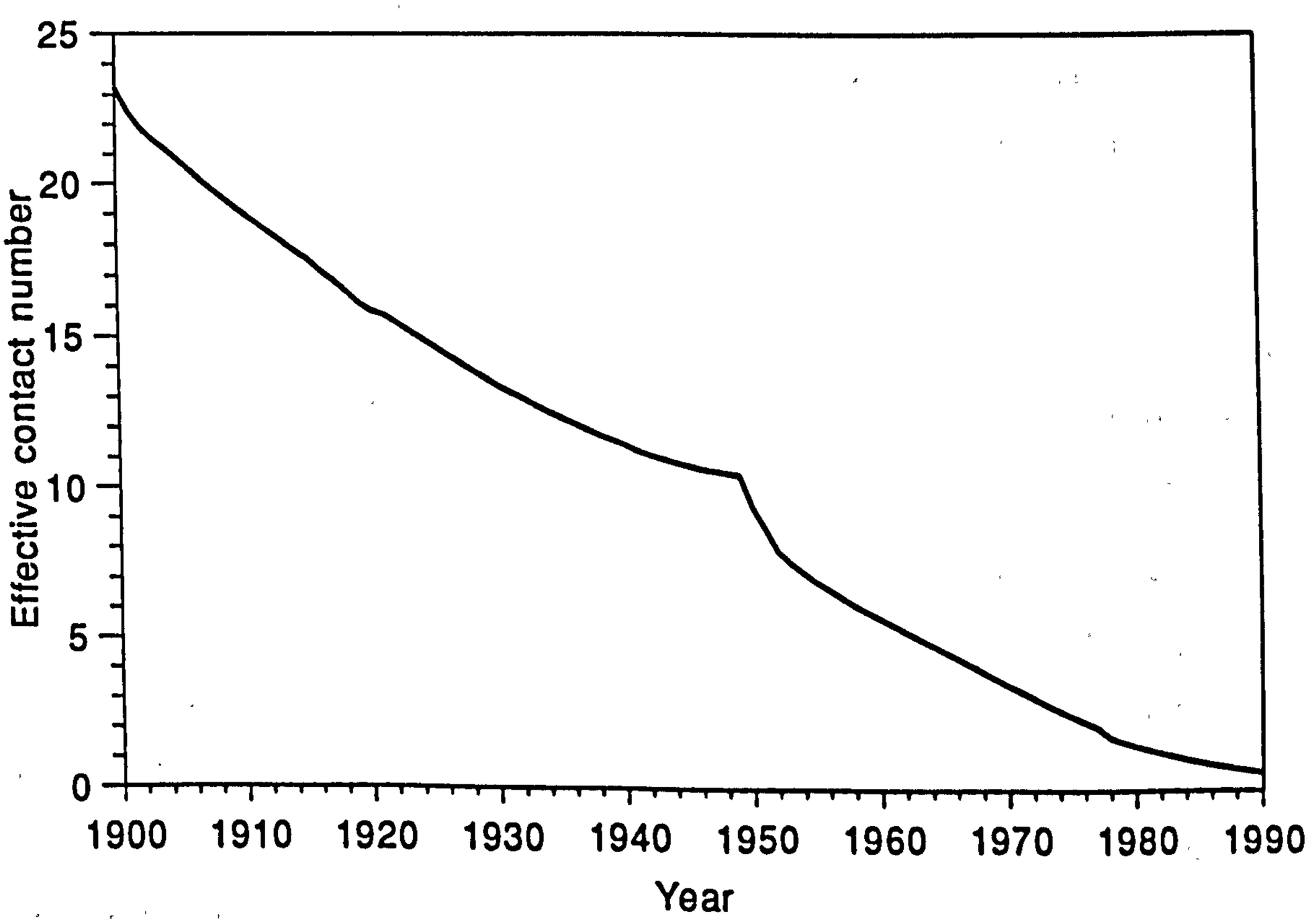


Figure 5.2: Annual effective contact number between sputum-positive and other individuals in the population in England and Wales during the period 1901–1990, as estimated using TBDYN3.

are sputum-positive for 2 years, and have a case fatality rate of 35% during the first year after disease onset during the pre-chemotherapy era. The relationship between the estimated sputum-positive disease incidence and prevalence changes after 1950, when we assume that the case-fatality decreased (see section 3.1.3).

Figure 5.1 also shows that in 1900, there were about 620 sputum-positive cases for every 100,000 individuals in the male population. The estimated annual risk of infection in 1900 was about 14,500 per 100,000 uninfected individuals (14.5%) (see section 3.2.1), and hence, on average, each sputum-positive case effectively contacted approximately 23 ($\approx$ 14,500/620) individuals. Figure 5.2 summarizes the effective annual contact numbers derived in the same way for the period 1900–1990. This shows that the estimated effective contact number declined steadily to about 10 by 1949, declined more rapidly for the subsequent three years, and then declined at about the pre-1950 rate thereafter.

5.2.2 Analyses of the basic and 'founder case' reproduction numbers

5.2.2.1 Definition of the founder case reproduction number

The secular decline in the effective contact number for tuberculosis is itself attributable to the effect of many 'interventions' (e.g. the reduction of crowding in living conditions, removal of infectious cases from the community to sanatoria (before 1950) or through treatment after 1950, and general improvements in hygiene — see section 5.4).

As we saw in section 5.1.1 the basic reproduction number for an infectious disease is interpreted as a parameter which describes the likelihood of continued transmission *in the absence of additional control or intervention.* On this basis, the definition of the basic reproduction number for tuberculosis given on page 190 is more appropriately rephrased as:

> **The average number of secondary infectious (pulmonary) cases resulting from a 'typical' infectious (pulmonary) case during its entire infectious period, following its introduction into a population where no individual has hitherto been infected with the tubercle bacillus, *assuming that environmental conditions and the effective contact number do not change further over time.***

Henceforth in this thesis, this definition will be used for the basic reproduction number for tuberculosis. It is important to recognize that, given this definition, the magnitude

of the basic reproduction number for tuberculosis in a given year does not necessarily reflect the likely trend in disease incidence subsequent to the introduction of a founder case into an uninfected population, as it does not take into account changes in epidemiological parameters, which actually did occur. The basic reproduction number for tuberculosis for a given year during the pre-chemotherapy era, for example, does not even take into account the fact that chemotherapy was introduced during the 1950s.

On this basis, a variant of the basic reproduction number, which is a function of future changes in the effective contact number and the introduction of chemotherapy, might be more appropriate for describing the likely trend in disease incidence subsequent to the introduction of a founder case into an uninfected population. In the following discussion, we refer to this parameter as the **founder case reproduction number**, and define it formally as:

> **The average number of secondary infectious (pulmonary) cases resulting from a 'typical' infectious (pulmonary) case during its entire infectious period, following its introduction into a population where no individual has hitherto been infected with the tubercle bacillus,** ***allowing for changes in the effective contact number.***

By definition, the magnitude of the founder case reproduction number changes over time as a function of changes in environmental and epidemiological parameters, and these changes determine the trend in disease incidence[5].

We describe the methods used to analyze the basic and founder case reproduction numbers for tuberculosis below.

5.2.2.2 Transmission dynamics in a 'totally susceptible' population

The potentially long time interval between infection and onset of infectious pulmonary disease and the fact that individuals can experience reinfection mean that it is difficult to

[5]We are aware of no published derivation of a parameter which describes the long-term persistence of an infectious disease for which environmental conditions or epidemiological parameters change continuously over time. Analogous issues have, however, been raised in the ecological modelling literature. Metz *et al* [128], for example, discuss the definition of a parameter describing the likelihood of long-term survival of a given species if mutations occur following changes to the environment. Metz *et al* define this parameter as the dominant Lyapunov exponent of a given matrix process. The parameter is negative if the population is destined for extinction, and is positive otherwise.

obtain analytic expressions for the basic and founder case reproduction numbers. On this basis, we estimated these parameters for a given time t by direct simulation.

To analyze the founder case reproduction number, we simulated the introduction of a founder case who was infectious for duration i_f into a population of uninfected individuals, which was identical to that in England and Wales at time t and in subsequent years with respect to its:

1. Age structure,

2. Age and time specific mortality rate,

3. Annual effective 'contact number' $c(t)$,

and tracking the individuals infected by the founder case, and summing the number who ultimately developed disease attributable to the initial infection ($F_{0i_f}(t)$). The basic reproduction number was analyzed in an analogous way, i.e. by deriving the number who ultimately developed disease attributable to the initial infection ($R_{0i_f}(t)$), assuming that the effective contact number and the case-fatality rate did not change subsequent to the introduction of the founder case. For simplicity, the founder case was assumed to be infectious for either one or two years. Given that we intend only to explore the implications of a declining effective contact number on the disease dynamics, individuals are not assumed to be vaccinated at any time. We related the magnitude of $F_{0i_f}(t)$ and $R_{0i_f}(t)$ for different values of i_f to the trend in disease incidence subsequent to the introduction of given founder cases in these simulations.

The statistic $F_{0i_f}(t)$ can be expressed in terms of the effective contact number as follows:

Suppose that a founder case is introduced into a population of uninfected individuals of size $N(t)$ in year t and effectively contacts $c(a,t)$ individuals of age a. Assuming homogeneous mixing (see page 99), and that there are $u(a,t)$ uninfected individuals of age a at time t, then of all the individuals effectively contacted, a proportion $\frac{u(a,t)}{N(t)}$ will be of age a. $c(a,t)$ is therefore given by

$$c(a,t) = c(t)\frac{u(a,t)}{N(t)} \tag{5.10}$$

Suppose that a proportion $p(a,t)$ of those infected at age a at time t develop sputum-positive disease *attributable to this infection episode* (i.e. without having been reinfected in the meantime). If the sputum-positive case is infectious for duration i_f, then out of all

those infected at age a during the infectious period (between time t and $t + i_f$), a total of

$$\sum_{j}^{i_f} c(a, t+j)p(a, t+j) \tag{5.11}$$

ultimately develop disease attributable to the initial infection by the founder case.

Hence summing over all the possible ages at infection, the number of secondary infectious cases resulting from a founder case introduced in year t who was infectious for duration i_f, $F_{0_{i_f}}(t)$, is given by

$$F_{0_{i_f}}(t) = \sum_{j}^{i_f} \sum_{a} c(a, t+j)p(a, t+j) \tag{5.12}$$

Analogous expressions hold for $R_{0_{i_f}}(t)$. We describe the extensions required to TBDYN3 to analyze the basic and founder case reproduction numbers below.

5.2.2.3 Extensions to TBDYN3

Hitherto, TBDYN3 has tracked the disease incidence in a population only within cohorts of individuals, i.e. among individuals of the same age. The approach used to analyze the basic and founder case reproduction numbers for tuberculosis was analogous, except that all individuals alive *in a given year* (and born thereafter) were followed up simultaneously. The system of equations describing the disease dynamics in these individuals is identical to that given in Appendix A.1, except that

1. we now assume that only 1 sputum-positive case was present at the start of the simulations in a given year, and

2. the risk of infection at a given time t is given by the product of the prevalence of sputum-positive cases at that time, and the effective contact number, as follows:

$$i(t) = c(t)\mathrm{D}_+(\mathrm{t}) \tag{5.13}$$

This follows from rearranging equation 5.9.

The whole system of equations was programmed in 'C' using forward Euler differencing [114] with time steps of 1 year in both age and time (i.e. secular time, and time since infection/disease)[6]. This required identical input data on the birth and the age and time-specific mortality rates to those used in the work described in the last two chapters.

[6]Note that individuals can be (re)infected and develop disease in the same year — see section 3.1.2. To avoid the problem of individuals contributing to the risk of infection in the same year as that in which they

To establish the number of cases resulting from a founder case, infected individuals have to be followed up over their entire lives. Thus individuals infected by a sputum-positive case when aged 20 years in 1950, for example, would have to be followed up until they were aged 100 years (i.e. a lifetime, as considered by TBDYN3) in the year 2030. The proportion of these individuals who developed disease without further reinfection after 1990, depends on the assumptions made about the subsequent trend in the effective contact number and in demography. For simplicity, it is assumed that the effective contact number does not change after 1990, and that the number of births and age and time-specific mortality rates for individuals born after 1990 remain at the 1990 levels.

5.2.3 Analyses of the net reproduction number

It is difficult to obtain analytic expressions for the net reproduction number for tuberculosis for a given year, given the potentially long time interval between infection and onset of infectious pulmonary disease and the fact that individuals can experience reinfection.

Given these complexities, we first describe the relationship between the net reproduction number, the effective contact number, and the lifetime risk of developing sputum-positive disease attributable to a given infection/reinfection.

Suppose that at time t each sputum-positive individual effectively contacts $c_i(a,t)$ individuals of age a in the uninfected class, and a proportion $p_i(a,t)$ of these ultimately develop sputum-positive disease attributable to their infection event. If the sputum-positive case is infectious for duration i_f, then out of all those infected during the infectious period (between time t and time $t+i_f$), a total of

$$\sum_{j}^{i_f} c_i(a,t+j)p_i(a,t+j) \tag{5.14}$$

will ultimately develop disease attributable to their initial infection. By an analogous argument, a total of

$$\sum_{j}^{i_f} c_r(a,t+j)p_r(a,t+j) \tag{5.15}$$

develop disease attributable to reinfection at age a during the infectious period of the sputum-positive case, where $c_r(a,t)$ denotes the number of individuals in the 'latent' class

are (re)infected (i.e. potentially (re)infecting themselves!), we set the risk of infection in year t to be the product of the contact rate in year t, and the prevalence of sputum-positive cases in year $t-1$.

effectively contacted by each sputum-positive case at age a and time t, and $p_r(a, t+j)$ denotes the proportion of these individuals who ultimately develop sputum-positive disease attributable to this reinfection event.

Summing these two components over all the possible ages at infection and reinfection, the number of secondary cases from a sputum-positive case who develops disease at time t and is infectious for duration i_f, $R_{n_{i_f}}(t)$, is given by

$$R_{n_{i_f}}(t) = \sum_j \sum_a^{i_f} \{c_i(a, t+j)p_i(a, t+j) + c_r(a, t+j)p_r(a, t+j)\} \qquad (5.16)$$

The *average number* of secondary cases resulting from individuals who developed disease at time t depends on the proportion surviving from year to year. As we saw in section 5.1.2, this depends on age, and is further complicated by the fact that individuals can be reinfected.

We obtained a rough estimate of the net reproduction number for individuals who developed sputum-positive tuberculosis at time t directly from TBDYN3 by considering the ratio between:

1. the total number of individuals who were infected or reinfected in a given year t, who ultimately developed disease attributable to this infection/reinfection event ($D_u(t)$), and

2. the total number of sputum-positive cases at time t ($D_+(t)$).

The relationship between these two statistics is described below.

Suppose that in year t there are $D_n(t)$ new sputum-positive cases. Given a total prevalence of sputum-positive cases in year t of $D_+(t)$, a proportion $\frac{D_n(t)}{D_+(t)}$ of the new infections and reinfections occurring in that year will be attributable to individuals who developed sputum-positive disease in year t.

Given that a total of $D_u(t)$ individuals ultimately develop sputum-positive disease *attributable* to their infection/reinfection event in year t, then

$$\frac{D_n(t)}{D_+(t)} . D_u(t) \qquad (5.17)$$

secondary cases result from the first year of infectiousness of individuals who developed sputum-positive disease in year t.

By an analogous argument, assuming that $D_s(t+1)$ of those who developed sputum-positive disease in year t *survive* until the subsequent year, $t+1$, then

$$\frac{D_s(t+1)}{D_+(t+1)}.D_u(t+1) \tag{5.18}$$

secondary cases result from the second year of infectiousness of these individuals.

TBDYN3 assumes that sputum-positive cases are infectious for 2 years, unless they die in the meantime (see section 3.1.3). Hence, summing these two components,

$$\frac{D_n(t)}{D_+(t)}.D_u(t) + \frac{D_s(t+1)}{D_+(t+1)}.D_u(t+1) \tag{5.19}$$

secondary cases result from individuals who developed sputum-positive disease in year t

Dividing through by $D_n(t)$, the average number of secondary cases attributable to individuals who developed sputum-positive disease in year t ($R_n(t)$) is given by

$$R_n(t) = \frac{1}{D_n(t)}\left\{\frac{D_n(t)}{D_+(t)}.D_u(t) + \frac{D_s(t+1)}{D_+(t+1)}.D_u(t+1)\right\} \tag{5.20}$$

This expression simplifies to:

$$R_n(t) = \frac{D_u(t)}{D_+(t)} + \frac{D_s(t+1)}{D_n(t)}.\frac{D_u(t+1)}{D_+(t+1)} \tag{5.21}$$

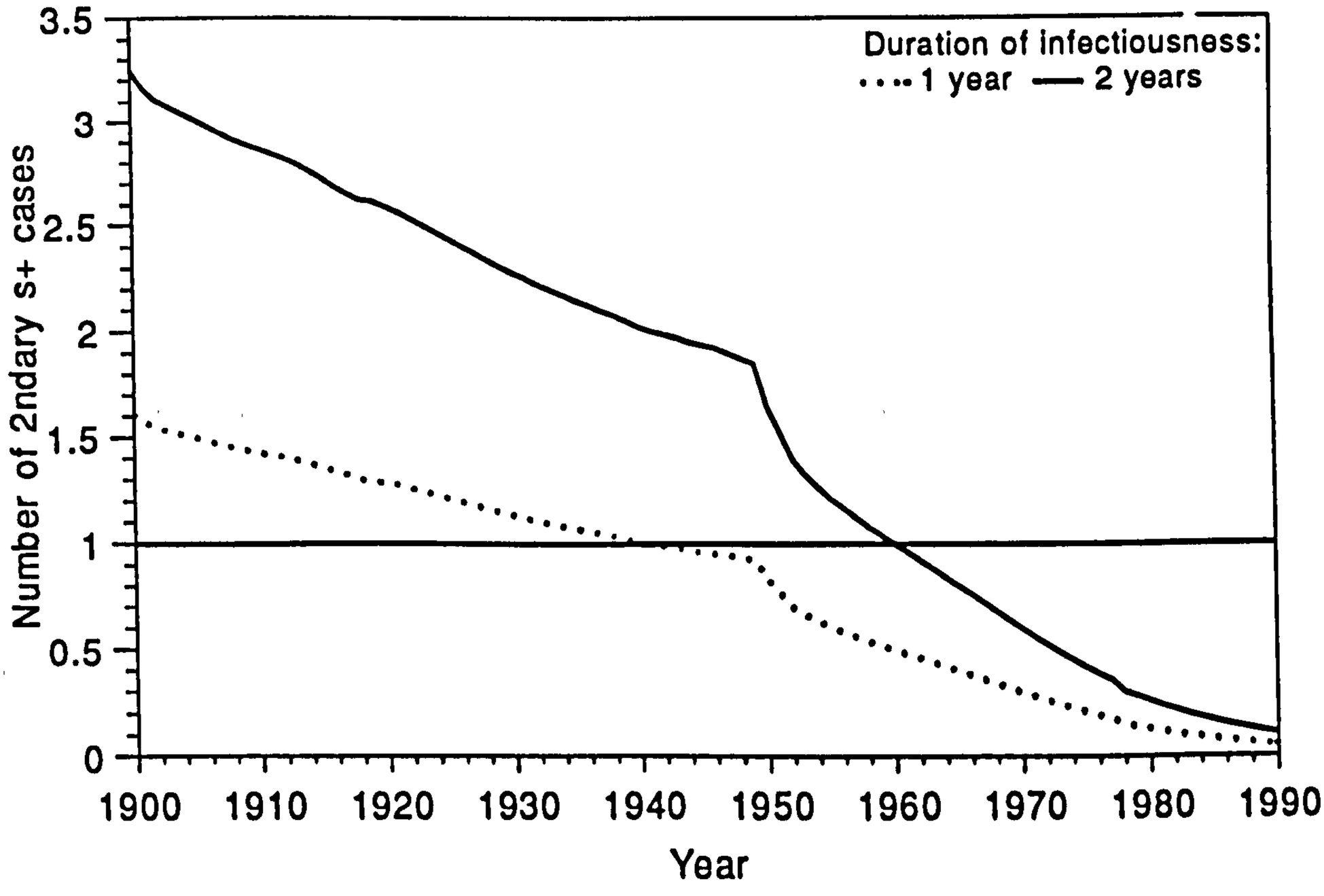


Figure 5.3: Number of secondary sputum-positive cases resulting from a founder case who was infectious for one or two years, introduced into an uninfected population similar in structure to that in England and Wales during the period 1900–1990, as estimated using TBDYN3.

5.3 Results

5.3.1 Analyses of the basic and founder case reproduction numbers

5.3.1.1 The founder case reproduction number

Figure 5.3 summarizes the number of secondary sputum-positive cases attributable to infection by a founder case who was introduced between 1900 and 1990 into a population similar to that in England and Wales, and was infectious for one or two years, assuming that the effective contact number declined subsequent to his/her introduction. This shows that the founder case who was infectious only for one year led to about 1.6 secondary cases if introduced in 1900, and first led to less than one other case if he/she was introduced after 1940.

In contrast, the founder case who was infectious for two years led to about 3.2 secondary cases if introduced in 1900, and led to less than one other case if introduced after 1960. We discuss how closely a founder case who was infectious for one or two years approximates a 'typical' founder case, and hence whether the number of secondary cases resulting from

these individuals approximates the founder case reproduction number below.

TBDYN3 assumes that sputum-positive cases are infectious for two years, unless they die in the meantime. During the pre-chemotherapy era, the case-fatality rate is assumed to be 35% during the first year after disease onset. Consequently, only 65% of sputum-positive cases are assumed to be infectious for 2 years during this time. Assuming that a negligible proportion of cases also develop further sputum-positive disease at some stage thereafter (which is not always realistic), then this implies that the number of secondary cases predicted to result from a founder case who is infectious for two years slightly *overestimates* the founder case reproduction number during the pre-chemotherapy era.

The case-fatality rate is assumed to decline after 1950, and is the same as the mortality rate in the general population after 1976 (see section 3.1.3). Hence a large (age-dependent) proportion of individuals who develop sputum-positive disease are assumed to be infectious for a full two years. Applying reasoning similar to that above, we can conclude that the number of secondary cases resulting from a founder case who is infectious for 2 years closely *approximates* the founder case reproduction number after 1950. Hence the founder case reproduction number probably first fell below one in about 1960.

Henceforth in this thesis, we use *the number of secondary cases resulting from an individual who is infectious for the full two years* to reflect the magnitude of the founder case reproduction number for tuberculosis. We discuss the implications of the magnitude of the founder case reproduction number in a given year for the disease dynamics subsequent to the introduction of a founder case into an uninfected population below.

Figure 5.4 summarizes the total expected number of cases occurring after the introduction of a founder case, who was infectious for two years into an uninfected population identical to that in England and Wales in terms of its demography and subsequent trend in the effective contact number in 1900, 1920, 1940 and 1960. Figure 5.5 shows these statistics on a log scale to facilitate comparison of their rates of increase for these years.

Figure 5.4 shows that, following the introduction of the founder case in 1900, the total number of sputum-positive cases increased for the first 40 years to about 200,000 in around 1940 and then started to decrease. The introduction of the founder case in 1920 and 1940, led to considerably lower peak numbers of sputum-positive cases, namely 4,000 and 13 respectively. Figure 5.5 shows that the incidence of sputum-positive cases increased more rapidly for the epidemic initiated in 1900, as compared with that for epidemics initiated in

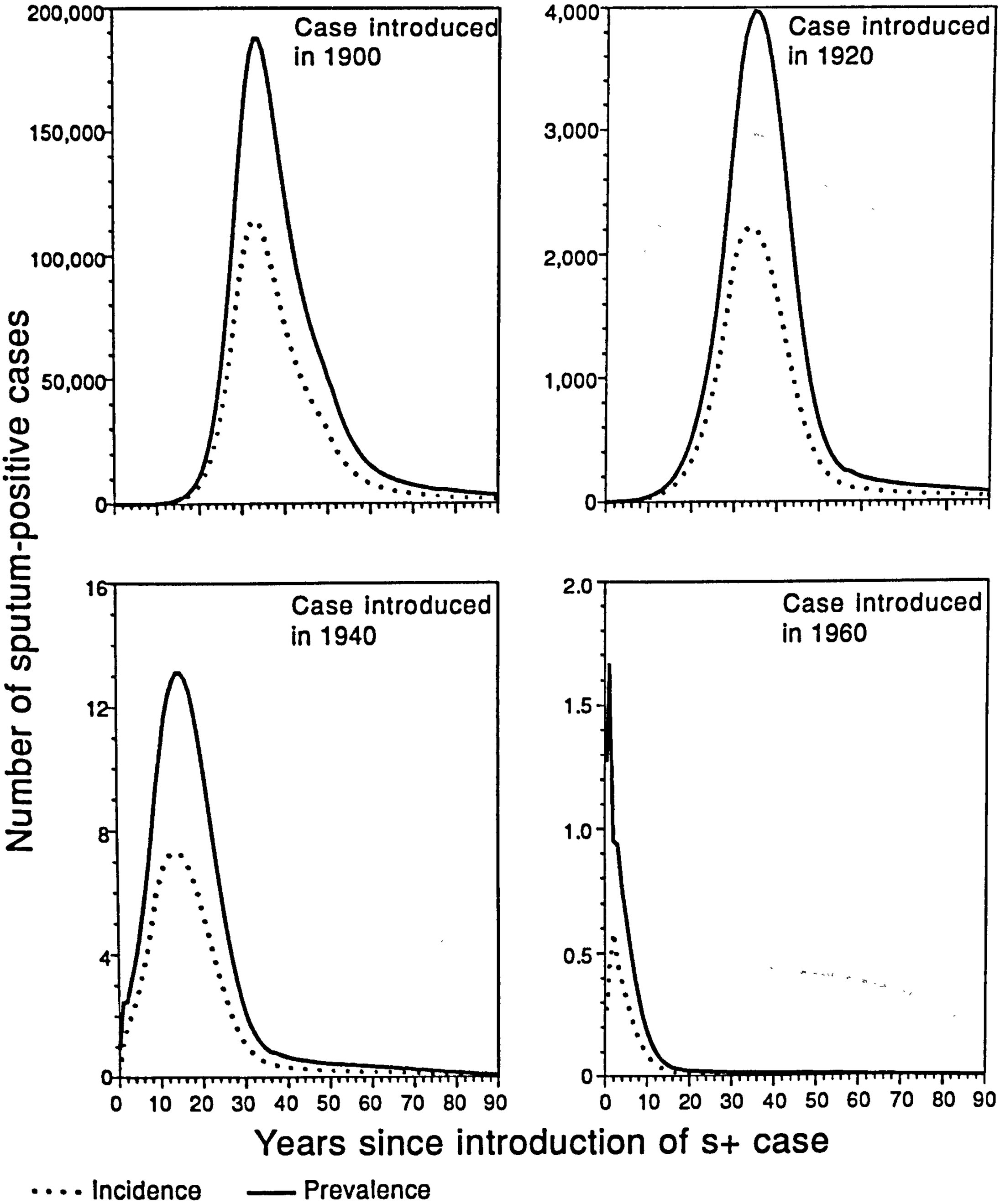


Figure 5.4: Number of incident and prevalent sputum-positive cases following the introduction of one sputum-positive individual into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TBDYN3. This assumes that the sputum-positive case was infectious for two years.

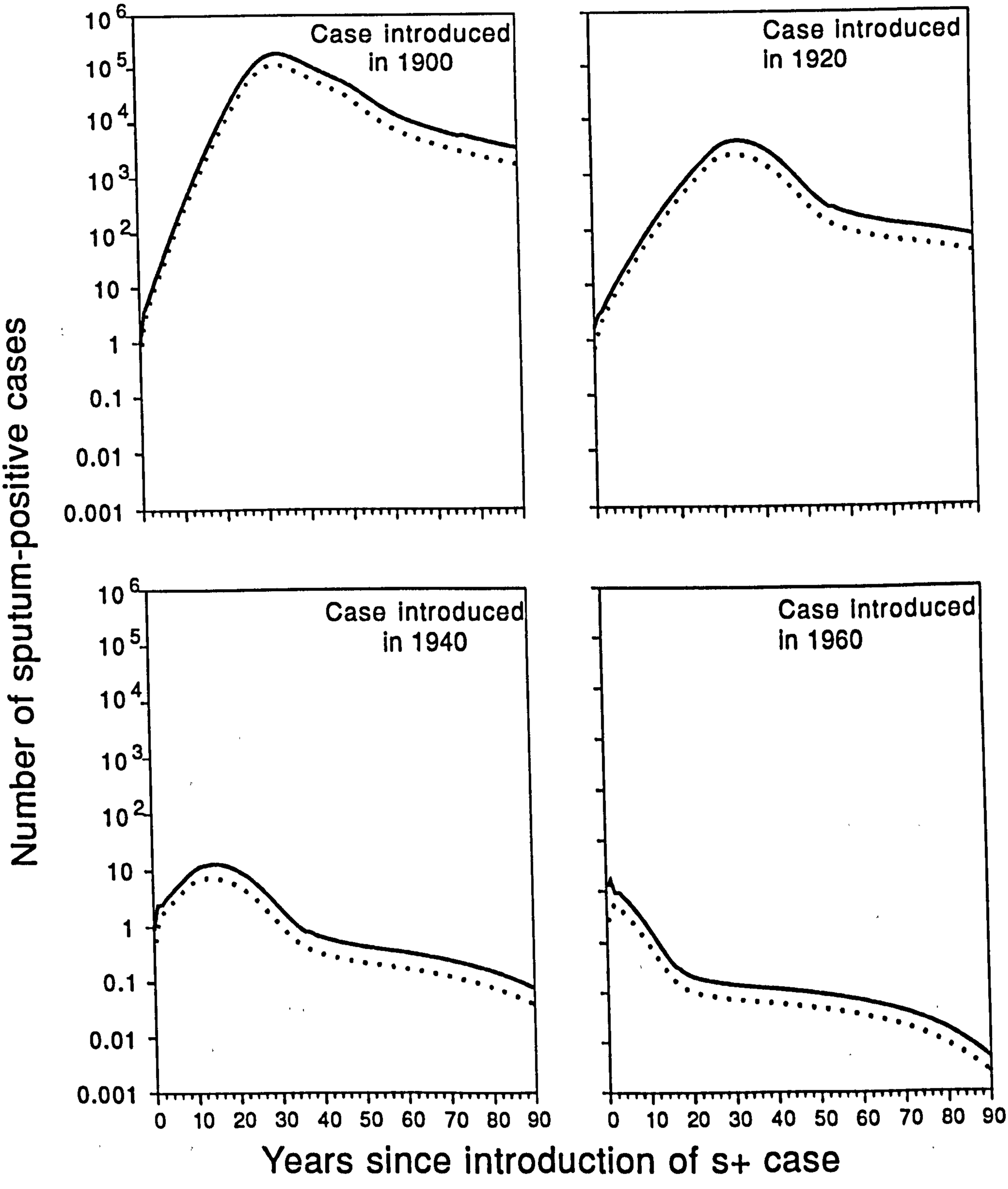


Figure 5.5: Total number of incident and prevalent sputum-positive cases following the introduction of one sputum-positive individual into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TBDYN3. This assumes that the sputum-positive case was infectious for two years.

subsequent years. This is consistent with the higher effective contact number estimated for 1900, than in later years (see Figure 5.2).

Figure 5.6 shows that the incidence rate for the epidemics initiated in 1920 and 1940 started to decline at the same time (i.e. in about 1954) *irrespective* of whether the founder case had been infectious for one or two years. The reason for this decline is interesting and important.

For simple infectious diseases in which infection confers protection against reinfection and for which the basic reproduction number exceeds one, the incidence of disease should initially increase following the introduction of a founder case into a 'totally susceptible' population, and then decrease as the proportion and number of 'susceptibles' decreases.

Figure 5.7 shows the corresponding crude prevalence of *infection* for the years subsequent to the initiation of the epidemics during the period 1900–1960. This shows that the peak prevalence of infection for the epidemic initiated in 1900 occurred in the same year as that in which the disease incidence peaked, i.e. 1940. Given the high prevalence of infection for this epidemic, the majority of the population alive during the years in which the epidemic peaked could develop disease only through endogenous reactivation or after reinfection.

This is illustrated in Figure 5.8, which shows the relative contribution of disease attributable to first primary episodes, exogenous and endogenous disease for these epidemics. Hence most of the disease incidence in the population during the first 40 years after the introduction of the founder case in 1900 is attributable to first primary episodes. Subsequently, an increasing proportion is attributable to exogenous disease, and then, 60 years after the introduction of the sputum-positive case, most is attributable to endogenous disease. The estimated peak in exogenous disease occurs about 10 years after the overall disease incidence started to decline. These results are generally consistent with the magnitude of the annual risk of infection (see Figure 5.9) subsequent to the introduction of the founder into the population, which peaks in about the same year as the disease incidence.

The estimated risks of developing both endogenous and exogenous disease are lower than those for developing the first primary episode for most age-groups (see Table 3.3 in section 3.3.1). Hence we can conclude that the ultimate decline in the epidemic initiated in 1900 *is attributable* in large part to a high prevalence of infection in the population, and hence the relatively low risks of developing (endogenous and exogenous) disease experienced by most individuals.

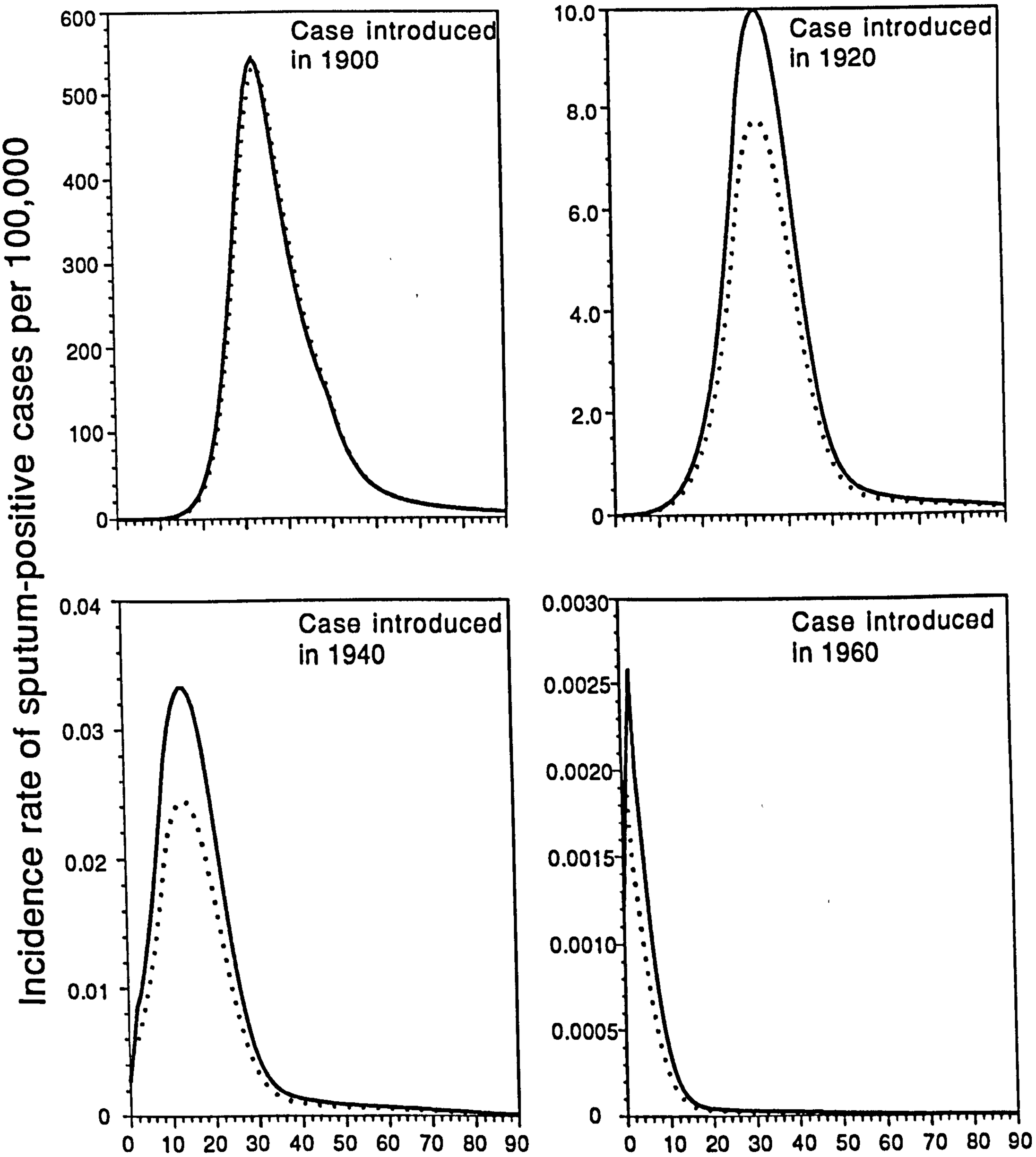


Figure 5.6: Predicted crude incidence rate of sputum-positive disease after the introduction of a founder case, who was infectious for either one or two years, into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TBDYN3.

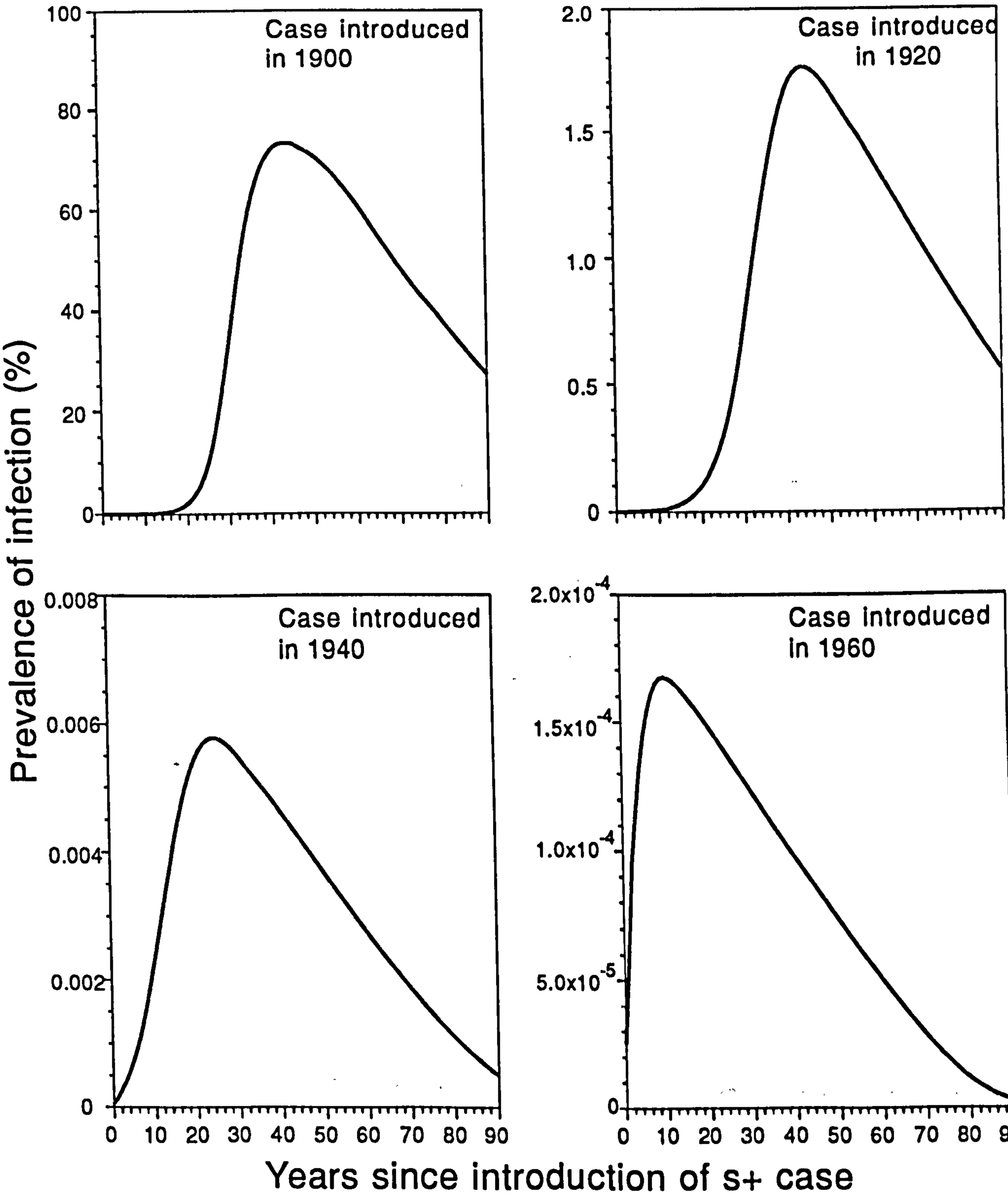


Figure 5.7: Predicted crude *prevalence of infection* after the introduction of a founder case into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TBDYN3. This assumes that the founder case was infectious for two years.

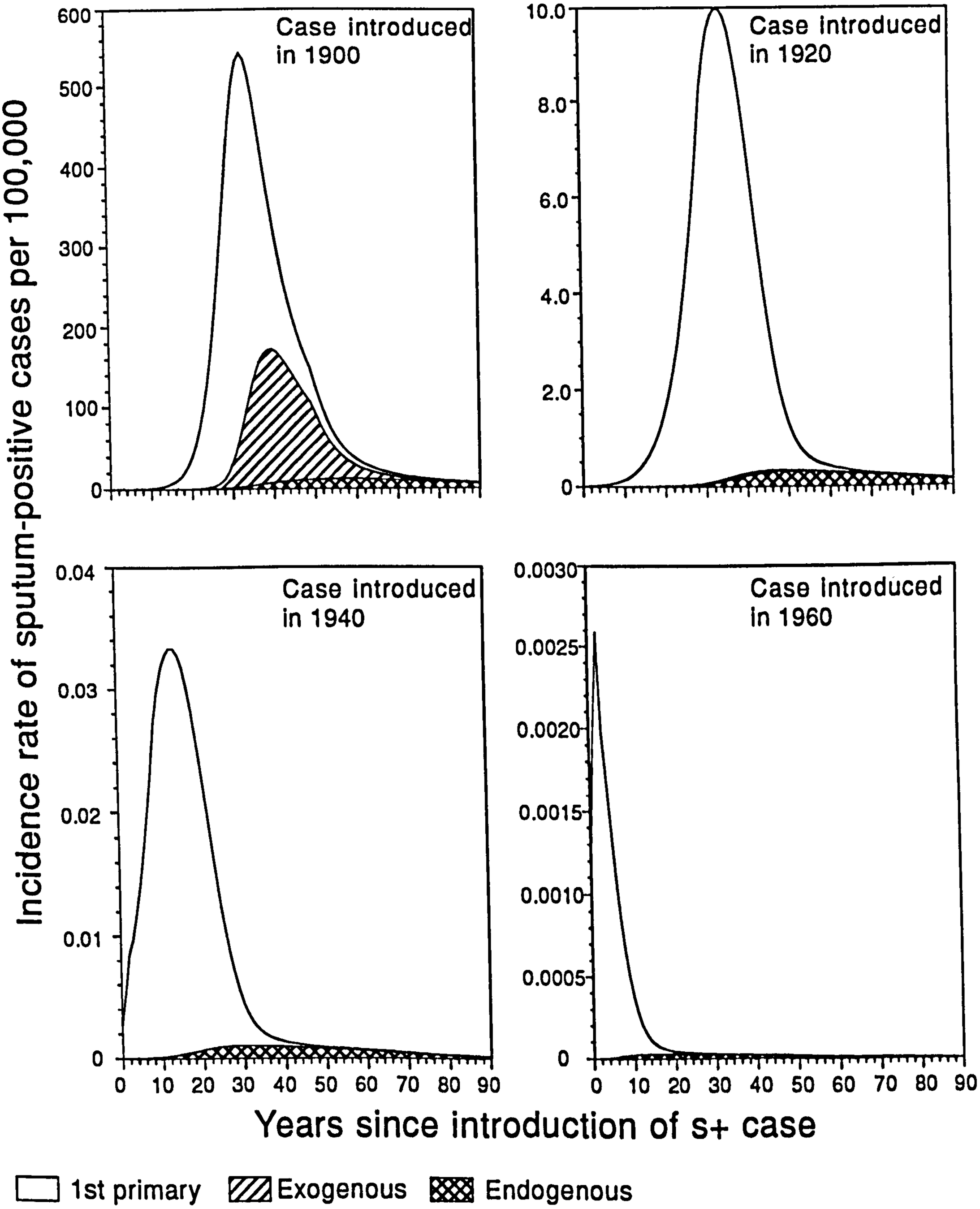


Figure 5.8: Relative contribution of '1st primary', endogenous and exogenous disease to the predicted incidence (per 100,000 general male population) of sputum-positive tuberculosis after the introduction of one founder case into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TBDYN3. This assumes that the founder case was infectious for two years.

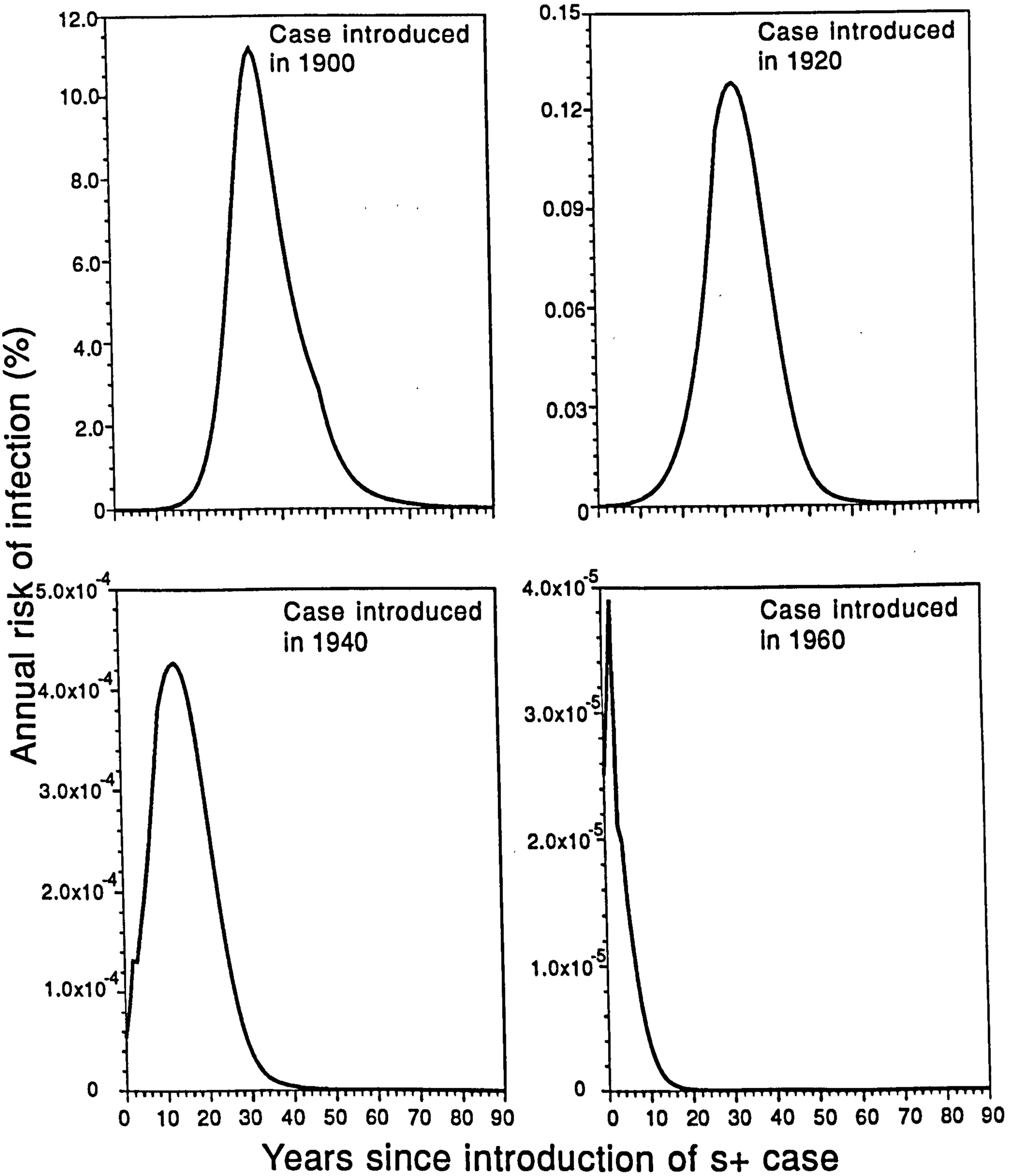


Figure 5.9: Predicted annual risk of infection after the introduction of one sputum-positive individual into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TBDYN3. This assumes that the sputum-positive case was infectious for two years.

Figure 5.7 shows that the peak prevalence of infection was about 1.8% and 0.006% respectively for the epidemics initiated in 1920 and 1940. These extremely low levels in the prevalence of infection indicates that the decline in these epidemics was *not* attributable to a depletion in the proportion of uninfected individuals in the population, and is attributable to other factors, which we discuss below.

The trend in disease incidence at a given time is a function of the rate at which secondary cases from infectious cases at that time, and at all previous times, develop disease. Virtually none of the disease incidence for the epidemics initiated in 1920 and 1940 is attributable to exogenous disease (see Figure 5.8). Hence the trend in the disease incidence for these epidemics at a given time is a function of

1. the rate at which individuals infected many years previously develop endogenous disease, and

2. the number of first primary episodes generated by each sputum-positive case, which depends on the magnitude of the effective contact number.

These factors depend on the age distribution of infected and uninfected individuals respectively in the population.

Figure 5.8 shows that epidemics initiated in both 1920 and 1940 declined *at the same time as the incidence of endogenous disease was increasing.* This suggests that at the time of decline of these epidemics, each infectious case was leading to less than one case (of first primary disease episodes) during the following year. This is attributable to a low effective contact number at this time.

5.3.1.2 The basic reproduction number

Figure 5.10 summarizes the number of secondary cases attributable to infection by a founder case who was introduced between 1900 and 1990 into a population similar to that in England and Wales, and was infectious for one or two years, assuming that the effective contact number *did not decline* subsequent to his/her introduction. This shows that such a founder case introduced in 1900 would have led to about 1.5 and 3.1 secondary cases if he/she was infectious for one and two years respectively. By applying a similar argument to that in section 5.3.1.1 on page 204, we will use the number of secondary cases resulting from an

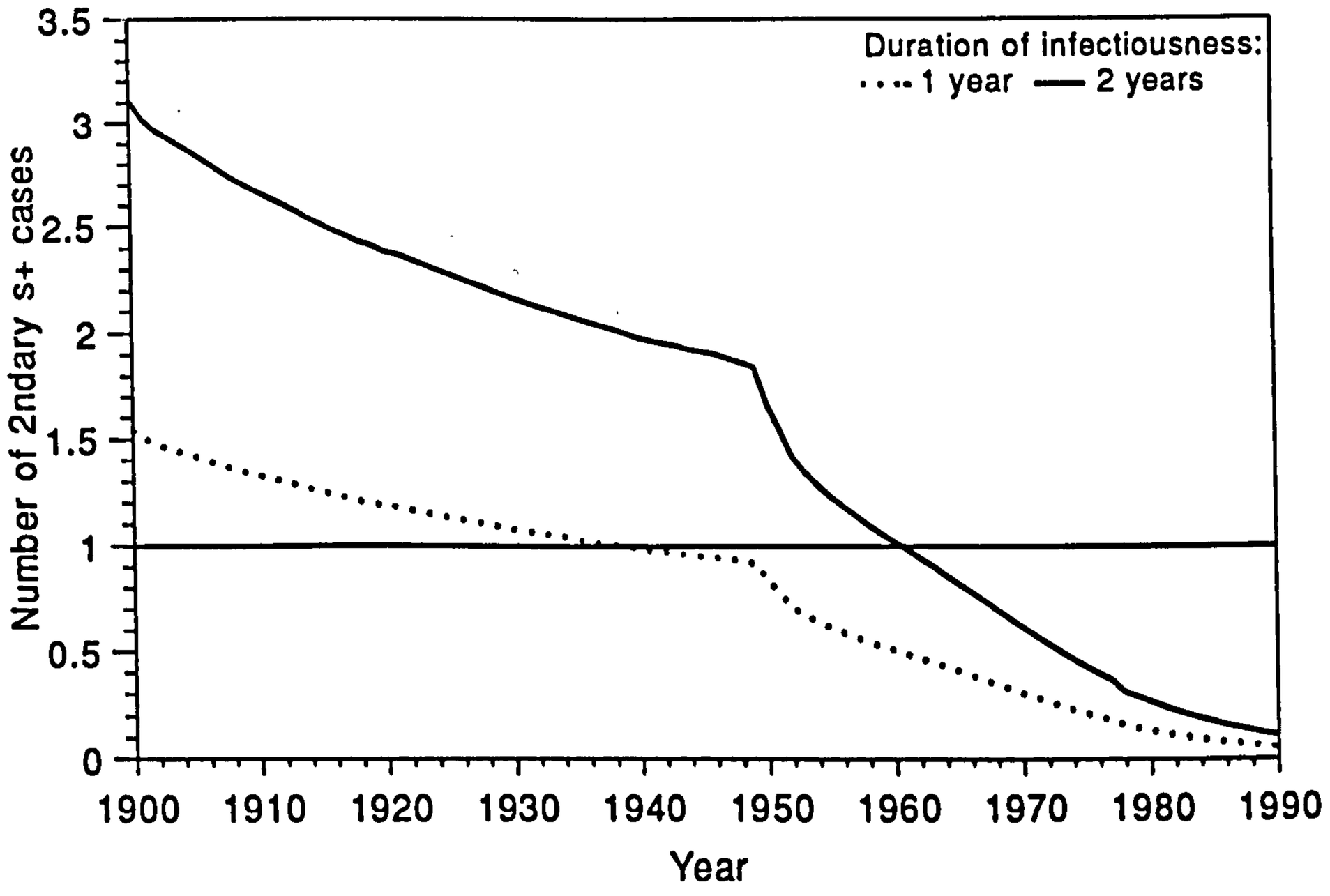


Figure 5.10: Number of secondary sputum-positive cases resulting from a founder case who was infectious for one or two years, introduced into an uninfected population similar in structure to that in England and Wales during the period 1900–1990, as estimated using TBDYN3, assuming that there was *no decline in the effective contact number* subsequent to his/her introduction.

individual who is infectious for the full two years, under the corresponding circumstances to reflect the magnitude of the basic reproduction number for tuberculosis.

Figure 5.11 compares the basic and founder case reproduction numbers for tuberculosis during the period 1900–1990. This shows that the founder case reproduction number slightly exceeded the basic reproduction number until 1950. This is paradoxical, given that a founder case leads to fewer new *infections* if the effective contact number declines in a population subsequent to his/her introduction, as compared with that when there is no such decline in the effective contact number. In practice, the decline in the effective contact number over time means that the risk of (re)infection will be *lower* subsequent to the introduction of a founder case than in the situation in which there is no decline in effective contact number. This means that individuals are less likely to be reinfected and hence *more likely to develop disease attributable to infection by a founder case*, as estimated using TBDYN3 if the effective contact number declines, than in the converse situation.

This is illustrated in Figure 5.12, which compares the annual risk of infection subsequent to the introduction of a founder case in England and Wales in 1900, 1920, 1940 and 1960,

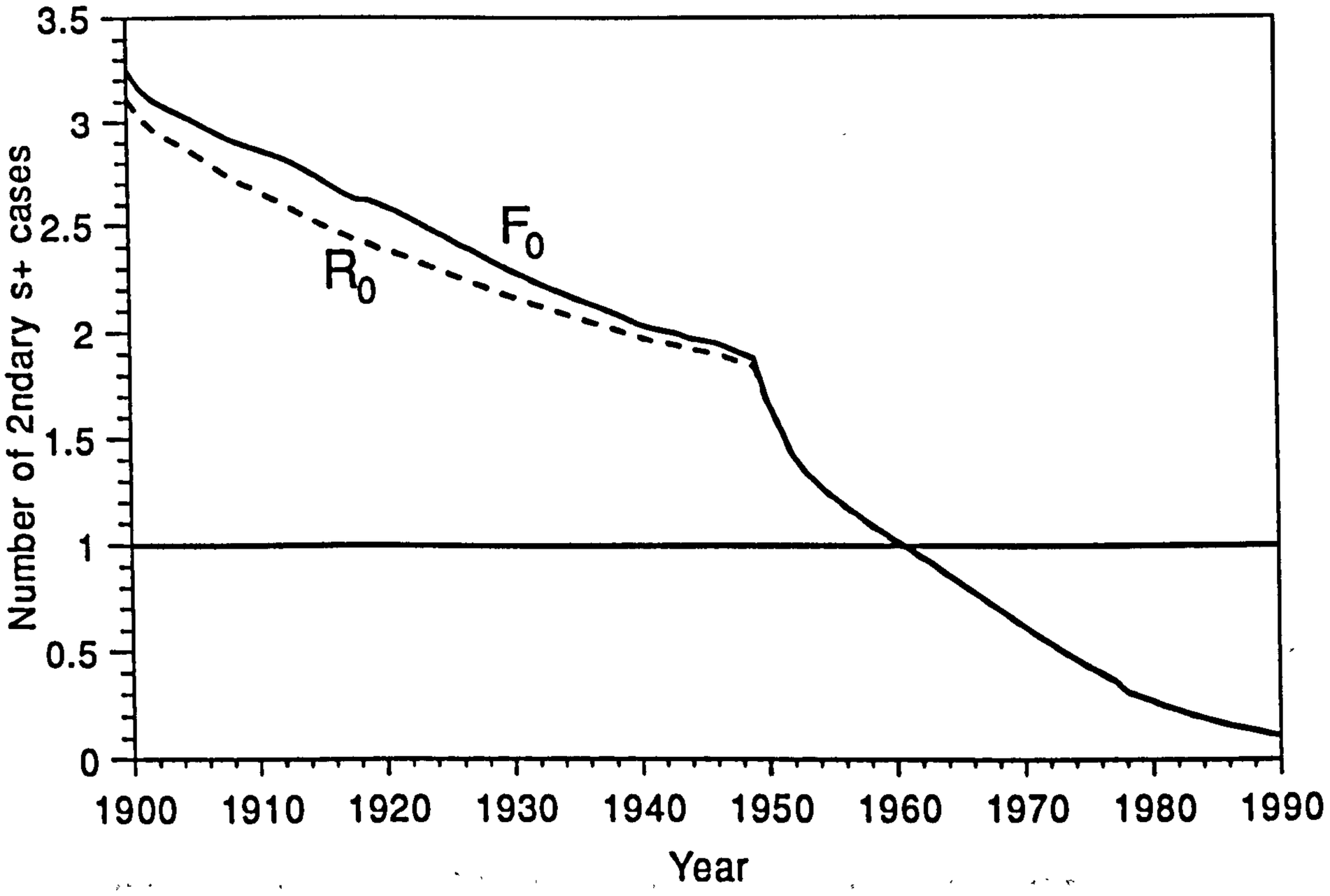


Figure 5.11: Comparison between the basic and founder case reproduction numbers for tuberculosis during the period 1900–1990, as estimated using TBDYN3.

assuming that the effective contact number did and did not subsequently decline. This shows that had there been no change in the effective contact number after 1900 under these circumstances, the annual risk of infection would have peaked at about 50% about 20 years thereafter, and ultimately would have remained relatively static thereafter at about 20%. In contrast, the decline in the effective contact number subsequent to the introduction of the founder case meant that the maximum risk of infection was about 10%, which is about the level of the estimated annual risk of infection in England and Wales in 1910.

The level of the risk of infection subsequent to the introduction of the founder case in 1900, assuming no decline in effective contact number, seems unrealistically high, and far exceeds that found in any other developing or developed country during this century (see section 2.3). The implications of such high risks of infection for the relative contribution of first primary, endogenous and exogenous forms to the disease incidence are illustrated in Figure 5.13. This shows that the peak disease incidence subsequent to the introduction of the case in 1900 would have been about 1,200 per 100,000, and that the disease incidence would have stabilized thereafter at about 500 per 100,000. Most of the disease incidence among individuals in the population for this epidemic would have been attributable to exogenous reinfection.

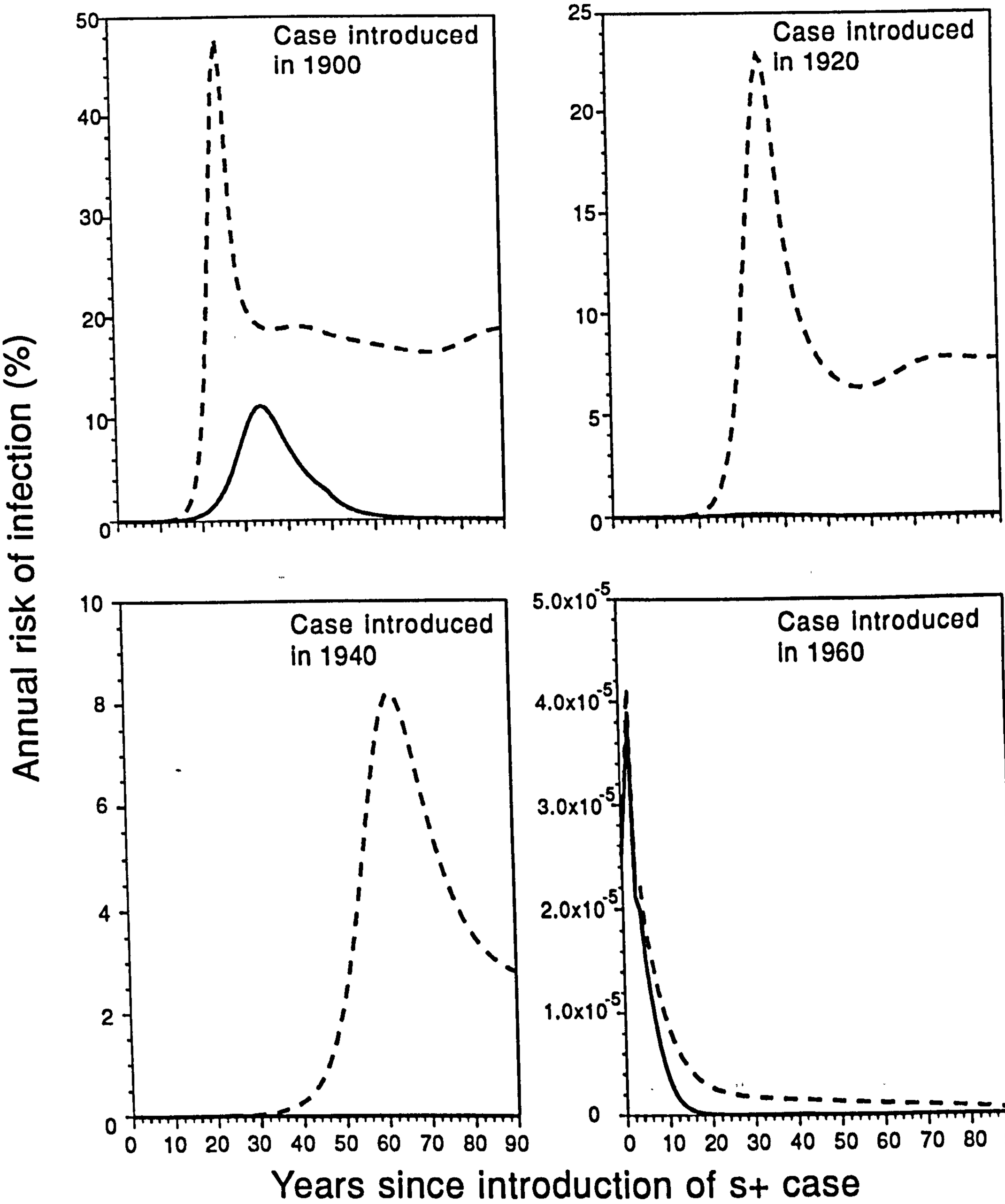


Figure 5.12: Comparison between the annual risk of infection subsequent to the introduction of a founder case into a population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, assuming that the effective contact number did and did not subsequently decline.

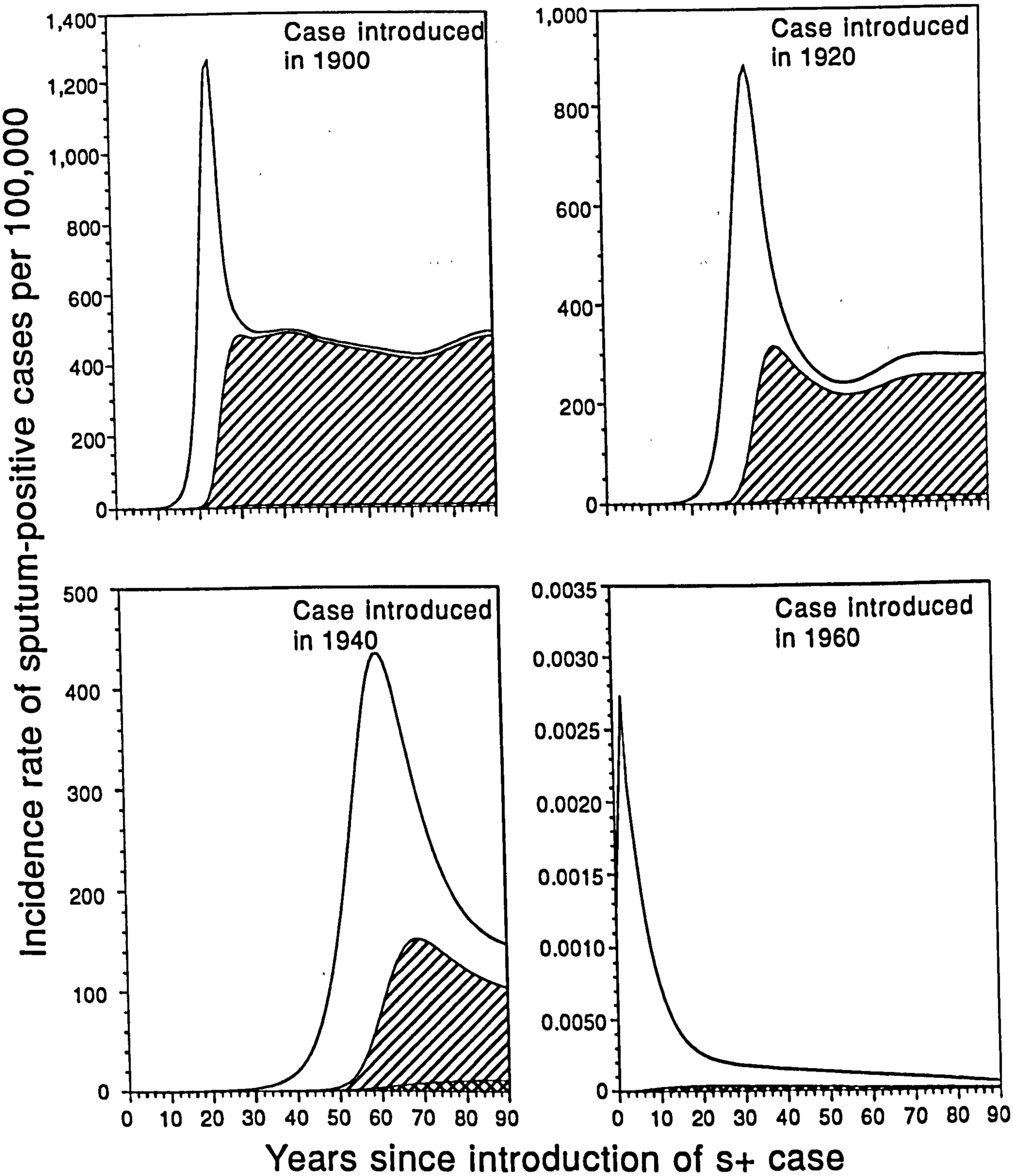


Figure 5.13: Relative contribution of '1st primary', endogenous and exogenous disease to the predicted incidence (per 100,000 general male population) of sputum-positive tuberculosis after the introduction of one founder case into an uninfected population similar in structure to that in England and Wales in 1900, 1920, 1940 and 1960, as estimated using TBDYN3. This assumes that the founder case was infectious for two years, and that *there was no subsequent change in the case-fatality rate and the effective contact number.*

Overall, these results show that, whereas the magnitudes of the basic and founder case reproduction number for tuberculosis differed only slightly until 1950, the implications for the disease dynamics subsequent to the introduction of a founder case into an uninfected population for corresponding assumptions relating to the effective contact number are substantially different. The epidemic initiated in 1940, for example, would have peaked during the early 1950s at about 0.03 sputum-positive cases per 100,000 if the effective contact number had declined subsequent to the introduction of the founder case, and the disease incidence would have subsequently declined dramatically. The peak incidence rate of about 450 sputum-positive cases per 100,000 would have occurred in the year 2000 if there had been no such decline in the effective contact number, and the disease incidence would have probably remained relatively static thereafter at over 100 sputum-positive cases per 100,000.

5.3.2 The net reproduction number

As given in equation 5.16, the number of secondary sputum-positive cases resulting from an individual who developed disease in a given year is determined by

1. the total number (and the age-distribution) of individuals first infected or reinfected by that case during the infectious period and,

2. the risk of developing sputum-positive disease attributable to this infection or reinfection.

Before describing the results obtained for the net reproduction number we first present results for these two statistics.

5.3.2.1 Number of new infections/reinfections per sputum-positive case

Figure 5.14 shows the numbers of individuals infected for the first time or reinfected by each sputum-positive case during the period 1900–1990, as estimated using TBDYN3. This is given by the ratio between the number of new infections or reinfections in that year, and the total number of sputum-positive cases respectively (see equation 5.5 in section 5.2.1).

This shows that of the 23 individuals effectively contacted by each sputum-positive case in 1900, about 12 were reinfected, 3 were infected and the contact did not lead to infection or (further) reinfection for the remaining 8. By 1950, each sputum-positive case effectively contacted 11 individuals: 6 were reinfected, 4 were infected for the first time and the contact did not lead to infection or reinfection in only 1 individual.

It is important to recognize that these results depend on the definitions of individuals assumed to be at risk of infection or reinfection. The comparatively large number of individuals for whom effective contact did not lead to infection or reinfection for the early 1900, for example, follows from

1. the fact that many individuals during this time were estimated to have recently experienced infection or reinfection, and

2. these individuals were not assumed to be at risk of further reinfection at that time (see sections 3.1.2 and 5.2.1).

The secular trend in the number of individuals infected or reinfected by each sputum-positive case is a function of the changing composition of 'uninfected' and 'latent' individuals

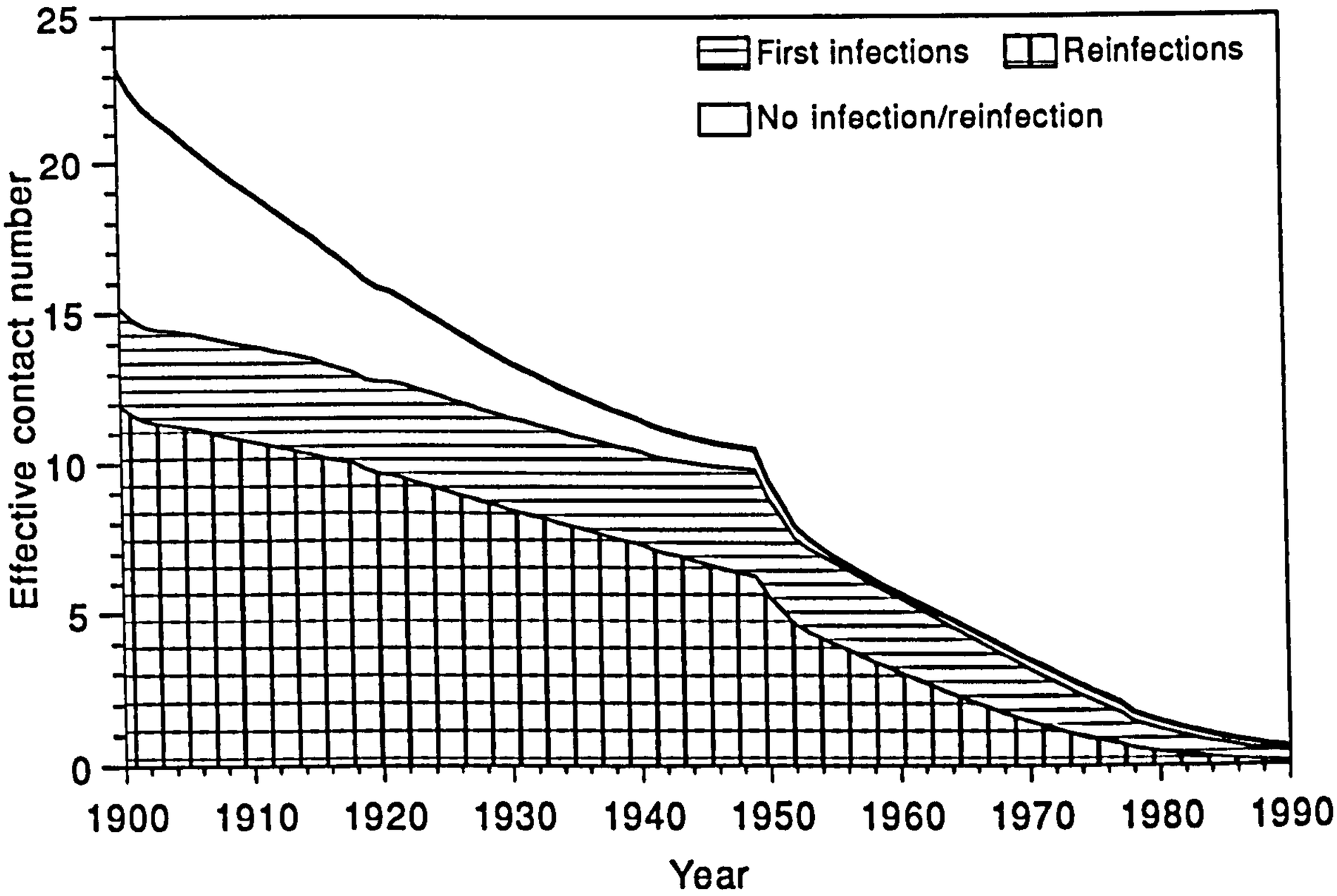


Figure 5.14: Comparison between the estimated effective contact number and estimated number of individuals infected for the first time or reinfected in England and Wales during the period 1900–1990

in the population during this time, and in the effective contact number. We see, for example, that the average number of new infections *resulting* from each sputum-positive case changed little over most of the century. This follows from the fact that the effective contact number decreased at the same time as the proportion of uninfected individuals in the population increased. Hence sputum-positive cases in 1960 were more likely to come into contact with uninfected individuals, as compared with such individuals in 1900.

Figures 5.15 and 5.16 summarize the age-distributions of uninfected individuals and 'latent' individuals for given years between 1900 and 1980 who, by definition, were at risk of infection or reinfection by a given sputum-positive case. Both these distributions in a given year are functions of past trends in the annual risk of infection. Figure 5.15 shows that uninfected 10 year olds comprised about 0.3% of the population in 1900, and that almost all uninfected individuals were aged under 20 years. This is consistent with the high risk of infection estimated for the early part of this century (see section 2.3), which meant that few individuals reached adolescence without having been infected.

By 1980, on the other hand, uninfected 10 year olds comprised about 1.5% of the population and over half of the uninfected population was aged over 20 years. Analogous shifts

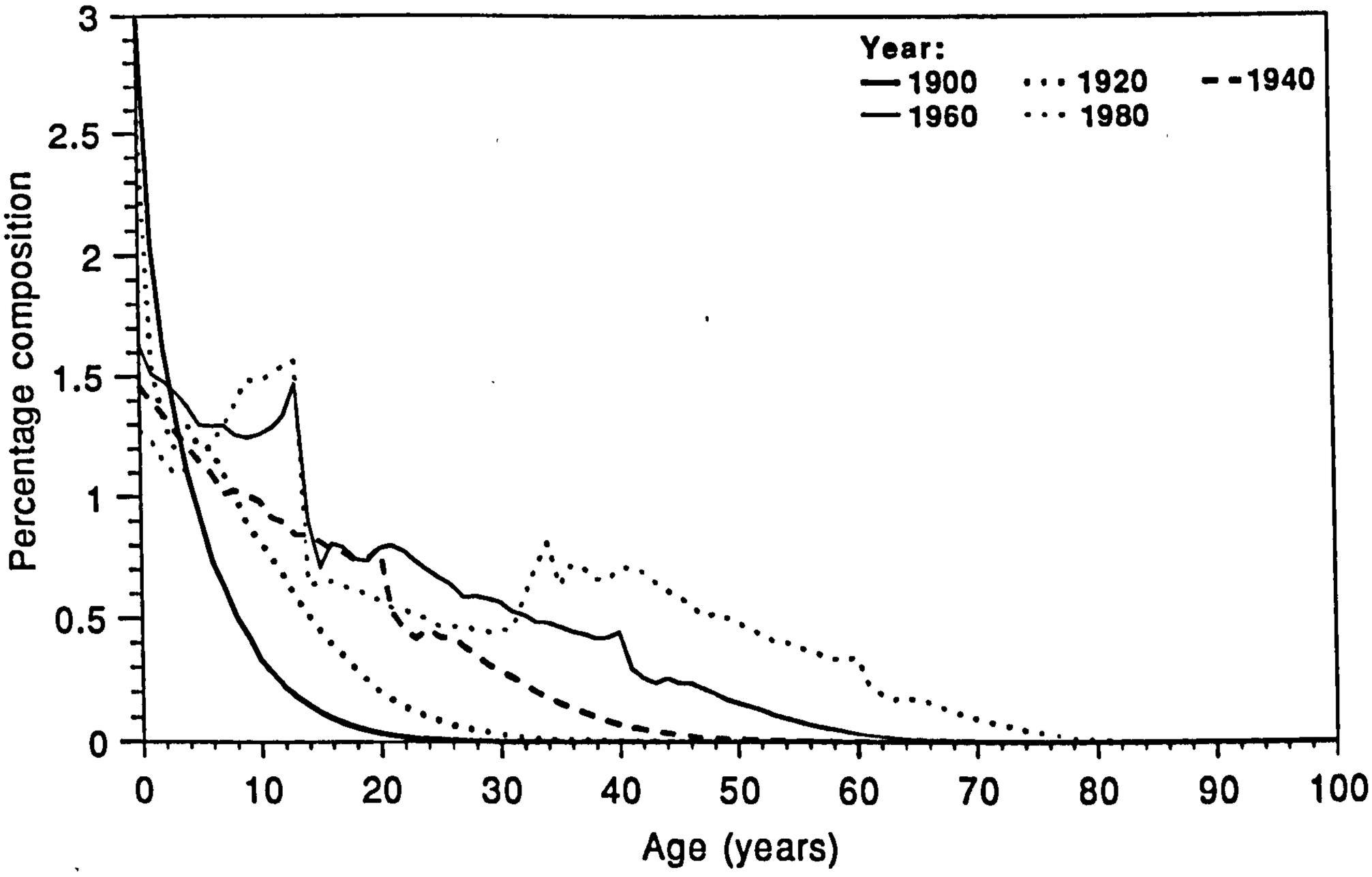


Figure 5.15: Percentage composition (by age) of *uninfected* individuals in the population in England and Wales between 1900 and 1980, as estimated using TBDYN3.

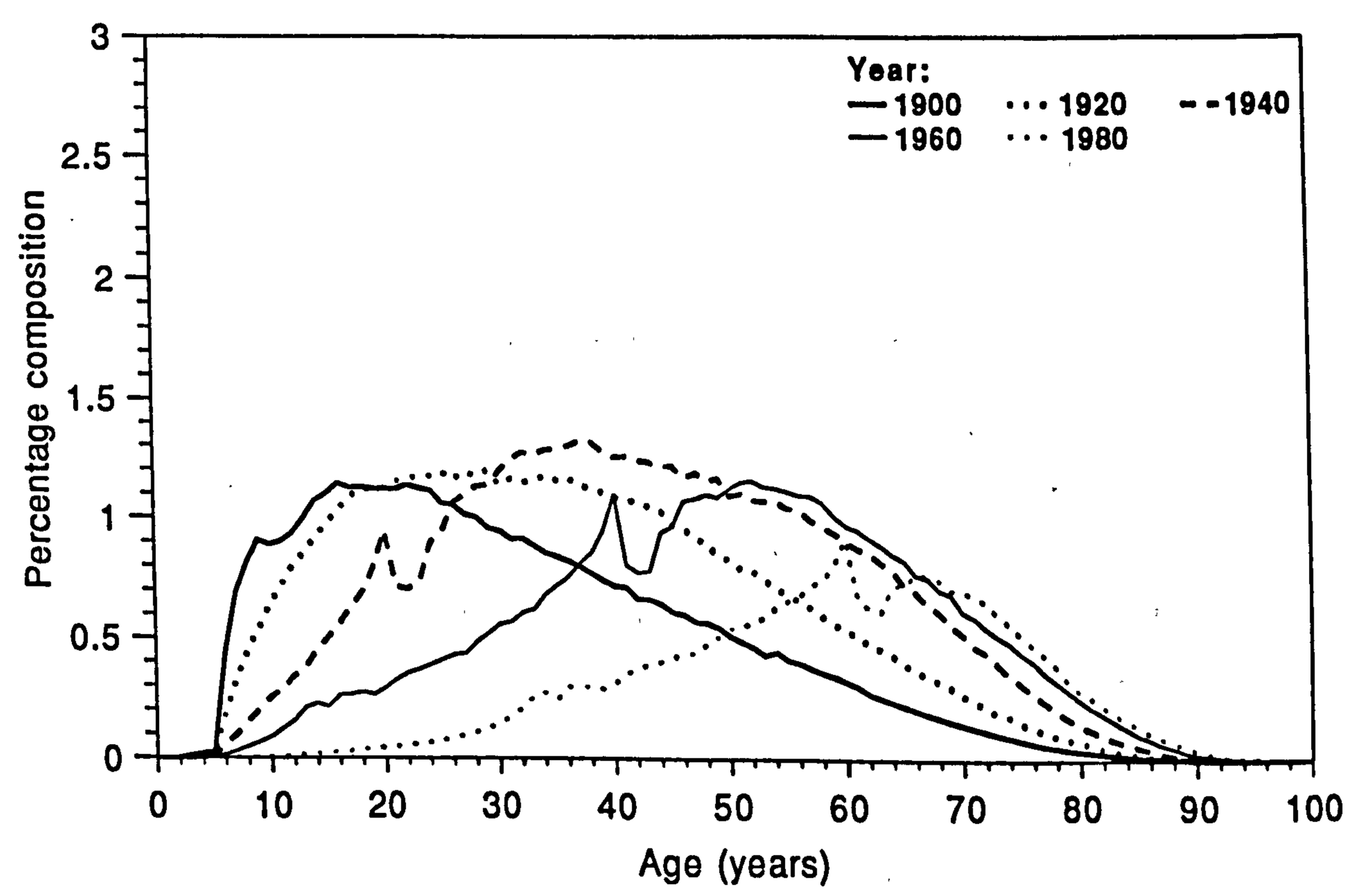


Figure 5.16: Percentage composition (by age) of individuals in the '*latent*' class in the population in England and Wales between 1900 and 1980, as estimated using TBDYN3.

occurred in the age-distribution of individuals in the 'latent' class, as shown in Figures 5.16. In deriving the proportion of individuals in both these categories, individuals assumed to be protected by BCG vaccination have been excluded from the uninfected population, but are included in the estimates of the size of the overall population.

5.3.2.2 Lifetime risk of developing sputum-positive disease attributable to a given infection/reinfection

Figure 5.17 shows the estimated proportion of individuals infected for first time during the period 1900–1990, who developed sputum-positive disease attributable to their initial infection. This shows that of the 30 year olds infected for the first time in 1900, about 10% developed sputum-positive disease attributable to this infection, as compared with about 11% of the corresponding individuals in 1980. Though this is a relatively small difference, it reflects the fact that a greater proportion of those infected in 1980 could potentially develop their first disease episode without being reinfected, as compared with similar individuals infected in 1900, whose disease was, according to TBDYN3, attributable to subsequent reinfection. We discussed these patterns in section 4.3.2.3, when we considered the lifetime risks of developing sputum-positive disease attributable to initial infection. Similar patterns are also seen in Figure 5.18, which shows the proportion of individuals reinfected during the period 1900–1990, who develop sputum-positive disease attributable to this reinfection event.

5.3.2.3 Estimates of the net reproduction number

The effect on the net reproduction number of changes in the age distribution of individuals infected and reinfected each year, and in the proportion subsequently developing disease during the period 1900–1990 is illustrated in Figure 5.19.

This shows that overall, there was very little change in the net reproduction number between 1900 and 1950, and that it was first below one during the late 1930s. This is paradoxical, given the decline in the tuberculosis mortality rate in England and Wales since the mid 1850s and the decline in the incidence predicted by TBDYN3 during this time. We discuss the reasons underlying this result below.

As described in section 5.1.1, a net reproduction number of greater than one for an infectious disease in which all disease parameters (such as the contact between individuals)

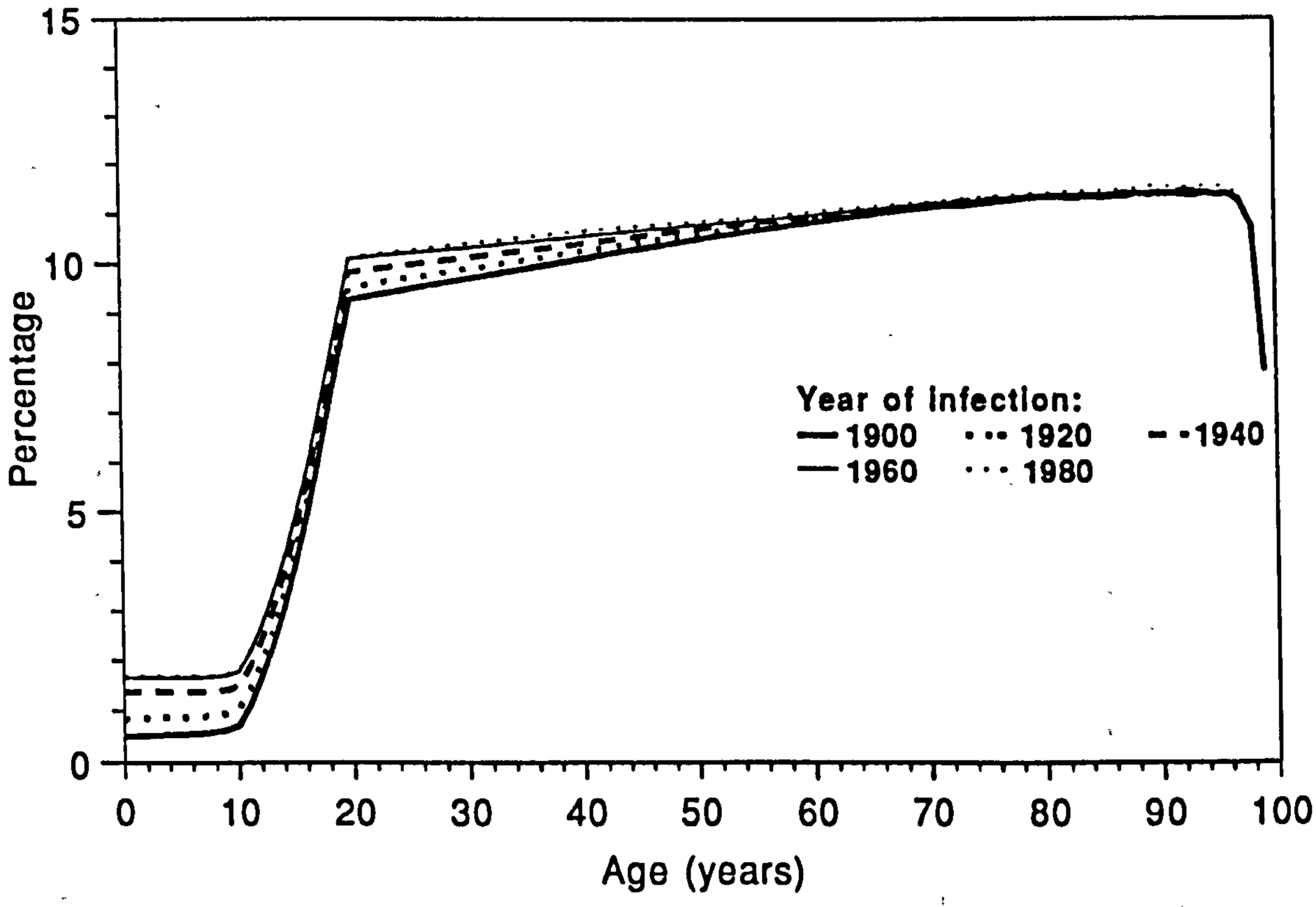


Figure 5.17: Percentage of individuals infected for the first time during the period 1900–1990 when aged 0–100 years *who develop sputum-positive disease attributable to their infection episode*, as estimated using TBDYN3.

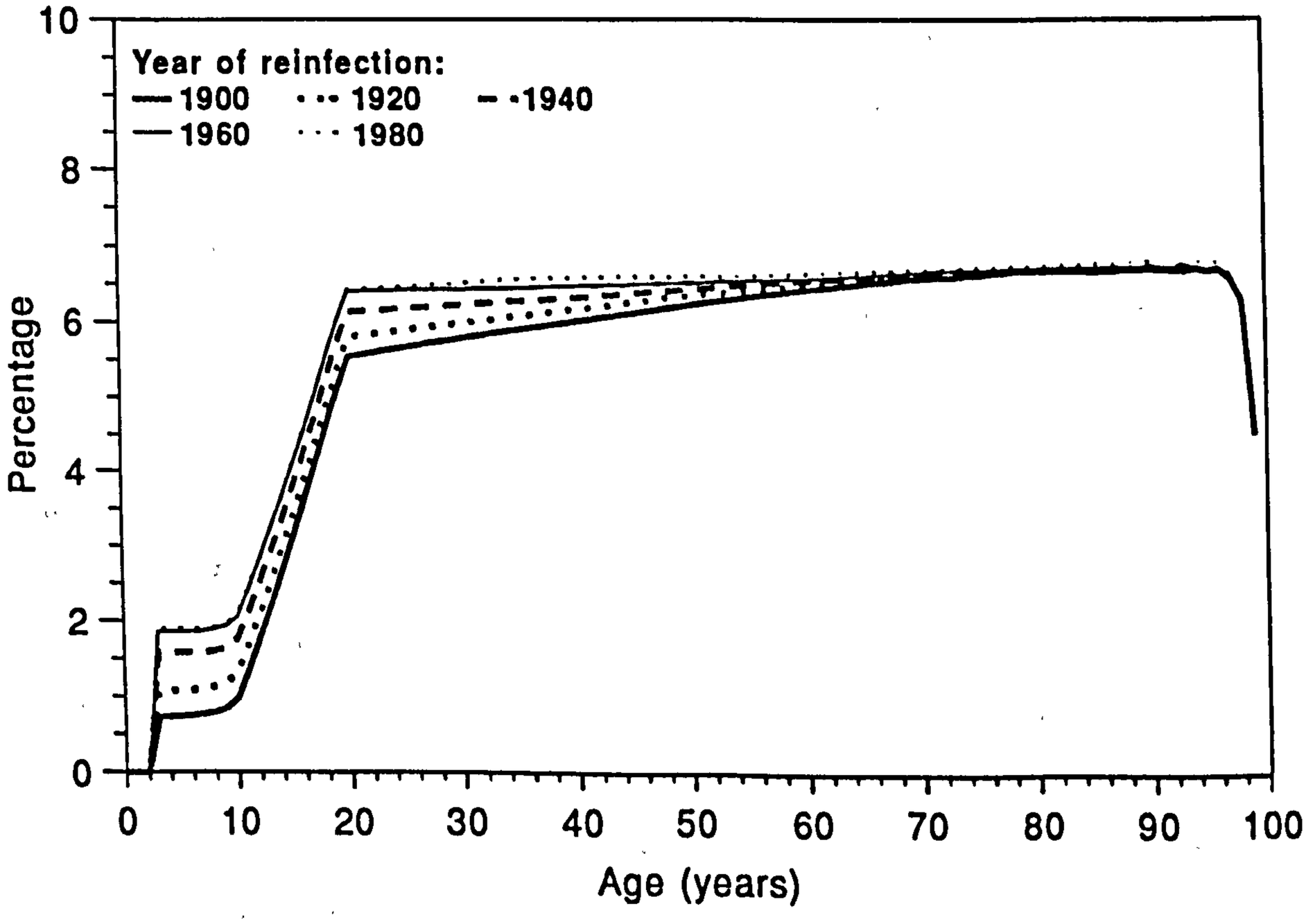


Figure 5.18: Percentage of individuals reinfected during the period 1900–1990 when aged 0–100 years *who develop sputum-positive disease attributable to this reinfection episode*, as estimated using TBDYN3.

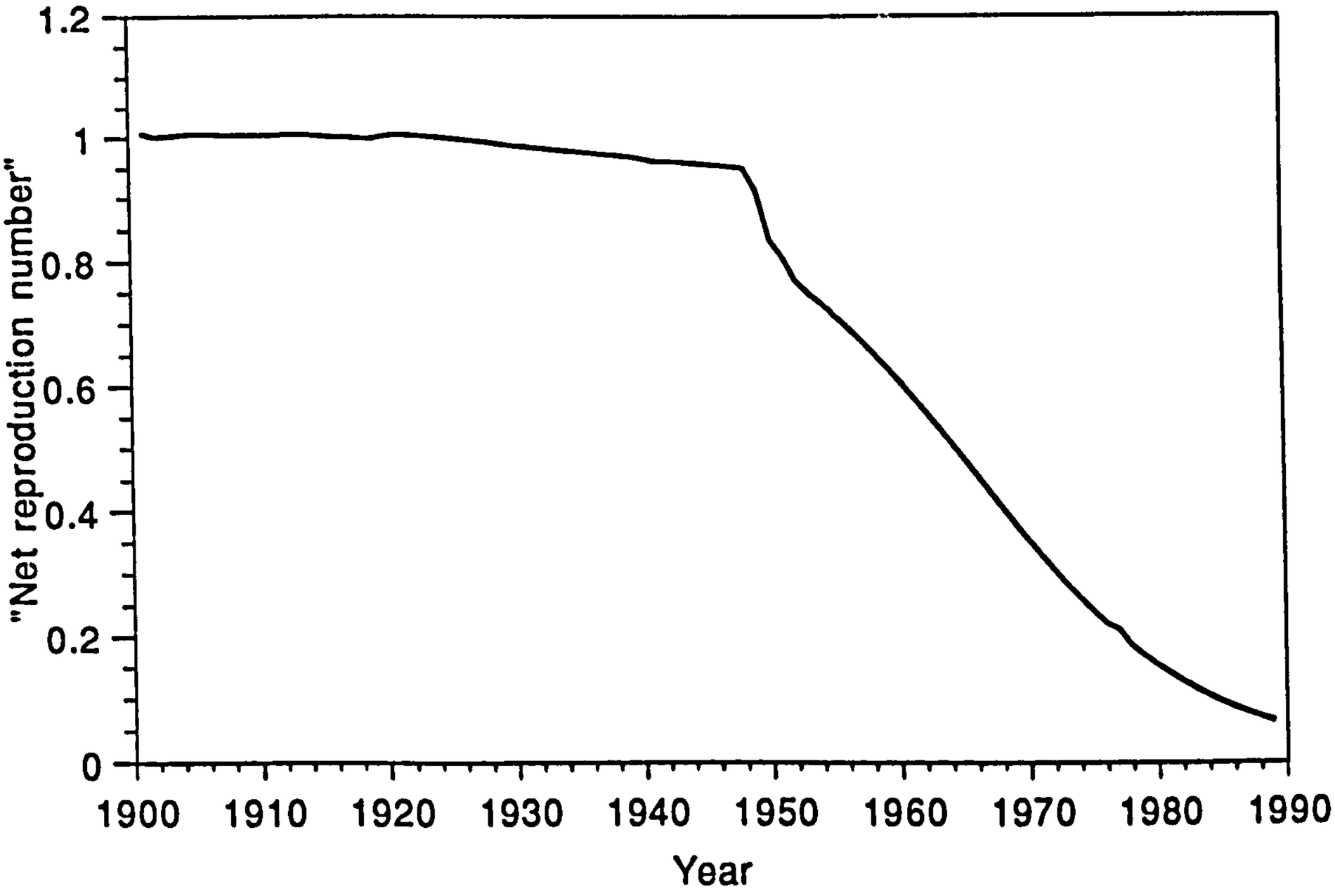


Figure 5.19: Estimates of the 'net reproduction number' for tuberculosis, as derived using TBDYN3

change little over time, typically leads to an increase in the disease incidence. The situation is complicated for tuberculosis, since the effective contact between individuals declined over time and the time interval between infection or reinfection and subsequent sputum-positive disease attributable to that infection/reinfection event (i.e. the 'true' serial interval — see definition in section 4.1) can be many years.

As we saw in section 4.3.2.3, the distribution of the time interval between infection or reinfection and sputum-positive disease attributable to the infection/reinfection event depends on both the age at infection/reinfection and on the subsequent risk of reinfection experienced. In 1900, for example, 0–10 year olds constituted most of the uninfected category (see Figure 5.15). These individuals faced low risks of developing (sputum-positive) disease immediately after infection, and hence first contributed to the sputum-positive incidence when they became young adults up to twenty years later. The same holds for 1920, when 0–20 year olds constituted most of the uninfected category.

Intuitively, if only a slightly greater proportion of individuals infected at age 0–10 years in 1900, for example, had developed disease earlier, when the number of secondary cases resulting *shortly after* initial infection/reinfection by a given sputum-positive case was just slightly smaller than one, then it is likely that the disease incidence would have increased

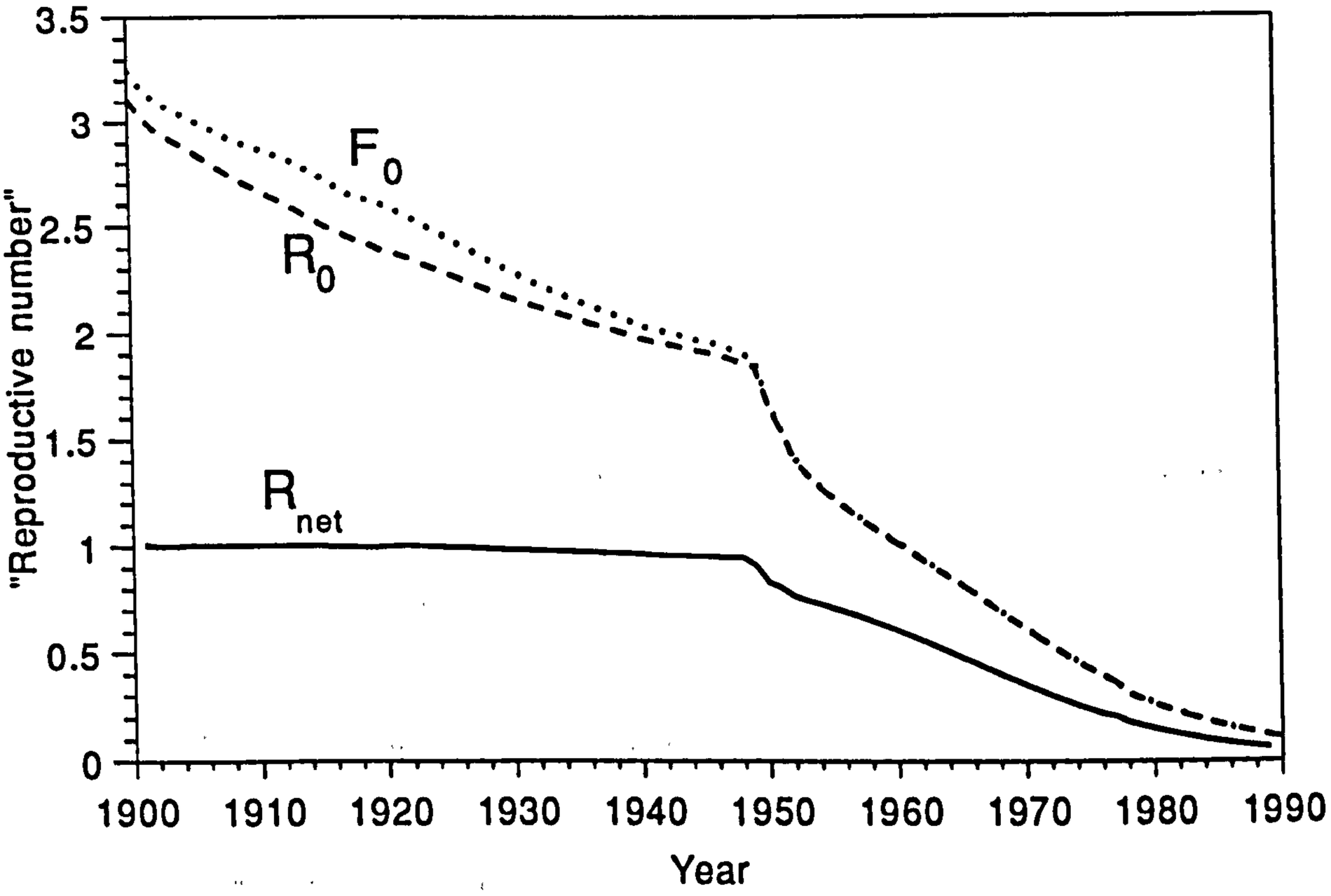


Figure 5.20: Comparison between the estimated net and founder case reproduction numbers for tuberculosis, as derived using TBDYN3.

during the 1900s. By this argument, the disease incidence did not increase during the period considered since only a proportion of secondary cases developed disease shortly after initial infection/reinfection, and the net reproduction number was *only slightly greater than one* until 1925 (i.e. ≈ 1.001).

Figure 5.20 compares the values obtained for the net reproduction number since 1900 against the estimated basic and founder case reproduction numbers for these years. We see that whilst the basic and founder case reproduction numbers declined over time, largely as a consequence of the decline in the effective contact number, the net reproduction number remained relatively static until about 1950, and then declined more rapidly. The comparatively small effect of the decline in the effective contact number on the net reproduction number, as compared with that on the basic and founder case reproduction numbers follows from the fact that *overall, the proportion of individuals developing sputum-positive disease attributable to a given infection increased over time.* This is attributable to the combination of two factors.

First, the high annual risk of infection (and effective contact number) estimated for the early 1900s meant that most of the individuals experiencing initial infection/reinfection at this time faced low risks of developing sputum-positive disease, given that

1. infants and young children comprised most of those newly infected, and

2. the estimated risk of developing exogenous disease after reinfection in adult life is relatively low (slightly over half the risk of developing the first primary episode — see Table 3.3 in section 3.3.1).

The decline in the effective contact number meant that by 1930, on the other hand, adolescents and adults comprised large proportions of those experiencing initial infections. The high risks of developing sputum-positive disease experienced by these individuals meant that, overall, *a greater proportion of all newly infected individuals went on to develop sputum-positive disease attributable to a given infection*, as compared with similar individuals in 1900. The situation is more complicated after 1950, given that a proportion of individuals in England and Wales are assumed to be immune to infection following BCG vaccination. This changes the age distribution of individuals defined to be at risk of infection, and hence the overall proportion of individuals who go on to develop disease attributable to initial infection.

Second, the decline in the effective contact number over time meant that individuals infected in 1960, for example, were less likely to experience further reinfection, as compared with counterparts infected in 1900. Any disease episode after an individual has experienced reinfection is attributed to the reinfection event, and this led to an increase over time in the proportion of individuals who developed sputum-positive disease attributable to an infection or reinfection event.

5.3.3 Effect of simplifying assumptions on the results obtained

It is important to recognize that the results obtained in this chapter depend on the simplifying assumptions incorporated into TBDYN3. We discuss the implications of the main simplifying assumptions below.

5.3.3.1 Effect of reinfection

TBDYN3 assumes that the risk of reinfection is identical to that of first infection, which may be unrealistic. As a result of this assumption, TBDYN3 estimates that about half of the individuals effectively contacted by sputum-positive cases during the early 1900s became reinfected, and less than one sixth experienced initial infections, which could well be overestimates and underestimates respectively.

In practice, it is unlikely that this assumption affected estimates of the founder case and net reproduction number over time, given that the estimated risk of developing exogenous disease after reinfection was estimated to be lower than that of developing the first primary episode. As discussed in section 3.4.1, reinfected individuals may face similar or greater risks of developing disease than individuals infected for the first time, given that the immune response must already be low, by definition, for reinfection to occur. Hence any *overestimate* in the number of secondary cases of resulting from a 'typical' infectious case, derived by assuming that the risk of reinfection is identical to that of first infection, is likely to have been balanced out by an *underestimate* in the 'true' risk of developing exogenous disease subsequent to reinfection.

5.3.3.2 Homogeneous mixing

Throughout these analyses, TBDYN3 has assumed that individuals mix randomly in the population, which is unrealistic, especially for more recent years.

In practice, is likely that the effective contact number, and the risk of developing disease, varies widely between different subgroups within the population. It is also plausible that for some subgroups, e.g. the homeless, the magnitude of the effective contact number may have changed little over time, and, in recent times, could even be as high as that estimated for the beginning of this century in England and Wales as a whole. This suggests that the estimates of the basic reproduction number for these individuals could, in turn, be of a

similar order of magnitude to that for 1900.

5.4 Discussion

We began our analyses of the basic and net reproduction numbers for tuberculosis by noting that the definitions and interpretation of both these measures should account for the facts that individuals can be reinfected, the serial interval can be long and variable, the risks of developing disease are age-dependent and that the effective contact number declined over time.

The decline in the effective contact number is intuitively reasonable and is attributable to several factors:

1. the reduction of crowding in living conditions,

2. removal of infectious individuals from the community either to sanatoria before the 1950s or through treatment after 1950,

3. an improvement in nutrition over time, which probably increased resistance to infection,

4. an improvement in general hygiene in the community (attributable to increasing health consciousness, which encouraged people to be careful when coughing, sneezing and spitting),

5. increased ventilation in buildings, which reduced the intensity of exposure of individuals to tubercle bacilli.

It is likely that, a few years after the introduction of chemotherapy, the proportion of sputum-positive cases who were treated and hence removed as sources of infection from the community remained comparatively static. Consequently, the prediction from TBDYN3 that the rate of the decline in the effective contact number only temporarily accelerated with the introduction of chemotherapy in 1950 is also realistic. For many developing countries today the effective contact number could be as high as that estimated for England and Wales during the pre-chemotherapy era. The implications for the corresponding basic and net reproduction numbers are unclear, given that the magnitude of the risks of developing disease may differ from those in England and Wales (see discussion in section 3.4.1).

The fact that the decline in the effective contact number is a consequence of many changes and interventions complicates the definition of the basic reproduction number for tuberculosis, given its interpretation as a parameter which describes the likelihood of continued transmission in the absence of any further control or intervention. As demonstrated in Figure 5.13, the time interval until the peak in disease incidence in the hypothetical situation in which there is no decline in the effective contact number subsequent to the introduction of a founder case into an uninfected population differs according to the magnitude of the effective contact number at the start. This peak would have occurred about 20 and 30 years after the introduction of the founder case in 1900 and 1920 respectively, i.e. over the same time period during which the actual effective contact number declined dramatically.

Given the potentially long and variable serial interval for tuberculosis, neither the basic nor the founder case reproduction numbers provide much information on the magnitude of the trend in disease incidence subsequent to the introduction of a founder case into an uninfected population. Analogous issues were raised recently by Levin in relation to the variable duration of infectiousness of HIV-positive individuals, the basic reproduction number and the rate of increase of HIV/AIDS incidence in a population [129]. The main use of the basic and founder case reproduction numbers for tuberculosis is that they provide insight into the 'potential for successful transmission of a sputum-positive case' (or the 'potential reproductive capacity', as described by Macdonald [125] — see page 187). Of the two measures, the founder case reproduction number is more appropriate for measuring this 'potential for successful transmission' than the basic reproduction number, given that it takes into account the decline in the effective contact number.

The secular decline in the basic and founder case reproduction numbers is intuitively reasonable, and occurred despite changes in the age-structure of the population over time. During the 1900s, for example, a large proportion of the population in England and Wales comprised infants and young children, as compared with that in later years, as illustrated in Figure 5.21. As demonstrated in section 4.3.2, infants and young children have lower lifetime risks of developing sputum-positive disease than adults. Hence, had there been no change in the effective contact number over time, a founder case would have led to fewer secondary cases if introduced during the early 1900s, as compared with a similar founder case introduced in 1930, for example.

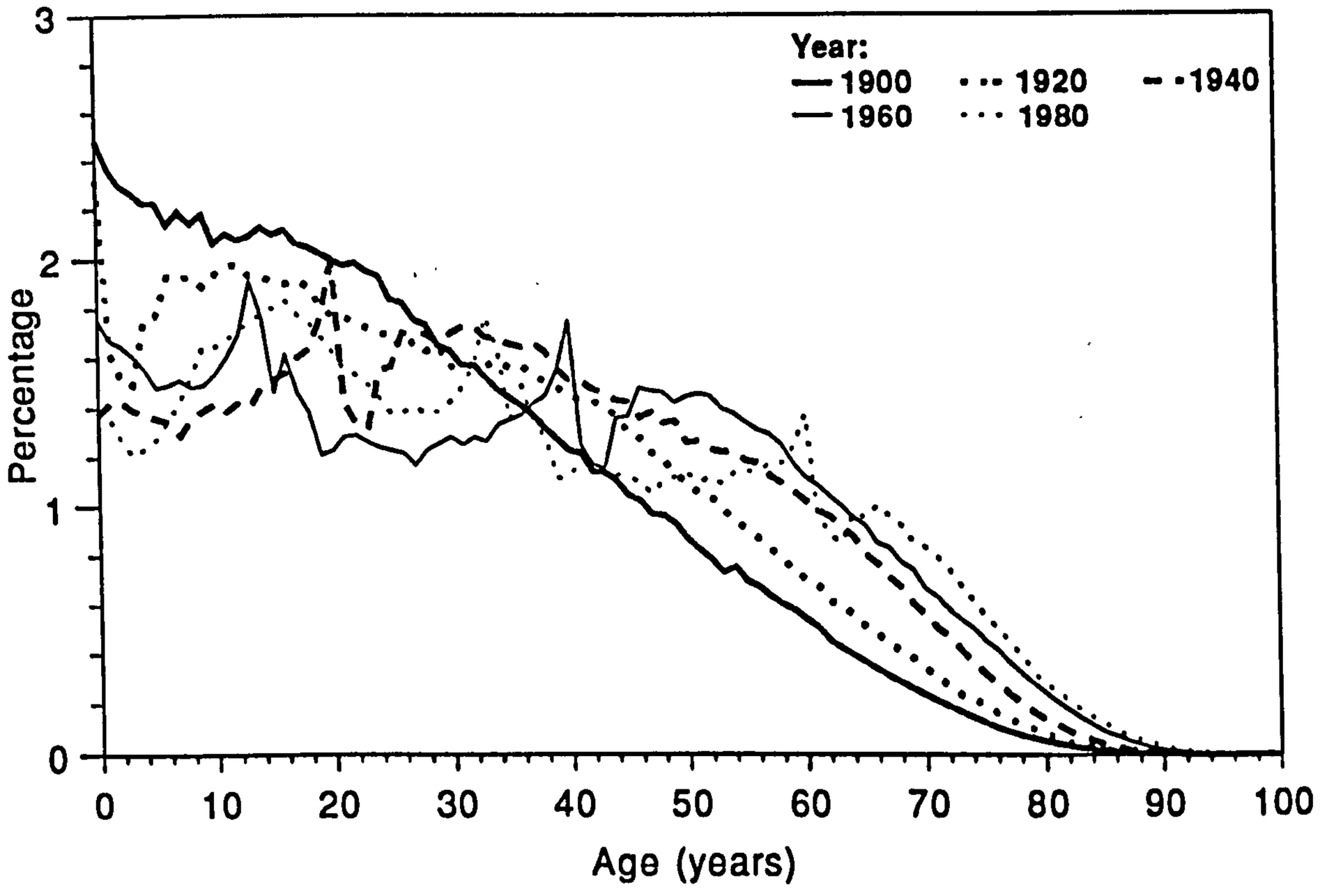


Figure 5.21: Age distribution of the male population in England and Wales during the period 1900–1990 in England and Wales.

In these analyses, the decline in the basic and founder case reproduction numbers is attributed to the decline in the effective contact number. In reality, such a decline could have also occurred as a consequence of a decline in the lifetime risks of developing sputum-positive disease, attributable to factors discussed in section 4.4.

Overall, the analyses suggests that neither the basic nor net reproduction numbers is ideal for analyzing the transmission dynamics of *M. tuberculosis*, given that fundamental epidemiological parameters, such as the effective contact number can change appreciably during a serial interval. This issue is also potentially relevant to other diseases with long serial intervals (e.g. AIDS, leprosy). On the other hand, we can obtain some insight into the *short-term* trend in disease incidence by studying parameters which are related to the basic and net reproduction numbers, but which describe the expected number of secondary cases arising from an infectious source *per unit time* in an uninfected population ($\tilde{R}_0(t)$) and in a population consisting of both infected and uninfected individuals ($\tilde{R}_n(t)$).

The relationship between $\tilde{R}_0(t)$ and $\tilde{R}_n(t)$ is analogous to that between the basic and net reproduction numbers for tuberculosis in a hypothetical population without age-structure and in which there is no change in the effective contact number. This relationship is

described by the following equation (see also Appendix C.1)

$$R_n = R_0(1 - pr_d) \tag{5.22}$$

where p denotes the prevalence of infected individuals in the population, and r_d denotes the protection imparted against reinfection and subsequent sputum-positive disease by having experienced infection in the past. Thus $\tilde{R}_0(t)$ and $\tilde{R}_n(t)$ are related in the following way:

$$\tilde{R}_n(t) = \tilde{R}_0(t)(1 - pr_d) \tag{5.23}$$

The magnitude of $\tilde{R}_n(t)$, as given by this expression, should reflect the likely magnitude of the *short-term* trend in the incidence of disease *attributable to recent infection or reinfection.* An $\tilde{R}_n(t)$ of less than 1 per year, for example, occurring when:

$$p > \frac{1}{r_d}\left(1 - \frac{1}{\tilde{R}_0(t)}\right) \tag{5.24}$$

would be associated with a decreasing incidence of disease attributable to recent infection/reinfection.

This logic could be extended to derive expressions for the minimum proportion of a population which would have to be immunized for the short-term disease incidence attributable to recent infection/reinfection to decrease. Though the practical utility of these expressions is complicated by the variable protection afforded by BCG vaccination, which may depend on geographical location or other factors [81], we describe the kind of reasoning which could be applied below.

Suppose that a proportion p_v in the population is totally protected against initial infection (in the short-term) by vaccination, then $\tilde{R}_n(t)$ and $\tilde{R}_0(t)$ are related in the following way (see Appendix C.2):

$$\tilde{R}_n(t) = \tilde{R}_0(t)(1 - p_v - pr_d) \tag{5.25}$$

For the disease incidence attributable to recent infection/reinfection to decrease in the short-term as a consequence of vaccination, $\tilde{R}_n(t)$ has to be less than 1 (secondary case per year), and hence p_v has to satisfy:

$$p_v > \left(1 - pr_d - \frac{1}{\tilde{R}_0(t)}\right) \tag{5.26}$$

If initial infection imparts 50% protection against reinfection and subsequent sputum-positive exogenous disease, the prevalence of infection is 60%, and $\tilde{R}_0(t)$ is 2 (secondary

cases per year), for example, then by this equation, the prevalence of immunized individuals should exceed 20% for the disease incidence attributable to recent infection/reinfection to decrease in the short-term. Although equations 5.25 and 5.23 reflect the relationship between $\tilde{R}_n(t)$ and $\tilde{R}_0(t)$ in a population without age-structure, and hence are over simplistic for describing the tuberculosis situation in a realistic setting, they nevertheless indicate the kind of measures and logic which, *if extended*, might be useful for assessing short-term trends in the disease incidence.

Chapter 6

Discussion and conclusions

> "However, for the eventual eradication of tuberculosis it is not necessary that transmission be *immediately* and *completely* prevented. It is necessary only that the rate of transmission be held permanently below the level at which a given number of infection-spreading (i.e. open) cases succeed in establishing an equivalent number to carry on the succession. If, in successive periods of time, the number of infectious hosts is continuously reduced, the end result of this diminishing ratio, if continued long enough, must be the extermination of the tubercle bacillus.
>
> Bearing in mind this principle, it is a fair inference that in this country as a whole we have already reached the stage at which the biological balance is against the survival of the tubercle bacillus, for year by year the mortality from tuberculosis is decreasing."
>
> Frost (1937) [2]

For tuberculosis, the issue of how and whether a given infectious case leads to subsequent cases is complicated, given that

1. only a small proportion of infected individuals develop (infectious pulmonary) disease during their lifetime,

2. the risk of developing disease, and hence the time interval between infection and disease onset, is a function of the age at infection, and

3. infection with or without an early disease episode does not confer solid immunity against subsequent reinfection and disease.

As a result of these complexities, the secular decline in the overall mortality rates over the past century in developed countries (in the absence of effective treatment until the 1950s) masks many changes in the underlying age and time-specific patterns in the overall disease incidence and in the incidence of disease attributable to either initial infection or subsequent reinfection. The objective of this thesis has been to explore these underlying changes on the basis that they should first be understood before we can make sound inferences about the necessary criteria for the eventual eradication of tuberculosis, and whether or not 'the biological balance' was 'against the survival of the tubercle bacillus' in the past. This work has provided insights in three areas of tuberculosis and infectious disease epidemiology, which we discuss below. The key findings are summarized on pages 246-249.

6.1 Insights obtained

6.1.1 The natural history of tuberculosis

One of the benefits of using a theoretical modelling approach to explore long-term disease dynamics is that it provides an opportunity to collate knowledge on the pathogenesis of disease and to identify the most important processes which require further understanding. In developing this model we have found that the conventional definitions of tuberculous disease, as being either 'primary' or 'post-primary', depending on whether it occurs within five years of initial infection or sometime thereafter [16], do not adequately distinguish between the different risks of developing disease experienced by individuals (see section 3.1.1). The model developed, TBDYN3, has therefore assumed that these risks differ by age and according to whether individuals are at risk of developing their first primary disease episode, endogenous and exogenous disease, where exogenous disease has been defined as the first disease episode within 5 years of reinfection, and the definition of endogenous disease is extended to include disease occurring within five years of infection/reinfection if individuals have experienced disease but no further reinfection.

The modelling work has highlighted the need for better understanding of the mechanism of reinfection. Although there is much evidence showing that reinfection can and does occur (see section 1.1), there are no data indicating whether the age and sex-specific risk

of reinfection is higher, lower or identical to that of first infection, and how soon after an initial infection it can occur. This uncertainty probably explains why most of the models of the transmission dynamics of *M. tuberculosis* created hitherto (all, excepting the work of Sutherland *et al* [7]) have neglected to consider the effect of reinfection on the overall disease incidence. In developing TBDYN3, we have assumed that it is important to incorporate this factor and to examine its implications, rather than avoid it.

Overall, the estimates of the age-specific risks of developing the first primary, endogenous and exogenous disease for England and Wales, as derived using TBDYN3, are of a realistic order of magnitude. The risk of developing the first primary episode among adults, for example, compares well against the estimate obtained from the UK MRC BCG trial during the 1950s [29], and the corresponding risk of developing endogenous disease is similar to that obtained by Styblo [6] and Horwitz [49] for the Netherlands and Denmark respectively (see section 1.2.1.4). The fact that the estimated risk of developing the first primary episode exceeds the risks of developing both endogenous and exogenous disease (for all age groups, excepting 0–10 year olds) suggests that initial infection and/or disease episode may confer some protection against subsequent disease. This finding is consistent with results from observational studies in developed countries, which have found a lower disease incidence among individuals presumed to be 'infected' (i.e. weakly tuberculin-positive), as compared with that among tuberculin-negative individuals (see section 1.2.1).

The disease risk estimates derived here lead to realistic patterns in the age-specific disease incidence among cohorts born during the pre-chemotherapy era, and compare well against those reflected in the observed mortality rates. The similarity between the predicted and observed mortality rates for this era is especially striking (see section 3.4.2).

The results suggest that the single most important factor which determined the magnitude of the disease incidence over the past century in England and Wales was the level and decline in the risk of infection. This determined

1. the age pattern of initial infection and hence the risk of developing disease, and

2. the risk of reinfection.

Model predictions suggest that most of the adult disease incidence during the pre-chemotherapy era, when the risk of infection was higher than that in many developing countries today, was attributable to exogenous disease and only a small proportion was

attributable to first primary episodes and to endogenous reactivation. Although these results were obtained by assuming that the risk of first infection is identical to that of reinfection, which may not be strictly realistic (see section 3.1.1), three observations indicate that reinfection must have been important in determining patterns of disease incidence in the past.

First, exclusion of exogenous disease from TBDYN3 led to unrealistically high estimates of the risks of developing the first primary episode for adults (i.e. 23% during the first five years after initial infection), and failed to match the decline in the *notification* rates especially for middle aged and elderly individuals (see section 3.3.3.1). Most individuals in this age-range during this period must have been infected during infancy, and hence the decline in the notification rates could not have been attributable to a reduction in the incidence of first primary episodes. Alternatively, if we assume that these individuals did not experience reinfection, this reduction could only have been achieved with a dramatic reduction in the risk of developing endogenous disease, which seems unlikely. By this logic, the reduction in the overall disease incidence for these individuals was most likely to have been attributable to a decline in the incidence of exogenous disease, which occurred as a consequence of the rapid decline in the risk of reinfection.

Second, the *mortality rates among 45-54 and 55-64 year olds* halved between 1901 and 1930 (e.g. from 313 to 157 per 100,000 between 1901 and 1930 for 45-54 year olds). By applying an argument similar to that above, we can conclude that it is unlikely that this was attributable either to a reduction in the incidence of first primary episodes, or to a dramatic reduction in the risk of developing endogenous disease. Consequently, it is most likely to have occurred as a consequence of a decline in the exogenous disease incidence.

The third reason relates to the *mortality rates among adolescents and young adults* during the pre-chemotherapy era, which declined more slowly than for infants. By definition, morbidity among infants is primary in origin and hence reflects recent transmission and the current trend in the annual risk of infection. The fact that the mortality among adolescents and young adults declined more slowly than for infants suggests that the disease incidence among adolescents was not entirely attributable to first primary episodes. Model predictions suggest that, as a consequence of the decline in the risk of infection, the proportion of individuals reaching adolescence and young adult life without experiencing infection *increased* during the first half of this century. According to TBDYN3, this should have led

to an increase in the incidence of first primary episodes among these individuals. Styblo predicted a similar increase in the incidence of first infections for adolescents in the Netherlands during this time [6]. If it is assumed that individuals did not experience reinfection, then given the higher risks of developing disease shortly after infection, as compared with those many years thereafter, this *increase in the incidence of first primary episodes would have translated into an increase in the overall disease incidence* for adolescents during this time, unless the risk of developing endogenous disease declined dramatically during this time, which seems unlikely. The fact that the mortality rates among adolescents actually declined suggests that another factor must have also affected the disease incidence. The most likely factor which could have contributed to this decline is the declining annual risk of infection, which implies that some adolescents and young adults in the early years of this century were reinfected and developed exogenous disease.

In his classic paper considering the decline in tuberculosis over time, Frost [2] also foresaw that a decline in the risk of infection would postpone infection until later in life or adolescence, when the risk of developing disease exceeds that in early childhood. Analogous issues have been discussed in relation to other diseases such as polio, mumps and rubella [119]. The decline in the risk of infection with the polio virus, for example, is believed to have led to an increase in the average age at infection, and hence an increase in paralytic polio among young adults in developed countries during the 1950s.

6.1.2 The incubation period, serial interval and lifetime risks of developing tuberculosis

The incubation period and the serial interval are two of the most important parameters underlying the dynamics of any infectious disease. The incubation period describes when, given infection, an individual is likely to develop disease and the serial interval indicates when that individual is likely to infect others.

Much of the literature on these statistics is based on observational studies of the transmission dynamics of 'simple' diseases, such as measles, in which

1. most diseased individuals are infectious,

2. infection induces a permanent solid immunity and,

3. the single disease episode experienced is attributable to one (initial) infection event.

The fact that individuals can develop disease following reinfection means that the conventional definitions of the incubation period and serial interval must be *modified* before they can be applied to tuberculosis. As demonstrated in this thesis, these definitions must account for age, secular trend in the risk of infection and mortality rates experienced.

Both the incubation period and the serial interval of tuberculosis can be analyzed either 'retrospectively' (as the time interval since initial infection for individuals diseased at a given age and in a given year) or 'prospectively' (as the time interval until the first disease or infectious episode for individuals infected at the same age in a given year). For both retrospective and prospective analyses, the ultimate distributions obtained also depend on whether we distinguish between disease attributable to an initial infection or to a subsequent reinfection. If we explore the time interval from initial infection at a given age in a given year until the first subsequent sputum-positive episode, for example, then the distribution obtained resembles an 'observed prospective' serial interval. If we explore the time interval from initial infection at a given age in a given year until the first subsequent sputum-positive episode *attributable* to that infection, then the distribution obtained corresponds to an 'actual prospective' serial interval. Both these distributions are also related to the overall lifetime risks of developing disease.

It is probably because of these complexities that the incubation period, serial interval and lifetime risks of developing tuberculosis have not previously been analyzed in detail. It is not even widely appreciated that the lifetime risks are age-dependent — the only detailed published discussion of this issue is that of Comstock [121], who, while recognizing that individuals face high risks of developing disease after infection during adolescence, suggested that the overall lifetime risk following infection in infancy might be higher as a consequence of a low annual risk operating over many years. In contrast, the estimated lifetime risk predicted by TBDYN3 for twenty year olds infected after 1950, for example, exceeds that for 10 year olds by a factor of about three (i.e. corresponding to lifetime risks of 17% vs 6% for 20 year olds and 10 year olds respectively). The corresponding risks of developing infectious pulmonary disease differ by a factor of about 4 (i.e. 12% vs 3%) (see section 4.3.2).

The fact that the distributions of the serial interval and incubation periods depend on both the age of individuals and on calendar year, and hence on assumptions as to the trend in the risk of (re)infection, complicates their utility. This also applies to the incubation

periods and serial intervals for other infectious diseases, for which the time interval between infection and disease onset can be very long. The analogy between leprosy and tuberculosis is obvious. The time interval between infection and onset of leprosy can be long and variable, ranging between 9 months and at least 20 years [130] although the mode of transmission is still poorly understood. The fact that both diseases are caused by mycobacterial pathogens enhances this analogy.

Herpes varicella-zoster (HVZ) also shares some similarities with tuberculosis. After first exposure to the HVZ virus, individuals usually develop chickenpox ('varicella'), and can subsequently experience shingles ('zoster') sometime later in their lives. The occurrence of zoster is typically associated with immunosuppression although it may also occur after reinfection with HVZ [131]. The estimated lifetime risk of developing zoster given infection with HVZ is about 23% [131], although this may be unreliable, given that the estimate has not been broken down by age, and is based only on reports from general practitioners. This overall lifetime risk is similar in magnitude to that of developing respiratory tuberculosis for 20 year olds in 1900 (see section 4.3.2).

6.1.3 The basic and net reproduction numbers

The basic reproduction number is generally defined as the average number of secondary cases resulting from a typical infectious case in a 'totally' susceptible population (see page 188). There have been two main areas in which the basic reproduction number has been used in the past to elucidate the dynamics of infectious diseases.

First, for 'simple' viral diseases, the relationship between the basic and net reproduction numbers (see definition on page 188) has been applied to assess the impact of interventions, such as vaccination, on the trend in disease incidence [119] (see section 5.1.1). Second, it has been used to explore the likely growth rate in disease incidence, and the 'epidemic doubling time' subsequent to the introduction of an infectious case into a 'totally susceptible population'. Such calculations have been applied in particular to the growth of the HIV/AIDS epidemic in developing and developed countries [119], and in particular subgroups of these populations.

As demonstrated in this thesis, the definitions and interpretation of the basic and net reproduction numbers for tuberculosis are complicated by three facts:

1. individuals can be reinfected,

2. the serial interval is long and variable, and

3. the effective contact number (see definition on page 193) appears to have declined dramatically over time.

The fact that individuals can be reinfected with *M. tuberculosis* and develop exogenous disease, for example, means that even *infected* individuals can be considered 'susceptible'. A 'totally susceptible' population in the definition of the basic reproduction number for tuberculosis is therefore more adequately described as one in which no individuals have hitherto been infected.

There has been little discussion of the basic reproduction number and its biological interpretation for diseases for which epidemiological parameters, such as the rates of contact between individuals, undergo long-term secular changes. Most analyses of the basic and net reproduction numbers for various diseases, have assumed that, in the absence of an obvious intervention, such parameters do not change over the time period considered. This is a reasonable assumption if we explore only the short-term dynamics spanning several years or 2–3 decades for a given disease. This assumption is not appropriate for analysing the dynamics underlying the decline in tuberculosis over many decades, or even centuries, given that the effective contact number appears to have declined appreciably over time.

Most analyses of the basic reproduction number for infectious diseases also assume that interventions, such as chemotherapy or behavioural changes, lead to *permanent* changes in epidemiological parameters, such as the duration of infectiousness, or the contact between individuals, and hence to fixed changes in the basic reproduction number. In these instances, the magnitude of this *modified* basic reproduction number determines the persistence of infection in the population [126]. It is important to recognize that few interventions lead to an instantaneous reduction in a given epidemiological parameter and hence in the basic reproduction number.

The secular decline in the effective contact number for tuberculosis is attributable to the effect of many 'interventions', including the reduction of crowding and improved ventilation in living conditions, removal of infectious cases from the community to sanatoria (before 1950) or through treatment after 1950, an improvement in nutrition, which increased individuals' resistance to infection (and probably, also to disease following infection), and an improvement in general hygiene. The fact that the effective contact number declined over

a long time period suggests that these 'interventions' for tuberculosis were continuously improving.

This complicates the relevance of the basic reproduction number for tuberculosis, given that it is interpreted as a parameter which describes the likelihood of continued transmission in a population *in the absence of any (further) control or intervention.* To be consistent with this interpretation and on the basis of the above discussion, for example, we have defined the basic reproduction number for tuberculosis as (see page 196):

> The average number of secondary infectious (pulmonary) cases resulting from a 'typical' infectious (pulmonary) case during its entire infectious period, following its introduction into a population where no individual has hitherto been infected with the tubercle bacillus, *assuming that environmental conditions and the effective contact number do not change further over time.*

The fact that the effective contact number declined over time means that the magnitude of the basic reproduction number for tuberculosis in a given year in the hypothetical situation in which a founder case is introduced into an uninfected population in that year, does not necessarily reflect the likely subsequent trend in the disease incidence. For this reason, the basic reproduction number is also less appropriate for interpreting the decline in tuberculosis during the past century than the founder case reproduction number (see definition on page 197), which we have defined to be a function of the declining effective contact number.

This thesis has also raised issues relating to the interpretability of the *net* reproduction number for tuberculosis. This was estimated to have fallen below one only after the late 1920s (see section 5.3.2.3), even though the observed mortality rates and the disease incidence predicted by TBDYN3 declined at least since 1900. This paradoxical outcome is a consequence of the long and variable serial interval for tuberculosis, depending on both the age of newly infected or reinfected individuals and the calendar year, which means that the disease incidence in a given year is a function of the net reproduction numbers over many years. Some secondary cases from an infectious case in 1900, for example, had onset only many years thereafter, when there were relatively few cases attributable to recent infection/reinfection. Hence a net reproduction number for tuberculosis exceeding one in 1900 did not translate into an increase in the *overall* disease incidence.

By a similar reasoning, it would also be possible for the net reproduction number to be less than one in a given year, but for the disease incidence to increase at the same time.

This leads to the question of how best to describe the "biological balance" of the tubercle bacillus or of any infectious disease-causing agent, as described by Frost (see quote on page 232). Intuitively there are two aspects to the biological balance of any infectious disease:

1. the biological balance at the level of an *individual* in the population. This reflects whether a given infectious source leads to more than one other case on average. The state of this balance is reflected by whether or not the net reproduction number exceeds one, and

2. the biological balance at the *population* level. This represents the *net effect* of all the net reproduction numbers *in the past (including that of the present) on the current disease* incidence in the population. This can be given by an overall weighted average, R_w, of the net reproduction numbers during preceding years. Intuitively, the magnitude of R_w at a given time t should reflect the *trend* in the disease incidence at that time.

For infectious diseases which have short serial intervals of duration s, *which do not vary substantially between individuals*, the trend in disease incidence at a given time t reflects the magnitude of the net reproduction number at time $t - s$. A declining disease incidence at time t indicates that the net reproduction number was less than one at time $t - s$, and an increasing disease incidence reflects a net reproduction number of greater than one at time $t - s$. Hence for these diseases, the biological balance at the individual level effectively reflects that at the population level.

As demonstrated in this thesis, there is a substantial difference between these two measures of the biological balance for tuberculosis. Until the mid 1920s, for example, the biological balance favoured the survival of the tubercle bacillus in the individual, but was against its survival in the population. Of the two measures, the decline in tuberculosis is therefore best described by the biological balance of the tubercle bacillus in the population.

Overall, we see that neither the basic nor net reproduction numbers is really appropriate for analysing the transmission dynamics of *M. tuberculosis*, given that fundamental epidemiological parameters, such as the effective contact number may undergo a secular

decline during a period of time as long as a serial interval. Our analyses of these measures have nevertheless raised other issues relating to the dynamics underlying the decline in tuberculosis. The fact that individuals can be reinfected and experience exogenous disease, and the decline in the effective contact number meant that the *mechanism* by which infectious cases led to secondary cases changed over time. The high risk of infection (and hence the high effective contact number) during the 19th century meant that, at the turn of the century, only infants and young children were still at risk of first infection. Hence each sputum-positive case alive in 1900 could generate secondary cases 'only' through *infecting infants and young children* or through *reinfecting infected adults.* The decline in the effective contact number after 1900 led to a gradual increase in the prevalence of uninfected adults. By 1970, sputum-positive cases on average led to secondary cases mainly by inducing primary infections rather than reinfections.

The reduction of crowding and improvements in general hygiene and nutrition may have also led to declines in the effective contact numbers, and in the basic reproduction numbers for other diseases, such as diphtheria, polio and measles, which were endemic in developed countries during the first half of this century. The proportion of children attending schools and crêches has increased over time, which implies that the decline in the effective contact number, and hence the basic reproduction numbers for these disease, which are associated with serial transmission in children, could have decelerated. This increased mixing among children is less important in determining the basic reproduction number for tuberculosis, given that only a small proportion of children develop sputum-positive pulmonary tuberculosis and are infectious. Hitherto, there have been no published analyses exploring whether long-term changes in the effective contact number have affected the basic reproduction number for any infectious disease, although various authors have speculated whether the basic reproduction number for polio could have declined as a consequence of improving standards of cleanliness [119].

Results from TBDYN3 suggest that the founder case reproduction number for tuberculosis was only slightly greater than 3, even at the start of the twentieth century. The basic reproduction number for scarlet fever, on the other hand, has been estimated to have been about 5–8 during the period 1908 and 1916, as compared with 4–5 for diphtheria and 11–17 for measles [119]. The results from TBDYN3 are therefore consistent with the general belief that tuberculosis is "less infectious" at some level than these other diseases.

6.1.4 Concluding remark

Overall, we see that the complicated natural history of tuberculosis, depending on the different risks of developing disease according to both age and time since infection, and on whether individuals experience reinfection, led to complex patterns and changes in the dynamics underlying the decline in mortality in developed countries. Given these complexities, the intuition of Frost's conclusion regarding the biological balance of the tubercle bacillus is all the more remarkable.

6.2 Directions for future work

Given that this thesis has provided new insights into the risks of developing disease, and into the changes in the age-specific patterns underlying the decline in tuberculosis in developed countries, there are several ways in which this work should be extended further.

Intuitively, similar estimates for the risks of developing disease should be obtained if the model were to be applied to other developed countries. As we saw in section 3.3.4, the estimated risk of developing the first primary episode for Dutch adults exceeded that for England and Wales by a factor of two. There are several reasons why this may have occurred (see section 3.3.4), and further insight into these factors could be obtained by applying the model to data from other developed countries.

Our analyses have focused only on the natural history of respiratory tuberculosis among males. There are important gender differences in the epidemiology of tuberculosis which have not been studied extensively. During the pre-chemotherapy era and after 1950 respectively, for example, the mortality rates and notification rates among adolescent and young adult females in England and Wales exceeded those among males. At older ages, these rates have been higher among males. Additionally, the rate of decline in the notification rates among older women since the mid 1950s has been *slower* than that among males. This could be attributable to many factors, such as differences between the risks of developing disease, in case-recognition, or in changing smoking patterns (see e.g. [122]). Alternatively, this suggests that the morbidity among elderly females was less sensitive to the decline in the risk of infection and hence a greater proportion was endogenous in origin, as compared with that in males. This might be attributable to a lower *risk of reinfection* among elderly females as compared with that among males.

This conclusion could well be consistent with results from tuberculin surveys in both developed and developing countries. The greater prevalence of tuberculin sensitivity among adult males, as compared with that among females, for example, could be attributable to the fact that a greater proportion of males had been reinfected, and hence their tuberculin response boosted, as compared with females. The mechanism behind these trends could be explored using the modelling approach developed in this thesis.

It is unclear whether the risks of developing disease obtained in this thesis apply also to extrapulmonary forms. In the US and England and Wales, for example, the notifications of extrapulmonary tuberculosis have been declining at a slower rate than those of respiratory tuberculosis even before the start of the HIV epidemic [96,132]. The reasons are not fully understood, although they are partly attributable to changes in immigration patterns from countries with a high prevalence of tuberculosis. These trends can also be analyzed using the approach developed in this thesis.

Most of the research of the epidemiology of tuberculosis has been carried out in developed countries, where tuberculosis has been declining since the beginning of this century, and at least since the middle of the 19th century in England and Wales. It is likely that the basic natural history of tuberculosis differs between developed and developing countries. In South India, for example, the risk of infection was about 2-6% during the 1970s [26], which is similar to that in England and Wales during the period 1920-1950, but there is little evidence of a secular decline since then. The few data that exist on the age-specific incidence of disease in developing countries (e.g. [25,26]), suggest that there is no peak in the incidence during young adult life and that BCG vaccination affords little protection against tuberculosis in many developing countries, as compared with the approximately 77% efficacy found for England and Wales [39]. Some of these differences in the epidemiology of tuberculosis between developing and developed countries may be attributable to differences in the prevalence of environmental mycobacteria, which are thought to impart some protection against infection (e.g. > 95% of 15–19 year olds in the South India/Chingleput vaccine trial population were strongly positive to *Mycobacterium avium-intracellulare* antigens [26]).

All these factors suggest that the age-specific risks of developing the first primary, endogenous and exogenous disease in developing countries must differ from those in England and Wales, and, given the high risks of infection, a large proportion may be exogenous in origin. They may also imply that the risks of reinfection differ between developing and

developed countries. The approach employed in this thesis could be adapted to study these differences and hence further the understanding of the natural history of tuberculosis.

The application of DNA fingerprinting to distinguish between strains of tubercle bacilli in sputum isolates potentially provides the opportunity to understand the natural history of tuberculosis in greater depth. These data can themselves be analyzed further using modelling techniques, such as those described in this thesis to yield further insight into the temporal change in the relative frequency of reinfection and reactivation disease and thus into the risks of being reinfected subsequent to initial infection.

The emphasis in the thesis has been on understanding the basic natural history of tuberculosis without the complication of HIV. Studies suggest that HIV-positive individuals face higher risks of developing tuberculosis than do those who remain HIV-negative [44], although this depends on the degree of immuno-suppression, e.g. as measured by CD4 count [133]. The likelihood of a disease episode being sputum-positive and hence the contribution of HIV-positive individuals to the overall risk of infection also depends on the stage of immuno-suppression at which infection is acquired. The modelling approach employed here could also be adapted to assess the relative contribution to the disease incidence and risk of infection in the past and how HIV is likely to shape the epidemiology of tuberculosis in the future.

Given secular trends and cohort patterns in age-specific disease incidence and the long duration of the incubation period and serial interval, the application of appropriate mathematical models should make an important contribution to our understanding of the natural history and control of tuberculosis in the future.

Summary

1. An age-structured deterministic model (TBDYN3) of the transmission dynamics of *M. tuberculosis* in England and Wales since 1900 has been developed, using infection risks estimated from tuberculin surveys and childhood meningitis statistics, and published demographic and vaccination data. The model extends the work of Sutherland *et al* [7], and is applied to estimate the age-specific risks of developing 'primary', 'endogenous' and 'exogenous' disease by fitting model predictions to notifications of respiratory tuberculosis from England and Wales over the period 1953–1988. TBDYN3 has been used to estimate the secular trend in the relative contribution of these disease forms to the overall incidence and to derive definitions and estimates of the incubation period, serial interval, basic and net reproduction numbers.

2. TBDYN3 assumes that the risks of infection are time-dependent (i.e. 14% in 1900, declining by 4% pa until 1950 and by 13% pa thereafter), but not age-dependent. The risk of reinfection is assumed to be identical to that of first infection. The risks of developing disease during the first five years after initial infection (the 'first primary episode') or reinfection (exogenous disease) depend on the age at infection and reinfection respectively (i.e. constant for 0–10 year olds, increasing linearly between the ages 10 and 20 years, and constant above age 20 years) and decline with time since infection and reinfection. The risk of developing endogenous disease depends only on the *current* age of an individual.

3. The estimated risks of developing the 'first primary episode' were about 4%, 9%, and 14% for individuals infected when aged 0–10, 15 and over 20 years respectively. For all age groups, except for 0–10 year olds, these risks exceeded those of developing exogenous disease. The annual risks of endogenous disease were negligible for 0–10 year olds, 0.014% and 0.029% for 15 year olds and those aged over 20 years respectively.

4. Disease risk estimates are also obtained by applying TBDYN3 to data from the Netherlands used by Sutherland *et al.* The estimates compare well against those of Sutherland *et al* [7], but the risks of developing the first primary episode exceed those obtained for England and Wales. Reasons for the apparent discrepancy are discussed.

5. The disease risk estimates predict a peak in both the disease incidence and mortality rates during young adult life for successive birth cohorts during the pre-chemotherapy era. This is consistent with the observed patterns in age-specific mortality in successive cohorts [3, 4, 27].

6. Estimates of the relative contribution of first primary, endogenous and exogenous disease to the overall age-specific disease incidence since 1900 are derived. These suggest that much of the disease experienced by individuals aged over 30 years, at least until 1940, was attributable to exogenous disease. The introduction of chemotherapy during the 1950s, which accelerated the decline in the risk of infection after 1950, led to a sharp decline in the overall incidence of disease for all age groups. For older individuals this was associated with a sharp reduction in the exogenous disease incidence.

7. Estimated lifetime risks of developing respiratory and sputum-positive tuberculosis depend on age and declined with calendar year of infection. Individuals first infected when aged 20 years faced higher lifetime risks of *respiratory* disease than did any other age group (e.g. 26% and 17% in 1900 and after 1950), which were three times greater than the risks for 10 year olds. The estimated lifetime risks of developing *sputum-positive* disease increased consistently with the age at infection and were about 19% and 12% for 20 year olds infected in 1900 and after 1950 respectively.

8. Definitions of the incubation period and serial interval are complicated for tuberculosis, as they must take into account the age of individuals, the secular trend in the risk of infection and the mortality experienced. Both can be analyzed 'retrospectively' (as the time interval since initial infection for individuals diseased at a given age and in a given year) and 'prospectively' (as the time interval until the first disease episode for individuals infected simultaneously at a given age and in a given year).

9. *Retrospective* analyses of the incubation period and serial interval suggest that about 3% of diseased 20 year olds in 1900 had experienced initial infections in the same year, and the greatest proportion ($\approx$ 25%) had been infected in their first year of life. The infection/reinfection event responsible for the disease episode among these individuals occurred during the preceding year for 60% and in the first year of life for less than

1%.

10. *Prospective* analyses of the *incubation period* show that about 25% of infants infected in 1900 who developed respiratory disease, did so during the first five years. The proportion doing so thereafter declined as the time since initial infection increased. Similar patterns are seen for individuals infected at other ages in 1900, except that the proportion developing disease during the first five years *increases* with the age at infection. Most of the disease occurring among individuals infected at any age in 1990 is estimated to have occurred within five years of infection.

11. *Prospective* analyses of the *serial interval* (by successive 5 year periods) indicate that the proportion of individuals infected in the first year of life in 1900 who developed sputum-positive disease sometime thereafter, was maximum between the 20th and 24th years after initial infection ($\approx$ 15%). The greatest proportion of such individuals infected in 1990 (23%) did so during the first five years after infection.

12. Definitions of the basic and net reproduction numbers for tuberculosis have to take into account the fact that both infected and uninfected individuals are 'susceptible', given that reinfection can occur.

13. Both the basic and net reproduction numbers depend on the 'effective contact number' (the number of individuals 'effectively contacted' by a sputum-positive case in a given year. An 'effective contact' by a sputum-positive case is here defined as one which would lead to infection if the contacted individual had never been infected in the past). Estimates of the effective contact number are derived for the period 1900–1990 by comparing the annual risk of infection against the prevalence of sputum-positive cases.

14. Estimates suggest that on average, each sputum-positive case effectively contacted 23 other individuals in 1900, of whom 12 were reinfected, 3 were infected and the contact did not lead to transmission in the remaining 8, because they were either already experiencing disease or were at high risk of developing disease attributable to a preceding infection/reinfection event. The effective contact number declined over time, and by 1950, each sputum-positive case effectively contacted 11 other individuals on average.

15. Estimates of the net reproduction number suggest that it was slightly greater than one between 1900 and 1930, declined slowly until 1950, after which the decline accelerated. This is interpreted in the context of the decline in mortality from tuberculosis at least since 1900.

16. Inferences about the biological definition of the basic reproduction number since 1900 are obtained by simulating the introduction of a sputum-positive case into a population identical in age-structure, demographic trends and in its effective contact number to that of England and Wales in given years and estimating the number of sputum-positive cases arising as a result of the 'founder case'. We define this measure as the 'founder case reproduction number'.

17. A 'founder' case would have led to about 3 other cases in 1900, to about 2 by 1950, and to less than 1 after 1960. The implications for the relationship between the basic and net reproduction numbers for tuberculosis are discussed.

18. Directions for future research stemming from the work in this thesis are described.

Appendix A

Formulation and results from TBDYN3

A.1 PDEs describing the model formulation for TBDYN3

$$\frac{\partial U(a,t)}{\partial a} + \frac{\partial U(a,t)}{\partial t} = -(i(t) + v(a,t) + m_g(a,t))U(a,t) \qquad (A.1)$$

$$\frac{\partial V(a,t)}{\partial a} + \frac{\partial V(a,t)}{\partial t} = v(a,t)U(a,t) - m_g(a,t)V(a,t) \qquad (A.2)$$

$$\frac{\partial I(a,t,s)}{\partial a} + \frac{\partial I(a,t,s)}{\partial t} + \frac{\partial I(a,t,s)}{\partial s} = -((d_p(a-s,s) + m_g(a,t))I(a,t,s)$$
$$-k_L(s)I(a,t,s) \qquad (0 < s \leq 5) \qquad (A.3)$$

$$\frac{\partial P(a,t,\hat{s})}{\partial a} + \frac{\partial P(a,t,\hat{s})}{\partial t} + \frac{\partial P(a,t,\hat{s})}{\partial \hat{s}} = \int_0^5 d_p(a-s,s)I(a,t,s)ds$$
$$-(m_+(t,\hat{s})d_+(a) + m_g(a,t)d_-(a))P(a,t,\hat{s})$$
$$-r(a,t,\hat{s})P(a,t,\hat{s}) \qquad (A.4)$$

$$\frac{\partial L(a,t)}{\partial a} + \frac{\partial L(a,t)}{\partial t} = (I(a,t,5) + I_r(a,t,5))k_L(5)$$
$$+r(a,t,2)(P(a,t,2) + E_n(a,t,2) + E_x(a,t,2))$$
$$-(i(t) + d_n(a) + m_g(a,t))L(a,t) \qquad (A.5)$$

$$\frac{\partial I_r(a,t,s)}{\partial a} + \frac{\partial I_r(a,t,s)}{\partial t} + \frac{\partial I_r(a,t,s)}{\partial s} = -(d_x(a-s,s) + m_g(a,t))I_r(a,t,s)$$
$$-k_L(s)I_r(a,t,s) \qquad (0 < s \leq 5) \tag{A.6}$$

$$\frac{\partial E_x(a,t,\hat{s})}{\partial a} + \frac{\partial E_x(a,t,\hat{s})}{\partial t} + \frac{\partial E_x(a,t,\hat{s})}{\partial \hat{s}} = \int_0^5 d_x(a-s,s)I_r(a,t,s)ds$$
$$-(m_+(t,\hat{s})d_+(a) + m_g(a,t)d_-(a))E_x(a,t,\hat{s})$$
$$-r(a,t,\hat{s})E_x(a,t,\hat{s}) \tag{A.7}$$

$$\frac{\partial E_n(a,t,\hat{s})}{\partial a} + \frac{\partial E_n(a,t,\hat{s})}{\partial t} + \frac{\partial E_n(a,t,\hat{s})}{\partial \hat{s}} = d_n(a)L(a,t) - r(a,t,\hat{s})E_n(a,t,\hat{s})$$
$$-(m_+(t,\hat{s})d_+(a) + m_g(a,t)d_-(a))E_n(a,t,\hat{s}) \tag{A.8}$$

Boundary conditions:

$$U(0,t) = B(t);$$
$$I(a,t,0) = i(t)U(a,t);$$
$$I_r(a,t,0) = i(t)L(a,t)$$

For notational convenience, we denote $1 - d_+(a)$ by $d_-(a)$.

A.2 Approximations for the partial derivatives of the age and time-specific disease incidence with respect to fitted parameter estimates

For notational convenience in this appendix, we denote the 6 parameters to be estimated in the fitting procedure, $\{d_p(10,0), d_p(20,0)\}$, $\{d_n(10), d_n(20)\}$ and $\{d_x(10,0), d_x(20,0)\}$ by:

$$\{d_{p10}, d_{p20}\},\ \{d_{n10}, d_{n20}\},\ \{d_{x10}, d_{x20}\}$$

respectively.

The risk of developing the first primary episode for an individual first infected at age a, who has been infected for duration s, $d_p(a,s)$ can be expressed as:

$$d_p(a,s) = \begin{cases} d_{p10} d_r(s) & \text{for } a \leq 10, \\ \left\{ \dfrac{(d_{p20} - d_{p10})}{10} a + 2d_{p10} - d_{p20} \right\} d_r(s) & \text{for } 10 < a < 20, \\ d_{p20} d_r(s) & \text{for } a \geq 20. \end{cases}$$

where $d_r(s)$ is the risk of developing disease s years after initial infection, relative to that experienced in the first year. The expressions for the risk of developing exogenous disease for an individual reinfected at age a, and who has been reinfected for duration s, $(d_x(a,s))$ are analogous.

Similarly, the risk of developing endogenous disease for an individual of age a, $d_n(a)$ is given by:

$$d_n(a) = \begin{cases} d_{n10} & \text{for } a \leq 10, \\ \dfrac{(d_{n20} - d_{n10})}{10} a + 2d_{n10} - d_{n20} & \text{for } 10 < a < 20, \\ d_{n20} & \text{for } a \geq 20. \end{cases}$$

The number of new cases of first primary episodes $(P(a,t,0))$, endogenous $(E_n(a,t,0))$ and exogenous disease $(E_x(a,t,0))$ among individuals of age a at time t, in terms of $I(a,t,s)$, $L(a,t)$ and $I_r(a,t,s)$ (see definitions in Table 3.1), using a discrete, rather than continuous-time formulation, are given by:

$$P(a,t,0) = \sum_{s=0}^{4} I(a,t,s) d_p(a-s,s) \qquad \text{(A.9)}$$

$$E_n(a,t,0) = L(a,t)d_n(a) \tag{A.10}$$

$$E_x(a,t,0) = \sum_{s=0}^{4} I_r(a,t,s)d_x(a-s,s) \tag{A.11}$$

The *total disease incidence* rate among individuals of age a at time t, $\mathbf{Y}(a,t)$, is given by:

$$\mathbf{Y}(a,t) = \frac{P(a,t,0) + E_n(a,t,0) + E_x(a,t,0)}{N(a,t)} \tag{A.12}$$

(see also expression on page 105).

The corresponding disease incidence rate in specific age-groups (e.g. 5–14 year olds) is given by:

$$\mathbf{Y}(a_{5-14},t) = \frac{\sum_{a=5}^{14}\{P(a,t,0) + E_n(a,t,0) + E_x(a,t,0)\}}{\sum_{a=5}^{14} N(a,t)} \tag{A.13}$$

We use the short-hand notation $N(a_{5-14})$ for $\sum_{a=5}^{14} N(a,t)$

The partial derivatives of the total disease incidence rate in specific age groups (e.g. 5–14 year olds), $\mathbf{Y}(a_{5-14},t)$, with respect to

1. d_{p10} and d_{p20} are approximately:

$$\frac{1}{N(a_{5-14})}\sum_{a=5}^{14}\frac{\partial P(a,t,0)}{\partial d_{p10}} \quad \text{and} \quad \frac{1}{N(a_{5-14})}\sum_{a=5}^{14}\frac{\partial P(a,t,0)}{\partial d_{p20}} \quad \text{respectively,} \tag{A.14}$$

2. d_{n10} and d_{n20} are approximately:

$$\frac{1}{N(a_{5-14})}\sum_{a=5}^{14}\frac{\partial E_n(a,t,0)}{\partial d_{n10}} \quad \text{and} \quad \frac{1}{N(a_{5-14})}\sum_{a=5}^{14}\frac{\partial E_n(a,t,0)}{\partial d_{n20}} \quad \text{respectively,} \tag{A.15}$$

3. d_{x10} and d_{x20} are approximately:

$$\frac{1}{N(a_{5-14})}\sum_{a=5}^{14}\frac{\partial E_x(a,t,0)}{\partial d_{x10}} \quad \text{and} \quad \frac{1}{N(a_{5-14})}\sum_{a=5}^{14}\frac{\partial E_x(a,t,0)}{\partial d_{x20}} \quad \text{respectively.} \tag{A.16}$$

We use the following approximations for the partial derivative of the age-specific incidence of *first primary episodes* with respect to

(i) d_{p10}:

$$\frac{\partial P(a,t,0)}{\partial d_{p10}} = \begin{cases} \sum_{s=0}^{4} I(a,t,s)d_r(s) & \text{for } a-s \leq 10, \\ \sum_{s=0}^{4} I(a,t,s)\left\{-\frac{(a-s)}{10}+2\right\}d_r(s) & \text{for } 10 < a-s < 20, \\ 0 & \text{for } a-s \geq 20. \end{cases}$$

and (ii) d_{p20}:

$$\frac{\partial P(a,t,0)}{\partial d_{p20}} = \begin{cases} 0 & \text{for } a - s \leq 10, \\ \sum_{s=0}^{4} I(a,t,s) \left\{ \frac{(a-s)}{10} - 1 \right\} d_r(s) & \text{for } 10 < a - s < 20, \\ \sum_{s=0}^{4} I(a,t,s) d_r(s) & \text{for } a - s \geq 20. \end{cases}$$

The expressions for the approximations used for the partial derivative of the incidence of *exogenous disease* with respect to d_{x10} and d_{x20} are analogous.

We use the following approximations for the partial derivative of the age-specific incidence of *endogenous disease* with respect to

(i) d_{n10}:

$$\frac{\partial E_n(a,t,0)}{\partial d_{n10}} = \begin{cases} L(a,t) & \text{for } a \leq 10, \\ L(a,t) \left\{ -\frac{a}{10} + 2 \right\} & \text{for } 10 < a < 20, \\ 0 & \text{for } a \geq 20. \end{cases}$$

and (ii) d_{n20}:

$$\frac{\partial E_n(a,t,0)}{\partial d_{n20}} = \begin{cases} 0 & \text{for } a \leq 10, \\ L(a,t) \left\{ \frac{a}{10} - 1 \right\} & \text{for } 10 < a < 20, \\ L(a,t) & \text{for } a \geq 20. \end{cases}$$

A.3 Methods used for sensitivity analyses of disease risk estimates

The method used to generate the notification rates for a given age group in the age range 0–54 years in a simulated data set is described by the following algorithm:

1. Randomly generate the number of cases of *all forms* of tuberculosis in 1953 and in each of the survey years, assuming that it follows a Poisson distribution with mean n_a, where n_a is the 'actual' number of cases of *all forms* of tuberculosis in the age groups in the survey year. Calculate the notification rate of all forms of tuberculosis for this age group and each survey year resulting from these random estimates.

   Note that the surveys were carried out for only part of a year, and reports of the surveys published only the *annual notification* rates. The 'actual' numbers of cases used in this algorithm refer to the numbers of cases which would have occurred during the entire year. Population estimates for most years were published in the national reports of the surveys; otherwise, we used sources described in Table 2.2.

2. Calculate the rate of decline since 1953 and between the survey years for these randomly generated notification rates of *all forms* of tuberculosis.

3. Randomly generate the number of cases of respiratory tuberculosis for 1953, assuming that this follows a Poisson distribution with mean n_r (where n_r is the 'actual' number of cases of *respiratory* tuberculosis in the age group in 1953. Note that for 45–54 year olds, n_r was derived assuming that it was 95% of n_a — see section 2.2.3). Calculate the corresponding notification rate of respiratory tuberculosis for 1953 using this randomly generated estimate.

4. Derive estimates of the notifications of respiratory tuberculosis for the period 1954–1988 assuming that they declined from the level in 1953 calculated in step 3 at the rates calculated in step 2 above.

The notification rate for a given year for 55–64 and 65–74 year olds in a simulated data set was generated more simply, i.e. by repeating step 3 for each year (between 1965 and 1988, and 1975 and 1988 respectively) using the published number of cases of respiratory tuberculosis for that year.

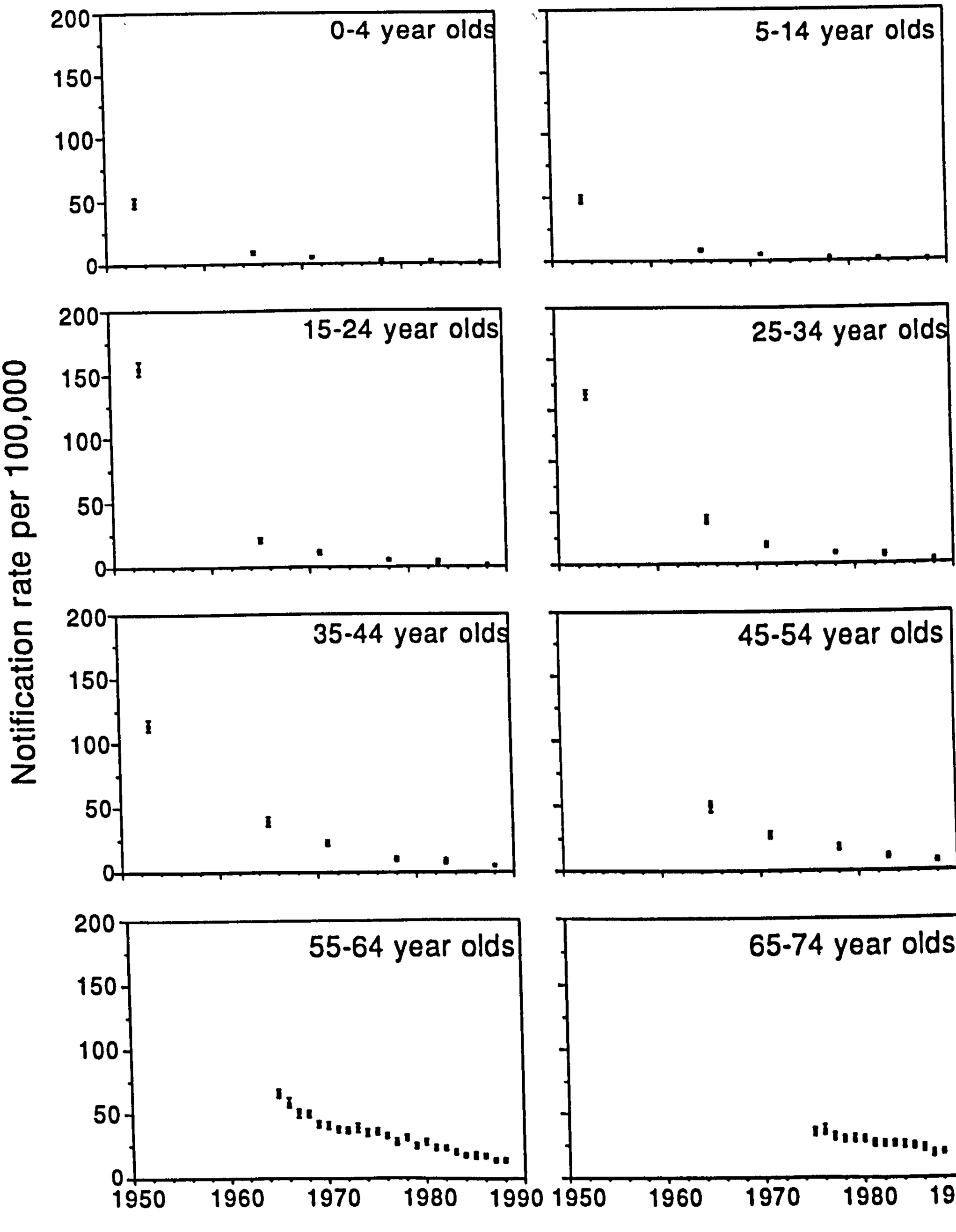


Figure A.1: Summary of the 95% range of the age-specific notification rates of respiratory tuberculosis during the survey years (for 0–54 year olds) and for all years (for 55–74 year olds) used in the 50 simulated data sets to analyze the sensitivity of the disease risk estimates.

See previous page for further details.

A.4 Results from applying TBDYN3 to Dutch data

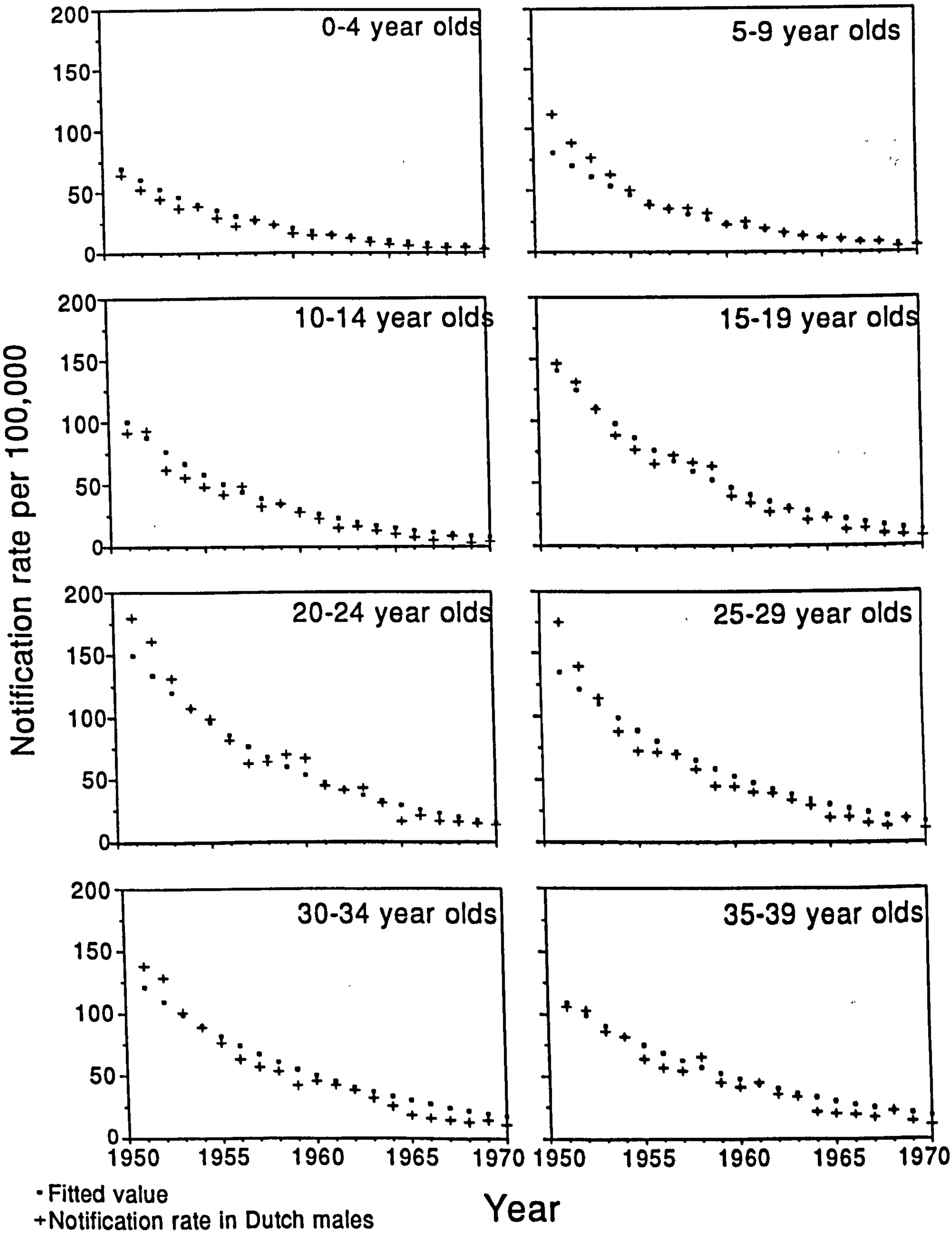


Figure A.2: Comparison between the incidence of respiratory tuberculosis in 0–39 year old males estimated using TBDYN3 for the *Netherlands* and the notifications used in the fitting process.
Notification data supplied by Dr. K. Styblo at the International Union against Tuberculosis.

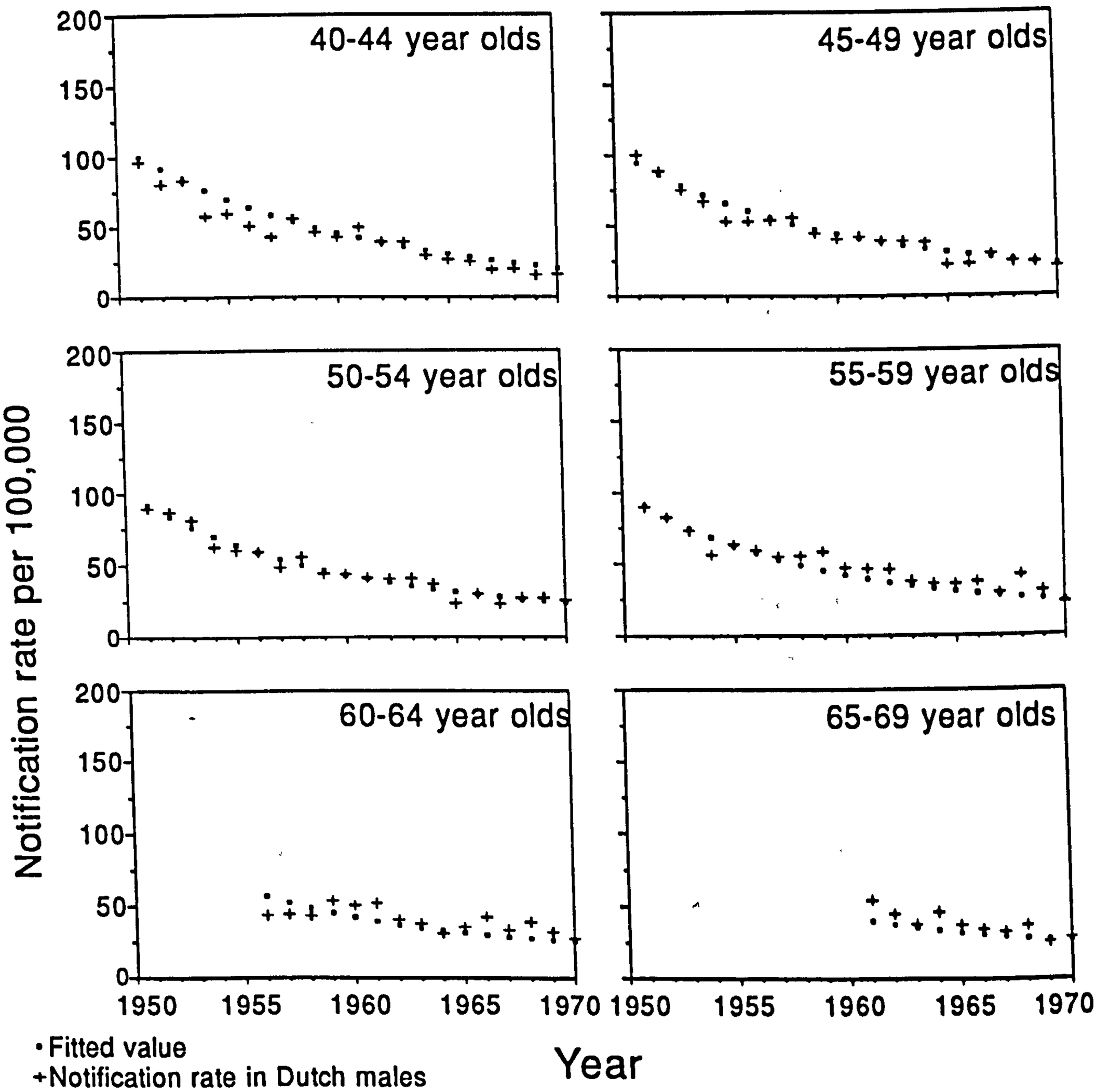


Figure A.3: Comparison between the incidence of respiratory tuberculosis in 40-69 year old males estimated using TBDYN3 for the *Netherlands* and the notifications used in the fitting process.
Notification data supplied by Dr. K. Styblo at the International Union against Tuberculosis.

Appendix B

Methods and results for the incubation period and serial interval for tuberculosis

B.1 PDEs illustrating the methods used to analyze the incubation period/serial interval using TBDYN3

We provide the equations used to derive the number of individuals who were infected at age A at time T, and who develop disease at time s thereafter without having experienced reinfection. These illustrate the kind of approach used to analyze aspects of the incubation period and serial interval of tuberculosis in chapter 4.

We use the subscript "A,T" to denote individuals who were infected at age A at time T (without having been reinfected). These individuals pass through the following classes:

Class	Definition
$I_{A,T}(s)$	the number of individuals have been infected for duration s,
$P_{A,T}(s,\hat{s})$	the number of individuals who have been infected for duration s, who have been experiencing their first primary episode for $\hat{s}$ years.
$L_{A,T}(s)$	the number of individuals in the 'latent' class who have infected for duration s.
$En_{A,T}(s,\hat{s})$	the number of individuals who have been infected for duration s, and who have been experiencing the endogenous disease for $\hat{s}$ years.

The definitions of these classes are analogous to those in table 3.1.

We use the following shorthand notation:

Parameter	Notation
$m_g(A+s, T+s)$	$\mathbf{m_g(s)}$
$m_+(T+s, \hat{s})$	$\mathbf{m_+(s)}$
$d_+(A+s)$	$\mathbf{d_+(s)}$
$d_-(A+s)$	$\mathbf{d_-(s)}$
$i(T+s)$	$\mathbf{i(s)}$
$r(A+s, T+s, \hat{s})$	$\mathbf{r_s(\hat{s})}$

(See Table 3.2 for the definitions of these parameters).

The disease dynamics among individuals are given by the following four equations:

$$\frac{dI_{A,T}(s)}{ds} = -(d_p(A,s) + \mathbf{m_g(s)} + k_L(s))I_{A,T}(s) \tag{B.1}$$

$$\frac{\partial P_{A,T}(s,\hat{s})}{\partial s} + \frac{\partial P_{A,T}(s,\hat{s})}{\partial \hat{s}} = -(\mathbf{m_+(s)d_+(s)} + \mathbf{m_g(s)d_-(s)} + \mathbf{r_s(\hat{s})})P_{A,T}(s,\hat{s}) + d_p(A,s)I_{A,T}(s) \tag{B.2}$$

$$\frac{dL_{A,T}(s)}{ds} = I_{A,T}(5)k_L(5) + \mathbf{r_s(2)}(P_{A,T}(2) + E_{n_{A,T}}(s,2)) - (\mathbf{i(s)} + d_n(A+s) + \mathbf{m_g(s)})L_{A,T}(s) \tag{B.3}$$

$$\frac{\partial E_{n_{A,T}}(s,\hat{s})}{\partial s} + \frac{\partial E_{n_{A,T}}(s,\hat{s})}{\partial \hat{s}} = d_n(A+s)L_{A,T}(s) - \mathbf{r_s(2)}E_{n_{A,T}}(s,2) - (\mathbf{m_+(s)d_+(s)} + \mathbf{m_g(s)d_-(s)})E_{n_{A,T}}(s,\hat{s}) \tag{B.4}$$

The total number of individuals experiencing disease having been infected for duration s, who were first infected at age A at time T, without having been reinfected in the meantime is given by:

$$P_{A,T}(s,0) + E_{n_{A,T}}(s,0) \tag{B.5}$$

B.2 Further results from analyses of the incubation period and serial interval

B.2.1 Retrospective analyses

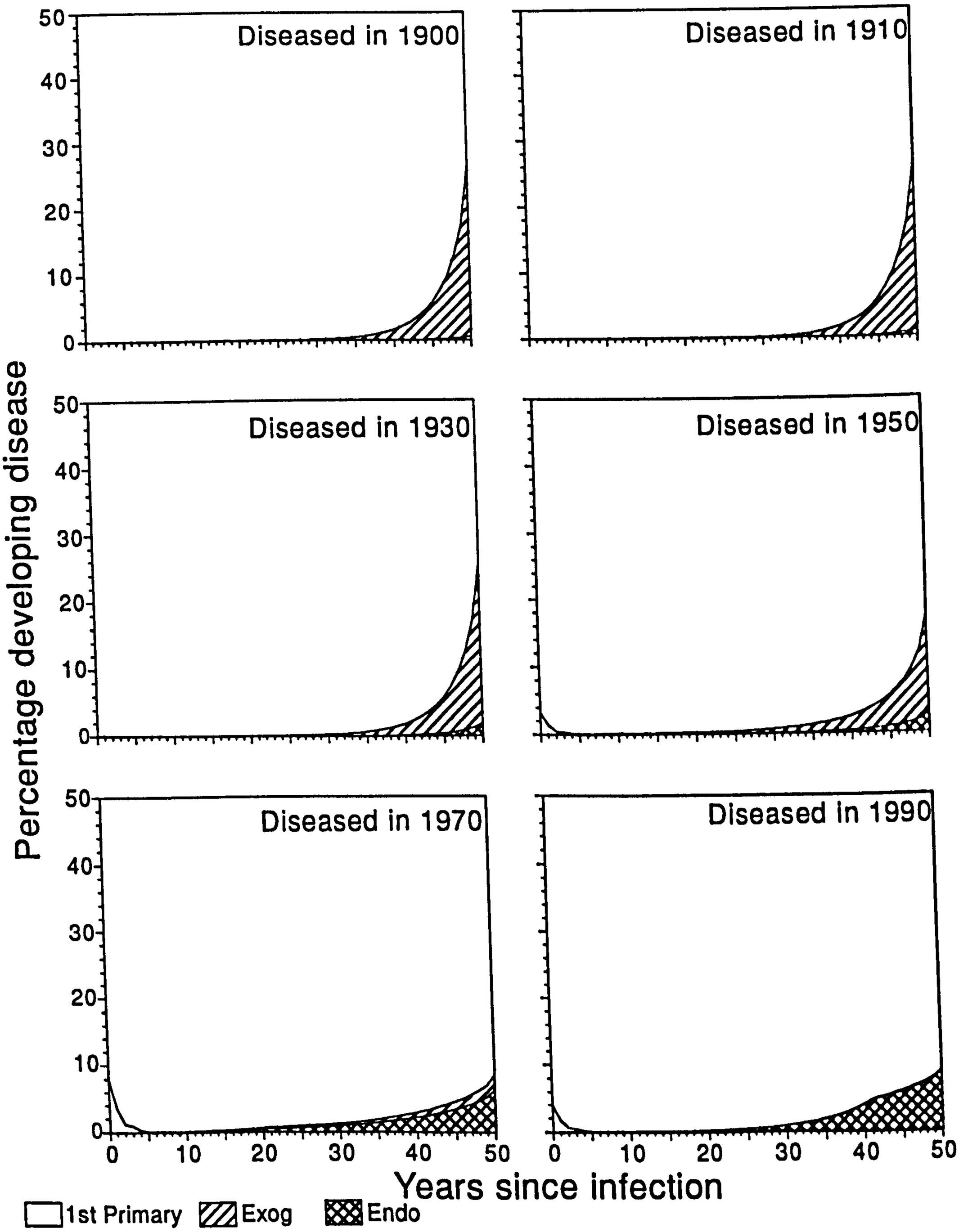


Figure B.1: Distribution of the time interval *since first infection* of individuals who developed respiratory tuberculosis when aged 50 years during the period 1900–1990, as estimated using TBDYN3.

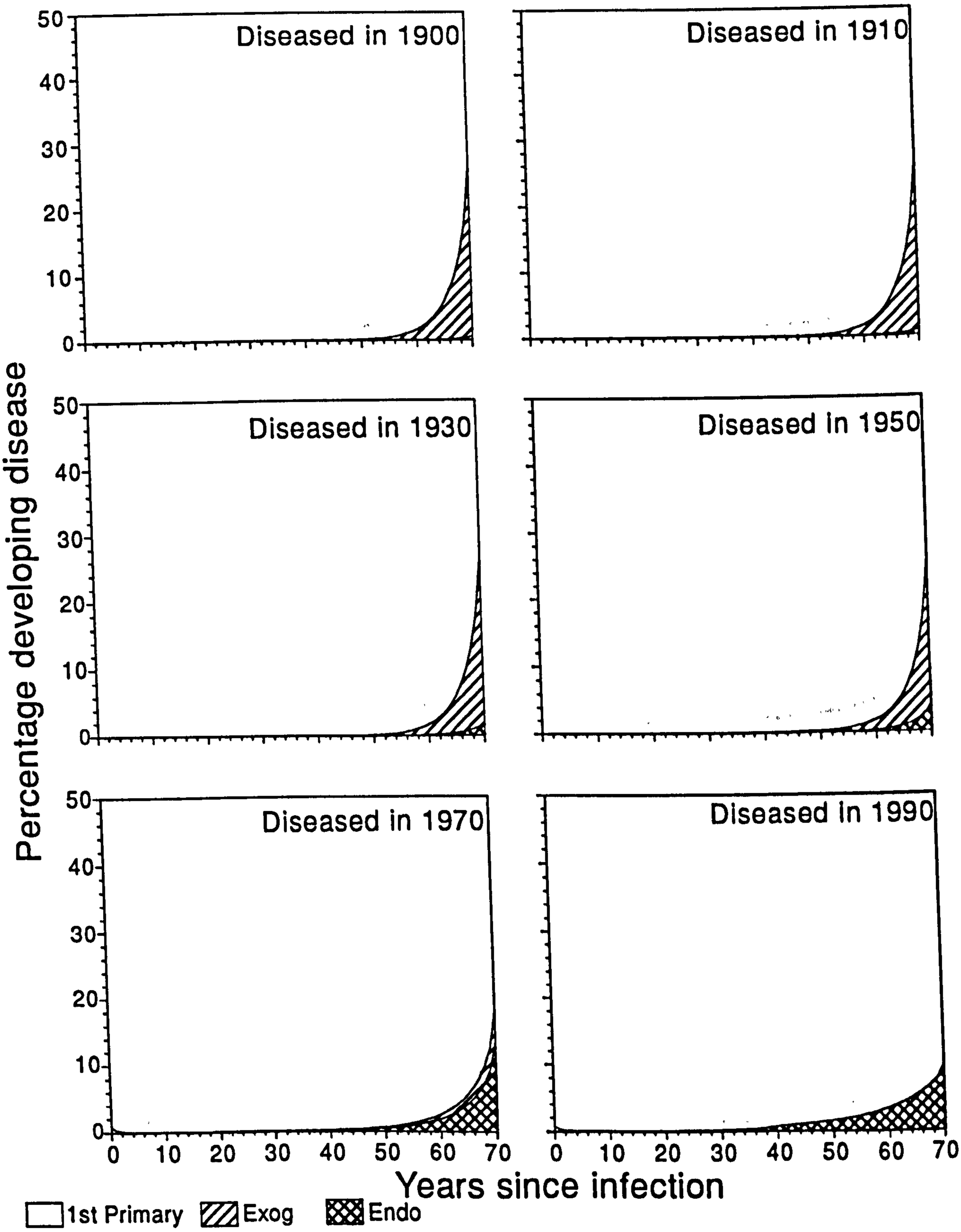


Figure B.2: Distribution of the time interval *since first infection* of individuals who developed respiratory tuberculosis when aged 70 years during the period 1900–1990, as estimated using TBDYN3.

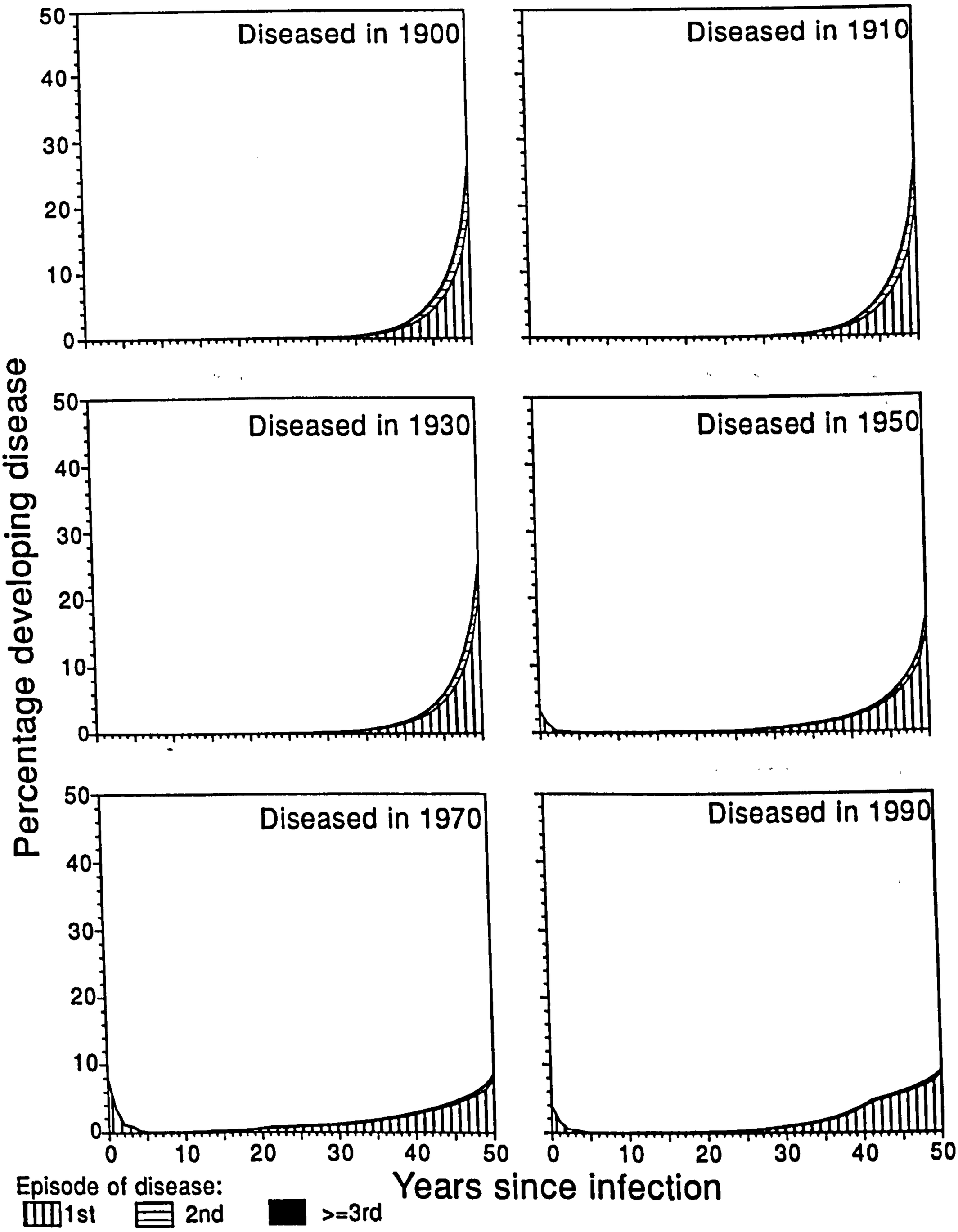


Figure B.3: Distribution of the time interval since first infection of individuals who developed respiratory tuberculosis when aged 50 years during the period 1900–1990, as estimated using TBDYN3. Shaded areas denote the proportion of individuals experiencing their first, second or subsequent episode.

B.2.2 Prospective analyses

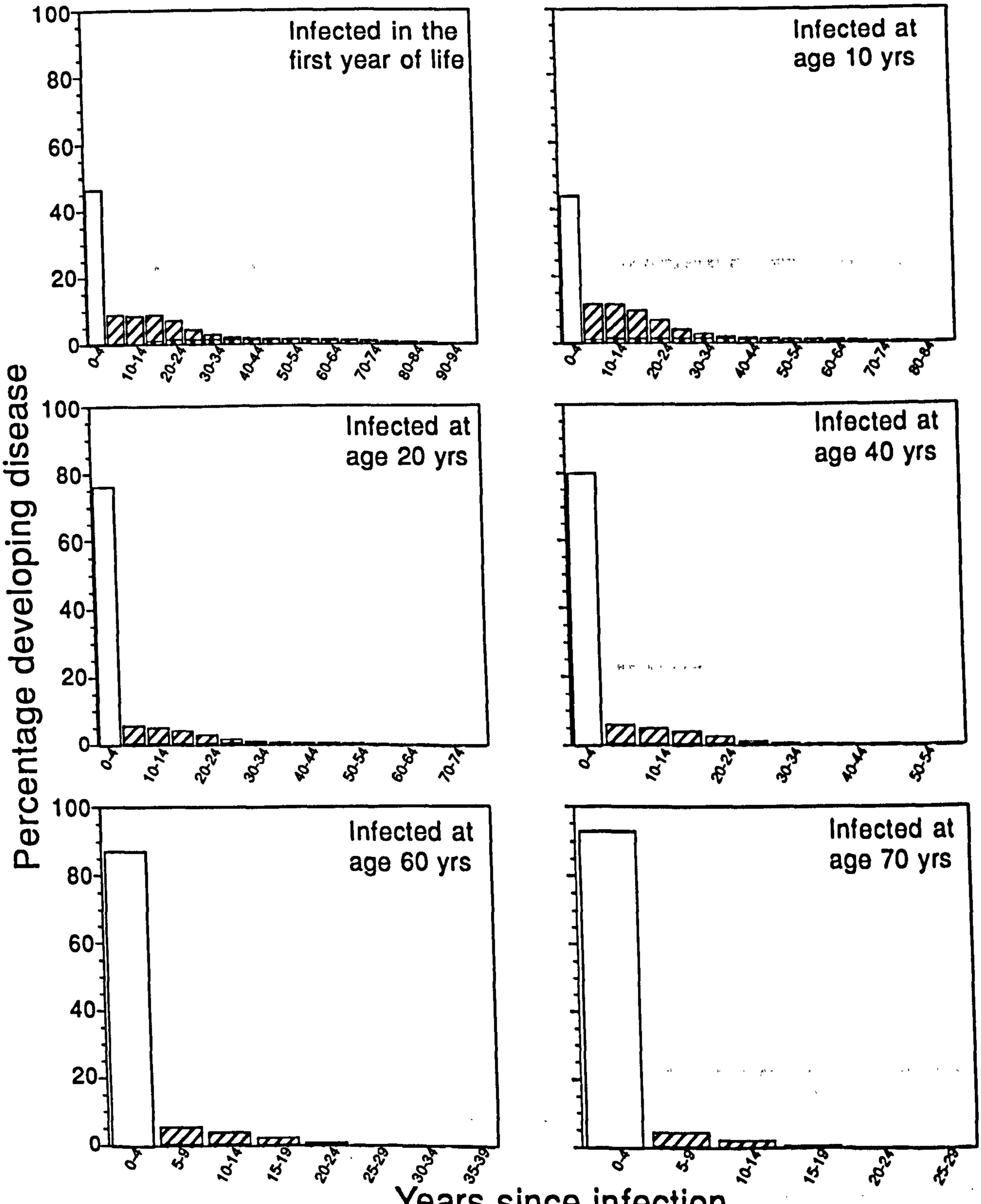


Figure B.4: Distribution of the time interval between first infection and the first episode of respiratory disease for individuals infected *at different ages in 1930*, as estimated using TBDYN3. These correspond to 'observed incubation periods'.

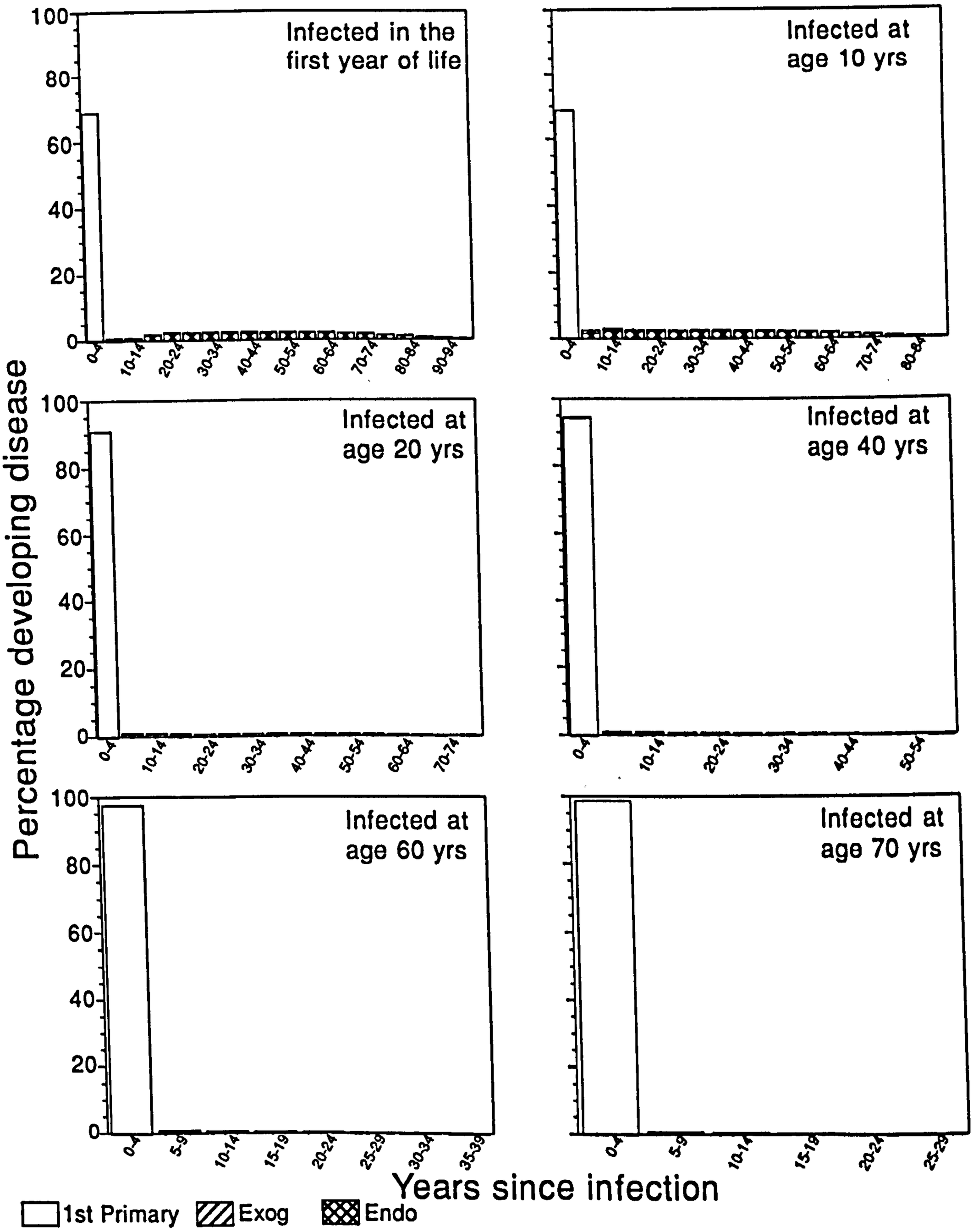


Figure B.5: Distribution of the time interval between first infection and the first episode of respiratory disease for individuals infected *at different ages in 1960*, as estimated using TBDYN3. These correspond to 'observed incubation periods'.

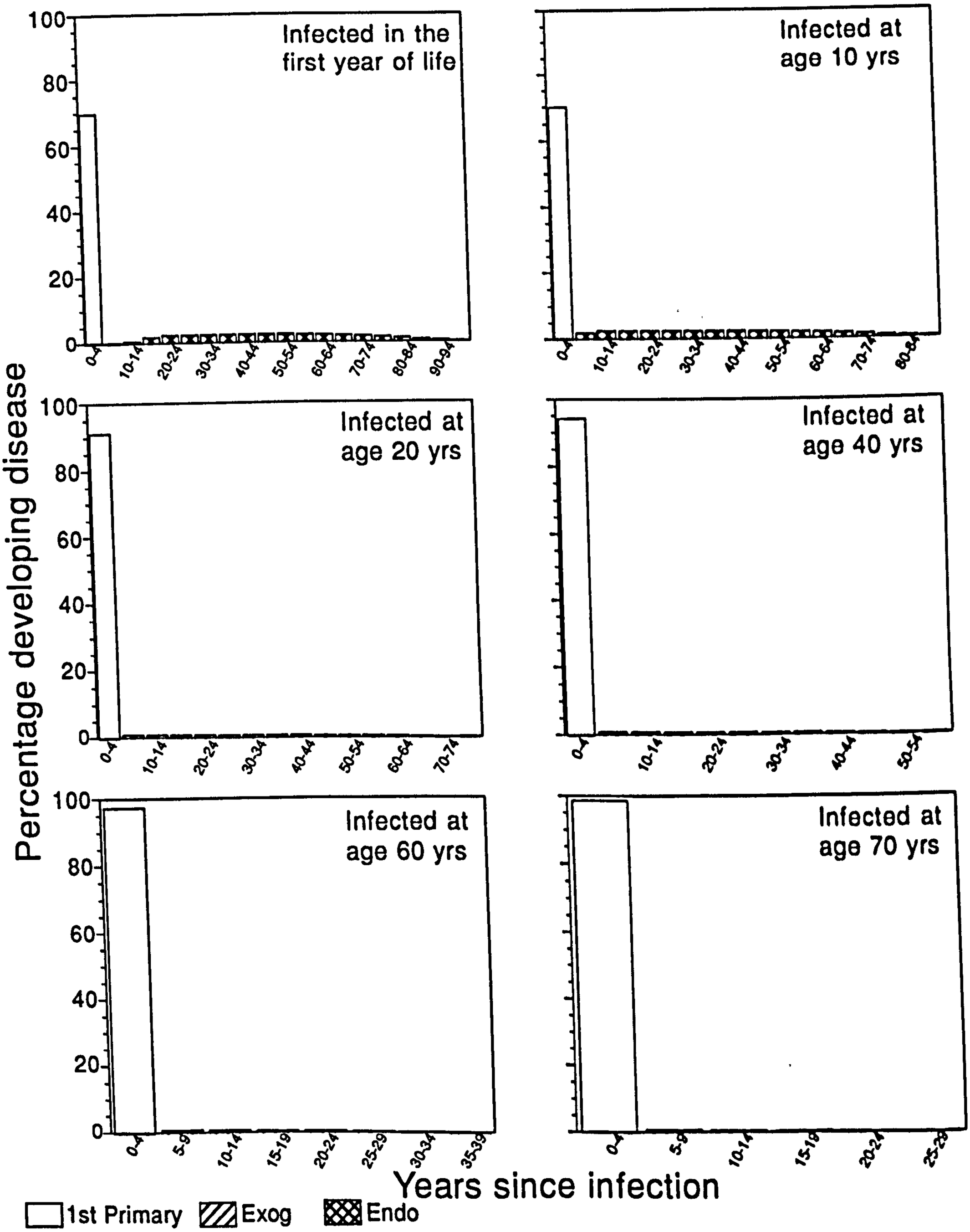


Figure B.6: Distribution of the time interval between first infection and the first episode of respiratory disease for individuals infected *at different ages in 1990*, as estimated using TBDYN3. These correspond to 'observed incubation periods'.

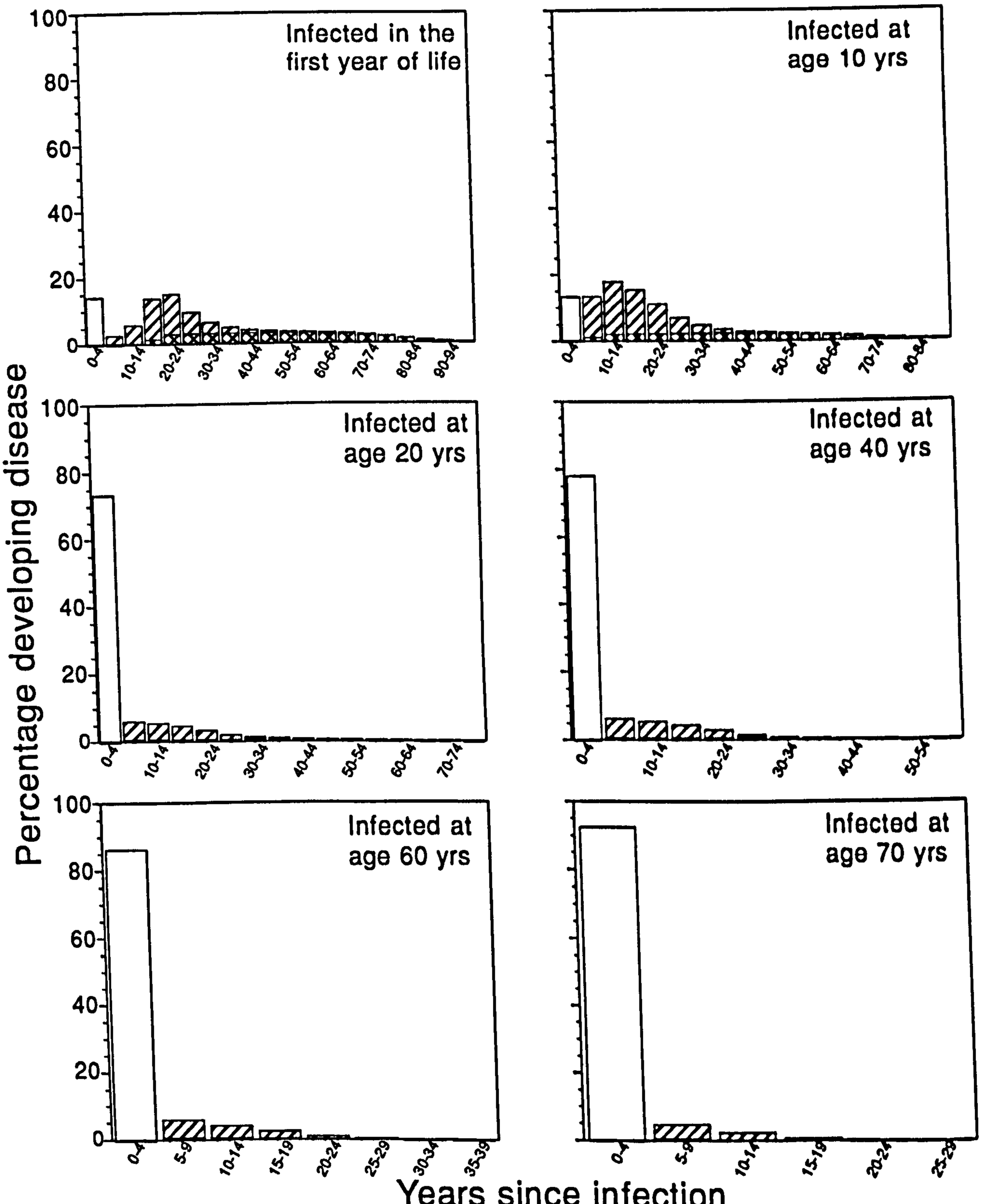


Figure B.7: Distribution of the time interval between first infection and the first episode of sputum-positive disease for individuals infected *at different ages in 1930*, as estimated using TBDYN3. These correspond to 'observed serial intervals'.

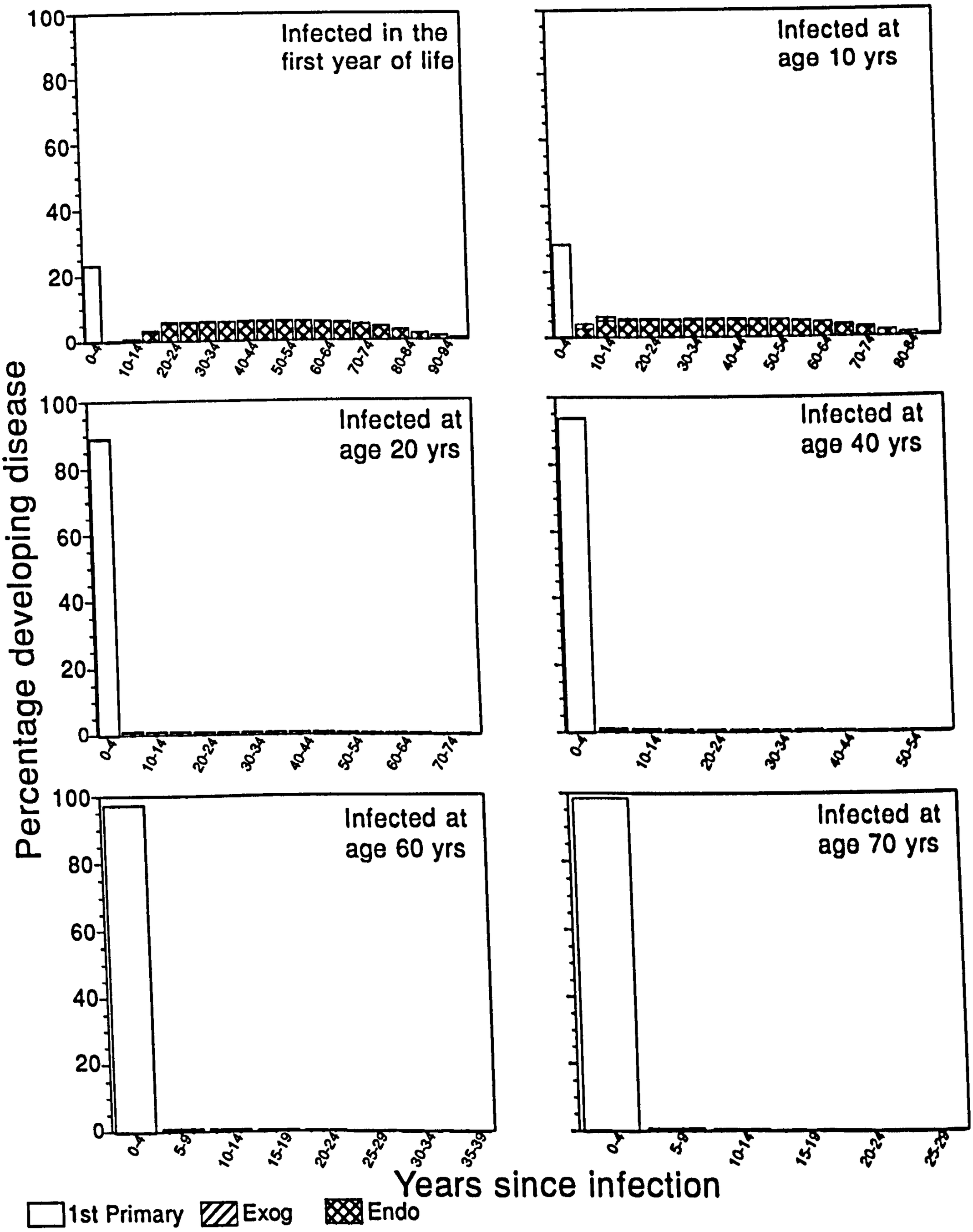


Figure B.8: Distribution of the time interval between first infection and the first episode of sputum-positive disease for individuals infected *at different ages in 1960*, as estimated using TBDYN3. These correspond to 'observed serial intervals'.

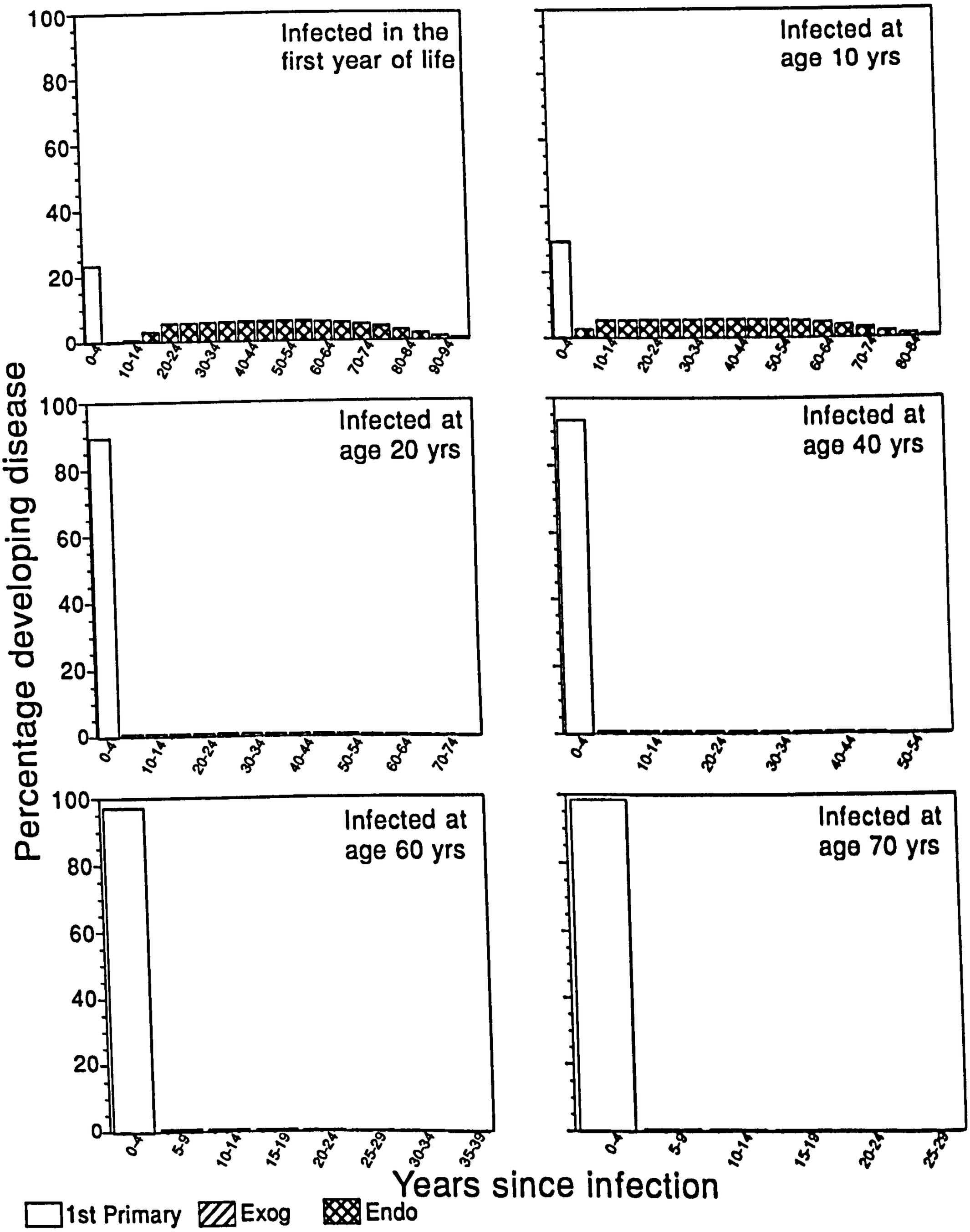


Figure B.9: Distribution of the time interval between first infection and the first episode of sputum-positive disease for individuals infected *at different ages in 1990*, as estimated using TBDYN3. These correspond to 'observed serial intervals'.

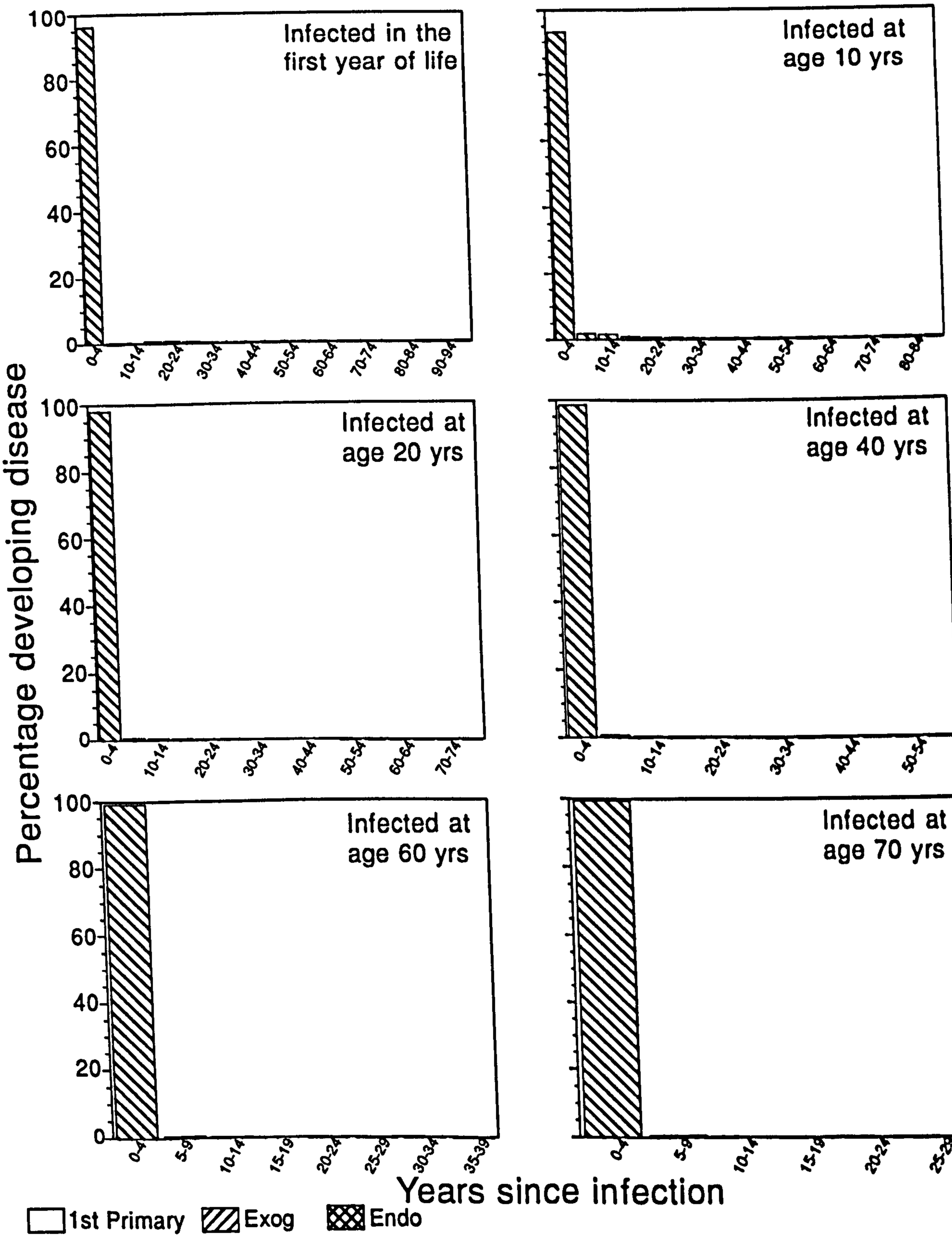


Figure B.10: Distribution of the time interval between first infection and the first episode of respiratory disease *attributable to initial infection* for individuals infected at different ages in 1900, as estimated using TBDYN3. These correspond to 'actual incubation periods'.

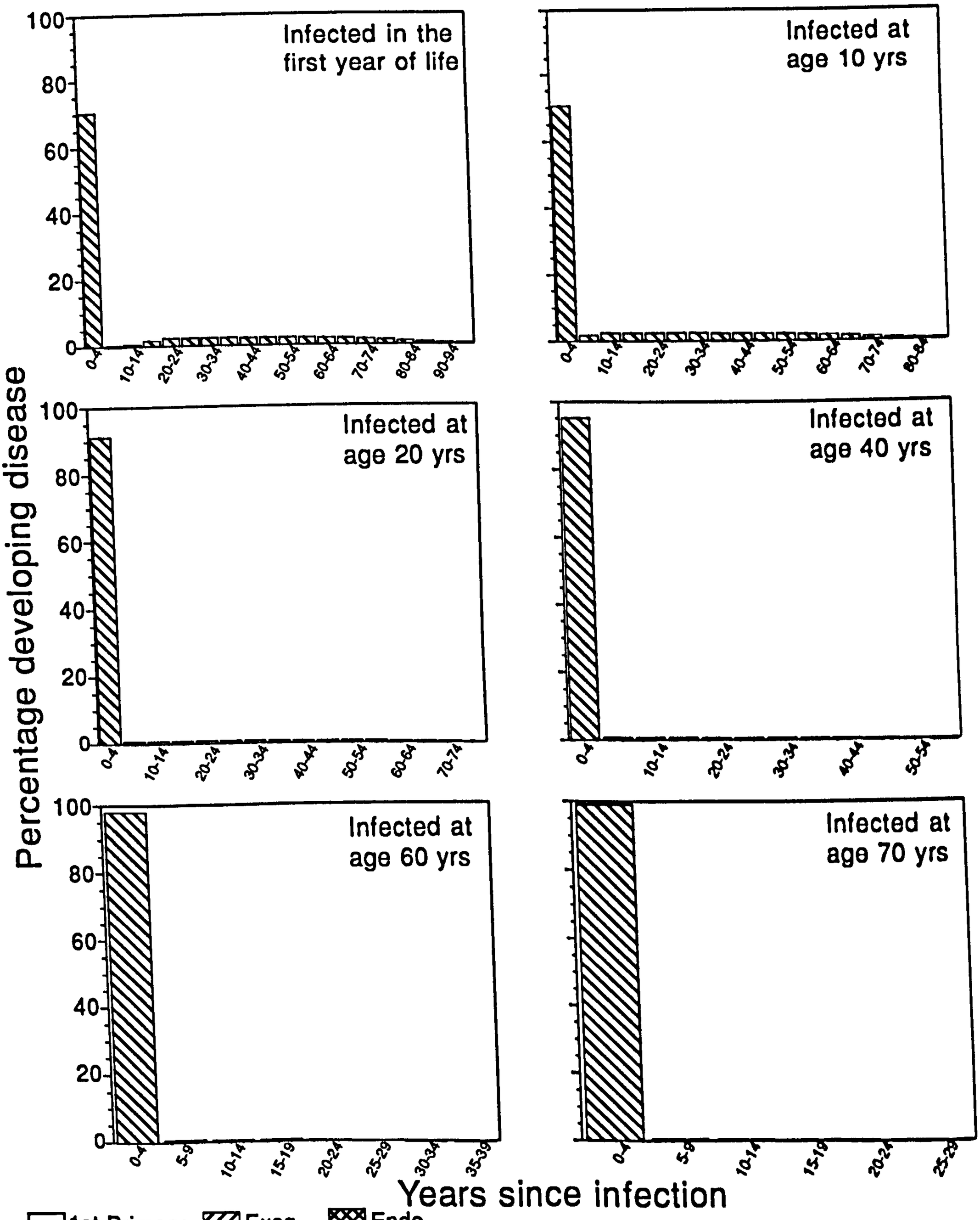


Figure B.11: Distribution of the time interval between first infection and the first episode of respiratory disease *attributable to initial infection* for individuals infected at different ages in 1960, as estimated using TBDYN3. These correspond to 'actual incubation periods'.

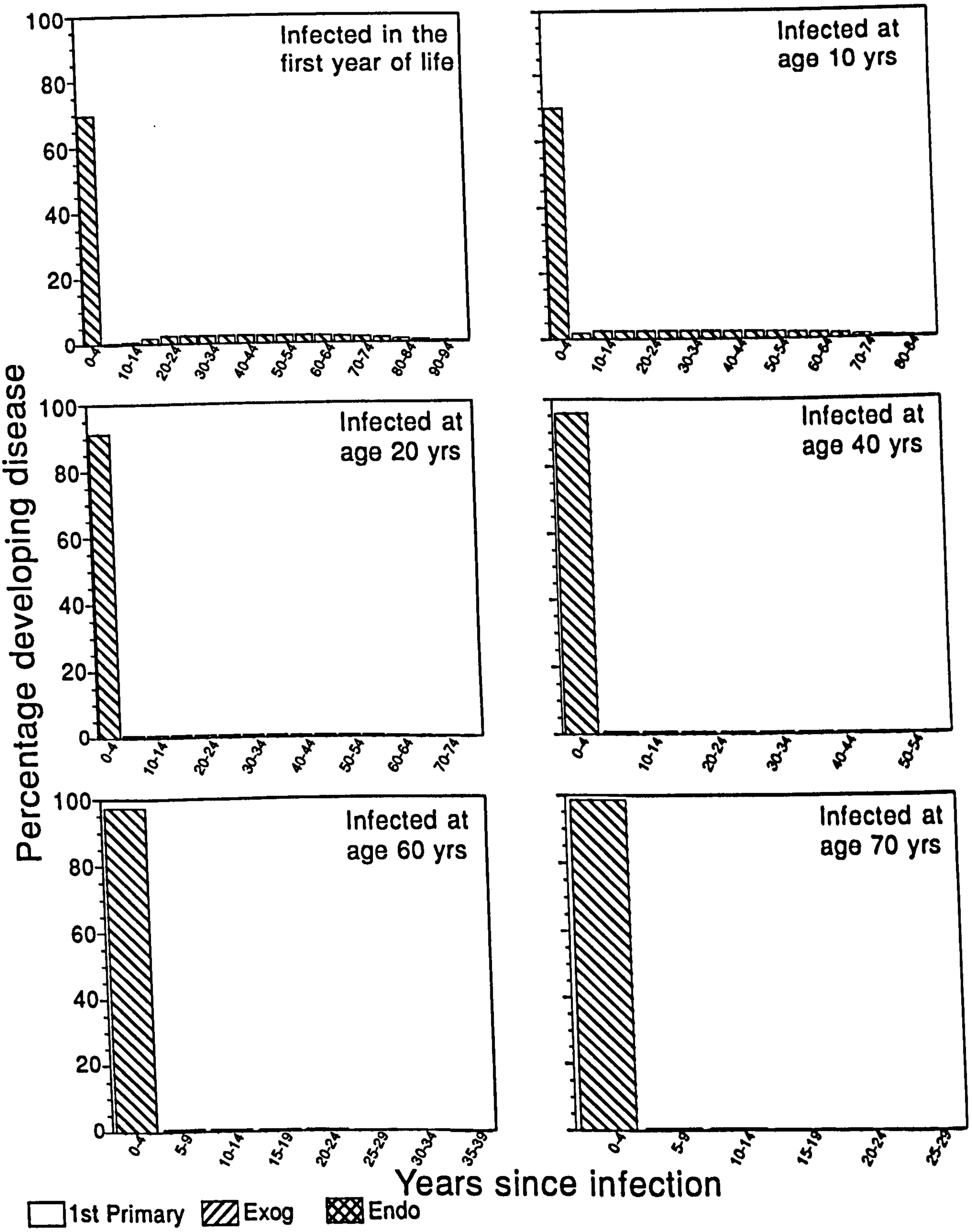


Figure B.12: Distribution of the time interval between first infection and the first episode of respiratory disease *attributable to initial infection* for individuals infected at different ages in 1900, as estimated using TBDYN3. These correspond to 'actual incubation periods'.

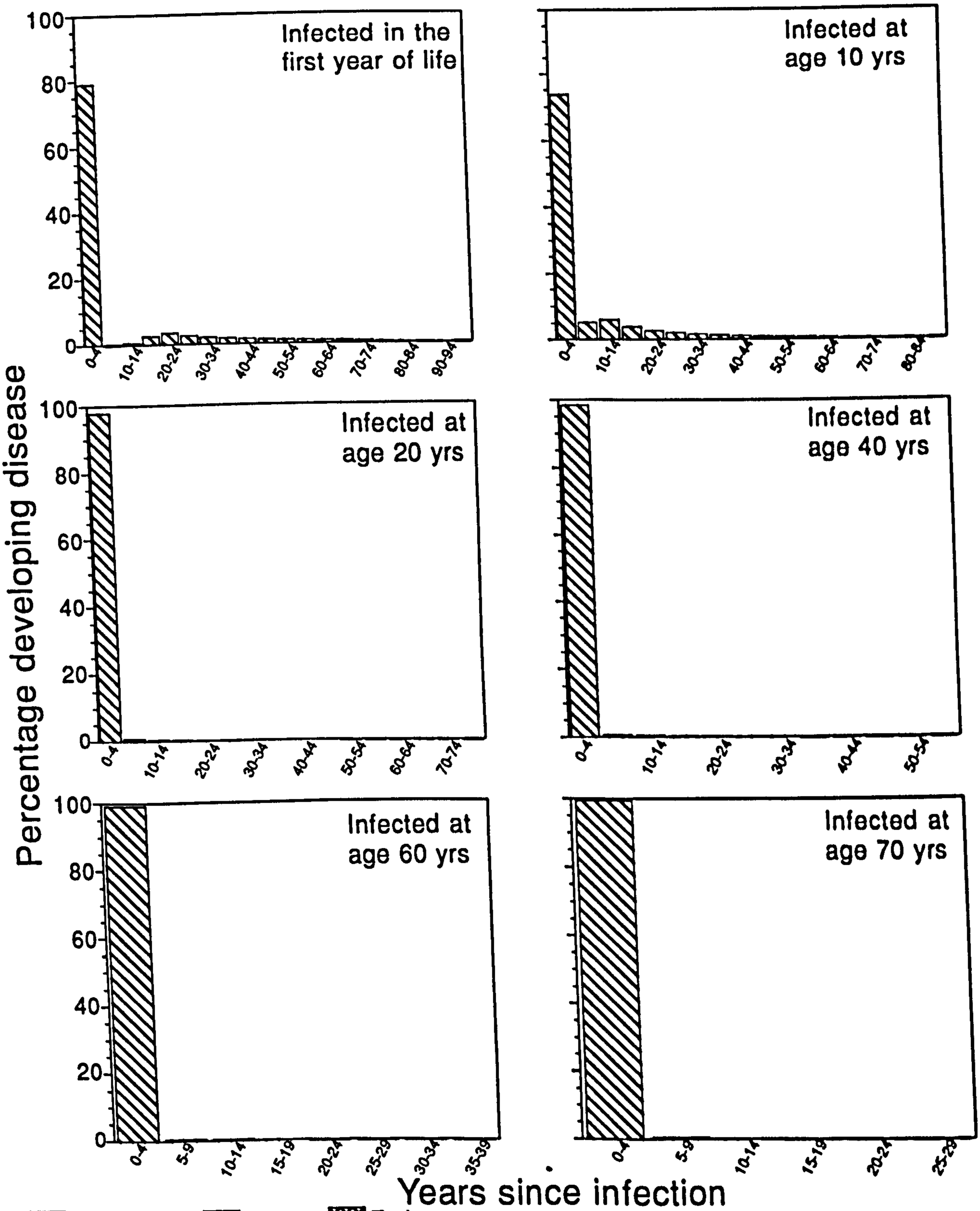


Figure B.13: Distribution of the time interval between first infection and the first episode of sputum-positive disease *attributable to initial infection* for individuals infected at different ages in 1900, as estimated using TBDYN3. These correspond to 'actual serial intervals'.

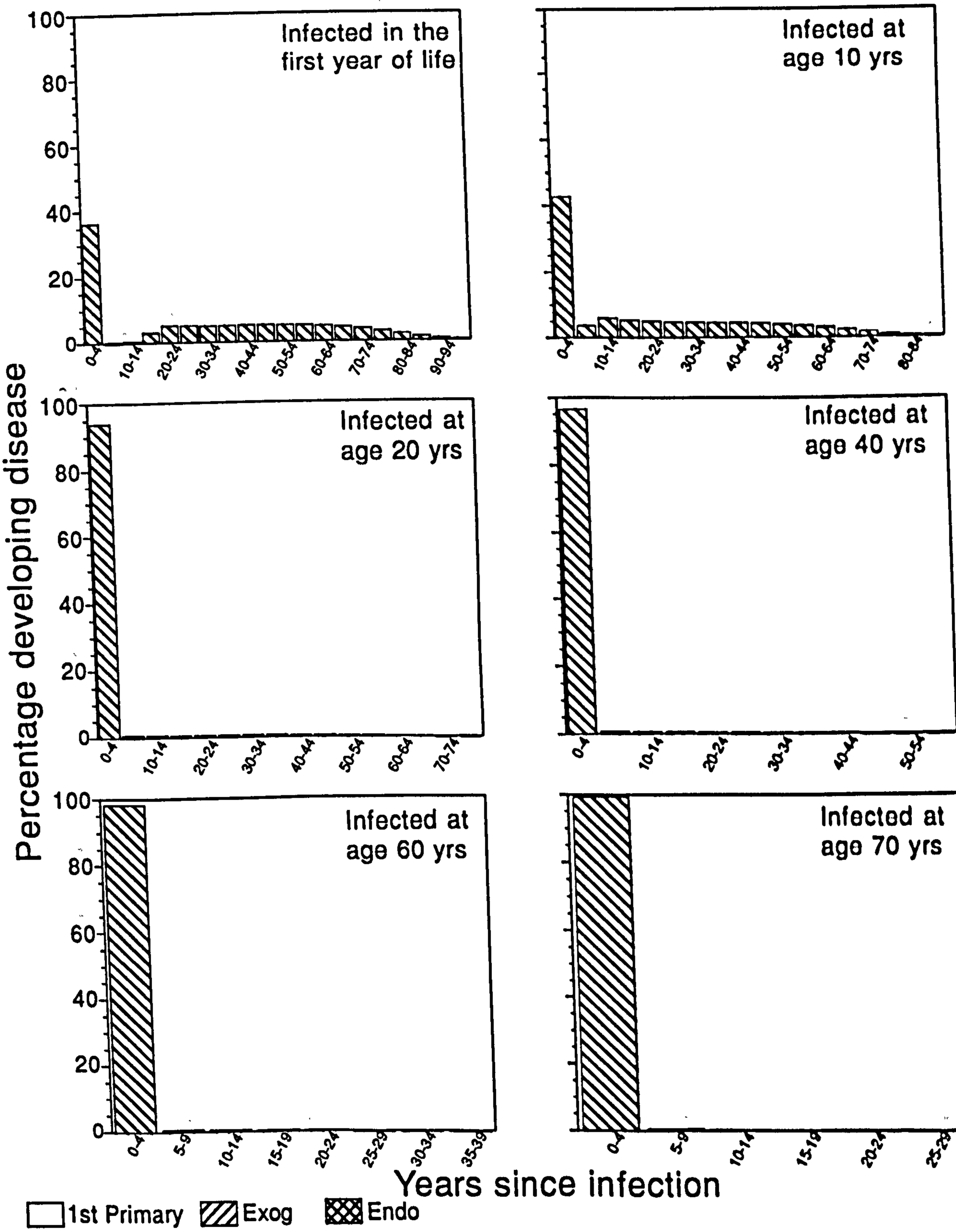


Figure B.14: Distribution of the time interval between first infection and the first episode of sputum-positive disease *attributable to initial infection* for individuals infected at different ages in 1930, as estimated using TBDYN3. These correspond to 'actual serial intervals'.

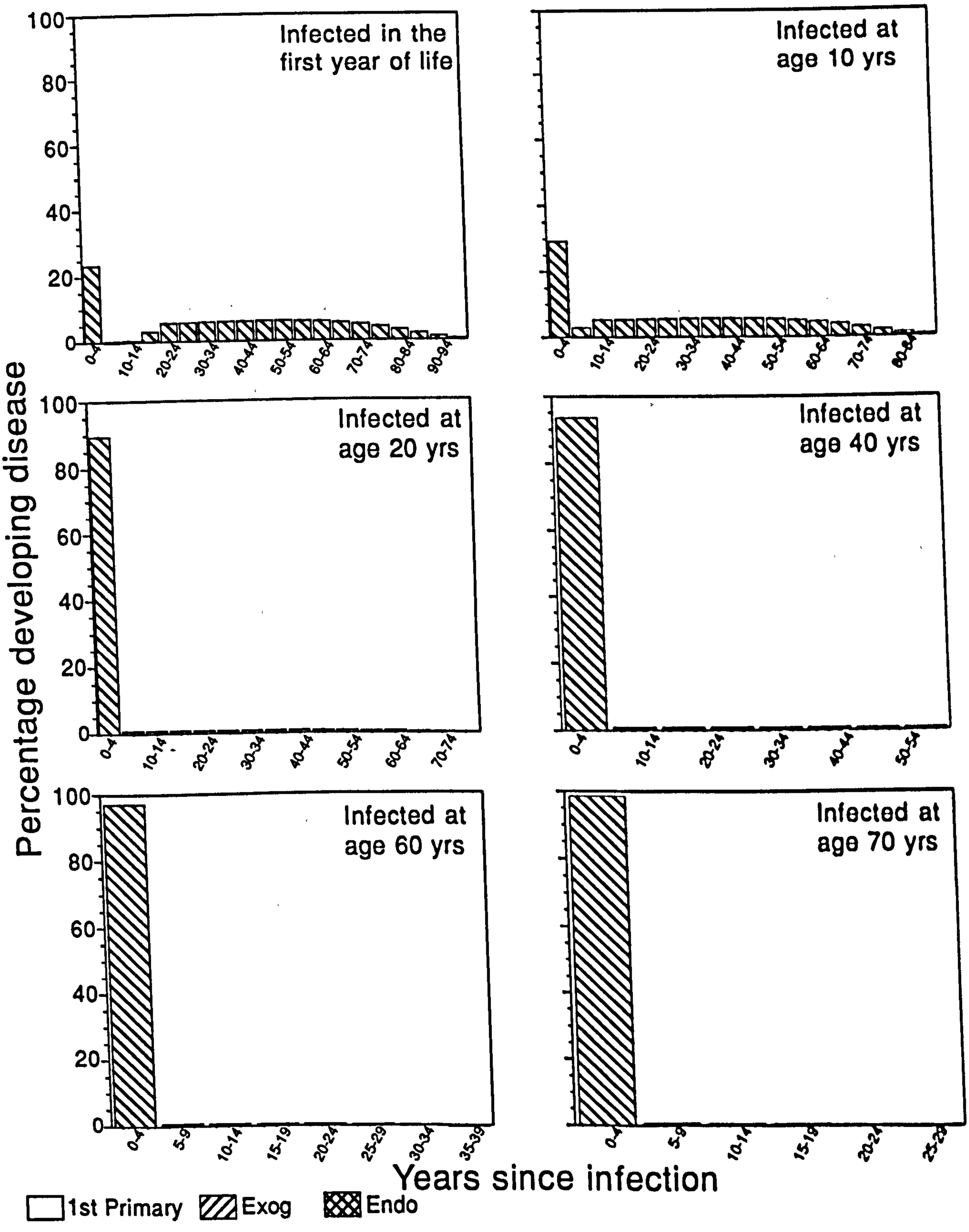


Figure B.15: Distribution of the time interval between first infection and the first episode of sputum-positive disease *attributable to the initial infection* for individuals infected at different ages in 1990, as estimated using TBDYN3. These correspond to 'actual serial intervals'.

Appendix C

The basic and net reproduction numbers for tuberculosis

C.1 Relationship between the basic and net reproduction numbers for tuberculosis under hypothetical situations

We consider a simplistic, hypothetical, non-age-structured, stable population for which a proportion p of individuals (who have been infected) face a risk d_x of developing sputum-positive (exogenous) disease if reinfected (i.e. effectively contacted), and the remainder $1-p$, who have not yet been infected, face a risk d_p of developing (the first primary) sputum-positive episode.

If a given sputum-positive case effectively contacts c other individuals, then the net reproduction number is given by:

$$R_n = cpd_x + c(1-p)d_p \tag{C.1}$$

and the basic reproduction number is given by:

$$R_0 = cd_p \tag{C.2}$$

The ratio between R_0 and R_n is thus given by:

$$\frac{R_0}{R_n} = \frac{cd_p}{cd_x + c(1-p)d_p} \tag{C.3}$$

Rearranging this expression, we see that the net reproduction number is given by:

$$R_n = R_0\left\{1 - p\left(1 - \frac{d_x}{d_p}\right)\right\} \tag{C.4}$$

The protective protection imparted against reinfection and subsequent sputum-positive disease by having experienced infection in the past r_d is given by:

$$r_d = 1 - \frac{d_x}{d_p} \tag{C.5}$$

Hence equation C.4 simplifies to:

$$R_n = R_0(1 - pr_d) \tag{C.6}$$

C.2 The relationship between $\tilde{R}_0(t)$ and $\tilde{R}_n(t)$ for tuberculosis

Suppose we consider a simplistic, hypothetical, non-age-structured population in which

1. a proportion p of individuals are infected and face a risk d_x of developing sputum-positive exogenous disease, if reinfected (i.e. effectively contacted by a sputum-positive case),

2. a proportion p_v in the population is protected in the short-term against initial infection by vaccination, and

3. the remainder $1-p-p_v$, who have not yet been infected or have not been immunized, face a risk d_p of developing the first primary sputum-positive episode if infected.

Suppose that at time t, a given sputum-positive case effectively contacts $c(t)$ other individuals, where effective contact is assumed to be equally likely for infected and uninfected individuals. $\tilde{R}_n(t)$ is therefore given by:

$$\tilde{R}_n(t) = c(t)pd_x + c(t)(1-p-p_v)d_p \tag{C.7}$$

and $\tilde{R}_0(t)$ is given by:

$$\tilde{R}_0(t) = c(t)d_p \tag{C.8}$$

The ratio between $\tilde{R}_n(t)$ and $\tilde{R}_0(t)$ is thus given by:

$$\frac{\tilde{R}_0(t)}{\tilde{R}_n(t)} = \frac{c(t)d_p}{c(t)pd_x + c(t)(1-p-p_v)d_p} \tag{C.9}$$

Cancelling out $c(t)$ and rearranging this expression, we see that the $\tilde{R}_n(t)$ is given by:

$$\begin{aligned} \tilde{R}_n(t) &= \tilde{R}_0(t)\left\{p\frac{d_x}{d_p} + 1 - p_v - p\right\} \\ &= \tilde{R}_0(t)\left\{1 - p_v - p\left(1 - \frac{d_x}{d_p}\right)\right\} \end{aligned} \tag{C.10}$$

The protection r_d conferred against reinfection and subsequent sputum-positive exogenous disease by initial infection is given by:

$$r_d = 1 - \frac{d_x}{d_p} \tag{C.11}$$

(see equation C.5).

Substituting for r_d into equation C.10, we see that:

$$\tilde{R}_n(t) = \tilde{R}_0(t)\{1 - p_v - pr_d\} \tag{C.12}$$

Bibliography

[1] Raviglione MC, Sudre P, Rieder HL, Spinaci S, and Kochi A. Secular trends of tuberculosis in Western Europe. *Bulletin of the World Health Organization*, 71(3/4):297–306, 1993.

[2] Frost WH. How much control of tuberculosis? *American Journal of Public Health*, 27:759–766, 1937.

[3] Adelstein A. Mortality from tuberculosis. *Health Trends*, pages 20–23, 1977.

[4] Donovan JW. A study in New Zealand mortality. *New Zealand Medical Journal*, 70:90–103, 1969.

[5] Andvord KF. Hvad kan vi laere ved å følge tuberkulosens gang fra generasjon til generasjon? *Norsk Magasin Laegevidenskapen*, 91:642–60, 1930.

[6] Styblo K. Epidemiology of tuberculosis. Selected Papers. KNCV, The Hague, The Netherlands, 1991.

[7] Sutherland I, Švandová E, and Radhakrishna SE. The development of clinical tuberculosis following infection with tubercle bacilli. *Tubercle*, 62:255–268, 1982.

[8] Comstock GW. Tuberculosis. In Evans AS and Feldman HA, editors, *Bacterial Infections in Man*, chapter 33, pages 605–632. Plenum Medical Books, New York and London, 1982.

[9] Blower SM, McLean AR, Porco TC, Small PM, Hopewell PC, Sanchez MA, and Moss AR. The intrinsic transmission dynamics of tuberculosis epidemics. *Nature Medicine*, 1(8):815–821, 1995.

[10] Grange JM. *Mycobacteria and Human Disease*. Edward Arnold (Publishers) Ltd, 1988.

[11] Wallgren A. The time-table of tuberculosis. *Tubercle*, 29:245–51, 1948.

[12] Thompson NJ, Glassroth JL, Snider DE, and Farer LS. The booster phenomenon in serial tuberculin testing. *American Review of Respiratory Disease*, 119(587–597), 1979.

[13] Felten MK and van der Merwe CA. Random variation in tuberculin sensitivity in school children. *American Review of Respiratory Disease*, 140:1001–1006, 1989.

[14] Grzybowski S and Allen EA. The challenge of tuberculosis in decline: a study based on the epidemiology of tuberculosis in Ontario, Canada. *American Review of Respiratory Disease*, 90:707–720, 1964.

[15] Rieder HL. Methodological issues in the estimation of the tuberculosis problem from tuberculin surveys. *Tubercle and Lung Disease*, 76:114–121, 1995.

[16] Holm J. Development from tuberculous infection to tuberculous disease. TSRU Progress Report, KNCV, The Hague, The Netherlands, 1969.

[17] Canetti G. Endogenous reactivation and exogenous reinfection. Their relative importance with regard to the development of non-primary tuberculosis. *Bulletin of the International Union against Tuberculosis*, 47:116–122, 1972.

[18] Grzybowski S, Styblo K, and Dorken E. Tuberculosis in Eskimos. *Tubercle*, Suppl. to Vol 57(4):1–58, 1976.

[19] Ormerod P and Skinner C. Reinfection tuberculosis: two cases in the family of a patient with drug-resistant disease. *Thorax*, 35:56–59, 1980.

[20] Raleigh JW, Wichelhausen R, Rado TA, and Bates JH. Evidence for infection by two distinct strains of *M. tuberculosis* in pulmonary tuberculosis: report of 9 cases. *American Review of Respiratory Disease*, 112:497–503, 1975.

[21] Nardell E, McInnis B, Thomas B, and Weidhaas S. Exogenous reinfection with tuberculosis in a shelter for the homeless. *New England Journal of Medicine*, 315:1570–5, 1986.

[22] Edlin BR, Tokars JI, Grieco MH, Crawford JT, Williams J, Sordillo EM, Ong KR, Kilburn JO, Dooley SW, Castro KG, Jarvis WR, and Holmberg SD. An outbreak of multi-drug resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. *New England Journal of Medicine*, 326:1514–1521, 1992.

[23] Coronado VG, Beck-Sague CM, Hutton MD, Davis BJ, Nicholas P, Villareal C, Woodley CL, Kilburn JO, Crawford JT, Frieden TR, and Sinkowitz RL aand Jarvis WR. Transmission of multi-drug resistant *Mycobacterium tuberculosis* among patients with human immunodeficiency virus infection in an urban hospital: epidemiologic and restriction fragment length polymorphism analysis. *The Journal of Infectious Diseases*, 168:1052–1055, 1993.

[24] Alland D, Kalkut GE, Moss AR, McAdam RA, Hahn JA, Bosworth W, Drucker E, and Bloom BR. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *New England Journal of Medicine*, 330(24):1710–1716, 1994.

[25] National Tuberculosis Institute, Bangalore. Tuberculosis in a rural population of South India: a five-year epidemiological study. *Bulletin of the World Health Organization*, 51:473–488, 1974.

[26] Madras Tuberculosis Prevention Trial. Trial of BCG vaccines in South India for tuberculosis prevention. *Indian Journal of Medical Research*, 72 (Suppl):1–74, 1980.

[27] Frost WH. The age selection of mortality from tuberculosis in successive generations. *American Journal of Public Health*, 30:91–96, 1939.

[28] Comstock GW, Livesay VT, and Woolpert SF. The prognosis of a positive tuberculin reaction in childhood and adolescence. *American Journal of Epidemiology*, 99(2):131–138, 1974.

[29] Sutherland I. The ten-year incidence of clinical tuberculosis following "conversion" in 2,550 individuals aged 14 to 19 years. TSRU Progress Report, KNCV, The Hague, The Netherlands, 1968.

[30] MMWR. Transmission of multidrug-resistant tuberculosis among immuno-compromised persons in a correctional system — New York, 1991. *MMWR*, 41(28):507–509, 1992.

[31] van Geuns HA, Meijer J, and Styblo K. Results of contact examination in Rotterdam, 1967–1969. *Bulletin of the International Union against Tuberculosis*, 50:107–121, 1975.

[32] Grzybowski S, Barnett GD, and Styblo K. Contacts of cases of active pulmonary tuberculosis. *Bulletin of the International Union against Tuberculosis*, 50:90–106, 1975.

[33] Barnett GD and Styblo K. Bacteriological and X-ray status of tuberculosis following primary infection acquired during adolescence or later. *Bulletin of the International Union against Tuberculosis*, 52:5–16, 1977.

[34] Murray C, Styblo K, and Rouillon A. Tuberculosis. *Global Burden of Disease: World Development Report*, pages 233–259, 1992.

[35] Berg G. The prognosis of open tuberculosis. A clinical-statistical analysis. *Acta Tuberculosea Scandinavica, supplementum*, IV:1–207, 1939.

[36] Drolet GJ. Present trend of case-fatality rates in tuberculosis. *American Review of Tuberculosis*, 37:125–151, 1938.

[37] Sutherland I and Lindgren I. The protective effect of BCG vaccination, as indicated by autopsy studies. *Tubercle*, 60:225–231, 1979.

[38] D'Arcy Hart P and Sutherland I. BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life. *British Medical Journal*, 2:293–295, July 30 1977.

[39] Fine PEM. BCG vaccination against tuberculosis and leprosy. *British Medical Bulletin*, 44(3):691–703, 1988.

[40] Springett VH. Ten-year results during the introduction of chemotherapy for tuberculosis. *Tubercle*, 52:73–87, 1971.

[41] British Thoracic Association. Short course chemotherapy in pulmonary tuberculosis. *The Lancet*, pages 1182–1183, May 31 1980.

[42] Pearson ML, Jereb JA, Frieden TR, Crawford JT, Davis BJ, Dooley SW, and Jarvis WR. Nosocomial transmission of nosocomial *Mycobacterium tuberculosis*. A risk to patients and health care workers. *Annals of Internal Medicine*, 117(3):191–196, 1992.

[43] Valway SE, Richards SB, Kovacovich J, Greifinger RB, Crawford JT, and Dooley SW. Outbreak of multi-drug-resistant tuberculosis in a New York state prison, 1991. *American Journal of Epidemiology*, 140(2):113–122, 1994.

[44] Selwyn PA, Hartel D, Lewis VA, Schoenbaum EE, Vermund SH, Klein RS, Walker AT, and Friedland GH. A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. *New England Journal of Medicine*, 320:545–50, 1989.

[45] Ferebee SH. *Controlled chemoprophylaxis studies in tuberculosis: a general review*, volume 17 of *Advances in Tuberculosis Research*, pages 28–106. Karger, Basel, 1970.

[46] Meyer SN. *Statistical investigations of the relationship of tuberculosis morbidity and mortality to infection.* Munksgaards Forlag, 1949.

[47] Zeidberg LD, Gass RS, and Dillon A *et al.* The Williamson County tuberculosis study. *American Review of Respiratory Disease*, 87(3):Part 2, 1963.

[48] Zeidberg LD, Dillon A, and Gass RS. Risk of developing tuberculosis among children of tuberculous parents. *American Journal of Hygiene*, 70:1009–1019, 1954.

[49] Horwitz O, Wilbek E, and Erikson PA. Epidemiological basis of tuberculosis eradication. 10. Longitudinal studies on the risk of tuberculosis in the general population of a low-prevalence area. *Bulletin of the World Health Organization*, 41:95–113, 1969.

[50] Diekmann O, Heesterbeek JAP, and Metz JAJ. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28:365–82, 1990.

[51] Heesterbeek H. R_0. Centrum voor Wiskunde en Informatica, Amsterdam, 1992.

[52] Styblo K. *Recent advances in epidemiological research in tuberculosis*, volume 20 of *Advances in Tuberculosis Research.* Karger, Basel, 1980.

[53] Styblo K. The relationship between the risk of tuberculous infection and the risk of developing infectious tuberculosis. *Bulletin of the International Union against Tuberculosis*, 60(3–4):117–119, 1985.

[54] Murray CJL, Styblo K, and Rouillon K. Tuberculosis in developing countries: burden, intervention and cost. *Bulletin of the International Union against Tuberculosis*, 65:6–26, 1990.

[55] Lincoln EM. *Epidemics of tuberculosis*, volume 14 of *Advances in Tuberculosis Research.* Karger, Basel, 1965.

[56] Hyge T. The efficacy of BCG vaccination. Epidemic of tuberculosis in a state school, with an observation period of 12 years. *Acta Tuberculosea*, 32:89–107, 1955.

[57] Kline S, Hedemark LL, and Davies SF. Outbreak of tuberculosis among regular patrons of a neighbourhood bar. *New England Journal of Medicine*, 333(4):222–227, 1995.

[58] Reves R, Blakey D, Snider DE, and Farer LS. Transmission of multiple drug-resistant tuberculosis: report of a school and community outbreak. *American Journal of Epidemiology*, 113(4):423–435, 1981.

[59] CDR Report. Outbreak of hospital acquired multi-drug resistant tuberculosis. *CDR Weekly*, 5(3):32, 1995.

[60] Barnes PF, Bloch AB, Davidson PT, and Snider DE. Tuberculosis in patients with human immunodeficiency virus infection. *New England Journal of Medicine*, 324:1644–1650, 1991.

[61] Hamer WH. Epidemic disease in England. *The Lancet*, i:733–739, 1906.

[62] Styblo K, Meijer J, and Sutherland I. The transmission of tubercle bacilli: its trend in a human population. *Bulletin of the International Union against Tuberculosis*, 42:5–104, 1969.

[63] Waaler HT, Geser A, and Andersen S. The use of mathematical models in the study of the epidemiology of tuberculosis. *American Journal of Public Health*, 52(6):1002–1013, 1962.

[64] Frimodt-Møller J. A community-wide tuberculosis study in a South Indian rural population, 1950–1955. *Bulletin of the World Health Organization*, 22:61–170, 1960.

[65] Azuma Y. A simple simulation model of tuberculosis epidemiology for use without large-scale computers. *Bulletin of the World Health Organization*, 52:313–322, 1975.

[66] Hock GE and Loy FK. A dynamic model of tuberculosis epidemiology for Singapore. *Annals Acad. Med.*, 10(1):40–49, 1981.

[67] Trefny J and Hejdova E. A model of the epidemiology of tuberculosis in the Czech Socialist Republic. *Bulletin of the International Union against Tuberculosis*, 57(3–4):206–211, 1982.

[68] ReVelle CS, Lynn WR, and Feldmann F. Mathematical models for the economic allocation of tuberculosis control activities in developing countries. *American Review of Respiratory Disease*, 96:893–909, 1967.

[69] Chorba RW and Sanders JL. Planning models for tuberculosis control programmes. *Health Services Research*, 6(2):144–164, 1971.

[70] Joesoef MR, Remington PL, and Tjiptoherijanto P. Epidemiological model and cost-effectiveness analysis of tuberculosis treatment programmes in Indonesia. *International Journal of Epidemiology*, 18(1):174–179, 1989.

[71] Schulzer M, Fitzgerald JM, Enarson DA, and Grzybowski S. An estimate of the future size of the tuberculosis problem in sub-Saharan Africa. *Tubercle and Lung Disease*, 73:52–58, 1992.

[72] Schulzer M, Radhamani MP, Grzybowski S, Mak E, and Fitzgerald JM. A mathematical model for the prediction of the impact of HIV infection on tuberculosis. *International Journal of Epidemiology*, 23(2):400–407, 1994.

[73] Heymann S J. Modelling the efficacy of prophylactic and curative therapies for preventing the spread of tuberculosis in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 87:406–411, 1993.

[74] Grigg ERN. The arcana of tuberculosis: with a brief epidemiologic history of the disease in the USA — Parts I, II and III. *American Review of Tuberculosis and Pulmonary Diseases*, 78(2):151–172,426–453,583–603, 1958.

[75] Brogger S. Systems analysis in tuberculosis control: A model. *American Review of Respiratory Disease*, 95:419–434, 1967.

[76] Waaler HT. A dynamic model for the epidemiology of tuberculosis. *American Review of Respiratory Disease*, 98:591–600, 1968.

[77] Waaler HT. Cost benefit analysis of BCG vaccination under various epidemiological considerations. *Bulletin of the International Union against Tuberculosis*, 41:42–52, 1968.

[78] Waaler HT and Piot MA. The use of an epidemiological model for estimating the effectiveness of tuberculosis control measures. Sensitivity of the effectiveness of tuberculosis control measures to the coverage of the population. *Bulletin of the World Health Organization*, 41:75–93, 1969.

[79] Waaler HT. Model simulation and decision-making in tuberculosis programmes. *Bulletin of the International Union against Tuberculosis*, 43:337–344, 1970.

[80] Waaler HT, Gothi GD, Baily GVJ, and Nair SS. Tuberculosis in rural South India. A study of population trends and the potential impact of antituberculosis programmes. *Bulletin of the World Health Organization*, 51:262–271, 1974.

[81] Fine PEM. The BCG story: lessons from the past and implications for the future. *Review of Infectious Diseases*, 11, Supplement 2:S353–S359, March–April 1989.

[82] Waaler HT and Piot MA. The use of an epidemiological model for estimating the effectiveness of tuberculosis control measures. Sensitivity of the effectiveness of tuberculosis control measures to the social time preference. *Bulletin of the World Health Organization*, 43:1–16, 1970.

[83] Balasangameshwara VH, Chakraborty AK, and Chaudhuri K. A mathematical construct of epidemiological time trend in tuberculosis — a fifty year study. *Indian Journal of Tuberculosis*, 39:87–98, 1992.

[84] Sutherland I, Bleiker MA, Meijer J, and Stýblo K. The risk of infection in the Netherlands from 1967 to 1979. *Tubercle*, 64:241–253, 1983.

[85] Sutherland I, Švandová E, and Radhakrishna SE. Alternative models for the development of tuberculous disease following infection with tubercle bacilli (Proceedings of the XXIIIrd International Tuberculosis conference, Mexico, 22–26 Sept 1975). *Bulletin of the International Union against Tuberculosis*, 51:171–179, 1976.

[86] Radhakrishna SE, Švandová E, and Sutherland I. The development of tuberculous disease following infection with tubercle bacilli. TSRU Progress Report, KNCV, The Hague, The Netherlands, 1975.

[87] Krishnamurthy VV and Chaudhuri K. Risk of pulmonary tuberculosis associated with exogenous reinfection and endogenous reactivation in a South Indian rural population — a mathematical estimate. *Indian Journal of Tuberculosis*, 37:63–67, 1990.

[88] Schulzer M, Enarson DA, Grzybowski S, Jong YP, Kim SJ, and Lin TP. An analysis of pulmonary tuberculosis data in Taiwan and Korea. *International Journal of Epidemiology*, 16(4):584–589, 1987.

[89] Medical Research Council Tuberculosis and Chest Diseases Unit. National survey of tuberculosis notifications in England and Wales 1978–9. *British Medical Journal*, 281:895–898, October 4 1980.

[90] Davies PDO, Darbyshire J, Nunn AJ, Byfield SP, and Fox W. Ambiguities and inaccuracies in the notification system for tuberculosis in England and Wales. *Community Medicine*, 3:108–119, 1981.

[91] Joint Tuberculosis Committee of the British Thoracic Association. Notification of tuberculosis: a code of practice for England and Wales. *British Medical Journal*, 284:1454–1456, 1982.

[92] Bradley BL, Kerr KM, Leitch AG, and Lamb D. Notification of tuberculosis: can the pathologist help? *British Medical Journal*, 297:595, 1988.

[93] Sheldon CD, King K, Cock H, Wilkinson P, and Barnes NC. Notifications of tuberculosis: how many cases are never reported? *Thorax*, 47:1015–1018, 1992.

[94] British Tuberculosis Association. Tuberculosis among immigrants to England and Wales: a national survey in 1965. *Tubercle*, 47:145–156, 1966.

[95] British Thoracic and Tuberculosis Association. A tuberculosis survey in England and Wales 1971: the influence of immigration and country of birth upon notifications. *Tubercle*, 54:249–260, 1973.

[96] Medical Research Council Tuberculosis and Chest Diseases Unit. National survey of tuberculosis notifications in England and Wales 1983. *British Medical Journal*, 291:658–661, 7 September 1985.

[97] Medical Research Council Cardiothoracic Epidemiology Group. National survey of notifications of tuberculosis in England and Wales in 1988. *Thorax*, 47:770–775, 1992.

[98] PHLS. Tuberculosis in 1993: preliminary results of national survey. *CDR Weekly*, 4(50):1, 1994.

[99] Fox W, Wiener A, Mitchison DA, Selkon JB, and Sutherland I. The prevalence of drug-resistant tubercle bacilli in untreated patients with pulmonary tuberculosis: a national survey, 1955–56. *Tubercle*, 38:71–84, 1957.

[100] Miller AB, Tall R, Fox W, Lefford MJ, and Mitchison DA. Primary drug-resistance in pulmonary tuberculosis in Great Britain: second national survey, 1963. *Tubercle*, 47:92–107, 1966.

[101] Springett VH, Darbyshire JH, Nunn AJ, and Sutherland I. Changes in tuberculosis notification rates in the white ethnic group in England and Wales between 1953 and 1983. *Journal of Epidemiology and Community Health*, 42:370–376, 1988.

[102] Springett VH. Tuberculosis notification rates in the elderly. *Communicable Disease Report*, 1(Review number 13):R149–R150, 1991.

[103] MRC. MRC national tuberculin survey 1949–50. *The Lancet*, pages 775–785, April 19 1952.

[104] Styblo K and Sutherland I. Epidemiology of tuberculosis in children. *Bulletin of the International Union against Tuberculosis*, 57(2):133–139, 1982.

[105] Sjögren I and Sutherland I. The risk of tuberculous infection in Sweden. *Tubercle*, 56:97–112, 1975.

[106] Christie PN and Sutherland I. A national tuberculin survey in Great Britain in 1971–1973. *Bulletin of the International Union against Tuberculosis*, 51:185–190, 1976.

[107] Rodrigues LC, Diwan VK, and Wheeler JG. Protective effect of BCG against tuberculous meningitis and miliary tuberculosis: a meta-analysis. *International Journal of Epidemiology*, 22(6):1154–1158, 1993.

[108] Joseph CA, Watson JM, and Fern KJ. BCG immunization in England and Wales: a survey of policy and practice in schoolchildren and neonates. *British Medical Journal*, 305:495–498, August 29 1992.

[109] Miller CL, Morris J, and Pollock TM. PHLS enquiry into current BCG vaccination policy. *British Medical Journal*, 288:564, February 18 1984.

[110] Department of Health and Social Security. *On the state of public health for the year.* HMSO, London, annually.

[111] Balasubramanian V, Wiegeshaus EH, Taylor BT, and Smith DW. Pathogenesis of tuberculosis: pathway to apical localization. *Tubercle and Lung Disease*, 75:168–178, 1994.

[112] Grange JM. The immunophysiology and immunopathology of tuberculosis. In Davies PDO, editor, *Clinical tuberculosis*, chapter 5, pages 55–71. Chapman & Hall Medical, London, 1994.

[113] Sutherland I and Fayers PM. The association of the risk of infection with age. *Bulletin of the International Union against Tuberculosis*, 50:71–81, 1975.

[114] Press WH, Teukolsky SA, Vetterling WT, and Flannery BP. *Numerical Recipes in C.* The Art of Scientific Computing. Cambridge University Press, second edition, 1992.

[115] Ratkowsky DA. *Nonlinear regression modelling.* Marcel Dekker, Inc, 1983.

[116] Armitage and Berry G. *Statistical methods in medical research.* Blackwell Scientific Publications, 1994.

[117] Weiss ES. Tuberculin sensitivity in Alaska. *Public Health Reports*, 68:23–27, 1953.

[118] Palmer SR, Caul EO, Donald DE, Kwantes W, and Tillet H. An outbreak of shingles? *The Lancet*, pages 1108–1111, November 16 1985.

[119] Anderson RM and May RM. *Infectious Diseases of Humans — Dynamics and Control.* Oxford University Press, 1992.

[120] Hope-Simpson RE. The period of transmission of certain epidemic diseases. *The Lancet*, Nov 13:755–760, 1948.

[121] Comstock GW. Frost revisited: the modern epidemiology of tuberculosis. *American Journal of Epidemiology*, 101(5):363–382, 1975.

[122] Lowe CR. An association between smoking and respiratory tuberculosis. *British Medical Journal*, 2(Nov 10):1081–1086, 1956.

[123] Ross R. *The prevention of malaria.* Murray, London, 1910.

[124] Macdonald G. The analysis of equilibrium in malaria. *Tropical Diseases Bulletin*, 49:813–29, 1952.

[125] Macdonald G. *The epidemiology and control of malaria.* Oxford University Press, London, 1956.

[126] Dietz K. The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research*, 2:23–41, 1993.

[127] Abbey H. An examination of the Reed-Frost theory of epidemics. *Human Biology*, 24:201–233, 1952.

[128] Metz JAJ, Nisbet RM, and Geritz SAH. How should we define 'fitness' for general ecological scenarios. *TREE*, 7(6):198–202, 1992.

[129] Levin BR, Bull JJ, and Stewart FM. The intrinsic rate of increase of HIV/AIDS: epidemiological and evolutionary implications. *Mathematical Biosciences*, 132:69–96, 1996.

[130] Benenson AS. *Control of communicable diseases in man.* American Public Health Association, Washington, fifteenth edition, 1990.

[131] Miller E, Marshall R, and Vurdien J. Epidemiology, outcome and control of varicella-zoster infection. *Reviews in Medical Microbiology*, 4:222–230, 1993.

[132] Farer LS, Lowell AM, and Meador MP. Extrapulmonary tuberculosis in the United States. *American Journal of Epidemiology*, 109:205–217, 1979.

[133] de Cock KM, Soro B, Coulibaly IM, and Lucas SB. Tuberculosis and HIV infection in Sub-Saharan Africa. *Journal of the American Medical Association*, 268:1581–1587, 1992.