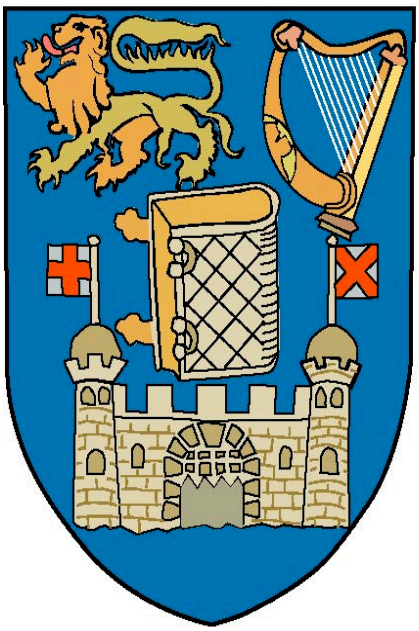
Mathematical Models of the spread of the Human Papillomavirus (HPV) and simulation of the impact of an immunisation programme in Ireland



Kathryn Marie Tobin

A thesis submitted to the University of Dublin

For the degree of Doctor of Philosophy

School of Nursing and Midwifery

University of Dublin, Trinity College

Under the supervision of Professor Catherine Comiskey

Declaration:

I hereby declare that the work described in this thesis is, except where otherwise stated, entirely my own work and has not been submitted as an exercise for a degree at this or any other university.

I agree that Trinity College Library may lend or copy this thesis upon request.

Signed,

_____

Katy Tobin

August 21st 2012.

Acknowledgements:

I would like to take this opportunity to thank my supervisor, Professor Catherine Comiskey for her support and guidance over the past three years. In particular I would like to thank her for providing me with endless opportunities to develop new skills and to continue to challenge myself. I would like to thank the School of Nursing and Midwifery for funding this project and also all of the staff and postgrad students in the school for allowing me to work with them in a supportive and inspiring environment. I would also like to thank Dr. Yoon Choi for his advice and guidance over the past few years.

I dedicate this thesis to my parents and brother for their endless love and support over the past 25 years.

Mathematical Models of the spread of the Human Papillomavirus (HPV) and simulation of the impact of an immunisation programme in Ireland

Kathryn Marie Tobin

It is now widely accepted that the Human Papillomavirus is the causative agent of cervical cancer. Disease control strategies such as vaccination programmes are often evaluated using mathematical models, as the carrying out of such experiments in a human population is unethical. Mathematical models can estimate the number needed to vaccinate in order to eradicate disease. Mathematical models of HPV dynamics and vaccine efficacy are used to project the impact of an immunization programme on the future dynamics of HPV and cervical cancer. The aim of this PhD research was to evaluate the impact of the HPV vaccine on the future dynamics of HPV in Ireland.

An SIR (Susceptible – Infectious – Recovered) model consisting of a system of ordinary differential equations (ODEs) was developed to represent HPV transmission dynamics in Ireland. Parameter values for the model equations were derived from publicly available data on HPV prevalence, sexual behaviour and population statistics. The model was solved using classical biomathematical techniques, and simulated using the modelling software packages MATLAB and Berkeley Madonna. Ethical approval was not required since this project does not involve human participants. The project only uses publicly available anonymised data.

The model was used to simulate the current dynamics of HPV in Ireland, and the results were found to be consistent with previously published models. Analytical and numerical solutions for the transmission parameter β and the WAIFW matrix have been evaluated using available data. Analysis of the model for the natural history of infection under three possible sexual mixing scenarios showed that the assumption of assortative mixing underestimated the prevalence of infection, and the assumption of proportionate mixing was most appropriate. Upon introducing a vaccination parameter to the model, and simulation of the model under various vaccination scenarios, it was found that a vaccine targeting males and females results in a more rapid reduction in prevalence than a female only vaccine. Detailed analysis of the case where 80% of young women are vaccinated each year showed that after 10 years of vaccination, prevalence is reduced by 84% in females. This reduction increases to 91% as the population reaches an endemic equilibrium after approximately 80 years of vaccination.

Abbreviations

AIN	anal intraepithelial neoplasia
CIN	cervical intraepithelial neoplasia
HPV	human papillomavirus
HSIL	high-grade squamous intraepithelial lesion
LSIL	low-grade squamous intraepithelial lesion
ODE	ordinary differential equations
PIN	penile intraepithelial neoplasia
SIR	susceptible - infected - recovered
WAIFW	who acquires infection from whom

Mathematical Models of the spread of the Human Papillomavirus (HPV) and simulation of the impact of an immunisation programme in Ireland

Contents

1 History of the Human Papillomavirus and Cervical Cancer Research 15

1.1 Introduction . . . 15

1.2 Viruses . . . 15

1.3 Human Papillomavirus . . . 18

1.3.1 Transmission . . . 20

1.3.2 HPV Infection in Males . . . 21

1.3.3 HPV Infection in Infants . . . 21

1.4 HPV and Cervical Cancer . . . 21

1.5 HPV Vaccine . . . 23

1.6 HPV Worldwide . . . 24

1.6.1 Australia . . . 25

1.6.2 United States . . . 26

1.6.3 Europe . . . 26

1.6.4 Ireland . . . 27

1.6.5 United Kingdom . . . 28

1.7 The History of Cervical Cancer Research . . . 29

1.7.1 Early Epidemiological Research . . . 29

1.7.2 Virological Research . . . 31

1.7.3 21st Century Research . . . 35

1.8 Conclusion . . . 35

2 Development of Mathematical Epidemiology and its Application to Sexually Transmitted Infections 37

2.1 Introduction . . . 37
2.2 Development of the Field of Epidemiology . . . 37
2.2.1 Early Epidemiology . . . 38
2.2.2 Classical Epidemiology . . . 39
2.2.3 Modern Epidemiology . . . 41
2.3 Development of Mathematical Epidemiology . . . 43
2.4 Population Models . . . 45
2.5 Review of Sexually Transmitted Infection Modelling and HPV Models . . . 50
2.6 General Deterministic Model for Endemic Disease . . . 53
2.6.1 Limitations of the Endemic Model . . . 59
2.7 Conclusion . . . 60

3 Simple Homogeneous Model: Analytical Solutions and Stability Analysis 61
3.1 Introduction . . . 61
3.2 SIR Model with Vaccination . . . 62
3.3 Model Equations . . . 63
3.4 Assumptions of the Basic Model . . . 64
3.5 Solving the Basic Model . . . 65
3.5.1 Stability Analysis of the Disease-Free Equilibrium . . . 67
3.5.2 Stability Analysis of the Endemic Equilibrium . . . 69
3.6 Natural History Model with Heterogeneity . . . 71
3.7 Conclusion . . . 72

4 Analysis of a Simple Homogeneous Model and the Introduction of Heterogeneity and Model Calibration 75
4.1 Introduction . . . 75
4.2 Simple Homogeneous Model . . . 76
4.2.1 General Model Equations . . . 76
4.2.2 Full System of Equations . . . 78
4.2.3 Epidemiological Parameters . . . 81
4.2.4 Threshold Estimates and Calibration of β . . . 83

4.3 Risk Heterogeneity in Sexual Behaviour 86
4.3.1 Parameter Input Values 90
4.3.2 Effective Contact Rate, $\hat{c}$ 92
4.3.3 Estimating the Transmission Parameter, β 93
4.3.4 Sensitivity Analysis 95
4.3.5 Basic Reproductive Number, R_0 97
4.4 Conclusion 99

5 Analysis of a Two Sex, Risk-Structured Model with Various Patterns of Sexual Mixing and Calculation of R_0 101

5.1 Introduction 101
5.2 Modelling sexual mixing patterns 102
5.3 WAIFW Matrices 103
5.4 Model Equations and Assumptions 107
5.5 Preliminary Results of Numerical Simulation and Sensitivity Analysis . . . 110
5.6 Calibration of the Model, adjusting μ 115
5.7 Numerical Analysis of a Heterogeneous Model with Various Sexual Mixing Patterns 118
5.7.1 Effects of Mixing Patterns on Sub-group Prevalence 121
5.8 Basic Reproductive Number 125
5.8.1 Basic Reproductive Number for the General Case 126
5.8.2 Basic Reproductive Number for the Male Population, Assuming Assortative Mixing 128
5.8.3 R_0 Assuming Disassortative Mixing 131
5.8.4 R_0 Assuming Proportionate Mixing 131
5.9 Discussion 133
5.10 Conclusion 134

6 Analysis of a HPV Vaccination Model 137

6.1 Introduction 137
6.2 Vaccination Model Development and Analysis 138

6.2.1 Vaccination Model Assumptions, Equations and Parameter Estimates 138
6.2.2 Vaccination Scenario Analysis 143
6.3 Reduction in Cervical Cancer Cases 151
6.4 Conclusion 152

7 Conclusions **153**
7.1 Introduction 153
7.2 Further Work 158

List of Figures

1.1 Influenza virus structure. 16
1.2 Electron Microscopy Image of Influenza Virus. 17
1.3 Electron Microscopy Image of Human Papillomavirus. 18
1.4 Division of HPV types. 19
1.5 Stages of HPV infection 22
1.6 Age standardised incidence and mortality rates, cervical cancer by region of the world 25
1.7 Picture of a mounted cottontail rabbit 32
1.8 Engraving from Tableau Encyclopedique et Methodique 33

2.1 John Snow Broad Street Map 40
2.2 Plot of the Malthusian parameter, r, for the Exponential Model 47
2.3 Plot of population size, varying the carrying capacity, K, for the Logistic Model Model 48
2.4 Compartmental diagram for an SIR Model 54
2.5 Phase portrait for the case when $i > 0$ and $\sigma < 1$ 58
2.6 Phase portrait for the case when $i > 0$ and $\sigma > 1$ 59

3.1 SIR model for endemic disease with vaccination 63

4.1 Beta parameter plot 84
4.2 Beta Calibration 94

5.1 Bar Chart of Sensitivity Analysis for the Assortative Mixing Case 112
5.2 Bar Chart of Sensitivity Analysis for the Disassortative Mixing Case 113

5.3 Bar Chart of Sensitivity Analysis for the Proportionate Mixing Case 114
5.4 Calibration of μ under the assumption of Assortative Mixing 115
5.5 Calibration of μ under the assumption of Disassortative Mixing 116
5.6 Calibration of μ under the assumption of Proportionate Mixing 117
5.7 Prevalence Plots under Three Sexual Mixing Scenarios 119
5.8 Risk group prevalence, with Assortative Mixing 121
5.9 Risk group prevalence, with Disassortative Mixing 122
5.10 Risk group prevalence, with Proportionate Mixing 124

6.1 Endemic prevalence of infection for a female only vaccine at 70% coverage . 145
6.2 Endemic prevalence of infection for a female only vaccine at 80% coverage . 146
6.3 Endemic prevalence of infection for a female and male vaccine at 70% coverage 146
6.4 Endemic prevalence of infection for a female and male vaccine at 80% coverage 147
6.5 Annual prevalence of infection assuming 80% vaccination coverage 148

List of Tables

4.1 Number of Partners per year, by gender. 90
4.2 Sensitivity Analysis 96
4.3 Adjusted Parameter Set 97

5.1 Parameter Values 106
5.2 Parameter Estimates 109
5.3 Total Prevalence of Infection in Males and Females 110
5.4 Sensitivity Analysis for the Assortative Mixing Case 112
5.5 Sensitivity Analysis for the Disassortative Mixing Case 113
5.6 Sensitivity Analysis for the Proportionate Mixing Case 114
5.7 Prevalence in Males and Females under various Mixing Scenarios 119
5.8 Endemic Prevalence in High and Low Risk Males under various Mixing Scenarios 125
5.9 Endemic Prevalence in High and Low Risk Females under various Mixing Scenarios 125
5.10 Parameter values for the Basic Reproductive Number 128
5.11 Basic Reproductive Number under various Mixing Scenarios 132

6.1 Vaccination scenarios and associated reduction in prevalence 144
6.2 Annual reduction in prevalence for a female only vaccine at 80% coverage . 149

Summary

This summary highlights the main aims of the research presented in this thesis. The Human Papillomaviruses (HPVs) are a group over 120 viruses known to infect humans (Moody & Laimins 2010). Approximately 40 strains of this virus are capable of invading mucosal surfaces of the body (Moody & Laimins 2010) and HPV is now regarded as the most common sexually transmitted infection (STI) (CDC 2011a). Two strains of HPV, known as HPV 16 and 18, are the cause of approximately 70% of cervical cancers (WHO 2007). Vaccination is one of the control measures that has been developed to prevent infection with HPV 16 and 18.

Mathematical models in healthcare research have been shown to provide an extensive platform for the advancement of the scientific understanding of infectious disease dynamics. Through the exploration of mathematical models, we can obtain quantitative predictions of the mechanisms of communicable diseases such as HPV. Appropriate and optimum strategies of disease control can be formed from this research without the need for expensive or ethically conflicting experiments.

This project aims to develop mathematical models to simulate the effects of a vaccination programme on the prevalence of HPV 16 and 18 in Ireland. Using classical mathematical modelling techniques a two-sex, risk structured Ordinary Differential Equation (ODE) transmission dynamic model for HPV 16 and 18 will be developed for the Irish population.

The specific research questions to be answered in this thesis are as follows:
Which epidemiological parameters have the most significant effect on the natural history

of HPV? Related to this are smaller research questions which must be addressed such as, which sexual mixing patterns fit the model outcomes best? What assumptions can be made on the gender differences in parameter values?
Another primary research question will be: What effect will a HPV vaccine have on prevalence in a population? Again, in answering this, smaller research questions must be answered such as, what are the most appropriate assumptions to be made about the vaccine characteristics? What effect will the vaccine have on HPV prevalence in the short term, for example in two, five or ten years? Also, what effect will the vaccine have on HPV prevalence in the longer term, for example in 20, 50, or 100 years?

The specific objectives of this study are:

- Study the classical techniques of mathematical modelling and develop a suitable ODE model for HPV in Ireland.

- Explore the epidemiologically significant factors contributing to the spread of HPV and estimate their numerical values.

- Using mathematical model simulations, explore the effects of the various contributing factors for HPV spread such as sexual behaviour and sexual mixing patterns.

- Use calibration techniques and sensitivity analyses to strengthen the validity of parameter estimates.

- Calculate the basic reproductive number for the natural history model.

- Explore the effects of a vaccination programme on the steady state endemic prevalence of infection.

Chapter 1 provides the history of the Human Papillomavirus (HPV) and how the virus affects humans. The association between HPV and cervical cancer is described. Details of infection control strategies that are in place in Ireland, Europe and the rest of the world are provided. This chapter also looks at how cervical cancer research has developed.

Chapter 2 is a detailed history of how the field of epidemiology developed and its progression to the analysis of sexually transmitted infections. This chapter also contains a literature review of mathematical models for sexually transmitted infections and HPV models.

In chapter 3 a simple ODE model for endemic disease is solved analytically and a stability analysis is carried out on the disease free and endemic equilibria to evaluate the dynamics of infection. This analysis provides valuable information about disease dynamics that cannot be achieved using a more complex model such as those in chapters 4, 5 and 6.

Chapter 4 introduces the first numerical simulation of the thesis. A simple homogeneous model for the natural history of HPV infection is simulated using estimated parameters from published data. A process of model calibration is carried out around the parameter β. A degree of heterogeneity is added to the model in the form of two risk groups based on sexual activity.

Chapter 5 develops the natural history endemic model from chapter 4 by analysing various sexual mixing patterns using WAIFW (Who Acquires Infection From Whom) matrices and carrying out detailed sensitivity analyses to guide the calibration of model parameters. An important epidemiological result in this chapter was the calculation of the basic reproductive number under the three sexual mixing scenarios: assortative, disassortative and proportionate.

And finally, chapter 6 introduces a vaccination parameter to the population which enables the simulation and epidemiological analysis of numerous vaccination scenarios such as the varying effects of a female only vaccine versus a male and female vaccine and the

relative reduction in HPV prevalence that could be achieved by these vaccines at various levels of population coverage.

Chapter 1

History of the Human Papillomavirus and Cervical Cancer Research

1.1 Introduction

This introductory chapter outlines the journey of discovery of the human papillomavirus (HPV) to the development of cervical cancer research and describes the most prominent research that has lead to the development of a life saving vaccine against the two main cancer-causing strains of HPV. This chapter also looks at the effects of HPV worldwide and the control measures that have been implemented in different countries.

1.2 Viruses

Viruses are infectious, non-living particles, which vary in structure and function. The field of virology, which is the study of viruses, emerged after the invention of the electron microscope in 1931 (Ruska 1986). Viral cells are generally smaller than 200 nm in diameter, which is smaller than bacterial cells, and consist of at least two parts: an external protein structure called a capsid, and an inner core of nucleic acid, either deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). The capsid of some, but not all viruses is surrounded by

an outer membrane called an envelope (Mader 1998). The envelope is constructed from the structural components of the host cell, a cell that is infected with the virus.

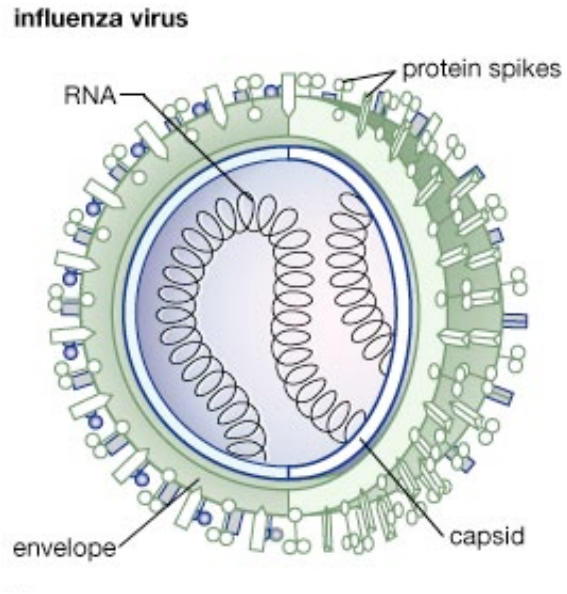


Figure 1.1: Influenza virus structure. Encyclopdia Britannica (2009).

A virus is classified according to: The type of nucleic acid in its core and whether it is single or double stranded; its size and shape; and whether it is enveloped or naked, as shown in Figure 1.1. Viruses are obligate intracellular parasites, meaning they are non-living and can only survive inside a living cell. They are capable of mutation, for example, the influenza virus mutates frequently, creating new strains to evade destruction from flu vaccines and so, new vaccines must be developed annually to combat new strains. Viruses are host specific, that is, they only infect specific host-cells. Portions of the capsid attach to host cell receptors using a lock-and-key mechanism. The viral nucleic acid then enters the host cell. Viral function typically requires the host's cellular enzymes, that is, the virus utilises the host cell's metabolic machinery for its own reproduction (Mader 1998).

Certain animal viruses are a cause of particular concern for humans. For example the papillomavirus, the herpes viruses, the hepatitis viruses and the adenoviruses cause disease and are also capable of altering the host cell's genome and creating cancerous cells (Mader 1998).

In humans, viral diseases are controlled by prevention techniques. This includes physical barriers such as condoms which can reduce the transmission of HIV (human immunodeficiency virus); frequent hand washing/ use of hand sanitisers can reduce transmission of an influenza virus. Vaccines have been developed to induce immunity to certain viruses such as polio, measles, mumps and more recently HPV (human papillomavirus). Antiviral drugs are currently used to treat certain forms of the influenza virus shown in Figure 1.2.

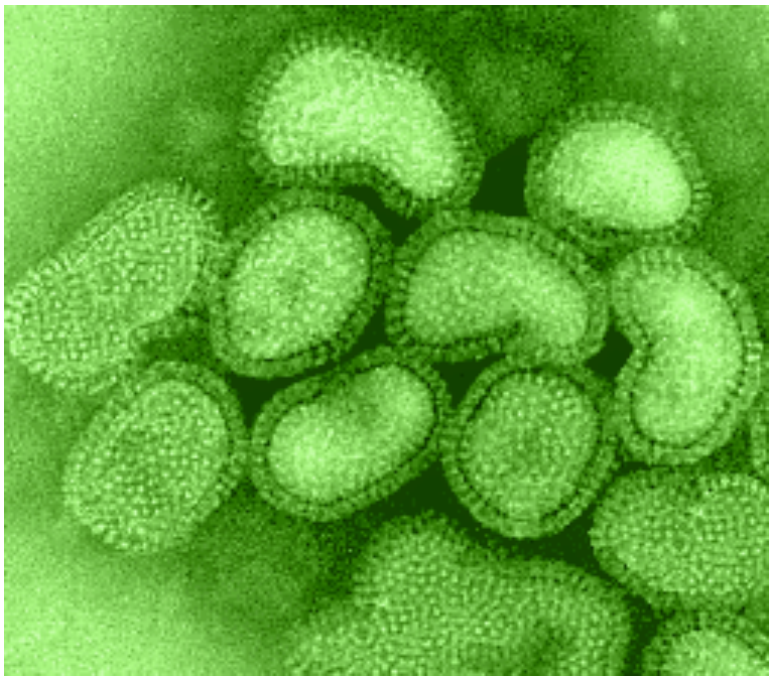


Figure 1.2: Electron Microscopy Image of Influenza Virus. (Stannard 1995)

1.3 Human Papillomavirus

Human Papillomaviruses (HPVs) are small, non-enveloped, double-stranded DNA viruses that infect epithelial cells of the skin and mucosal membranes (WHO 2009). An electron microscopy image of the virus is shown in Figure 1.3. The epithelium is the outer layer of cells covering all surfaces of the body including the skin and mucous producing membranes such as the respiratory and genital tracts. Papillomaviruses are prevalent amongst animal species and are species-specific. HPV infections of the skin may cause the growth of warts (verrucae) typically on the hands and feet, or genital warts, which are a contagious and recognisable symptom of genital HPV infection (WHO 2009). Infection is often subclinical and in most cases will resolve itself without the host knowing it existed, although less commonly, some HPV types may progress to cancer, most notably to cervical cancer.

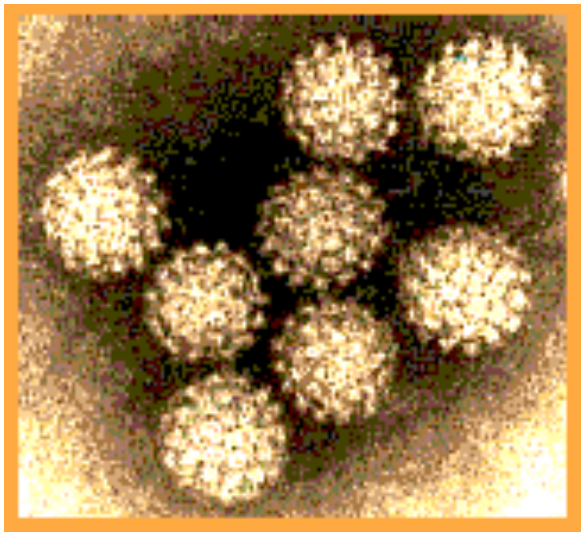


Figure 1.3: Electron Microscopy Image of Human Papillomavirus. (Stannard, 1995).

More than 120 HPV genotypes have been identified, of which approximately 40 are capable of invading mucosal surfaces (Moody & Laimins 2010). Figure 1.4 shows the distribution of these strains according to their associated mechanism of causing infection. These

are categorised as high risk strains (oncogenic or cancer causing) and low risk strains (non-oncogenic) relative to their association with cancer. Oncogenic HPV types are pathogenically linked to intraepithelial neoplasia (tumour growth).

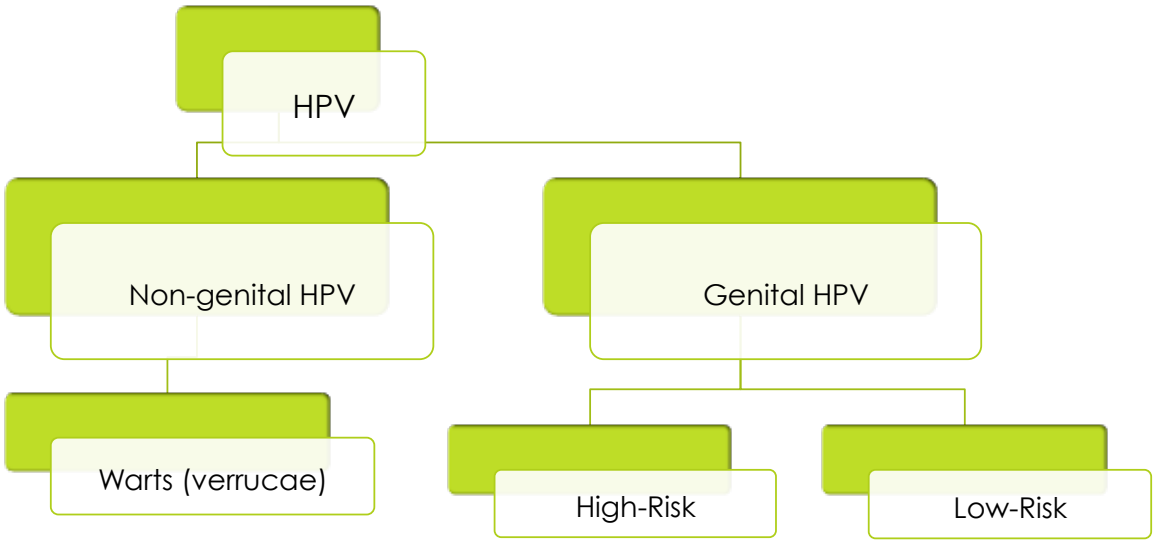


Figure 1.4: Division of HPV types.

Low-risk forms of HPV commonly cause genital warts or benign changes in the cells of the cervix but are not associated with invasive cancers. The most common low-risk strains are 6, 11, 40, 42, 43, 44, 54, 61, 72, 73 and 81 (CDC 2007). Of these low-risk strains, 6 and 11 cause between 90% and 100% of all cases of genital warts (Greer *et al.* 1995). The prevalence of genital warts is highest at 20 to 29 years of age for females and males; female incidence falls after this time, but incidence remains high in males up to approximately 40 years of age (Koshiol *et al.* 2004).

The most common strains of high risk HPV are 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 66 (WHO 2007). Types 16 and 18 are responsible for approximately 70% of all cervical cancers (WHO, 2007). The body may eliminate infection with a high-risk strain of HPV, particularly if the host has a strong immune system. However, cases of cervical cancer and precancerous cervical states are almost exclusively related to infection with these high-risk strains. In its early stages, the virus can cause cell changes called cervical dysplasia, which can lead to cancer if untreated. Although less common, the virus is also associated with cancers of the vulva, penis and anus, as well as other sites.

1.3.1 Transmission

HPV is highly transmissible and is now regarded as the most common sexually transmitted infection (STI) (CDC 2011a). It is estimated that over half of all sexually active males and females will be infected with HPV at some time (CDC 2007). HPV is generally transmitted via skin-to-skin contact during sexual intercourse, and less commonly through other forms of non-penetrative genital contact. Sexual behaviour is directly related to the probability of acquiring a HPV infection. Prevalence of cervical HPV infection is highest amongst women under the age of 25 (Woodman *et al.* 2007) and lowest amongst women who have never had sex. Increased risk of exposure to HPV is proportionally linked to infection, therefore abstaining from sexual activity ensures the lowest risk. A monogamous sexual relationship with a partner who has had no or few previous partners decreases the risk of contracting an infection, as does the correct use of physical barriers such as condoms.

It is important to emphasise that infection is usually eliminated spontaneously. Less than 10% of cases suffer from persistent high-risk infection and are at high risk of developing neoplastic lesions (abnormal cell growth) of the anogenital tract (CDC 2007). There are contributing factors which may increase the risk of this progression, such as smoking cigarettes, multiparity (multiple births), long-term use of hormonal contraceptives (Castellsagué & Munoz 2003), immuno-suppression associated with conditions such as HIV, and

possibly co-infection with Chlamydia trachomatis or Herpes simplex virus (WHO 2007). However, those at highest risk, regardless of other cofactors are women who do not attend regular screening.

1.3.2 HPV Infection in Males

Low risk HPV infection is associated with genital warts in men. Infection with high-risk types is associated with some pre-invasive squamous lesions of the penis and anus (penile intraepithelial neoplasia - PIN, and anal intraepithelial neoplasia AIN) and penile and anal cancer. Invasive penile cancer is very rare, and cases of anal cancer are significantly higher in men who have sex with men (MSM), particularly when coupled with HIV (CDC 2011b).

1.3.3 HPV Infection in Infants

It is possible for low-risk HPV infections to be transmitted from mother to baby during birth, though this is very rare (CDC 2011a). This results in juvenile-onset recurrent respiratory papillomatosis (JORRP) and causes warts on the respiratory tract of the infant.

1.4 HPV and Cervical Cancer

Extensive epidemiological and clinical studies have illustrated that high-risk (or oncogenic) strains of HPV are associated with virtually 100% of cervical cancer cases, opening the debate of whether cervical cancer may occur without HPV (Castellsagué 2008). Infection is in general transient, about 90% are cleared within two years. The median duration for infection is eight months (Ho *et al.* 1998). A persistent high-risk infection (for example

HPV 16) is uncommon but is necessary for progression to invasive disease. According to a Genital HPV Infection Fact Sheet published by the CDC (Centers for Disease Control and Prevention) approximately 10% of infected women will experience persistent infection (CDC 2007). It is possible that in cases where infection appears to be cleared from the body, HPV is not eliminated but remains in a dormant, non-detectable state and may resume reproduction later in life. This would explain why seemingly new infections have appeared in women who were thought to have HPV DNA clearance without being exposed to a new strain of the virus, that is, they have remained in a mutually monogamous relationship.

Cervical carcinogenesis can be described in three stages as shown in Figure 1.5: The majority of infections are transient, meaning they clear spontaneously. A small proportion of infections are persistent in nature and may lead to cervical cancer if left untreated.

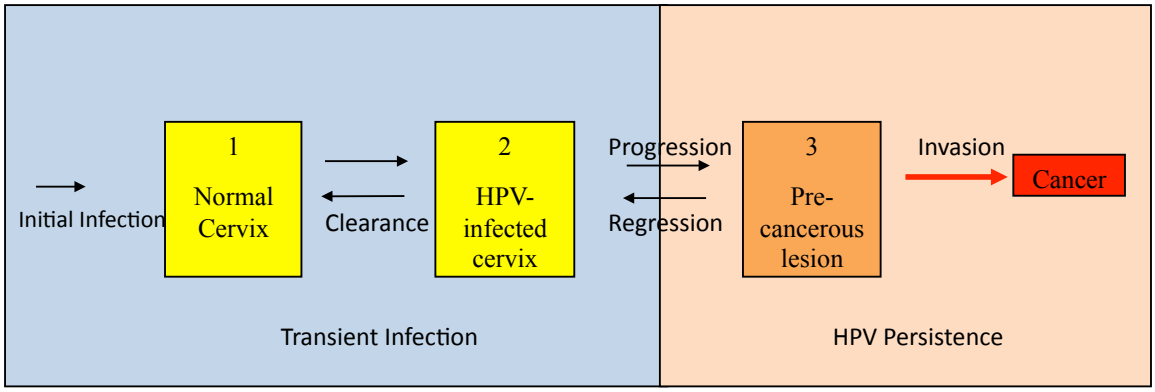


Figure 1.5: Stages of HPV infection

HPV infects the innermost layer of the epidermis, the basal layer. It is thought to evade the natural barrier of the skin by entering through micro-abrasions in the cervical epithelial layer to reach the basal cells. Infection can then cause cellular dysplasia, a pre-cancerous condition which can be categorised according to its severity and risk of becoming cancerous. An abnormal growth of cervical squamous cells (outer surface cells of the epidermis, which have a flattened appearance) can be classified cytologically as a low-grade squamous intraepithelial lesion (LSIL) or a high-grade SIL (HSIL), depending on the abnormality of

the cells and how much of the cervix is affected. Abnormal cervical cell biopsies can also be defined by histological examination as CIN (Cervical Intraepithelial Neoplasia). CIN are graded according to the degree of neoplasia and how much of the cervix it covers; CIN1 = mild, CIN2 = moderate, CIN3 = severe. In general CIN1 lesions are cleared by the body within a few months without treatment, cases of persistent infection can lead to CIN2/3 lesions and a higher risk of cancer formation.

1.5 HPV Vaccine

The HPV vaccine is designed to prevent infection from the most common strains of HPV and acts to complement regular cervical screening rather than replace it. Vaccinated women are advised to continue to be screened according to national guidelines since cervical cancer can be caused from less common high-risk strains not included in the vaccine.

There are currently two prophylactic HPV vaccines on the market; Gardasil which is manufactured by Merck (marketed in Europe by Sanofi Pasteur MSD) and Cervarix by Glaxo Smith Kline (Health Service Executive 2010). Gardasil is a quadrivalent vaccine, meaning it protects against four strains of the virus, the two most common low-risk strains HPV 6 and 11, and the two most common high-risk strains HPV 16 and 18. Cervarix is a bivalent vaccine and protects against HPV 16 and 18.

These vaccines do not contain any viral DNA and are non-infectious. They are prepared using virus-like particle (VLP) technology. The HPV capsid which encloses the HPV genome is made up of two proteins, L1 and L2. Purified L1 protein can self-assemble in vitro, forming hollow shells which act very like HPV and with the aid of an adjuvant, an immune response is initiated which produces much more antibody than would be induced during a natural HPV infection. These vaccines are not therapeutic; they are designed to prevent

infections from the HPV types contained in the vaccine in individuals who have not been previously infected with these types. The vaccine is administered over a period of six months, with three separate 0.5 ml intramuscular doses. Gardasil doses are injected at months 0, 2 and 6, while Cervarix is administered at 0, 1 and 6 months. The duration of immune protection is currently unknown. The longest follow-up study for HPV 16 has been up to 9 years post-vaccination and shows lasting antibody persistence and protection from HPV persistent infection (Jit *et al.* 2011). Future studies will be necessary to evaluate the duration of protection offered by this vaccination program. Side effects of the vaccine have been minimal, with pain at the injection site being the most common complaint (CDC 2011c). At this stage the main arguments against the vaccine are the uncertainty of duration of immunity, the high cost of the vaccine and the fear that it may lead to a false sense of security and encourage women to stop attending regular cervical screening.

1.6 HPV Worldwide

Cervical cancer affects about 530,000 women worldwide every year; about 275,000 women lose their lives to the disease (Ferlay *et al.* 2008). More than 85% of cases occur in developing countries where screening and treatment are extremely limited (American Cancer Society 2011). Cervical cancer is the number one cause of cancer-related death in women in the majority of developing countries. Figure 1.6 shows the world estimates of incidence and mortality from cervical cancer in 2002. Incidence and mortality are highest in Africa. Incidence of cervical cancer is higher in Western Europe than in Northern America, Northern Europe and Australia.

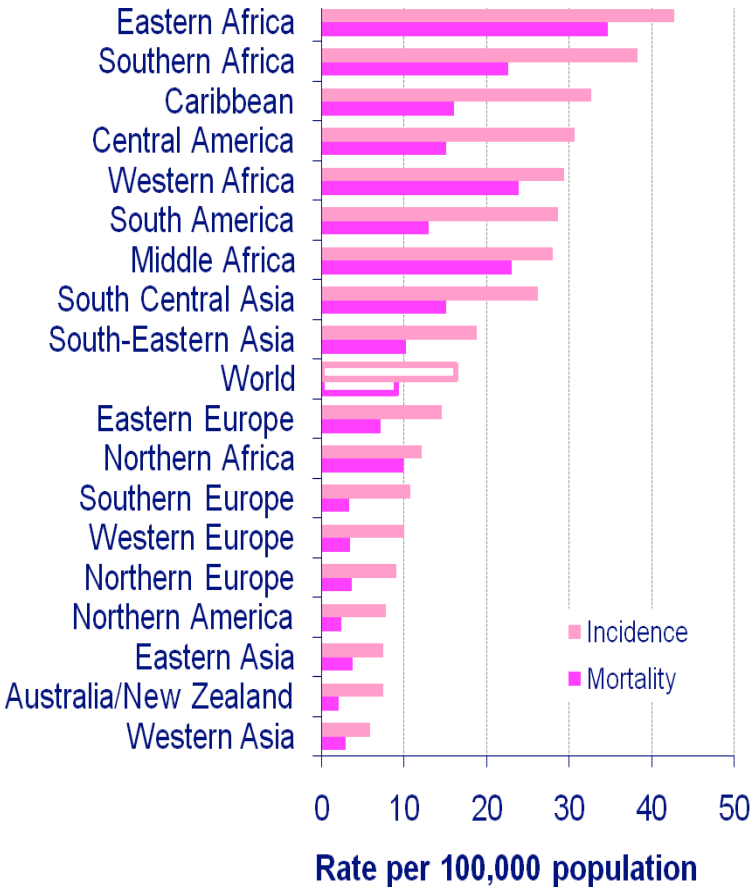


Figure 1.6: Age standardised incidence and mortality rates, cervical cancer by region of the world, 2002 estimates (Cancer Research UK 2008).

1.6.1 Australia

Australia commenced a very successful National Cervical Cancer Screening Program in 1991. At the introduction of this program the incidence of cervical cancer in women aged 20-69 was 17.1 new cases per 100,000 and mortality was at 4.0 deaths per 100,000. In 2004, the number of new cases had dropped to 8.9 per 100,000 and mortality fell to 2.0 per 100,000 in 2005 (Australian Institute of Health and Welfare 2008).

On 29th November 2006, the Australian Government announced funding for a National HPV Vaccination Programme using Gardasil. This is funded under the National Immuni-

sation Programme and consists of two separate projects. One is an ongoing school-based program to vaccinate 12 and 13 year old girls and also girls in schools aged 13 up to 18 in 2007 and 2008 as part of a catch-up programme. They also funded a community-based HPV catch-up program to vaccinate women aged 18-26. This catch-up program ended on 30th June 2009.

1.6.2 United States

The age-standardised rate of cervical cancer on the entire continent of America is about 18.8 per 100,000 with a mortality rate of 8.1. In North America the rate of incidence is 7.7 per 100,000 and mortality is about 2.3 per 100,000 (WHO/ICO, 2007). Gardasil received FDA (United States Food and Drugs Administration) approval in June 2006. Vaccination protocol varies from state-to-state in the U.S. Most large private insurance companies cover the cost of recommended vaccines such as Gardasil. Children aged 18 or younger who do not have private healthcare may be eligible for the vaccine through the Vaccines for Children (VFC) program. Some states also provide free or low-cost vaccines at public health department clinics for those without private health insurance.

1.6.3 Europe

Approximately 15,000 European women die each year from cervical cancer, with about 33,000 diagnosed cases. This is the second most common cancer in European women aged 15-44 years (second to breast cancer). The most common strains of HPV associated with cervical cancer in Europe are HPV 16/18 which together account for 73% of cases. As much as 80% of cases of this disease can be prevented by maintaining an adequate screening programme and treating precancerous lesions (ECCA 2009). This screening procedure began in the 1960s, and brought about a significant decrease in cervical cancer cases in many EU countries in the 1970s. However this form of prevention is not without its flaws. Reliable results are dependent on an adequate amount of sample being taken and correct

slide preparation and analysis by a trained laboratory technician.

Organised cervical screening programmes are in place in nine EU member states: Denmark, Finland, Slovenia, Sweden, the Netherlands, the United Kingdom, Ireland, Poland, and parts of Italy. Although screening does take place in all EU member states, without an organised program which includes the widespread provision of information on the importance of screening, uptake is opportunistic and will tend to be lower, particularly among low socio-economic groups.

According to a survey carried out by the European Cervical Cancer Association (ECCA 2009), only 9 countries of the 40 European countries who participated in the survey offer the HPV vaccine free of charge to females of at least one age cohort (ECCA, 2009). These countries are Denmark, Germany, Greece, Italy, Luxembourg, Netherlands, Portugal, Spain and the UK, and a further three countries (Belgium, France and Sweden) offer partial funding for the vaccine. Since the publication of this report, Ireland has also introduced a fully funded HPV vaccination programme. All of these countries are in Western Europe where healthcare budgets are high and better able to cope with the high cost of the vaccine. Incidence of cervical cancer is generally already lower in these countries than in Eastern Europe due to successful cervical screening practices. The UK and the Republic of Ireland have introduced school-based programmes nationally, which is known to reach a high proportion of the target population (Health Protection Agency 2012, Health Service Executive 2011). This is in comparison to introducing the vaccine through healthcare providers, which can have suboptimal coverage rates as people from low socioeconomic groups and minorities may miss out. This is also the group most likely to miss out on screening opportunities when they are offered opportunistically.

1.6.4 Ireland

Approximately 170 new cases of cervical cancer occur annually in Ireland with about 76 deaths. Cervical Check is Ireland's national screening programme. This is a population

based, quality assured programme which was launched in September 2008 and is managed by the National Cancer Screening Service (NCSS). This service is free to women aged 25-60 years and is funded by the Department of Health and Children. Screening is provided every 3 years for women aged 25-44 and every 5 years for women aged 45-60 in accordance with guidelines set out by the World Health Organisation (WHO). The implementation of this programme can potentially reduce the number of cervical cancer cases in Ireland by up to 80%.

According to the National Cancer Registry (2010), the age specific rate of cervical cancer in 2009 for women aged 20-24 years was 2.56 per 100,000 and has been generally increasing in recent years, the rate was 1.94 in the year 2000 (National Cancer Registry 2010). This opens the debate of whether the age of screening should be reduced to 20 years of age. However, these figures may seem low when compared to the peak incidence rate for women aged 40-44 which was 33.55 per 100,000 in the year 2009 (National Cancer Registry 2010).

1.6.5 United Kingdom

The UK have implemented a screening programme similar to Cervical Check in Ireland. Both of these programmes offer a population based service for women from the age of 25 years. However, in Northern Ireland, Scotland and Wales, the age at which women begin to be screened is 20 years. This was also the case in England until 2004 when the age was raised to 25 years in line with WHO guidelines. The WHO, International Agency for Research on Cancer (IARC) and European guidelines recommend that screening should begin at age 25 (Ferlay *et al.* 2010). Evidence shows that though HPV infection is very common in women <30 years, the number of cases developing into cancer is very low. The WHO states that screening younger women will detect many lesions that will never develop into cancer, which will lead to considerable over treatment and will not be cost effective. A national vaccination program using Cervarix was introduced in the UK in September 2008 for girls aged 12-13, funded by the government (Health Protection Agency 2012). This included a 2 year catch-up programme for young women up to the age of 18.

Cervical cancer was the third most commonly diagnosed cancer worldwide in females in 2008 (American Cancer Society 2011). As sections 1.6.1 - 1.6.5 show, control strategies are being implemented in developed countries to reduce cancer incidence. The burden of cervical cancer is greatest in developing countries mainly due to the lack of screening services in these populations (American Cancer Society 2011).

1.7 The History of Cervical Cancer Research

1.7.1 Early Epidemiological Research

It is now a well known fact that cervical cancer is caused by an infectious agent. Many authors attribute the first report of this relationship to the work of Italian surgeon Domenico Antonio Rigoni-Stern (as noted by Griffiths 1991 and Skrabanek 2000). However, this work is mostly misquoted (as noted by Griffiths 1991 and Skrabanek 2000). At the time of the research, in 1842, the diagnosis of "cancer of the womb" did not differentiate between carcinoma of the cervix and other cancers of the uterus, meaning Rigoni-Stern could not make any accurate observations on the incidence of deaths from cervical cancer.

The paper was written in Italian, but was translated to English and published in 1987. Speaking on the subject of cervical cancer in 1984, Drife stated that *"it is now well documented that the disease is rare in nuns and common in prostitutes"* and furthermore he claimed that *"a connection between intercourse and cervical cancer was apparently first suggested in 1842"* (as cited by Griffiths 1991). These statements are false, but commonly believed to be true. The findings of Rigoni-Stern are clearly presented in his work and do not postulate a relationship between cervical cancer and sexual activity, a discovery he is often credited for.

In 1950, French-Canadian gynaecologist Fabien Gagnon produced evidence for the popular claim that cervical cancer is rare in nuns (Gagnon 1950). Gagnon provided care to a number of Roman Catholic convents and through this position he gained access to the medical records of a number of religious orders going back many years. He studied "medical files of an annual average of 13,000 women, covering a twenty-year period, in the archives of many different convents of nuns". He found no cases of cervical cancer, but admitted that data from 3500 files were missing (1500 had been destroyed following the death of each individual and a further 2500 files could not be verified). Gagnon searched further using records from pathological laboratories in the same area. These statistics identified three cases of cervical cancer in nuns but the writer admits that some cases may have escaped attention. Gagnon concluded that although it is rare, cancer of the cervix can occur in virgin women.

Janet Towne, a gynaecologist based in Chicago, explored the role of childbirth in cervical cancer (Towne 1955). Studying cases from a radiography unit over a twenty-one year period, she concluded that 16.3% of cervical carcinomas occurred in nulliparous women (women who have never given birth), 6.4% in unmarried women and 0.52% in nuns (3 cases). Towne extended her study to include analysis of medical records from convents and found a further 3 cases. From this she concluded, "cervical malignancy can arise in women irrespective of virginity or parity".

A study carried out by Fraumeni *et al.* (1969) sought to clarify the role of marital status in human carcinogenesis. They reviewed 5893 death certificates covering 41 religious orders between the years 1900 and 1954. The study showed that of a total of 1021 cancer deaths, 102 were cases of uterine cancer and 11 of these cases were cancers of the cervix.

Petr Skrabanek joined the Department of Community Health in Trinity College Dublin in 1984 where his work earned him a reputation as a convincing critic, particularly in the field of preventative medicine. In a paper entitled "Cervical cancer in nuns and prosti-

tutes: A plea for scientific continence" first published in 1988, Petr Skrabanek reviews the epidemiological research carried out on the subject of cervical cancer in the 19^{th} and 20^{th} centuries. In his conclusion, Skrabanek states

"The epidemiological evidence on the prevalence of cervical cancer in nuns and prostitutes is of very poor quality and neither supports nor contradicts the belief that cervical cancer is a venereal disease" (Skrabanek, 2000).

These epidemiological studies were imperative to the discovery of the aetiology of cervical cancer, but more research was needed since, as stated by Skrabanek (2000), *"Epidemiological research cannot prove causation"*. Epidemiological research led to the conclusion that cervical cancer was infectious and could be transmitted to others through sex, the next step was to isolate and identify the infectious agent.

It is clearly evidenced here that the link between cervical cancer and sexual contact was postulated long before strong evidence for the association became available. The studies present here show some fundamental flaws, for example, Gagnon (1950) found three cases of cervical cancer in nuns and concluded that although rare, cervical cancer can occur in virgin women. He came to this conclusion without showing rigorous proof of the fact that these were virgin women, other than stating that they were nuns.

The following section outlines the history of virological research which has helped to develop scientific knowledge of HPV infection and associated disease.

1.7.2 Virological Research

This section describes the prominent work that lead to the discovery that HPV is the causative agent of cervical cancer, and the development of control strategies for this disease.

Figure 1.7: This is a picture of a mounted cottontail rabbit at the Museum of Natural History at the University of Kansas. The animal has a severe infection of the Shope papillomavirus. The picture was taken by Heather York.

Peyton Rous

Peyton Rous was born in Texas in 1879 and grew up in Baltimore. In 1909, after completing medical school, Peyton turned his attentions to medical research. He joined the staff of the Rockefeller Institute in New York where he was put in charge of the cancer research laboratory. Rous won the Nobel Prize in Physiology or Medicine in 1966 for his discovery of tumour-inducing viruses. He proved that some tumours in chickens are initiated and driven by viruses known as Rous Sarcoma viruses (Rous 1911).

Richard Shope

In 1934, Rous was asked to work on a new virus discovered by a close friend in the Rockefeller Institute. Dr. Richard Shope had already achieved considerable success in the field of virology, being the first to isolate influenza virus A from pigs (Shope 1931).

"Our attention was recently called to a disease occurring in wild cottontail rabbits in northwestern Iowa. Rabbits shot there by hunters were said to have numerous horn-like protuberances on the skin over various parts of their bodies. The animals were referred to popularly as "horned" or "warty" rabbits" (Shope, 1933) (Figure 1.7).

Shope concluded that he had found a papilloma in wild cottontail rabbits that is transmissible to wild and domestic animals. This virus was named the Shope papillomavirus, which later became known as the cottontail rabbit papillomavirus (CRPV).
In a paper published from the Laboratories of the Rockerfeller Institute for Medical Research, Rous states that the papillomas induced in domestic rabbits do progress to malignancies under certain conditions (Rous and Beard 1935).
It is thought that the presence of these papillomas in cottontail rabbits gave rise to the legend of the "Jackalope"; a mythical animal described as being a cross between a rabbit and an antelope, goat or deer, as shown in Figure 1.8.
The causative link between Human Papillomavirus and cervical cancer was not postulated until the 1970s.

Figure 1.8: Engraving from Tableau Encyclopedique et Methodique, 1789.

Harald zur Hausen

German virologist Harald zur Hausen qualified as a doctor in 1960 and developed a keen interest in infectious diseases and microbiology. After working as a researcher in Philadelphia, zur Hausen returned to his native Germany where he ran his own laboratory at the Institute of Virology in Würzburg.

While working at Würzburg, he began to question the popular hypothesis that the herpes simplex virus (HSV-2) was the causative agent of cervical cancer. As Professor of Virology at the University of Erlangen-Nuremberg in Bavaria, zur Hausen began testing other possible etiological agents for cervical cancer. During this process he discovered that papilloma was a group of many viruses, not just one (McIntyre 2005).

zur Hausen took the Chair at the Institute of Virology at Freiburg in 1977. His team successfully isolated HPV 6 and 11. They went on to isolate HPV 16, which they proved was present in about 50% of cervical cancers and HPV 18, present in a further 18-20% of cases. This research was viewed as very controversial at a time when most scientists favoured herpes simplex type 2 (HSV-2) as the etiological agent for cervical cancer (Nahmias *et al.* 1970). zur Hausen approached a number of drugs companies in Germany about the possibility of creating a vaccine for this form of cancer, but his idea was rejected as the companies did not see a market for such a vaccine. Nearly a decade passed before the results of zur Hausen's research were accepted and research began on the development of a vaccine. zur Hausen won the Nobel Prize in Physiology or Medicine in 2008 for his discovery that the human papillomaviruses are the causative agents of cervical cancer.

George Papanicolaou

Another prominent name in the history of cervical cancer research is George Papanicolaou; a physician and researcher, born in Greece. Papanicolaou worked as a researcher in

the pathology laboratory in Cornell Medical College for 47 years. While studying vaginal smears from guinea pigs, he noted changes in the epithelial cells, which he related to the menstruation cycle of the animal. In 1923 Papanicolaou began working on human vaginal smears from women who had cervical cancer. He found cancer cells in these smears and proposed that analysing smears of vaginal fluid microscopically could diagnose cancer and reduce the need for biopsies. He published a definitive explanation of his findings in 1943 showing how cervical cancer could be diagnosed before the onset of symptoms (Thoms 1943). His work was widely accepted and the Pap smear (named after Papanicolaou) became a routine screening technique saving thousands of womens lives in ensuing decades.

1.7.3 21st Century Research

The development of two prophylactic vaccines for cervical cancer has opened the door to a new wave of research. In the field of epidemiology, work is being carried out worldwide to establish what effect this vaccine will have on future incidence and mortality rates for this disease. Epidemiological findings complement virological research to increase our understanding of the trends of this disease and make informed decisions that utilise the vaccine's full efficacy. For example, epidemiological research can predict the ideal age for a patient receiving the vaccine, or analyse the cost effectiveness of vaccinating males.

1.8 Conclusion

This chapter detailed the history of HPV and cervical cancer research. It also looked at the various control strategies being put in place around the world to reduce disease incidence and outlined the prominent researchers involved in the development of our knowledge of

the virology of HPV and cervical cancer. The next chapter describes the evolution of the field of epidemiology which has become a crucial tool in the field of infectious disease. The application of epidemiological techniques to sexually transmitted infection research is also introduced.

Chapter 2

Development of Mathematical Epidemiology and its Application to Sexually Transmitted Infections

2.1 Introduction

This chapter details the landmark studies that have shaped the evolution of the field of epidemiology and mathematical modelling in healthcare. Some of the classical mathematical models such as the exponential and logistic models are explored. Using all of this knowledge as a foundation, and a review of previously published HPV models, a structure is chosen and presented for the mathematical models which will be explored in subsequent chapters of this thesis.

2.2 Development of the Field of Epidemiology

Numerous definitions have been proposed to accurately describe the subject of epidemiology; one comprehensive definition states that epidemiology is

"The study of the occurrence and distribution of health-related states or events in specified populations, including the study of the determinants influencing such states, and the application of this knowledge to control the health problems." (Porta M., 2008).

Epidemiology is often used as the starting point in developing and assessing treatment/control programmes for infectious disease and to describe patterns of disease whether they are endemic, epidemic or pandemic, as defined below.

A disease outbreak can be described in terms of its incidence and prevalence in a given population. Incidence is the number of new cases of disease in a given population over a specific time period. Point prevalence is the total number of cases of a specific disease within a population at a given point in time. Period prevalence is calculated over a specific period of time.

Endemic infections have a continuous, steady-state prevalence in a given population, for example, HPV and chickenpox.
An epidemic is a single source outbreak of an infection in excess of the expected level of incidence in a given population, for example, the mumps epidemic in 15-34 year olds in Ireland 2009 (HSE 2009).
Pandemic is a widespread epidemic where clusters of infection occur in many populations across a country, continent or worldwide, for example, HIV/AIDS and H1N1 (Swine) Influenza.

2.2.1 Early Epidemiology

Although epidemiological research has become well known in recent decades, the origins of the discipline date back to approximately 400 B.C. In an essay entitled On Airs, Waters, and Places, Hippocrates suggested that the development of disease might be influenced by

environmental and host factors such as social behaviour (Dicker *et al.* 2006).

John Graunt, a haberdasher from London published the first analysis of mortality data in 1662. His work looked at patterns of births, deaths and disease occurrence and noted inconsistencies associated with geographical location, seasonal variation and also gender (Dicker *et al.* 2006).

2.2.2 Classical Epidemiology

The work of Graunt was built upon by British Epidemiologist William Farr. In 1838, as Compiler of Extracts for the Registrar General's Office, Farr introduced the first National vital statistics system. Farr analysed this data on an annual basis and used it to compare causes of death at different occupational levels (Page *et al.* 1995). The International Classification of Diseases (ICD) used by epidemiologists and vital statisticians today was developed from a disease classification system that Farr created to support his own work (Lilienfeld, 2007).

In 1849, an outbreak of cholera killed about 15,000 people in London. London was an industrialised city with a large population and the River Thames was heavily polluted with untreated sewage. In his now famous epidemiological study of the 1854 cholera epidemic in the Golden Square of London, British physician John Snow compiled a spot map showing the geographical distribution of cholera cases and the local water sources, as he believed water was a source of infection for cholera, Figure 2.1 (Dicker *et al.* 2006).

To confirm his theory, Snow gathered information on the water sources used by those who had cholera and found that use of the Broad Street pump was a common factor among these patients. The handle of this pump was subsequently removed and the outbreak of cholera ended. Snow went on to investigate two water supply companies that served districts with a high death rate from cholera. His investigation took the form of epidemiological research still used today; beginning with descriptive epidemiology and progressing to the

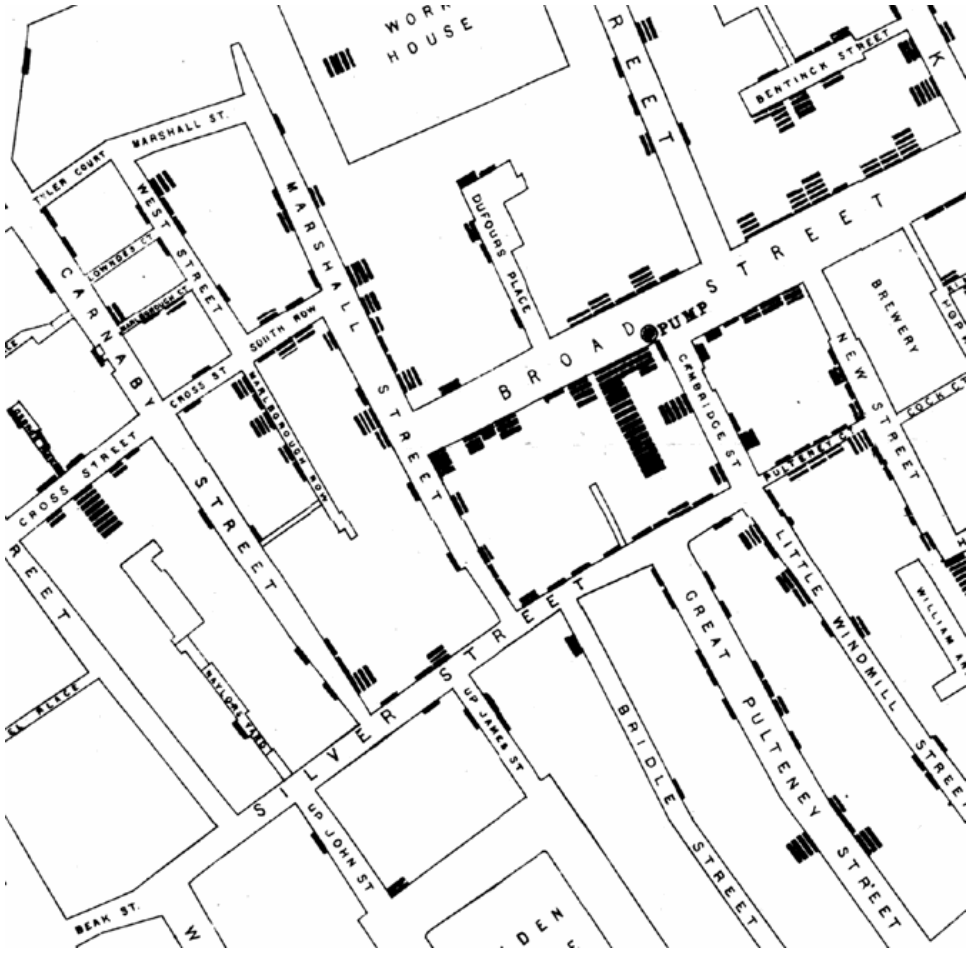


Figure 2.1: John Snow Broad Street Map (Snow J. 1855) Available from http://www.ph.ucla.edu/epi/snow/html.

generation and testing of a hypothesis, and application of a solution.

Both water supply companies took their water from the River Thames. The Southwark and Vauxhall Company's intake point was downstream of London city, while the Lambeth Company's intake came from upstream of London to avoid sewage contamination. Snow hypothesized that water taken from the Thames downstream of London was a source of cholera. To test this hypothesis he studied comparable districts served by both water companies and gathered information about water supply from each household where a death from cholera occurred over a 7-week period. He found an increased death rate associated with use of the water from the Southwark and Vauxhall Company, thus strengthening his hypothesis (Dicker *et al.* 2006).

The work of Snow and the evolution of germ theory by biologists such as Pasteur, Henle and Koch revolutionised epidemiological research around the turn of the 19^{th} century. Germ theory was the idea that microbes such as bacteria cause disease.

2.2.3 Modern Epidemiology

Epidemiological research in the late 19^{th} and early 20^{th} centuries evolved concurrently with laboratory experiments in microbiology. The field of epidemiology has developed rapidly since World War II with the introduction of epidemiological studies for numerous health-related outcomes and non-infectious diseases affecting communities, such as cancer and cardiovascular disease (Page *et al.* 1995). Epidemiologists began investigating the concept of risk factors associated with disease rather than the idea of single causative agents.

Shortly after the Second World War, physicians noted an increase in the incidence of lung cancer. Cigarette smoking was presumed to be the primary cause of this increase. The most famous study (though not the first) to examine the relationship between smoking and lung cancer was carried out by Sir Richard Doll and Sir Austin Bradford Hill. Doll and Hill performed two studies representative of epidemiological studies carried out today.

One was a case-control study and the other was a cohort study (Doll and Hill, 1950, 1954, 1956). A case-control study looks at the factors associated with one or more diseases. In a cohort study, diseases which are associated with one or more factors are determined. More specifically, in the context of epidemiology, a case-control study looks at one sample of individuals who have a given disease, the "case" group, while the "control" group does not have the disease. A cohort study has a longitudinal aspect in which you follow two groups prospectively and look for characteristics such as an increased incidence of a disease.

In their case-control study, Doll and Hill aimed to determine whether lung cancer patients differed from other persons in respect to their smoking habits. The study included patients with cancer of the lung, stomach, colon or rectum, and also patients without cancer. Detailed information on smoking history was examined and a strong positive association was found between smoking and lung cancer (as noted by Ahrens & Pigeot, 2004).

Convinced by the strong evidence produced from the case-control study, Doll and Hill began a cohort study in 1951 called the British Doctors Study. They produced a prospective study to determine the frequency with which the disease appeared, in the future, among groups of persons whose smoking habits were already known. Twenty thousand British male physicians were included in the study to further investigate the relationship between smoking and lung cancer. Doll and Hill viewed lung cancer mortality data in terms of the smoking habits of the patient and reported that.

"Though the numbers of deaths at present available are small the resulting rates reveal a significant and steadily rising mortality from deaths due to cancer of the lung as the amount of tobacco smoked increases" (Doll and Hill, 1954).

2.3 Development of Mathematical Epidemiology

Mathematics is a valuable tool in epidemiology. A mathematical model is a predictive representation of a real world situation, used to answer a specific question. In a healthcare setting, mathematical models are used to conduct experiments that would be unethical to carry out in a human population. These models have been used in recent years to predict the course of epidemics such as SARS and avian influenza (Meyers, 2007). Equally, mathematical models allow for the evaluation of disease control strategies such as vaccination or quarantine.

Daniel Bernoulli, a member of the famous European family of mathematicians, created one of the first mathematical models for an infectious disease in 1760. Variolation is an inoculation technique whereby a healthy person is intentionally exposed to smallpox from the scab of a person infected with a mild form of the disease. The idea was that a mild infection of smallpox would induce immunity against further infection from the disease (Meyers, 2007). Although variolation did reduce the probability of mortality from smallpox, the identification of suitable strains of the virus was not an exact science and deaths did occur when inoculation progressed to serious infection (Meyer 2007). Bernoulli developed a mathematical model to demonstrate the efficacy of this inoculation technique. This mathematical model was probably the first used to show the advantages of a vaccination control programme (Murray 2002). He showed that the increase in life expectancy that would be achieved if smallpox were eradicated, far outweighed the risk associated with the controversial procedure (Blower and Bernoulli, 2004).

Following Bernoulli's model, deterministic mathematical models only began to appear in epidemiological studies in the early 20^{th} century. A deterministic model is one in which each variable state is determined by defined model parameters rather than probability distributions. Hamer introduced the mass-action principle to infectious disease dynamics (Meyer 2007). This principle states that the rate at which infection spreads is proportional to the product of the densities of susceptibles times infectious individuals (Anderson and May, 1991). Ross constructed a well-structured mathematical model which analyzed basic pa-

rameters associated with transmission of malaria (Bailey 1975). He developed a differential equation model for malaria as a host-vector disease (Hethcote 2000). Ross' malaria model paved the way for subsequent researchers to use mathematical theory in epidemiological research (Bailey 1975). Pioneering work of Kermack and McKendrick had a profound influence on the development of modern mathematical models for disease spread. In their first paper they introduced the famous compartmental model to describe disease transmission (Kermack and McKendrick 1927). Their most outstanding result was the famous Threshold Theorem (Kermack and McKendrick 1927), which states that for an epidemic outbreak to occur, the density of susceptibles must exceed a certain critical value (Bailey 1957, Hethcote 2000). Briefly, this theorem applies to a closed population of size N which can be given by the set of equations for an SIR model below:

The population N is divided into three disjoint classes x, y and z defined as follows:

The *Susceptible* class, x, consists of individuals who can contract infection but are not yet infective.

The *Infective* class, y, are the individuals who are infected and transmitting infection to others.

The *Removed* class, z, consists of those individuals who have recovered or become immune to infection.

$$\begin{aligned} \frac{dx}{dt} &= -bxy, \\ \frac{dy}{dt} &= bxy - gy, \\ \frac{dz}{dt} &= gy \end{aligned}$$

where b is the rate of infection, g is the removal rate of infections from the population. The structure and definitions of the SIR model are further discussed in section 2.6.

The *Threshold Theorem* result states that an epidemic will spread in the population only

if $x(0) > g/b$, that is, if the initial proportion of susceptibles is greater than the removal rate divided by the rate of infection.

The SIR model described above ignores births and deaths in the population. The following section describes two classic population models which include the effects of reproduction.

2.4 Population Models

Population models are mathematical models that are applied to population dynamics. This section describes the structure of two classic simple population models. Living organisms must reproduce for a population to survive. Thus, a population model must include reproduction, represented in the form of birth and death rates. The well known Exponential Model and the Logistic Model are two important models based on patterns of reproduction and are described presently.

We begin by defining a population of size N and analyse its dynamics over a period of time t. The *rate of change* of $N(t)$ given by $\frac{dN(t)}{dt}$, where $N(t)$ is the population size at time t can be expressed as

$$\frac{dN(t)}{dt} = births - deaths + migration$$

Exponential Model: This model is often associated with Thomas Robert Malthus, famous for his work on population growth models in the 19th century. The simplest form of this model has no migration, and births and deaths are proportional to N. This takes the form

$$\frac{dN}{dt} = bN - dN \tag{2.1}$$

where b and d are positive constants (b = births, d = deaths) and the initial population when $t = 0$ is given by $N(0) = N_0$. This has the explicit solution

$$N(t) = N_0 e^{(b-d)t} \tag{2.2}$$

The *intrinsic growth rate* or *Malthusian parameter*, r, is equal to the birth rate, b, minus death rate d, that is, $r = b - d$. This model has 3 possible outcomes which have been plotted in Figure 2.2:

$r > 0$ Population exponentially increases

$r < 0$ Population exponentially declines

$r = 0$ Population does not change

This model does not take seasonality in reproduction rates into account, which is seen in the animal kingdom. It also assumes all organisms are identical and ignores age structure. Finally, this model assumes that resources are unlimited. An example of exponential growth is the introduction of a virus such as SARS into a population where there is no immunisation programme in place. Each infected person can infect multiple individuals and the rate of infection rises exponentially. Another example is human reproduction in the absence of environmental constaints, that is, assuming resources such as food and space are unlimited. The human population, or equivalently an animal population in the absence of predators, rises exponentially when births are greater than deaths. However, in reality, many populations are restricted in their rate of growth by environmental constraints. The following model addresses this scenario.

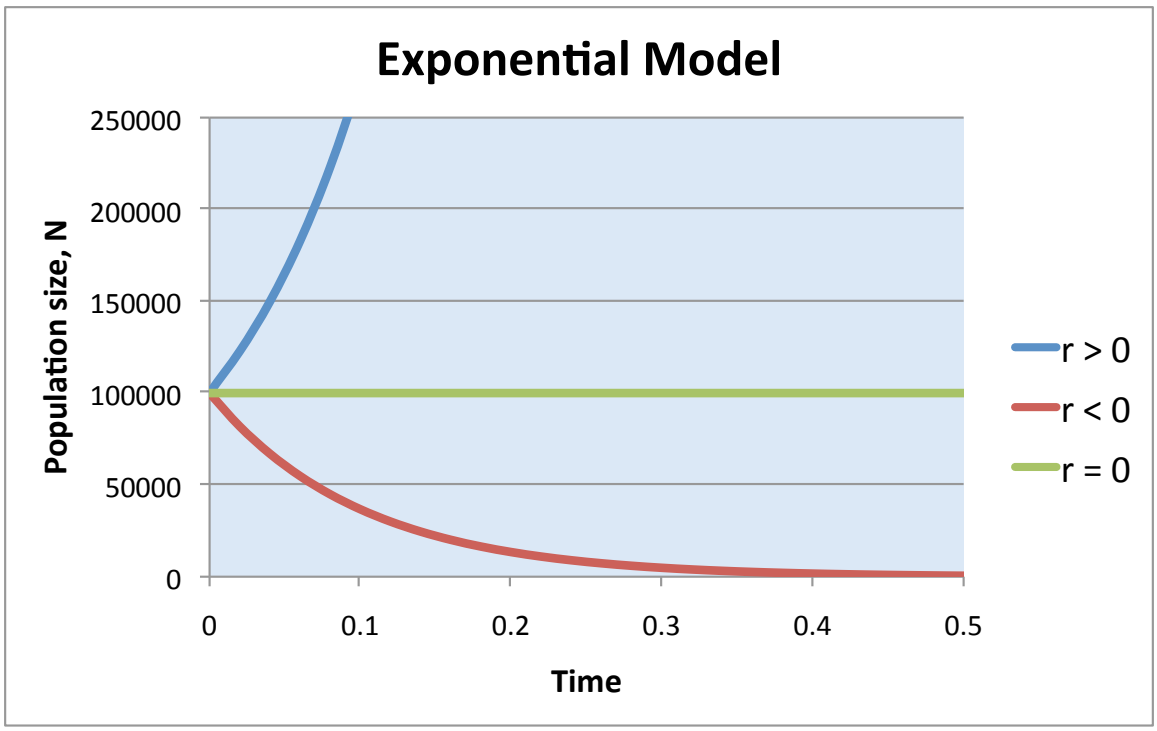


Figure 2.2: Plot of the Malthusian parameter, r, for the Exponential Model

Belgian mathematician Pierre Verhulst modified Malthus' model in 1838 and presented the Logistic Model for population growth (Murray 2002).

Logistic Model: Verhulst's logistic growth model suggested that the rate of population increase may be self limiting and depend on population density and availability of resources. He proposed that

$$\frac{dN}{dt} = rN(1 - \frac{N}{K}) \tag{2.3}$$

where r is as before and K is the *carrying capacity* of the environment, the maximum population that the environment can support. Equation (2.3) has the solution

$$N(t) = \frac{N_0 K e^{rt}}{[K + N_0(e^{rt} - 1)]} \tag{2.4}$$

when initial condition $N(0) = N_0$. If food supplies are plentiful, the population should grow exponentially. As the population N grows towards the *carrying capacity* K, the *per capita* growth rate $r(1 - \frac{N}{K})$ moves towards 0, that is, $N(t) \to K$ as $t \to \infty$.

This model has 3 possible outcomes:

$K > N_0$ The logistic curve. Population increases and reaches a plateau

$K < N_0$ Population decreases and reaches a plateau

$K = N_0$ Population does not change

These three solutions are shown in Fig 2.3.

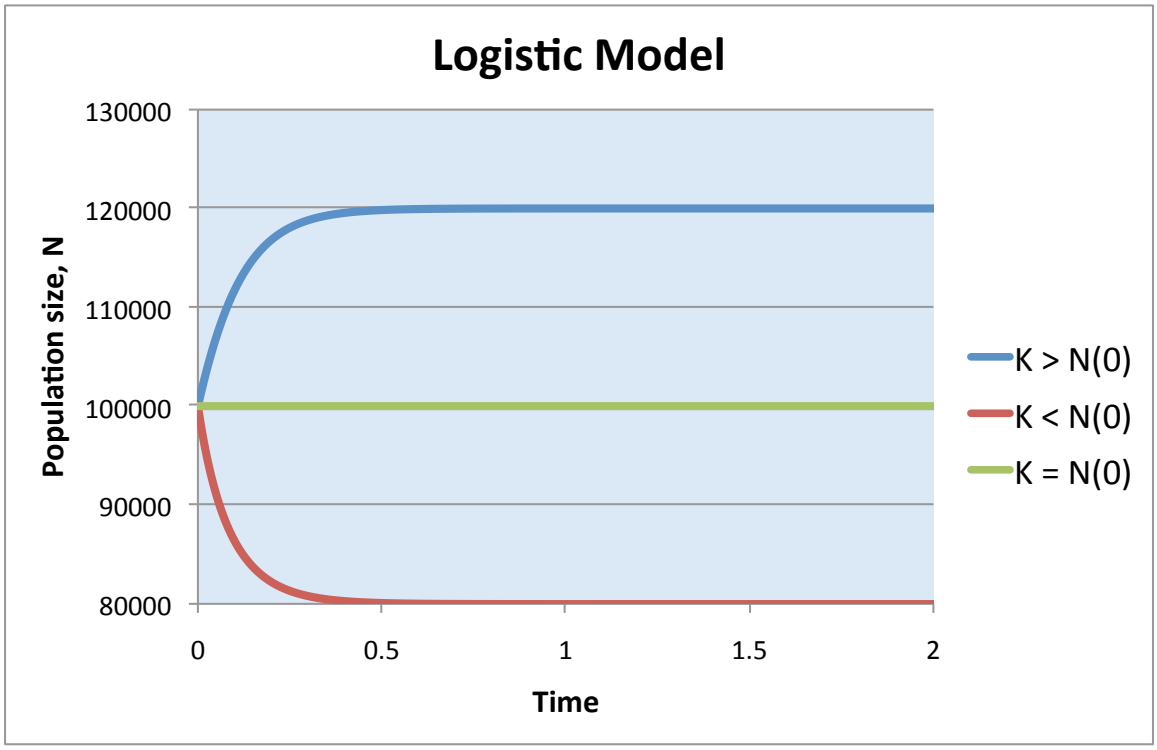


Figure 2.3: Plot of population size, varying the carrying capacity, K, for the Logistic Model Model

Given that this thesis focusses on HPV, a cancer causing virus, it seems appropriate to mention a third population model that has seen considerable success in modelling cancer cell growth.

Gompertzian Model: Modelling tumour growth has become a leading area of research given the significant number of cancer related deaths. Most studies follow population growth models consisting of differential equations (Lo 2009). The Gompertzian model has been used to describe avascular tumour growth (Behera and O'Rourke 2008). These deterministic models have been successfully applied to fit experimental data (Behera and O'Rourke 2008). The model can be described as follows. If $x(t)$ is tht evolume of the tumour at time t, the deterministic growth is defined by the following equation:

$$dx = \{A_1 x - A_2 x \ln(x)\}dt$$

Here A_1 is the intrinsic growth rate of the tumour, A_2 is the rate at which growth decelerates. The solution of the equation is given by the following sigmoidal function. A sigmoidal function is a type of mathematical model for a time series, where growth is slowest at the start and end of a time period.

$$x(t) - \exp\{\frac{A_1}{A_2} - [\frac{A_1}{A_2} - \ln(x_0)]\exp(-A_2 t)\}$$

The equilibrium point $x(\infty) = \exp(A_1/A_2)$ represents the *carrying capacity*, which is the largest tumour size that an organism can tolerate.

Moving on from the evolution of mathematical epidemiology and some of the classic models, we can now focus on developing an understanding of the various model structures that can be used to describe HPV. The following section reviews the available literature on modelling sexually transmitted infections and HPV with the aim of selecting a model structure for this thesis.

2.5 Review of Sexually Transmitted Infection Modelling and HPV Models

The process of creating a model for a sexually transmitted infection (STI) is quite different than for other infections in a population such as measles or rubella. The epidemiology of many STIs are influenced by social, biological and behavioural factors. Firstly, we can only consider the sexually active individuals of a population. Papers by Hethcote and Yorke on the transmission dynamics of gonorrhea defined the future development of mathematical models for STIs (Hethcote and Yorke 1984). In their research they established concepts which have become standard in STI transmission modelling. These include the notion of tracing infectors (individuals who are infected and spreading infection) rather than infectees which proved most effective in disease control strategies. They introduced the concept of a "core group" which has a high partner change rate and therefore contributes a disproportionately large amount to the prevalence of infection in the population. They showed that tracing cases of infection in the core group could effectively control an STI in a population (Hethcote and Yorke 1984).

Epidemiologic equations for a sexually transmitted infection differ from other infectious diseases in the general population, for example, if we consider an STI in a homogeneously mixing population, the basic reproductive number R_0, which is the average number of secondary infections that occur when one infective is introduced into a completely susceptible population (Hethcote 2000), is given by

$$R_0 = \beta c D$$

where c denotes the average rate of partner change, β is the probability of transmission per partnership, and D is the average duration of infectiousness (Vynnycky and White 2010). This equation is not affected by population density as it would with other diseases such as measles, where an increase in population density will be likely to increase the average number of contacts an infective individual has. In STI dynamics, increased population density would not normally result in increased promiscuity among individuals in a population. This fact is discussed in greater detail in chapter 5, section 5.2.

In developed countries, the persistence of an STI in a population is often dependent on a small number of asymptomatic individuals since most symptomatic individuals seek treatment relatively quickly. This is known as the carrier phenomenon (Anderson and May 1991). Another important feature in STI dynamics is that most STIs result in little or no acquired immunity (Anderson and May 1991, Garnett 2002). An individual with an infection such as chlamydia or gonorrhea revert back to the susceptible class following infection. This is not the case for hepatitis B infection, where individuals acquire immunity following infection, HPV where the vaccine confers immunity before infection or AIDS where remission does not occur.

The HPV vaccine is a recently developed technology and its longterm effects are yet unknown. Mathematical models are being developed to bridge the gap between the initial administration of the vaccine today and its effects over the decades to come. A number of models have been used to project the benefits and cost effectiveness of HPV vaccination programmes and various questions related to cervical cancer. The models constructed so far differ in the assumptions and parameters that are included and also the model structure, of which there are currently three types: Markov/cohort model, dynamic population model and hybrid model. These three model structures and previously published studies are discussed below:

Cohort Models: This type of model examines the progression of a single cohort in the population whose structure is typically linear and based on probability. Hughes *et al.* (2002) published a cohort model which looked at the impact that a HPV vaccination program would have on HPV prevalence and how a reduction in HPV incidence might affect the incidence of cervical cancer . This cohort of women studied became susceptibles at time 0 (assumed individuals become sexually active at age 16). Risk of infection varied according to age and sexual activity. In 2003, Sanders and Taira published a HPV cohort model that attempted to evaluate the effectiveness and cost effectiveness of a prophylactic vaccine (Sanders and Taira 2003). The hypothetical cohort comprised of 12 year old girls. The

cohort was divided into girls who received the vaccine at age 12, about 70 percent of the cohort, and girls who did not receive the vaccine but received standard care. Vaccine efficacy was set to 75 percent with ten years protection. Other cohort models include Kulasingam and Myers (2003) and two models by Goldie *et. al* (2003, 2004).

Population Dynamic Models: Dynamic models track a changing population over time. These models include vital dynamics, that is, individuals are allowed to enter and leave the model in the form of births and deaths. Dynamic models include the force of infection, λ, the rate at which susceptibles become infected, which changes over time. A similar parameter is used in a cohort model, but it is a fixed rate and does not change as the prevalence of infection is reduced over time which results in herd immunity. Cohort models do not take account of the benefits of herd immunity.

Hughes *et al.* (2002) published the first dynamic model for HPV. They concluded that HPV prevalence could be reduced by 44 percent in women by vaccinating males and females (a vaccine of 75% effectiveness, 10 years duration). In 2005, Elbasha and Galvani modelled the changes in HPV type distribution in a population where mass vaccination was implemented (Elbasha and Galvani 2005). They used a system of ordinary differential equations and simulated progression of infection using a nine-compartment model. The model considered synergistic versus antagonistic interactions between two HPV strains and the resulting effect on HPV prevalence. They considered an imperfect vaccine that can protect against one or both HPV strains. Other models include Barnabas *et al.* (2007) and Jit *et al.* (2008) who carried out an economic evaluation of vaccination strategy in the UK and sorted the population by HPV strain, age, sex and sexual activity based risk group. They also included three groups for HPV type (16, 18 and other) and a further two groups for anogenital warts (6, 11). Usher *et al.* (2008) constructed a transmission dynamics model for the cost effectiveness of a vaccination program in Ireland.

Hybrid models: These are a combination of cohort and dynamic models. Taira *et al.* (2004) created a hybrid model where they combined a transmission model with their previous cohort model. The main advantage of a hybrid model is that the transmission dynamics

estimates the HPV infection rate over time (λ) for the cohort of interest.

After reviewing the available literature, the decision was made to create a population transmission dynamic model for HPV in Ireland. A transmission dynamic model, though perhaps more labour intensive in terms of computation and returning results, offers more advantages over a static cohort model. With the inclusion of vital dynamics, the herd immunity effect can be assessed. A dynamic model does require more information about behaviour in a population and history of infection, therefore is subject to greater uncertainty than a static model. The next section outlines the crucial structures of a deterministic model for an endemic disease, defines the key parameters and describes the phase plane portrait for the model system.

2.6 General Deterministic Model for Endemic Disease

We begin as before with a population of size N which we divide into disjoint classes which change with time t. Following Hethcote's influential papers on the mathematical biology of sexually transmitted disease (Hethcote 1989, 2000) we consider an infection which divides N into three classes whose dynamics can be analysed in the widely discussed SIR model (Anderson and May 1991, Vynnycky and White 2010). The three classes are defined as follows:

The *Susceptible* class, S, consists of individuals who can contract infection but are not yet infective.

The *Infective* class, I, are the individuals who are infected and transmitting infection to

others.

The *Removed* class, R, consists of those individuals who have been removed from $S - I$ interactions as a result of recovery with conferred immunity, isolation or death.

This SIR model can be represented by the following compartmental diagram, Figure 2.4, showing the direction of movement of individuals in the population through time.

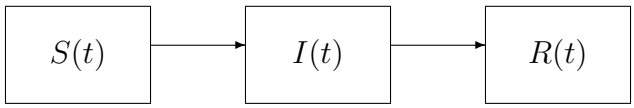


Figure 2.4: Compartmental diagram for an SIR Model

This model is confined by some epidemiological assumptions.

- We allow N to be sufficiently large such that each class can be considered continuous variables.
- The population mixes homogeneously. This means that for a model of a sexually transmitted infection, individuals choose sexual partners completely randomly.

Since HPV is an *endemic* disease which will remain present in a population over a long time period, the model must include vital dynamics, allowing births into the susceptible class and natural deaths to leave the population.

- Rates of natural births and deaths are equal so the population remains stationary and newborns entering the population are all susceptible.
- The latent period for the infection is assumed to be 0. The latent period is the time between exposure to infection and subsequent illness. A mechanism for HPV latency

has not been established (Woodman *et al.* 2007). Therefore it is unknown whether individuals can infect others when they are in a state of latency. To overcome this uncertainty it is assumed here that an individual is capable of spreading HPV infection as soon as they have contracted HPV, thus the latent period is set to zero.

- β is the contact rate, the average number of adequate contacts of a person per unit time. An adequate contact is sufficient for the transmission of infection to occur if the recipient is susceptible to the infection.

- Individuals move from I to R at the rate $I\gamma$, where γ is the recovery rate per unit time.

- Individuals are removed from each class by death at the removal rate per unit time, μ, which is proportional to the class size.

- The contact number σ is the average number of adequate contacts of a typical infective during the infectious period. This value remains constant throughout the infectious period and is equal to the basic reproductive rate, defined below.

- The basic reproductive rate, $R_0 = \frac{\beta}{\gamma+\mu}$ is the average number of secondary infections that occur when one infective is introduced into a completely susceptible population (Hethcote 2000). This value is equal to the contact number, $R_0 = \sigma$, and is calculated as the contact rate β times the average death-adjusted infectious period $\frac{1}{\gamma+\mu}$ (Hethcote 2000).

The Initial Value Problem (IVP) for this model is given by the following system of equations (2.5). An IVP is a differential equation with a specified value, the initial condition.

$$
\begin{aligned}
\frac{dS}{dt} &= \mu N - \frac{\beta IS}{N} - \mu S, \qquad S(0) = S_0 \geq 0 \\
\frac{dI}{dt} &= \frac{\beta IS}{N} - \gamma I - \mu I, \qquad I(0) = I_0 \geq 0 \\
\frac{dR}{dt} &= \gamma I - \mu R, \qquad R(0) = R_0 \geq 0 \\
S+I+R &= N
\end{aligned}
\tag{2.5}
$$

where S, I and R represent the number of individuals in each class and N is the total number of individuals in the population. This system of equations is similar to that for an epidemic model except for the inclusion of births given by the term μN in the first equation of the system, and natural deaths μS, μI, μR, the last term in each equation. Dividing each equation in (2.5) by N, the IVP becomes

$$
\begin{aligned}
\frac{ds}{dt} &= -\beta is + \mu - \mu s, \qquad s(0) = s_0 \geq 0 \\
\frac{di}{dt} &= \beta is - (\gamma + \mu)i, \qquad i(0) = i_0 \geq 0
\end{aligned}
\tag{2.6}
$$

Here, s, i and r represent the proportion of the total population in each group, and n is the total population and has a value of 1. So, all three classes sum to n, that is, $s + i + r = n = 1$.

To investigate infection dynamics in the population, it is sufficient to analyse the IVP in the si plane since $r(t)$ can be found easily from $s(t)$ and $i(t)$ by the equation $r(t) = 1 - s(t) - i(t)$.

A phase plane portrait is a graphical display of the characteristics of a differential equation. The interaction between the variables s and i through time, t, can be displayed in a phase plane plot and the analysis of this plot is used to determine the stability of the solutions to the system.

Phase Plane Analysis: We define a triangle in the si plane to be the region of epidemiological significance

$$T = \{(s, i) | s \geq 0, i \geq 0, s + i \leq 1\}$$

Theorem 2.0: *Let $(s(t), i(t))$ be a solution of (2.6) in T. If $\sigma \leq 1$ or $i_0 = 0$, then the solution paths starting in T approach the disease-free equilibrium given by $s = 1$ and $i = 0$. If $\sigma > 1$, then all solution paths with $i_0 > 0$ approach the endemic equilibrium given by $s_e = \frac{1}{\sigma}$ and $i_e = \frac{\mu(\sigma-1)}{\beta}$.* (Hethcote 2000)

This theorem says that if the contact number σ is less than one, or if the initial proportion of infected individuals in the population is zero, then the disease will not survive in the population, and will approach the disease-free equilibrium through time, t. If the contact number is greater than one, $\sigma > 1$ and the initial proportion of infected individuals in the population is greater than zero, $i_0 > 0$, then the infection will be maintained in the population. The solution path will spiral towards the endemic equilibrium through time.

Figures 2.5 and 2.6 are phase plane portraits for the cases described in theorem 2.0. The susceptible class, s, is plotted against the infective class, i, through time. The phase plane plot shows how s changes with respect to i through time.

If the contact number σ, which is the *threshold quantity* is < 1, the disease will die out, since each infective infects less than 1 susceptible. Over time as the infective class approaches 0, the recovered class slowly dies off and the susceptible class increases and approaches 1 due to new births, until the population is full of susceptibles at the disease-free equilibrium with $s = 1$ and $i = 0$. If $R_0 = \sigma > 1$ and the initial infective fraction $i_0 > 0$, then $s(t)$ will decrease as $i(t)$ increases to a peak, followed by a decrease as would

be seen in epidemic dynamics. However, the susceptible fraction grows again after the infective fraction reaches a low level. New births increase the susceptible fraction to a level which induces another smaller epidemic. This cycle continues and results in a spiral path toward the equilibrium point. At this endemic point the replacement number $\sigma s_e = 1$, this is the actual proportion of secondary cases from a typical infective individual, and disease is neither increasing nor decreasing.

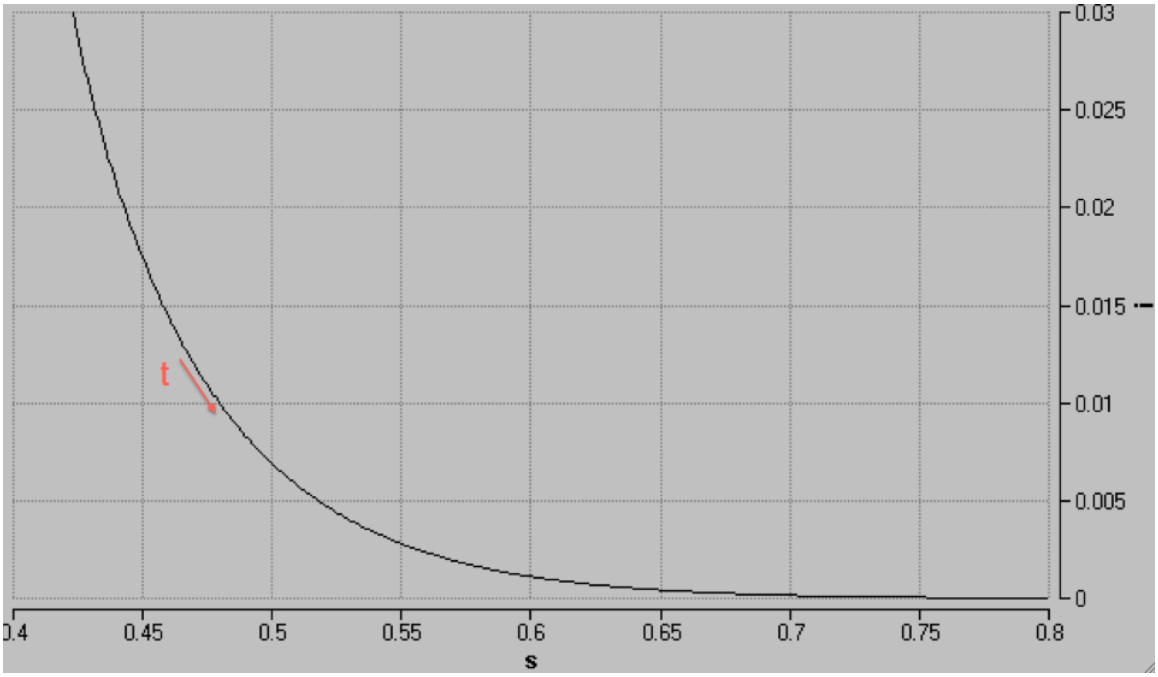


Figure 2.5: Phase portrait for the case when $i > 0$ and $\sigma < 1$

Figure 2.5 shows the disease free case. Although i is initially greater than zero, the contact number, σ, is less than one and therefore is not large enough to maintain the infection in the population, and the prevalence of infection, i declines to zero.

Figure 2.6 is the endemic equilibrium case. The infective proportion of the population is greater than zero, and the contact number, σ, is greater than one, which means that the rate at which individuals acquire infection, given by β, is greater than the combined removal

rate of $\gamma + \mu$ infection and the infection maintains a constant presence in the population. The prevalence of infection, i, spirals towards the endemic equilibrium with time.

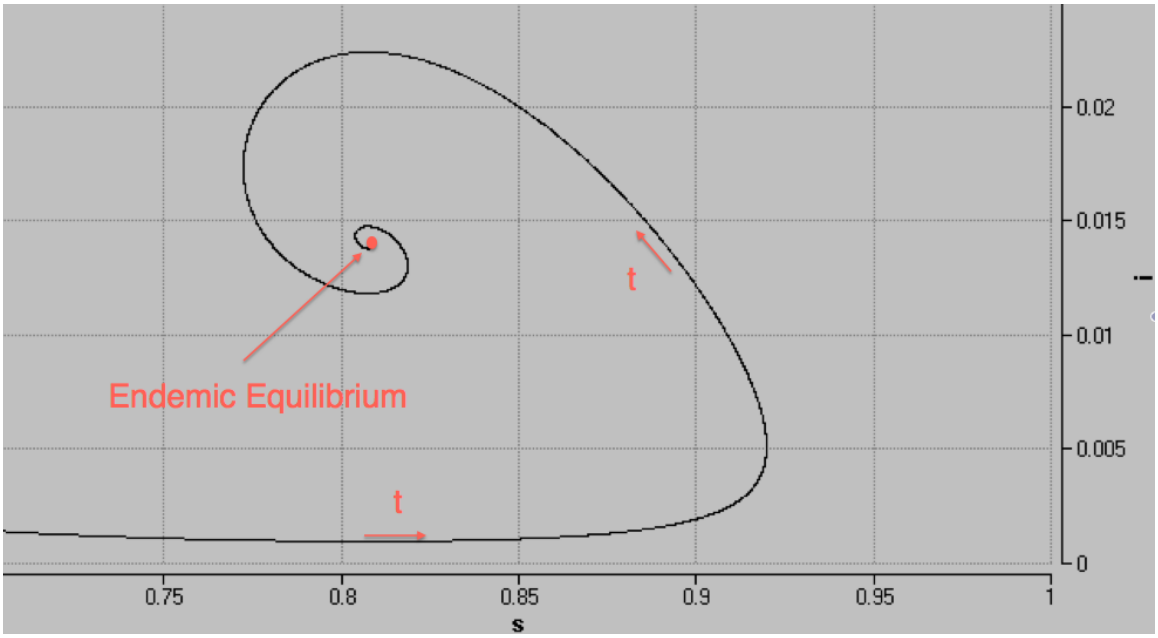


Figure 2.6: Phase portrait for the case when $i > 0$ and $\sigma > 1$

2.6.1 Limitations of the Endemic Model

This model assumes that the population size, N, remains constant, that is, births = deaths. This is perhaps an oversimplification of a typical population for endemic disease. The assumption of homogeneous mixing is not an accurate reflection for all endemic diseases in all populations, for example, in sexually transmitted infection analysis we cannot assume homogeneous sexual mixing patterns across all age groups. Contact rates may also differ across different geographical regions and among socioeconomic groups (Layte *et al.* 2006, p 174).

2.7 Conclusion

This chapter explored classical mathematical models and various model types used to describe HPV dynamics. Following this review of the literature, the deterministic ODE model for endemic disease was chosen as the model structure for this thesis. The model was then presented and will be built upon in subsequent chapters. This model takes births and deaths into account through the parameter μ, and also contains the parameter β representing the probability of transmission. Infected individuals recover at the rate γ. Preliminary assumptions for these model parameters will be made in chapter 3 and developed in subsequent chapters.

As outlined in the study objectives at the beginning of this thesis, chapter 2 studies classical mathematical modelling theory. The following chapter describes the structure of a simple transmission dynamic model for HPV. The model is initially outlined and defined. The system of equations is then solved analytically and a stability analysis is carried out on the equilibrium points of the model.

Chapter 3

Simple Homogeneous Model: Analytical Solutions and Stability Analysis

3.1 Introduction

This chapter introduces a simple unstructured model to describe HPV dynamics. This model is over-simplified and does not accurately represent a real population, but acts as a good starting point to explore the prominence of each parameter contributing to disease dynamics. The model presented in sections 3.2 - 3.4 is a simple SIR model for endemic disease with vaccination. An analytical solution for the system and stability analysis is provided in section 3.5 to determine whether small changes in Ordinary Differential Equation (ODE) conditions (initial conditions and parameter values) lead to changes in the solution.

3.2 SIR Model with Vaccination

The basic model illustrated in Figure 3.1 represents a deterministic model for an endemic disease such as HPV. The model consists of three distinct compartments which sum to give the total population. These compartments represent groups of the population with specific characteristics.
These are:

$s(t)$: The proportion of the population who are susceptible to infection at time t. For this HPV model, individuals entering the sexually active population are assumed to be susceptible unless they have become immune through vaccination.

$i(t)$: The proportion of the population who are infected with HPV at time t and can spread infection to others.

$r(t)$: The proportion of the population who are recovered or immune from HPV at time t. Immunity can be achieved in one of two ways; by moving to the immune class following vaccination, or from the infective class to the recovered class upon recovery from infection.

The model has five key parameters. They are defined as follows:

μ : The birth/removal rate. To keep the population size stationary, births are assumed to equal deaths in the population. Individuals enter the susceptible class at a rate μ. Individuals are removed from the population through death or by cessation of sexual mixing. Individuals are removed at a rate proportional to the class size given by the terms μs, μi, μr.

β : The probability of transmission per partnership.

c: The average annual partner change rate. This parameter ignores many influential factors which affect HPV transmission. For example, c does not take account of different levels of sexual activity in the population.

v : The vaccination rate. The rate at which individuals entering the model are vaccinated and moved into the immune class. In the case of a HPV vaccination, a proportion of 12-13 year old females are vaccinated each year before they become susceptible. Therefore, when they enter the sexually active population, the vaccinated proportion move straight to the immune/recovered class given by the term μv in Figure 3.1, and the remainder (individuals who did not receive the vaccine) move into the susceptible class, given by the term μ in Figure 3.1.

γ : The rate of recovery. This parameter represents the rate at which infected individuals recover and achieve conferred immunity. Its reciprocal, $1/\gamma$, represents the duration of infection.

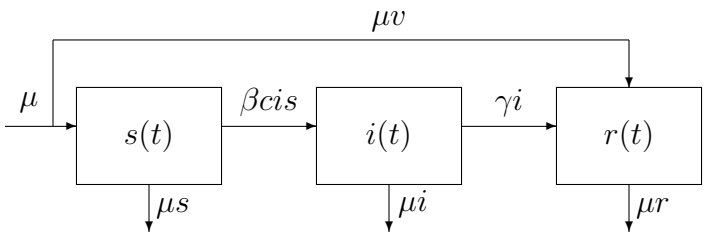


Figure 3.1: SIR model for endemic disease with vaccination

3.3 Model Equations

The following set of Ordinary Differential Equations (ODEs) (3.1) represent the compartmental model illustrated in Figure 3.1.

$$\begin{aligned}
\frac{ds}{dt} &= \mu(1-v) - \beta cis - \mu s, \\
\frac{di}{dt} &= \beta cis - \gamma i - \mu i, \\
\frac{dr}{dt} &= \gamma i + \mu v - \mu r,
\end{aligned} \tag{3.1}$$

$$s(0) = s_0 \geq 0, \qquad i(0) = i_0 \geq 0, \qquad r(0) = r_0 \geq 0$$

3.4 Assumptions of the Basic Model

The model described by the system of equations (3.1) makes the following assumptions:

- n = s + i + r, where n is the total population and s, i and r represent the proportion of the population in each class.

$$s + i + r = n = 1$$

- The population size remains constant such that

$$\frac{ds}{dt} + \frac{di}{dt} + \frac{dr}{dt} = 0$$

- This is a deterministic model, therefore it is assumed that the population size is large enough such that stochastic effects can be ignored.

- Probability of transmission is constant, that is, transmission from infected males to susceptible females equals transmission from infected females to susceptible males.

- HPV infection leads to conferred immunity

- The vaccine is a perfect vaccine which means that susceptibles who are vaccinated have life-long immunity. The assumption of a perfect vaccine is a simplification of the model. However studies show lasting antibody persistence and protection from HPV persistent infection 9 years post-vaccination (Jit *et al.* 2011). Vaccinating infected individuals is assumed to have no effect.

- This is a homogeneously mixing population and all members are assumed to be equally susceptible to contracting a HPV infection.

3.5 Solving the Basic Model

To determine the stability of the system of equations (3.1), we must first evaluate the equilibria or steady state points of the system of ODEs. An equilibrium or equilibrium point of a system of ODEs is a solution that does not change with time. The system of equations (3.1) has two equilibria. These points are the disease-free equilibrium where $i = 0$ and the endemic equilibrium where $i \neq 0$.

We set $\frac{ds}{dt}$ and $\frac{di}{dt}$ equal to zero and solve for s and i. An expression for r can be calculated from s and i since we know that $s + i + r = 1$.

Setting $\frac{di}{dt} = 0$ we obtain

$$\begin{aligned} \beta cis - \gamma i - \mu i = 0, \\ i(\beta cs - \gamma - \mu) = 0 \end{aligned} \tag{3.2}$$

There are two possible solutions:

$$i = 0 \qquad \text{or} \qquad s = \frac{\gamma+\mu}{\beta c} = \frac{1}{R_0}$$

where R_0 is the Basic Reproductive Number, which is defined as the expected number of secondary infections caused by a single infectious individual in an entirely susceptible population.

- for $i = 0$,
  from $\frac{ds}{dt}$ in equation (3.1)

$$\begin{aligned} \mu(1-v) - \beta c(0)s - \mu s = 0, & \\ s &= \frac{\mu(1-v)}{\mu} \\ s &= 1 - v \end{aligned}$$

  This solution gives the Disease-Free Equilibrium:

$$(s, i) = (1 - v, \ \ 0) \tag{3.3}$$

- for $s = \frac{\gamma+\mu}{\beta c} = \frac{1}{R_0}$,
  from $\frac{ds}{dt}$ in equation (3.1)

$$\begin{aligned} \mu(1-v) - \beta ci(\frac{1}{R_0}) - \mu(\frac{1}{R_0}) = 0, & \\ -\beta ci(\frac{1}{R_0}) &= \mu(1-v) - \mu(\frac{1}{R_0}), \\ i &= R_0\frac{\mu(1-v)}{\beta c} - \frac{\mu}{\beta c}, \\ &= \frac{\mu(1-v)}{\gamma+\mu} - \frac{\mu}{\beta c}, \\ &= \frac{\mu}{\gamma+\mu} - \frac{\mu v}{\gamma+\mu} - \frac{\mu}{\beta c}, \end{aligned}$$

This solution gives the following Endemic Equilibrium:

$$(s, i) = (\frac{\gamma + \mu}{\beta c} \quad , \quad \frac{\mu}{\gamma + \mu} - \frac{\mu v}{\gamma + \mu} - \frac{\mu}{\beta c}) \tag{3.4}$$

3.5.1 Stability Analysis of the Disease-Free Equilibrium

Having found a solution to the model, a stability analysis can be carried out to test whether small changes in the defined ODE conditions such as parameter values or initial conditions have a large effect on the model solution. In general, if small changes to ODE conditions have only a small effect on the model solution, we conclude that the solution is stable. To carry out a stability analysis of the disease-free equilibrium, the system of ODEs (3.1) must be linearised at $(s, i) = (1 - v \ , \ 0)$ by forming the Jacobian matrix. The stability of typical equilibria of ODEs is determined by the sign of the real part of the eigenvalues of the Jacobian matrix. These eigenvalues are often referred to as the eigenvalues of the equilibrium. In order to determine stability we must re-define the system of ODEs (3.1) as

$$\begin{aligned} f(s, i) &= \mu(1 - v) - \beta cis - \mu s \\ g(s, i) &= \beta cis - \gamma i - \mu i \end{aligned} \tag{3.5}$$

The Jacobian of f and g with respect to s and i is given by the following first-order partial derivatives:

$$J = \begin{pmatrix} \frac{\partial f}{\partial s} & \frac{\partial f}{\partial i} \\ \frac{\partial g}{\partial s} & \frac{\partial g}{\partial i} \end{pmatrix}$$

$$J = \begin{pmatrix} -\beta ci - \mu & -\beta cs \\ \beta ci & \beta cs - \gamma - \mu \end{pmatrix}$$

J_0 at the disease free equilibrium $(s, i) = (1 - v\ ,\ 0)$ is

$$J_0 = \begin{pmatrix} -\mu & -\beta c(1-v) \\ 0 & \beta c(1-v) - \gamma - \mu \end{pmatrix}$$

This is an upper triangular matrix, which means that all entries below the main diagonal are zero. The eigenvalues of an upper triangular matrix are the diagonal entries. Therefore the eigenvalues are:

$$\lambda_1 = -\mu \quad , \quad \lambda_2 = \beta c(1-v) - \gamma - \mu.$$

λ_2 has the largest magnitude, we call this the dominant eigenvalue.
An equilibrium is asymptotically stable if all eigenvalues have negative real parts (Braun, 1975).

From the Jacobian, we can see that $\lambda_1 < 0$,

The second eigenvalue, λ_2, is less than zero under the following conditions:

$\lambda_2 < 0$ when

$$\begin{aligned} \beta c(1-v) - \gamma - \mu &< 0 \\ \beta c(1-v) &< \gamma + \mu \\ 1 - v &< \frac{\gamma + \mu}{\beta c} \\ -v &< -1 + \frac{1}{R_0} \\ v &> 1 - \frac{1}{R_0} \end{aligned}$$

or equivalently $\lambda_2 < 0$ when

$$\begin{aligned} \beta cs - \gamma - \mu &< 0 \\ \beta cs &< \gamma + \mu \\ s &< \frac{\gamma + \mu}{\beta c} \\ s &< \frac{1}{R_0} \end{aligned}$$

where $s = 1 - v$

These results can be interpreted as saying that the disease free equilibrium is stable when the vaccination rate v is greater than $1 - \frac{1}{R_0}$. Disease cannot invade the population when $s < \frac{1}{R_0}$.

3.5.2 Stability Analysis of the Endemic Equilibrium

To carry out a stability analysis of the endemic equilibrium, the Jacobian matrix at the endemic equilibrium point must be constructed. As in section 3.5.1, we re-define the system

of ODEs from equation (3.1) as

$$\begin{aligned} f(s,i) &= \mu(1-v) - \beta cis - \mu s \\ g(s,i) &= \beta cis - \gamma i - \mu i \end{aligned} \tag{3.5}$$

The Jacobian of f and g with respect to s and i is given by the following first-order partial derivatives:

$$J = \begin{pmatrix} \frac{\partial f}{\partial s} & \frac{\partial f}{\partial i} \\ \frac{\partial g}{\partial s} & \frac{\partial g}{\partial i} \end{pmatrix}$$

$$J = \begin{pmatrix} -\beta ci - \mu & -\beta cs \\ \beta ci & \beta cs - \gamma - \mu \end{pmatrix}$$

J_E at the endemic equilibrium $(s,i) = (\frac{\gamma+\mu}{\beta c} \quad , \quad \frac{\mu}{\gamma+\mu} - \frac{\mu v}{\gamma+\mu} - \frac{\mu}{\beta c})$ is

$$J_E = \begin{pmatrix} -\beta c(\frac{\mu}{\gamma+\mu} - \frac{\mu v}{\gamma+\mu} - \frac{\mu}{\beta c}) - \mu & -\beta c(\frac{\gamma+\mu}{\beta c}) \\ \beta c(\frac{\mu}{\gamma+\mu} - \frac{\mu v}{\gamma+\mu} - \frac{\mu}{\beta c}) & \beta c(\frac{\gamma+\mu}{\beta c}) - \gamma - \mu \end{pmatrix}$$

$$= \begin{pmatrix} \beta c\mu(\frac{v}{\gamma+\mu} - \frac{1}{\gamma+\mu}) & -\gamma - \mu \\ \mu(\frac{-\beta cv+\beta c}{\gamma+\mu} - 1) & 0 \end{pmatrix}$$

The eigenvalues of the Jacobian matrix can be found by solving for λ in the following equation, known as the characteristic equation for J_E.

$$(\beta c\mu(\frac{v}{\gamma+\mu} - \frac{1}{\gamma+\mu}) - \lambda)(0-\lambda) - (-\gamma-\mu)(\mu(\frac{-\beta cv+\beta c}{\gamma+\mu} - 1)) = 0$$

$$\lambda^2 - \frac{\beta c\mu(v-1)}{\gamma+\mu}\lambda + \frac{(1-v)(\gamma\beta c\mu+\beta c\mu^2)}{\gamma+\mu} - \mu(\gamma+\mu) = 0$$

The solution of the characteristic equation was found using the standard formula for solving quadratic equations. The complexity of the solution does not allow any simple conclusion to be drawn on the qualitative significance of the stability of the endemic equilibrium.

We have

$$\lambda_{1,2} = \frac{\beta c\mu(v-1)}{\gamma+\mu} \pm \sqrt{\frac{(\frac{\beta c\mu(v-1)}{\gamma+\mu})^2 - 4(\frac{(1-v)(\gamma\beta c\mu+\beta c\mu^2)}{\gamma+\mu} - \mu(\gamma+\mu))}{2}}$$

3.6 Natural History Model with Heterogeneity

The model presented in equation (3.1) is a general, unstratified SIR model for endemic disease with vaccination. This model makes the assumption that the entire population

mixes homogeneously. This simplification permitted the detailed exploration of the dynamics of the equilibria of the system, which is applicable to a more complex, stratified model, but is easier to interpret through an unstructured model such as the one presented in equation (3.1).

In order to assess the impact that a vaccination programme will have on the endemic prevalence of a population, it is appropriate to initially analyse a natural history model. In chapters 4 and 5, the model presented in equation (3.1) will be developed to include host heterogeneity. The model will initially introduce two genders, and will be further developed to include heterogeneity in sexual mixing. In these next two chapters, the vaccination parameter included in equation (3.1) will be set a value of zero, representing the case where nobody in the population is vaccinated. This scenario is the natural history of infection.

3.7 Conclusion

This chapter detailed the analytical analysis of a simple unstructured model for a general endemic disease such as HPV. The model parameters were introduced and defined along with the basic assumptions of the model. The birth/removal rate μ and recovery rate γ initially introduced in chapter 2 were re-introduced in this chapter along with a newly defined transmission parameter β, which is accompanied by the parameter c the partner change rate. The vaccination parameter v is also introduced here. These parameter definitions carry through to the more complex models presented in subsequent chapters. The model presented here assumes a homogeneously mixing population and does not differentiate between genders. This limiting assumption will be explored in chapters 4 and 5. This chapter takes an initial step to achieving the first objective outlined in the thesis summary. A suitable ODE model for HPV in Ireland will continue to be developed in subsequent chapters.

Given that the model in this chapter was unstructured, it was possible to solve the system of equations analytically. Solutions for the disease free and endemic equilibria were found, along with the Jacobian matrices and eigenvalues.

In Chapter 4, model (3.1) is developed in two stages and analysed numerically to simulate the natural history endemic prevalence of infection. A vaccination parameter will be re-introduced into the model in chapter 6.

Chapter 4

Analysis of a Simple Homogeneous Model and the Introduction of Heterogeneity and Model Calibration

4.1 Introduction

This chapter provides a numerical analysis of the simple model presented in chapter 3 using Irish specific data. Analytical solutions of a simple SIR model for HPV were explored in section 3.5. As previously stated, analytical solutions of ODE models are generally very complicated and provide little insight into disease dynamics and population behaviour. These trends are explored through numerical analyses of the model.

This chapter details the analysis of two models. The first model (4.6) introduces the numerical analysis of the simple homogeneous model presented in chapter 3 equation (3.1) and demonstrates how the simplicity of this model leads to an error in finding a numerical estimate for β. The second model, introduced in section 4.3, introduces heterogeneity in sexual behaviour, and following on from the lesson learned

in model (4.6) concerning the calibration of β, a sensitivity analysis is carried out on the model parameters related to R_0. This analysis was carried out to overcome validity issues surrounding the parameter values, and to discover which parameters have the greatest effect on the simulated population prevalence, and hence, which parameters require the greatest level of precision in estimation.

The following section introduces the model equations and key assumptions. The data sources for each model parameter and the rationale behind the assumptions relating to the selection of model inputs are explored.

4.2 Simple Homogeneous Model

The following sections describes the model equations and assumptions, the definitions and initial estimates for the epidemiological parameters and the process of calibrating the model around the parameter β.

4.2.1 General Model Equations

The general model is given by,

$$\begin{aligned}
\frac{ds_k}{dt} &= \mu n_k - \beta_k c_k s_k i_{k'} - \mu s_k, \\
\frac{di_k}{dt} &= \beta_k c_k s_k i_{k'} - \gamma i_k - \mu i_k, \\
\frac{dr_k}{dt} &= \gamma i_k - \mu r_k,
\end{aligned} \tag{4.1}$$

where the subscript k represents gender, $k = f$ or m representing females and males respectively. The subscript k' represents the opposite gender, for example when $k = f$, $k' = m$.

n_k is the proportion of the population in group k and $n_k + n_{k'} = n = 1$.
s_k is the proportion of the population who are susceptible and in group k.
i_k is the proportion of the population who are infectious and in group k.
r_k is the proportion of the population who are immune/recovered and in group k.

The total population can be given by n where:

$$\begin{aligned}
s_k + i_k + r_k &= n_k \\
s_{k'} + i_{k'} + r_{k'} &= n_{k'} \\
n_k + n_{k'} &= 1 = n
\end{aligned} \tag{4.2}$$

The model parameters are defined as follows:

μ: The birth/removal rate. This parameter consists of two components, the natural death rate in the population and the rate at which individuals leave the sexually active population. μ is assumed to be constant in the population and universal for all individuals. The birth rate is set equal to the removal rate to maintain a constant

population size. This is a standard assumption in deterministic mathematical modelling. This parameter μ is not gender specific.

γ: The recovery rate from HPV infection. The average rate of recovery is calculated as 1/duration of infection, where time is measured in years. For simplicity, the duration of infection is assumed to be equal for both genders.

β: The probability of transmission of HPV 16 and/or 18 per partnership. The term β_f represents the probability of transmission per partnership from males to females, while β_m is the probability of transmission per partnership from females to males.

c: The average number of sexual partners per year.

4.2.2 Full System of Equations

The equations for the female population are:

$$\begin{aligned} \frac{ds_f}{dt} &= \mu n_f - \beta_f c_f s_f i_m - \mu s_f, \\ \frac{di_f}{dt} &= \beta_f c_f s_f i_m - \gamma i_f - \mu i_f, \\ \frac{dr_f}{dt} &= \gamma i_f - \mu r_f, \end{aligned} \tag{4.3}$$

The equations for the male population are:

$$\begin{aligned}
\frac{ds_m}{dt} &= \mu n_m - \beta_m c_m s_m i_f - \mu s_m, \\
\frac{di_m}{dt} &= \beta_m c_m s_m i_f - \gamma i_m - \mu i_m, \\
\frac{dr_m}{dt} &= \gamma i_m - \mu r_m,
\end{aligned} \tag{4.4}$$

The total population can be given by,

$$\begin{aligned}
s_f + i_f + r_f &= n_f \\
s_m + i_m + r_m &= n_m \\
n_f + n_m &= 1
\end{aligned} \tag{4.5}$$

The model simulates an open, constant population. That is, individuals are allowed to enter and leave the model through birth, death or recovery/immunity from infection. Births are assumed to be equal to the death/removal rate such that the population remains constant.

This is a simple model that does not take into account an individual's age or sexual activity class and the impact these sub-divisions have on parameter values in the model. Although this is a simple model, many insights can be gained from studying its dynamics, which will prove useful in the more complex model presented in chapter 5.

In the following section, some simplifying assumptions are applied to model 4.1 to facilitate numerical analysis. Based on these assumptions, no distinction is made between sexes in terms of sexual dynamics, infection transmission, or population distribution. This is a limitation of the model, but will facilitate closer investigation

of infection dynamics in the population. Gender differences will be re-introduced in chapter 5. This renders the subscript k unnecessary and results in a new set of equations (4.6).

Specifically, the following simplifying assumptions are applied to model (4.1):

- All individuals in the population are assumed equal in terms of sexual behaviour, that is, we assume that the partner change rate is equal for males and females, $c_k = c_{k'}$, or $c_f = c_m$. These parameters are replaced by the global parameter c.
- Virulence of HPV is assumed to be independent of gender allowing for the relation $\beta_f = \beta_m$. The transmission parameter for HPV is simply β.
- In terms of population distribution, the proportion of males is set equal to the proportion of females, that is $n_f = n_m$, which is replaced by n.
  Similarly, $s_f = s_m$, $i_f = i_m$ and $r_f = r_m$. These variables are replaced by s, i and r, respectively.

These assumptions reduce the system from two groups and six equations (equations (4.4) and (4.5)), to one group with just three equations given by:

$$\begin{aligned} \frac{ds}{dt} &= \mu - \beta csi - \mu s, \\ \frac{di}{dt} &= \beta csi - \gamma i - \mu i, \\ \frac{dr}{dt} &= \gamma i - \mu r, \end{aligned} \tag{4.6}$$

The total population for the model can now be given by equation (4.7) and compared to equation (4.5).

$$s + i + r = n = 1 \tag{4.7}$$

4.2.3 Epidemiological Parameters

Prevalence rates for HPV are from the ARTISTIC study (Kitchener *et al.* 2006), a randomised trial of HPV testing carried out in the UK and published in 2009. Although Irish specific data would have been preferable for the model, it was decided that the sample size used in Ireland's HPV prevalence study published by Keegan *et al.* (2007) was too small and not statistically reliable, as noted by Usher *et al.* (2008). It is worth noting however, that the results from the Keegan *et al.* study were quite similar to the ARTISTIC trial, providing some confidence in the assumption that the Irish and UK populations are comparable in terms of prevalence.

The epidemiological parameters inputted to the model are as follows:

μ: The removal rate/birth rate. Individuals enter the model at age 18 and are removed at age 64, which coincides with the study age range of Layte *et al.* (2006). This is a limitation of the model since it assumes that all individuals are sexually inactive until age 18. According to the ISSHR (Layte *et al.* 2006) 21% of males and 12% of females first experienced vaginal sex before the age of 17. However there is currently no detailed information on the average annual partner change rates for individuals who are sexually active before age 18. Also, the age range of the study populations for the two HPV prevalence studies comparable to the Irish population does not include those under the age of 18, hence we do not know the prevalence of infection in this age group, or have detailed information of their sexual behaviour. Therefore, rather than make an unfounded assumption about the average annual

partner change rate for the fraction of individuals under the age of 18 who are sexually active, the model presented here, and those in subsequent chapters assume that all individuals are sexually inactive until age 18, and have at least one partner per year from the age of 18. The study by Layte *et al.* (2006) provided the data for the parameter c and also for the proportions of the population in each risk group. Individuals remain in the sexually active population for 40 years (Elbasha, Galvani 2005). So, $\mu = \frac{1}{46} + \frac{1}{40} = 0.047$. The birth rate is set equal to the removal rate to maintain a constant population size.

As mentioned in chapter 1, there are over 120 strains of HPV, and about 15 are known to cause cervical cancer. Of these 15 strains, just two (HPV 16 and 18) are the causative agents of approximately 70% of cervical cancers. To limit complexity, only these two strains are included in the model and are treated as one infection. Individuals can be infected with either, or both strains. The prevalence rate published by the ARTISTIC trial for HPV 16 and/or 18 in the female population is 4.4% (Kitchener *et al.* 2006). This study estimated prevalence of infection in females. Large scale studies on HPV prevalence such as the ARTISTIC trial are generally confined to the female population. For this reason, and to limit model complexity, the prevalence of infection in males is assumed to equal 4.4%.

γ: The recovery rate from HPV infection. The prevalence of HPV 16 is approximately two and a half times that of HPV 18 (Kitchener *et al.* 2006). Since this model limits complexity by not differentiating between these two strains, the published rate of recovery for the more prevalent HPV 16 was used in the model to represent recovery from HPV 16 and/or 18. The average duration of HPV 16 infection is 20 months (Insinga *et al.*, 2010). The average rate of recovery is calculated as 1/duration of infection, therefore $\gamma = 1/1.667 = 0.599$ given that time is measured in years.

c: The average number of sexual partners per year. International comparisons show that Irish people have fewer partners on average, over all age groups than people

in other countries (Layte *et al.* 2006). The average annual partner change rate is approximately 1.2 (Layte *et al.* 2006).

4.2.4 Threshold Estimates and Calibration of β

An analytical solution for the endemic equilibrium for the model presented in equation (3.1) was found in section 3.5.

We had,

$$(s_e, i_e) = (\tfrac{\gamma+\mu}{\beta c} \quad , \quad (\tfrac{\mu}{\gamma+\mu} - \tfrac{\mu v}{\gamma+\mu} - \tfrac{\mu}{\beta c}))$$

Following Hethcote's seminal work (Hethcote 2000), the *contact number*, σ, in an infectious disease model is defined as the average number of adequate contacts of a typical infective individual during the infectious period, as previously defined in section 2.5. In an endemic HPV model, $\sigma = R_0 = \frac{\beta c}{\gamma+\mu}$ throughout time since there is no change in the infectivity of the virus after the initial invasion to the population, as previously discussed in section 2.5. The replacement number, R, is the actual number of secondary cases from a typical infective. At the endemic equilibrium, $R = \sigma s_e = (\frac{\beta c}{\gamma+\mu})(\frac{\gamma+\mu}{\beta c}) = 1$. This is an intuitive result since, if R were greater than 1, infection would increase, if $R = 1$, the population remains at a stable equilibrium.

The parameter R_0 is the threshold value for the model. If $R_0 < 1$, the model approaches the disease free equilibrium, as shown in Figure 2.5. If $R_0 > 1$, the model approaches the endemic equilibrium, as shown in Figure 2.6. Since HPV is a sexually transmitted virus, transmission is based on β and c, as discussed in section 2.5.

Model (4.6) was initially solved in Berkeley Madonna (Macey *et al.* 2000) using the parameter values defined above. Appendix A outlines the method used by Berkeley Madonna to solve the system of equations. All parameter values were set as described above. In simulating the model it was found that varying the initial conditions for s, i and r had no effect on the model solution. All simulations where i was not equal to zero resulted in the solution reaching the same endemic equilibrium. Therefore, it was unnecessary to define an initial condition. The model was calibrated around the unknown transmission parameter β. The process of model calibration involved simulating the model for a range of 30 values for the input parameter β, while all other parameters were held constant. A parameter plot of these simulations was produced which showed the output result, the prevalence of infection, plotted against its associated estimate of β. The aim of this process was to find a value for β that would produce the known prevalence of infection of 4.4% when all other parameters were held at the values described above. Figure 4.1 depicts this process and was produced using the Parameter Plot feature in Berkeley Madonna, which plots the final value of the selected variable, β as a function of the prevalence, i. The plot shows how the prevalence of infection, i, on the y-axis increases as β increases.

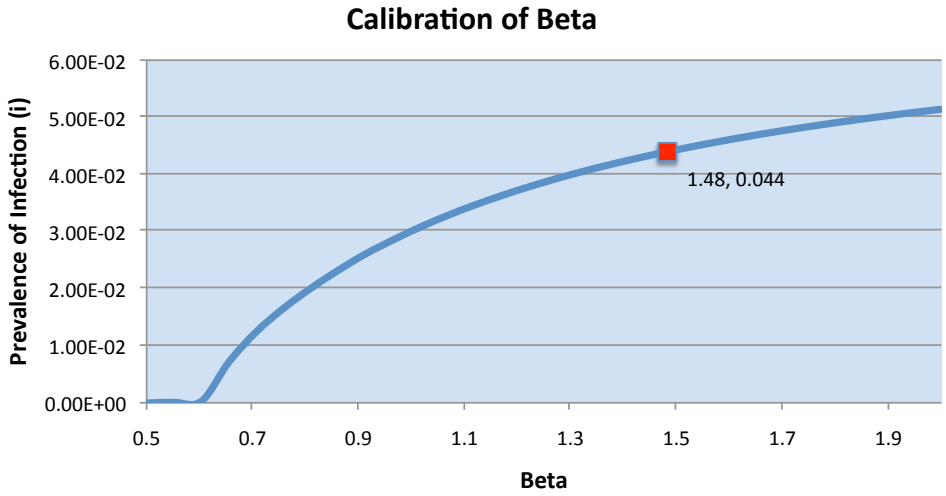


Figure 4.1: Beta parameter plot

The value for β which matches the prevalence rate of 4.4% was approximately 1.5. This value is considerably higher than those published in the literature, which set β at a value of 0.7 - 0.8 (Hughes *et al.* 2002). This value of $\beta = 1.5$ is also illogical since it represents the probability of transmission of infection per partnership, and the maximum possible value for a probability is 1. The effect we are seeing here is due to the fact that the prevalence of infection is determined by the previously defined Basic Reproductive Number for the infection, R_0, given by $\frac{\beta c}{\gamma+\mu}$. The model is forced to produce a known prevalence rate for the population and R_0 is the key parameter in producing this result. The unusually high value for β produced by the model calibration is brought about as a result of the necessity to balance R_0 with the infection prevalence. All parameters making up R_0 were held constant in the calibration, except for β, so this variable β contains the short-fall of the other parameters in the equation, namely c, μ and γ in keeping R_0 at a value which produces the known prevalence of infection. So, reducing β to a logical value less than 1, for example 0.8, would require the other parameters to change, either c must in turn increase, or the denominator of R_0 must decrease by decreasing either parameters γ or μ, or an equivalent change in a combination of two or all three of these parameters that produces a value for R_0 that results in the known prevalence of infection.

It is also likely that this model is too simple to accurately estimate the epidemiological parameters for this infection, and so a more complex, risk-structured model is more appropriate. This idea will be explored in the following section, 4.3.

The value of R_0 may change as the parameter values are altered to produce the prevalence value, since the parameters are multiplied by prevalence and the proportion of susceptibles, given by s and i within the model equations, thus increasing a parameter within the equation for R_0 results in a change in the prevalence of infection, i, and this change is not proportional for the parameters making up R_0. Further investigation of the possible values for R_0 and its associated parameters is disregarded for this simple model, further analysis is carried out in subsequent sections on more complex models.

The value for R_0 is typically higher in an SIR model than it would be in an SIS

(Susceptible - Infectious - Susceptible) model for the same infection dynamics. This is due to the fact that some of the contacts made by an infectious individual will be essentially "wasted" in terms of the spread of infection, on individuals who are already infectious or those who are immune (Garnett *et al.* 2006). This reduces R_0 to an effective reproductive rate, which only considers contacts with susceptible individuals. The effective reproductive rate is given by $R_t = R_0 x$, where x is the proportion of the population who are susceptible (Garnett *et al.* 2006). Thus, R_0 will be higher in an SIR model than if the same infection was described by an SIS model, to account for the proportion of contacts "wasted" on immune individuals. This situation arises in constructing a model for HPV since there is uncertainty as to whether infection with HPV leads to conferred immunity or not.

Increased variance in risk between groups within a model population can also allow for a higher value of R_0 (Garnett *et al.* 2006) since, as will be shown in section 4.3.5 the basic reproductive number in a heterogeneous population model combines the basic reproductive numbers for each group. The following section develops model (4.6) by introducing two risk groups based on annual partner change rates.

4.3 Risk Heterogeneity in Sexual Behaviour

Results from the ISSHR report (Layte *et al.* 2006) show that over a period of time, most people have relatively low numbers of sexual partners, but a small fraction have much higher numbers of partners. This heterogeneity of sexual behaviour has been well documented in the literature, most notably by Hethcote and Yorke (1984) who introduced the concept of a small "core group" of individuals in the sexually active population. This core group have a higher average annual partner change rate than the population average. This variation in sexual behaviour within the population

may have a profound effect on STI spread and control as demonstrated by Hethcote and Yorke (1984) in their model for the spread of gonorrhea. Though the core group only represents a small proportion of the population, their high partner change rate can contribute a disproportionate amount to the spread of infection when compared to the non-core group who have fewer partners per unit time.

Model (4.6) was adapted to demonstrate heterogeneous sexual behaviour, giving a new system of equations (4.8).
The population is split into two groups based on the frequency at which individuals change their partners. Mixing between groups is assumed to be completely proportionate for now (partners are chosen with a probability that is proportional to the number of partnerships that they generate). Low risk individuals can infect/become infected by other low risk individuals or by high risk individuals. A high partner change rate is associated with high risk of infection. The adapted model equations are given by:

$$\begin{aligned} \frac{ds_j}{dt} &= \mu n_j - \lambda_j s_j - \mu s_j, \\ \frac{di_j}{dt} &= \lambda_j s_j - \gamma i_j - \mu i_j, \\ \frac{dr_j}{dt} &= \gamma i_j - \mu r_j, \end{aligned} \tag{4.8}$$

where j can be L or H representing Low and High partner change rates respectively.

n_j is the proportion of the population in group j and $\sum n_j = 1$.
s_j is the proportion of the population who are susceptible and in group j.
i_j is the proportion of the population who are infectious and in group j.
r_j is the proportion of the population who are immune and in group j.

The full set of equations for the above system of equations (4.8) is as follows:

Low Risk Group:

$$\begin{aligned}
\frac{ds_L}{dt} &= \mu n_L - \lambda_L s_L - \mu s_L, \\
\frac{di_L}{dt} &= \lambda_L s_L - \gamma i_L - \mu i_L, \\
\frac{dr_L}{dt} &= \gamma i_L - \mu r_L,
\end{aligned} \tag{4.8a}$$

$$\lambda_L = c_L \, \beta \, p$$

$$p = g_L \times i_L + g_H \times i_H$$

High Risk Group:

$$\begin{aligned}
\frac{ds_H}{dt} &= \mu n_H - \lambda_H s_H - \mu s_H, \\
\frac{di_H}{dt} &= \lambda_H s_H - \gamma i_H - \mu i_H, \\
\frac{dr_H}{dt} &= \gamma i_H - \mu r_H,
\end{aligned} \tag{4.8b}$$

$$\lambda_H = c_H \, \beta \, p$$

$$p = g_L \times i_L + g_H \times i_H$$

The force of infection is given by $\lambda_j = c_j \, \beta \, p$. The parameter p is the probability that a chosen partner is infectious and is calculated as the sum of the probabilities

that a partner chosen according to proportionate mixing is in either the low activity or high activity group (Vynnycky and White, 2010). This is given by:

$$p = g_L \times i_L + g_H \times i_H$$

Here, g_L and g_H are the probabilities that a partner chosen according to proportionate mixing is in either the low activity or high activity group respectively (Vynnycky and White, 2010) and i_L and i_H are defined as i_j above.
The probabilities g_L and g_H are given by:

$$g_L = \frac{c_L \times n_L}{c_L \times n_L + c_H \times n_H} \quad , \quad g_H = \frac{c_H \times n_H}{c_L \times n_L + c_H \times n_H} \tag{4.9}$$

Model Assumptions:

- The model simulates an open, continuous population

- As in model (4.6), the model shown in equations (4.8a and 4.8b) makes the simplifying assumption that there is no differences in sexual behaviour between genders. Therefore the parameters n_j, s_j, i_j, r_j and c_j are assumed to be equal for males and females. Hence, both genders will have the same infection dynamics and it is only necessary to model one gender.

- Mixing is assumed to be entirely proportionate.

- It is assumed that all individuals have at least one sexual partner in their lifetime. This assumption is a simplification of the model. However, data published in

the ISSHR (Layte *et al.* 2006 - appendix 5, table 5.1) shows that 93.7% of males and 94.2% of females have at least one sexual partner in their lifetime. Therefore, although this is a simplifying assumption, it is close in representing the model population.

4.3.1 Parameter Input Values

Stratifying the population by risk group requires more detailed data on sexual behaviour. Table 4.1 contains data published in the ISSHR (Layte *et al.* 2006) which was used to calculate parameter values for the distribution of partnerships formed in the model population.

Partners/year	% Males	% Females	%Total
0	15.4	19.7	17.5
1	70.3	74.9	72.6
2	6.2	3.3	4.8
3-4	5.1	1.5	3.3
5-9	2.2	0.5	1.4
10+	0.8	0.1	0.4
Mean	1.3	0.9	1.1

Table 4.1: Number of Partners per year, by gender. Data taken from Appendix Table 5.3, page 330 in the ISSHR Report (Layte *et al.* 2006).

Data in Table 4.1 shows the gender specific distribution of the number of partners in one year taken from the ISSHR report (Layte *et al.* 2006).

Based on the data displayed in Table 4.1, the population was separated as follows:

Low Activity: The low activity group consists of those individuals having the population mean number of partners in one year, or less. The percentage of the population in this group is the cumulative proportions of those with zero or one partners in the last year. We have, $17.5 + 72.6 = 90.1\%$. The average annual partner change rate for this low activity group is taken as the population mean number of partners, which gives $c = 1.1$.

This calculation is based on the final assumption defined in section 4.3, that all individuals have at least one partner in their lifetime. This means that although 17.5% of the study population had no partners in the past year, it is assumed that they have at least one partner in their lifetime. This is a reasonable assumption given that the ISSHR study reports that 94% of participants had at least one partner in their lifetime (Layte *et al.* 2006). This figure would also be expected to increase as the younger participants get older.

High Activity: The high activity group consists of the cumulative proportions of individuals having an annual partner change rate that is greater than the population average. That is, the proportion of the population having two or more partners per annum. The proportion of the population is calculated as $4.8 + 3.3 + 1.4 + 0.4 = 9.9\%$. The average annual partner change rate for the high activity group is taken as the median number of partners represented in this group, which is six, $c = 6$.

This median value is based on the assumption that the max partner change rate per annum in the population is ten. Table 4.1 shows that the proportions of males and females having ten or more partners in one year are quite low, 0.8% and 0.1% respectively. In calculating c for the high activity group, it was assumed that any outliers in the data, that is, a small proportion of individuals having a comparatively high partner change rate to the rest of the population, had a minimal effect on the

population mean and could therefore be neglected from calculations.

Using these parameters, we can now calculate the previously defined probabilities g_L and g_H from equation (4.9).

$$g_L = \frac{1.1 \times 0.901}{1.1 \times 0.901 + 6 \times 0.099} = 0.625, \qquad g_H = \frac{6 \times 0.099}{1.1 \times 0.901 + 6 \times 0.099} = 0.375 \tag{4.10}$$

4.3.2 Effective Contact Rate, $\hat{c}$

In contrast to the homogeneous model (4.6) where the partner change rate, c, was simply the mean number of partners per unit time, this heterogeneous model incorporates a measure for the variance of c in the population. The appropriately adjusted effective partner change rate, developed by Anderson and May in 1979 (Anderson and May 1991, Vynnycky and White 2010) is given by

$$\hat{c} = m + \frac{\sigma^2}{m} \tag{4.11}$$

where m is the mean number of contacts per unit time and σ^2, the variance of sexual behaviour is

$$\sigma^2 = \frac{\sum \omega_j (m_j - \mu_c)^2}{V_j} \tag{4.12}$$

where μ_c in this case is the population average contact rate per unit time, ω_j is the proportion of the population in group j, and $V_i = 1$ for normalised weights. For normalised weights the weighted means sum to one, that is, $\sum_{j=1}^{n} \omega_j = 1$.

Estimating c as simply the mean number of partners in a group can lead to an underestimate of R_0 if variance is high since $R_0 = \frac{\beta c}{\gamma+\mu}$. It follows from the definition of the basic reproductive number R_0, that the smaller this value is, the easier it will be to eradicate the disease.

Using the above formulae, 4.11 and 4.12, the variance in sexual behaviour in the population, σ^2 was calculated as 2.39 and the appropriately adjusted effective partner change rate $\hat{c} = 1.1 + 2.39/1.1 = 3.27$

4.3.3 Estimating the Transmission Parameter, β

The process of model calibration, as described in section 4.2.4 was repeated for the heterogeneous model to obtain an estimate for β which corresponds to a population prevalence of infection of 4.4%. The parameter plot in Figure 4.2 shows that the optimum value for β is 1.17. This value is lower than the estimate obtained in section 4.2.4, and is considerably higher than the previously published estimates of 0.7 - 0.8 for this transmission parameter (Hughes *et al.* 2002). Also, as was the case in section 4.2.4, a value > 1 for the parameter β is illogical since β represents a probability, which can never be > 1. The model was simulated using this β estimate and all other parameter values previously defined in sections 4.2 and 4.3.

This value for β is invalid as previously discussed in section 4.2.4, since the maximum possible value for probability is 1. Therefore, previously defined parameter

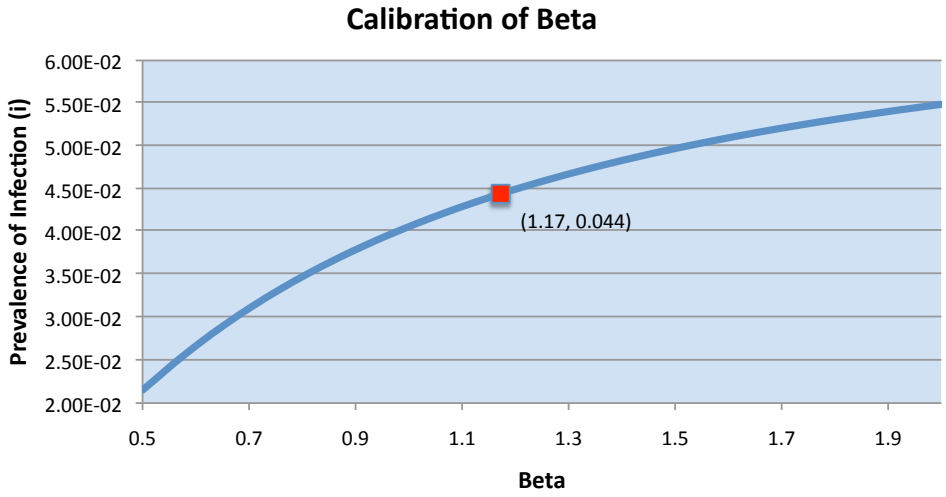


Figure 4.2: Beta Calibration

values which make up the basic reproductive number R_0 must be re-evaluated to allow for a reasonable value for β which will match the known prevalence of infection .

As discussed in section 4.2.4, the transmission parameter β is closely related to R_0, the basic reproductive rate. R_0 is sensitive to changes in the rate of transmission of infection, duration of infectiousness, the removal rate and also the partner change rate. Given that the model is calibrated around the known population prevalence of infection of 4.4%, changing any of the parameters contained in the equation for R_0 must be compensated for by the other parameters to keep the population prevalence of infection at the same value. For example increasing the variance in sexual behaviour between the risk groups such that c_H is higher will result in an increase in the effective partner change rate, $\hat{c}$ since we saw in equation (4.11) that $\hat{c}$ takes account of the variance in partner change rates among risk groups. In order to keep the population prevalence at the same level, either β must be decreased, or one or both of γ and μ must increase.

Again, as was seen in the previous model in section 4.2.4, setting the duration of infection, contact rate and removal rate as constants and calibrating the model

around the known prevalence of infection (4.4%) has led to an invalid result for the probability of transmission, $\beta = 1.17$.

4.3.4 Sensitivity Analysis

In order to decide which of the previously defined parameters should be adjusted in order to rectify the invalid result for β, a sensitivity analysis was carried out on the model. A sensitivity analysis evaluates how the variation or uncertainty in the output of a model can be attributed to variations in model inputs. This method involves systematically changing variables and evaluating the effects of these changes on the model output. The model was tested for its sensitivity to the parameters β, c_L, c_H, γ and μ. A detailed example of the method for calculating the model sensitivity to changes in the parameter μ is given as follows:

- The model was run in Berkeley Madonna with all parameters set as described above, specifically $\mu = 0.047$. The prevalence produced by this parameter set is referred to as Prev1(t).

- The parameter μ is amended slightly by adding an amount Δ which equals $0.001 \times \mu$ to test the sensitivity of the model output to a small adjustment in the variable μ.

- The model was run again using the amended value for μ, the results of this simulation are referred to as Prev2(t).

- The sensitivity Sen(t) is calculated using the equation

$$Sen(t) = \frac{Prev2(t) - Prev1(t)}{\Delta}$$

Parameter	Sen(t) (+ 0.1%)
μ	0.8400
β	0.0197
γ	-0.1043
c_H	$3.38e^{-4}$
c_L	0.0191

Table 4.2: Sensitivity Analysis. These values for Sen(t) show the change in equilibrium prevalence per year divided by 0.1% of the value of the relevant parameter.

The value Sen(t) represents the change in the equilibrium prevalence rate divided by the change in the value of the relevant parameter. Table 4.2 shows the result of the sensitivity analysis for the previously mentioned parameters.

The results in Table 4.2 show the sensitivity of the model output to a 0.1% increase in each parameter. For example, increasing μ by 0.1% increases prevalence by 0.8400. The sensitivity analysis revealed that the model is most sensitive to changes in the removal rate, μ. The model was re-simulated using appropriately adjusted values for the parameters μ, β and γ, based on the results of the sensitivity analysis. The value for β was reduced to 0.8, matching the estimate for the probability of transmission from males to females used by Hughes *et al.* (2002). The parameters μ and γ were also adjusted to make up for the reduction in β to achieve the end result of a prevalence of infection of 4.4%. As described in section 4.2.1, μ is the product of two rates, the death rate and the duration in the sexually active and mixing population, that is, the number of years for which the parameter $c > 0$. The average number of years in the sexually active and mixing population is not a well defined variable, and an estimate for this variable is not available within the sexual mixing data published by Layte *et al.* (2006) used to populate the model. It is likely that the average length of time spent in the sexually active and mixing population is less than the previous estimate of 40 years, and so it was decided to allow this rate to be reduced from 40 years

Parameter	Previous Estimate	New Estimate
μ	0.045	0.058
β	1.3	0.8
γ	0.599	0.58
c_H	6	6
c_L	1.1	1.1

Table 4.3: Adjusted Parameter Set

to 27 years. The parameter γ was adjusted by a small amount, since this estimate was not well defined given that this value represents the duration of infectiousness for two separate strains of HPV, HPV 16 and 18 as discussed in section 4.2.3. The initial estimate of 20 months for the duration of infectiousness was taken from the published duration of infection for HPV 16 infection by Insinga *et al.* (2010). This value was adjusted following the sensitivity analysis to a value of 20.6 months, which is still in line with the published value which had a 95% confidence interval of 18.4 - 22 months. The new set of parameter values are given in Table 4.3. The model was run in Berkeley Madonna using the new parameter values and successfully predicts a prevalence of 4.4%.

4.3.5 Basic Reproductive Number, R_0

Following on from the sensitivity analysis on the model and subsequent adjustment of the parameters related to the basic reproductive number $R_0 = \frac{\beta c}{\gamma+\mu}$, it is possible to calculate R_0 for this simple heterogeneous mixing model. The Basic Reproductive Number, R_0, for the system can be calculated using the following formula:

$$R_0 = g_H R_H + g_L R_L$$

where R_H is the number of secondary infections caused in a totally susceptible pop-

ulation by an infected high-activity individual and R_L is the number of secondary infections caused in a totally susceptible population by an infected low-activity individual. These parameters are given by:

$$R_H = \frac{\beta c_H}{\gamma + \mu} \quad , \quad R_L = \frac{\beta c_L}{\gamma + \mu}$$

So, using the new parameter estimates from Table 4.3 we have

$$R_H = \tfrac{0.8\times 6}{0.58+0.058} = 7.52 \text{ and } R_L = \tfrac{0.8\times 1.1}{0.58+0.058} = 1.38$$

Hence, R_0 is calculated as

$$R_0 = 0.375 \times 7.52 + 0.625 \times 1.38 = 3.68$$

The correct interpretation of this Basic Reproductive Number is, the average number of secondary infections caused by a single infective individual in an entirely susceptible population is 3.68. This is an important epidemiological parameter as it gives an indication as to how rapidly an infection can spread in a population. The effects of heterogeneity on R_0 will be explored in the following chapter.

4.4 Conclusion

This chapter explored the limitations of a simple homogeneous model for HPV which led to the development of a heterogeneous model with proportionate mixing. This model was simulated and calibrated for the parameter β, and a sensitivity analysis was carried out on the model parameters. Although the introduction of risk structure increases the validity of the model in representing a real population, the assumptions made are still over-simplified and a more complex model is required before any conclusions about the natural history of HPV infection in the Irish population can be made. This chapter introduced gender structure to model dynamics. Parameters c and n were estimated using gender specific values taken from Layte *et al.* (2006). Estimates for the prevalence of infection were taken from the ARTISTIC trial (Kitchener *et al.* 2006). This is a study carried out in the UK. Statistically reliable Irish specific data was not available, but the UK study population was determined to be comparable to the Irish sexually active population as noted by Usher *et al.* (2008). Large scale studies on HPV prevalence have so far focussed on females. Hence no estimate for male prevalence currently exists. For this reason, this model makes the limiting assumption that prevalence of HPV in males is equal to that of females. In the absence of relevant data, assuming that prevalence is equal for both genders is more appropriate than making an unfounded estimate for male prevalence. A further limiting assumption in the model is the grouping of HPV 16/18. The model considers these strains as one infection, thus an assumption must be made about the recovery rate of infection. Given that HPV 16 is much more prevalent than HPV 18, the recovery rate was set to equal the published rate for HPV 16 (Insinga *et al.* 2010). This data is difficult to collect and there is considerable uncertainty around the true value for the rate of recovery from HPV, thus, although grouping HPV 16 and 18 is a limiting factor for the model, a consistent estimate for the parameter is currently unknown.

This chapter further developed the ODE model for HPV in Ireland as outlined in the first objective in the thesis summary. Here, the next four objectives were

also addressed and will be looked at in a more complex model in the next chapter. Specifically, the significant effects of β on the spread of HPV and numerical estimates of model parameters were investigated. The effects of sexual risk groups within the population were explored through model simulations. Calibration techniques and sensitivity analyses were used to strengthen model validity and the basic reproductive number for the natural history model was calculated. These objectives will be repeated in chapter 5 for a more complex model structure. The model presented here is still relatively simple in structure. For this reason, no application of the simulation results will be made to real-world infection dynamics. These conclusions will be drawn following the analysis of the complex model in chapter 5. In the next chapter, model (4.8) is developed to include a more complex gender structure, as well as exploring various sexual mixing patterns which will strengthen the validity of the model and allow for a more detailed exploration of HPV dynamics.

Chapter 5

Analysis of a Two Sex, Risk-Structured Model with Various Patterns of Sexual Mixing and Calculation of R_0

5.1 Introduction

Chapter 4 introduced a heterogeneous model with a high and low risk group which assumed proportionate mixing. In this chapter model (4.8) is developed further to analyse the effects of various possible sexual mixing patterns in the population. The model is also stratified by gender, allowing for different values for β between males and females.

5.2 Modelling sexual mixing patterns

Generally, the transmission of an infection can be classified as being either "density dependent" or "frequency dependent" according to the relation between the number of cases of infection and population size. Density dependent transmission or "pseudo mass action" refers to a situation where an increase in population size results in a linear increase in the number of cases of infection, for example measles, influenza or many respiratory infections. The force of infection is given by the equation $\lambda(t) = \beta I(t)$, where β is the rate at which two specific individuals come into effective contact per unit time and $I(t)$ is the number of infectious individuals at time t.
However, in the case of a sexually transmitted infection, a sudden rise in population size is unlikely to increase the number of sexual contacts an individual has. Thus, transmission of infection is dependent on the frequency of contacts with infected individuals, termed "frequency dependent" transmission, or "true mass action". The force of infection is given by $\lambda(t) = c_e \frac{I(t)}{N(t)}$, where c_e is the average number of individuals effectively contacted by each person per unit time (Vynnycky and White 2010) and $\frac{I(t)}{N(t)}$ is the prevalence of infection at time t.

When modelling an STI, the transmission parameter is usually expressed as the product of two components affecting transmission of infection, c and β. The specific definition of the parameters c and β depends on whether the infection is modelled at the per-partnership or per-act level. The force of infection is written as $\lambda(t) = c\beta \frac{I(t)}{N(t)}$
Sexually transmitted infections are commonly modelled at the per-partnership level where c represents the partner change rate and β is the transmission probability per-partnership. At the per-act level, c represents coital frequency per unit time and β is the probability of transmission per sex-act.

The following section introduces various sexual mixing matrices, which are described by WAIFW (Who Acquires Infection From Whom) matrices.

5.3 WAIFW Matrices

In Chapter 4, a natural history model (4.8) with two risk groups was presented. This model assumed random or proportionate mixing between risk groups. In this chapter, the appropriateness of this sexual mixing assumption is investigated through the introduction of mixing matrices which allow various mixing patterns between the two risk groups ranging from assortative to disassortative, which are defined below. The influence that these mixing patterns have on endemic prevalence of infection will also be explored.

In contrast to equation (4.6) and (4.8) this model is stratified by gender and risk class. The term k represents gender and is either f or m representing female and male individuals respectively, i and j represent the risk classes Low and $High$ which are dependent on an individual's sexual behaviour. Patterns of sexual mixing between the strata are described by the mixing matrix, ρ_{kij}, i and j represent the "chosen" and "choosing" individuals respectively.

The mixing matrix is given by:

$$\rho_{ij} = \begin{array}{c} \\ H \\ L \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{pmatrix} \rho_{HH} & \rho_{HL} \\ \rho_{LH} & \rho_{LL} \end{pmatrix} \end{array}$$

where for example, ρ_{LH} represents the proportion of the total contacts made by high risk individuals that were with low risk individuals.

Model (4.8) assumed proportionate mixing, but this is only one possible scenario. Mixing between strata can vary between two extremes, fully assortative mixing where individuals mix purely with other individuals in the same risk class as themselves, or the opposite situation where mixing is fully disassortative and individuals only mix with individuals not in their own risk group. These two scenarios can be represented

in matrix form as:

$$\begin{array}{c} \\ Assortative \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{array}{c} H \\ L \end{array} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}$$

$$\begin{array}{c} \\ Disassortative \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{array}{c} H \\ L \end{array} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

This model will be simulated under the assumption of fully assortative mixing, fully disassortative mixing and will also be simulated under the assumption of proportionate mixing as in chapter 4. By the definition of proportionate mixing, individuals randomly choose sexual partners in proportion to the number of partnerships that the group generates.

The matrix is given by:

$$\begin{array}{c} \\ Proportionate \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{array}{c} H \\ L \end{array} \begin{pmatrix} g_{HH} & g_{HL} \\ g_{LH} & g_{LL} \end{pmatrix} \end{array}$$

where g_{ij} represents the probability that someone in group j forms a partnership with someone in group i (Vynnycky and White, 2010).

Given this assumption of proportionate mixing, the structure of the WAIFW matrix is such that the probability of an individual forming a partnership with an individual from either the low or high risk group is equal. Therefore $g_{HH} = g_{HL} = g_H$ and $g_{LL} = g_{LH} = g_L$.

Also, the sum of the probabilities in each column must equal one since a chosen partner must be either from the high or low risk group. This is written as

$$\sum_j g_{jk} = 1$$

The matrix for proportionate mixing requires the calculation of two probabilities, g_H and g_L. These probabilities are given by:

$$g_H = \frac{c_H \times n_H}{c_H \times n_H + c_L \times n_L} \quad , \quad g_L = \frac{c_L \times n_L}{c_H \times n_H + c_L \times n_L} \tag{5.0}$$

As previously stated, the proportions of the population in each risk group, n_i are gender specific. Table 5.1 shows the values for the partner change rates, c_{ki} and the proportion of the population in each gender specific risk group, n_{ki}. As in chapter 4, data was taken from the ISSHR study (Layte *et al.* 2006).

Here, $c = c_H n_H + c_L n_L$ for each gender.
Using the above information, we obtain the numerical values for the two gender specific WAIFW matrices representing proportionate mixing in the population. They are:

Parameter	Gender and Risk Class			
	Male		Female	
	H	L	H	L
n_i : the proportion of the population in each sub-group	0.143	0.857	0.054	0.946
c_i : the average annual partner change rate of each sub-group	6	1.1	6	1.1
c : the average annual partner change rate for each gender	1.8		1.4	

Table 5.1: Parameter Values

$$Male \begin{array}{c} \\ H \\ L \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{pmatrix} 0.48 & 0.48 \\ 0.52 & 0.52 \end{pmatrix} \end{array} \quad , \quad Female \begin{array}{c} \\ H \\ L \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{pmatrix} 0.24 & 0.24 \\ 0.76 & 0.76 \end{pmatrix} \end{array} \tag{5.1}$$

Row-wise inspection of these matrices shows that high risk partnerships account for almost half of the total number of partnerships formed in the male population and account for about one quarter of the total partnerships formed in the female population. This is due to the fact that a higher proportion of males have a high average annual partner change rate. So, although only a small proportion of the population practise high risk sexual behaviour, for example 14.3% of males as shown in Table 5.1, they account for almost half of the total partnerships formed in the male population. This may have an impact on disease transmission and will be explored in the simulation analysis.

5.4 Model Equations and Assumptions

The heterogeneous model equations are given by:

$$\begin{aligned} \frac{ds_{ki}}{dt} &= \mu n_{ki} - s_{ki}\lambda_{ki} - \mu s_{ki}, \\ \frac{di_{ki}}{dt} &= s_{ki}\lambda_{ki} - \gamma i_{ki} - \mu i_{ki}, \\ \frac{dr_{ki}}{dt} &= \gamma i_{ki} - \mu r_{ki}, \end{aligned} \tag{5.2}$$

where $\lambda_{ki} = \beta_k c_{ki} \sum_{j=1}^{n} (\rho_{kij})(\frac{i_{k'j}}{n_{k'j}})$

The full set of equations for this system are as follows:

Female, Low Risk Model

$$\begin{aligned} \frac{ds_{fL}}{dt} &= \mu n_{fL} - s_{fL}\lambda_{fL} - \mu s_{fL}, \\ \frac{di_{fL}}{dt} &= s_{fL}\lambda_{fL} - \gamma i_{fL} - \mu i_{fL}, \\ \frac{dr_{fL}}{dt} &= \gamma i_{fL} - \mu r_{fL}, \end{aligned} \tag{5.2a}$$

where $\lambda_{fL} = \beta_f \; c_{fL} \; (\rho_{LL} \; \frac{i_{mL}}{n_{mL}} + \rho_{LH} \; \frac{i_{mH}}{n_{mH}})$

Female, High Risk Model

$$\begin{aligned} \frac{ds_{fH}}{dt} &= \mu n_{fH} - s_{fH}\lambda_{fH} - \mu s_{fH}, \\ \frac{di_{fH}}{dt} &= s_{fH}\lambda_{fH} - \gamma i_{fH} - \mu i_{fH}, \\ \frac{dr_{fH}}{dt} &= \gamma i_{fH} - \mu r_{fH}, \end{aligned} \tag{5.2b}$$

where $\lambda_{fH} = \beta_f \; c_{fH} \; (\rho_{HL} \; \frac{i_{mL}}{n_{mL}} + \rho_{HH} \; \frac{i_{mH}}{n_{mH}})$

Male, Low Risk Model

$$\begin{aligned} \frac{ds_{mL}}{dt} &= \mu n_{mL} - s_{mL}\lambda_{mL} - \mu s_{mL}, \\ \frac{di_{mL}}{dt} &= s_{mL}\lambda_{mL} - \gamma i_{mL} - \mu i_{mL}, \\ \frac{dr_{mL}}{dt} &= \gamma i_{mL} - \mu r_{mL}, \end{aligned} \tag{5.2c}$$

where $\lambda_{mL} = \beta_m \; c_{mL} \; (\rho_{LL} \; \frac{i_{fL}}{n_{fL}} + \rho_{LH} \; \frac{i_{fH}}{n_{fH}})$

Male, High Risk Model

$$\begin{aligned} \frac{ds_{mH}}{dt} &= \mu n_{mH} - s_{mH}\lambda_{mH} - \mu s_{mH}, \\ \frac{di_{mH}}{dt} &= s_{mH}\lambda_{mH} - \gamma i_{mH} - \mu i_{mH}, \\ \frac{dr_{mH}}{dt} &= \gamma i_{mH} - \mu r_{mH}, \end{aligned} \tag{5.2d}$$

where $\lambda_{mH} = \beta_m \; c_{mH} \; (\rho_{HL} \; \frac{i_{fL}}{n_{fL}} + \rho_{HH} \; \frac{i_{fH}}{n_{fH}})$

Parameter definitions remain as previously defined in section 4.2.1 with the relevant subscripts for different gender and risk group combinations. Parameter values are as outlined in Table 5.1, all other parameter values are given below in Table 5.2: where β_m is the probability of transmission per partnership from females to males, and β_f is the probability of transmission per partnership from males to females. In line with published literature (Hughes *et al.* 2002), β_f is assumed to be higher than

Parameter	Estimate
β_m	0.7
β_f	0.8
μ	0.057
γ	0.58

Table 5.2: Parameter Estimates

β_m. For simplicity, the birth/removal rate μ, and recovery rate γ, are assumed to be equal for both genders, and are initially set equal to the values calculated following the sensitivity analysis in chapter 4, these are given in section 4.3.4, Table 4.3.

This model is constrained by the following assumptions:

- The model is assumed to have a constant population, that is, births are equal to deaths. To satisfy this condition the following terms from the set of ODEs (Ordinary Differential Equations) 5.2 must be equal: $\mu n_{ki} = \mu s_{ki} + \mu i_{ki} + \mu r_{ki}$.
- Individuals enter the sexually active population at age 18 into either the low or high risk group and remain in that risk group until they leave the model at age 64.
- The model assumes exclusively heterosexual mixing.
- The following model parameters are not gender specific: $c_L m = c_L f, \quad c_H m = c_H f, \quad \gamma, \quad \mu$.
- The population is 50% females and 50% males.
- HPV is a non-fatal disease. Deaths in the model are not HPV-related.
- This is a natural history model for HPV, and it assumes that no control strategies are in place in the population. Therefore, there is no cervical screening programme or vaccination programme.

5.5 Preliminary Results of Numerical Simulation and Sensitivity Analysis

The model was simulated using MATLAB software under the three sexual mixing scenarios outlined in section 5.3, using the parameter values as outlined above. Appendix B outlines the method used by MATLAB to solve the model. Table 5.3 shows the endemic prevalence of infection in males and females for the three mixing scenarios produced by the model simulations. The model was run for a sufficient time to allow the population to reach the endemic equilibrium. This occurred after approximately 150 years.

	Prevalence in Males	Prevalence in Females
Assortative	0.014	0.012
Disassortative	0.024	0.024
Proportionate	0.022	0.026

Table 5.3: Total Prevalence of Infection in Males and Females

All three of these simulations underestimate the prevalence of infection in the population, which is known to be approximately 4.4% or a proportion of 0.044. So, these parameter values do not accurately represent HPV infection in the irish population and need to be adjusted. As in chapter 4, a sensitivity analysis was carried out to deduce which parameters have the most profound influence on the model outcome, the prevalence of infection. The results of the sensitivity analysis are outlined in Tables 5.4, 5.5 and 5.6 below and are graphically depicted in Figures 5.1, 5.2 and 5.3.

As was the case in the previous sensitivity analysis in section 4.3.4, varying the parameter μ has a large impact on the prevalence of infection in the population under all three mixing scenarios. This rate, μ, measures the rate of natural death in the population, which is assumed to be age 64 for this model,which corresponds to the study age range of the sexual behaviour data and HPV prevalence data used to

populate the model (Layte *et al.* 2006, Kitchener *et al.* 2006), and also the average number of years spent in the sexually active population, which is not well defined in the literature. Given these facts, it is appropriate that the model is calibrated to match the known prevalence of infection using an adjusted value for the parameter μ.

Parameter	$Sen(t)(+0.1\%)$ - Males	$Sen(t)(+0.1\%)$ - Females
β_m	0.025	0.021
β_f	0.016	0.024
c_{mL}	0.015	0.014
c_{mH}	$1.38e^{-4}$	$7.15e^{-6}$
c_{fL}	0.012	0.017
c_{fH}	$1.86e^{-5}$	$4.8e^{-5}$
n_{mL}	0.044	$-8.67e^{-5}$
n_{mH}	0.076	0
n_{fL}	$-6.65e^{-5}$	0.021
n_{fH}	0	0.077
γ	-0.069	-0.072
μ	0.176	0.139

Table 5.4: Sensitivity Analysis for the Assortative Mixing Case

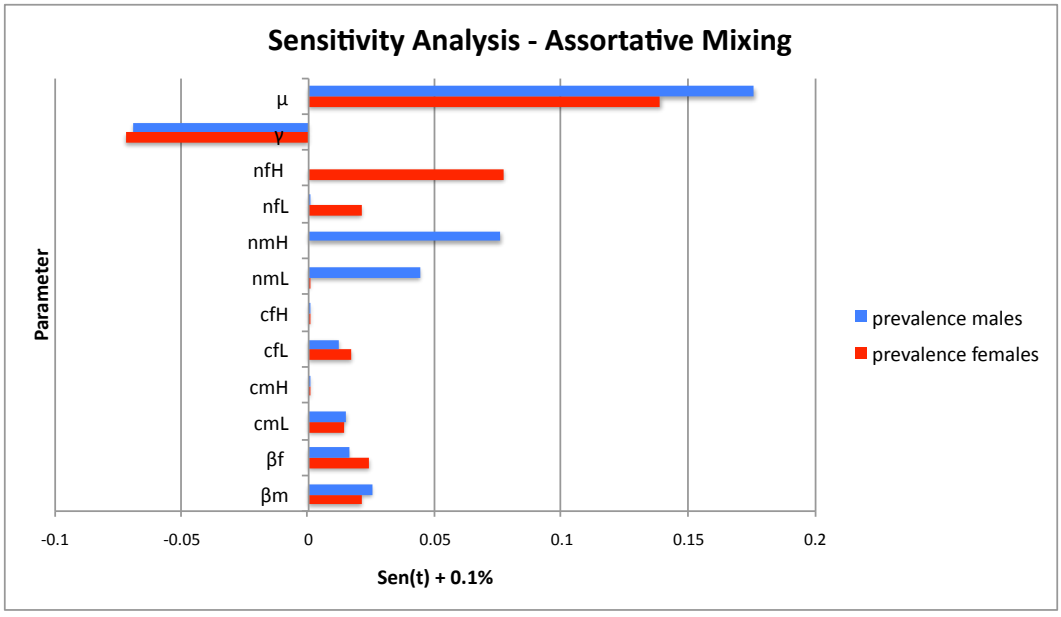


Figure 5.1: Bar Chart of Sensitivity Analysis for the Assortative Mixing Case

Parameter	$Sen(t)(+0.1\%)$ - Males	$Sen(t)(+0.1\%)$ - Females
β_m	0.017	0.004
β_f	0.004	0.015
c_{mL}	0.0098	$2.14e^{-4}$
c_{mH}	$2.10e^{-4}$	$4.49e^{-4}$
c_{fL}	$5.53e^{-4}$	0.011
c_{fH}	$3.84e^{-4}$	$7.6e^{-5}$
n_{mL}	0.044	0
n_{mH}	0.069	0
n_{fL}	0	0.046
n_{fH}	0	0.070
γ	-0.061	-0.061
μ	0.354	0.355

Table 5.5: Sensitivity Analysis for the Disassortative Mixing Case

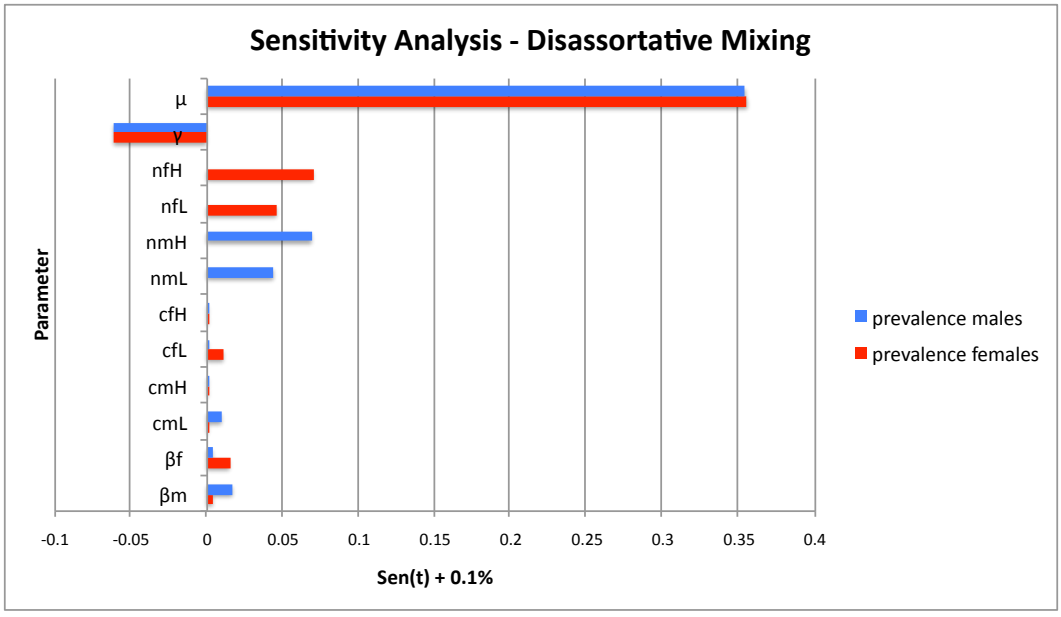


Figure 5.2: Bar Chart of Sensitivity Analysis for the Disassortative Mixing Case

Parameter	$Sen(t)(+0.1\%)$ - Males	$Sen(t)(+0.1\%)$ - Females
β_m	0.006	0.017
β_f	0.015	0.005
c_{mL}	0.002	0.009
c_{mH}	$2.64e^{-4}$	$2.67e^{-4}$
c_{fL}	0.010	0.002
c_{fH}	$3.36e^{-4}$	$2.01e^{-4}$
n_{mL}	0.04	0
n_{mH}	0.072	0
n_{fL}	0	0.051
n_{fH}	0	0.062
γ	-0.066	-0.060
μ	0.383	0.326

Table 5.6: Sensitivity Analysis for the Proportionate Mixing Case

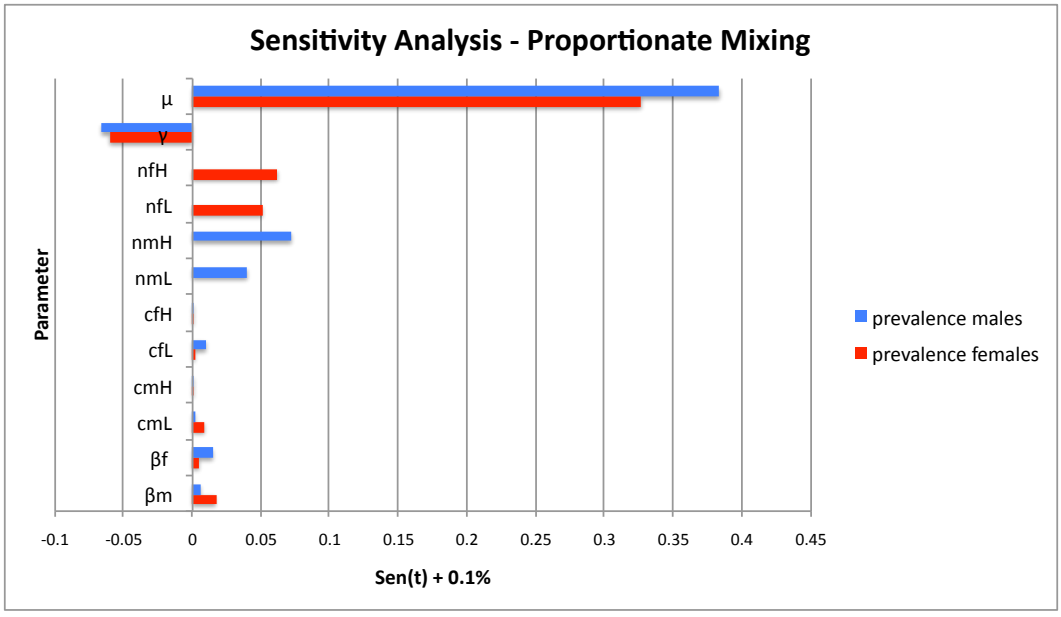


Figure 5.3: Bar Chart of Sensitivity Analysis for the Proportionate Mixing Case

5.6 Calibration of the Model, adjusting μ

The model was calibrated using Berkeley Madonna. A parameter plot was produced for a range of appropriate values of μ, under all three mixing scenarios, while all other parameters were held constant. For each mixing scenario, the model was simulated 30 times, each time varying μ along a range of values from 0.05 to 0.15. A parameter plot for each group of 30 simulations was produced which plotted μ on the x-axis against the associated prevalence of infection for each gender on the y-axis. The following plots show the results of these investigative simulations.

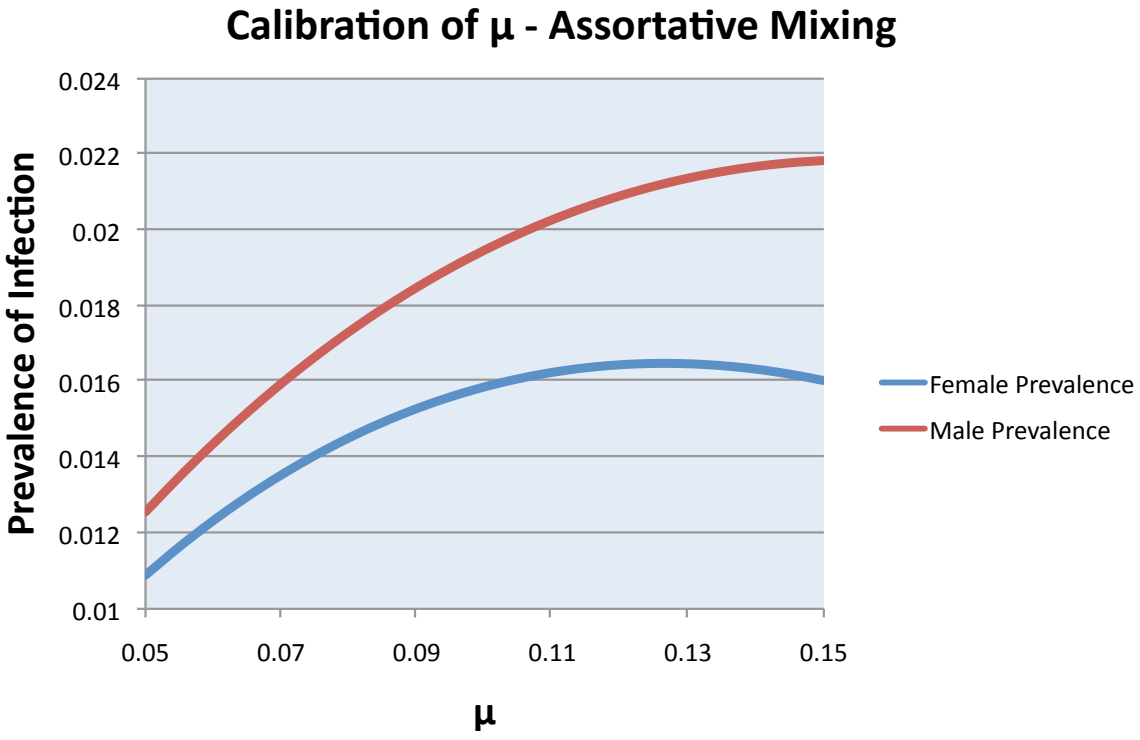


Figure 5.4: Calibration of μ under the assumption of Assortative Mixing

Under the assumption of assortative mixing, varying μ from 0.05 to 0.15 failed to reach the known prevalence of infection in the population of 0.044, or 4.4% as

depicted in Figure 5.4. This was not an unexpected result, since preliminary results for the model shown in Table 5.3 showed that the assumption of assortative mixing underestimated prevalence by a large amount. Therefore, no new estimate for μ can be taken from this calibration attempt.

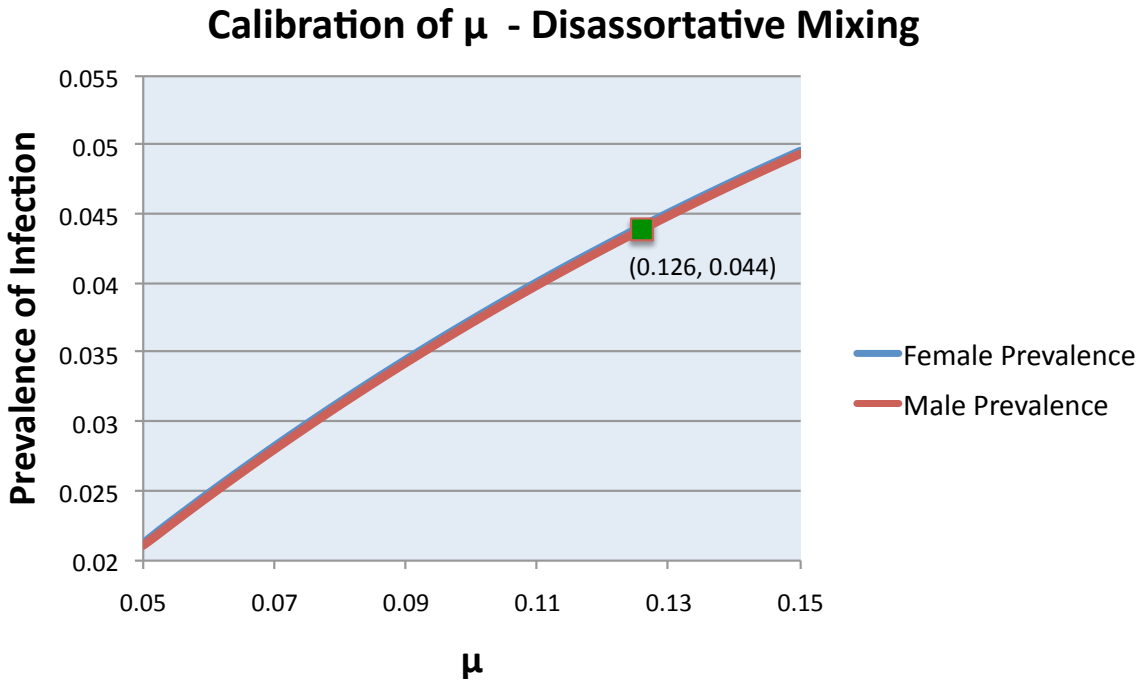


Figure 5.5: Calibration of μ under the assumption of Disassortative Mixing

In the case of disassortative mixing, prevalence increases linearly as μ increases, with no difference between genders, as shown in Figure 5.5. The value of $\mu = 0.126$ accurately predicts the known prevalence of infection of 0.044 in the population. This data value, labelled in Figure 5.5, equates to an average duration of 9.6 years in the sexually active population for all individuals, that is, an average duration of 9.6 years during which the partner change rate $c > 0$. The parameter μ is calculated as $\frac{1}{46} + \frac{1}{9.6} = 0.126$, which takes account of the two aspects of the the birth/removal rate which are: individuals entering the model aged 18 and leaving aged 64 which equates

to a removal rate of $\frac{1}{64-18}$, and secondly, these individuals are allowed to remain in the sexually active and mixing population for 9.6 years.

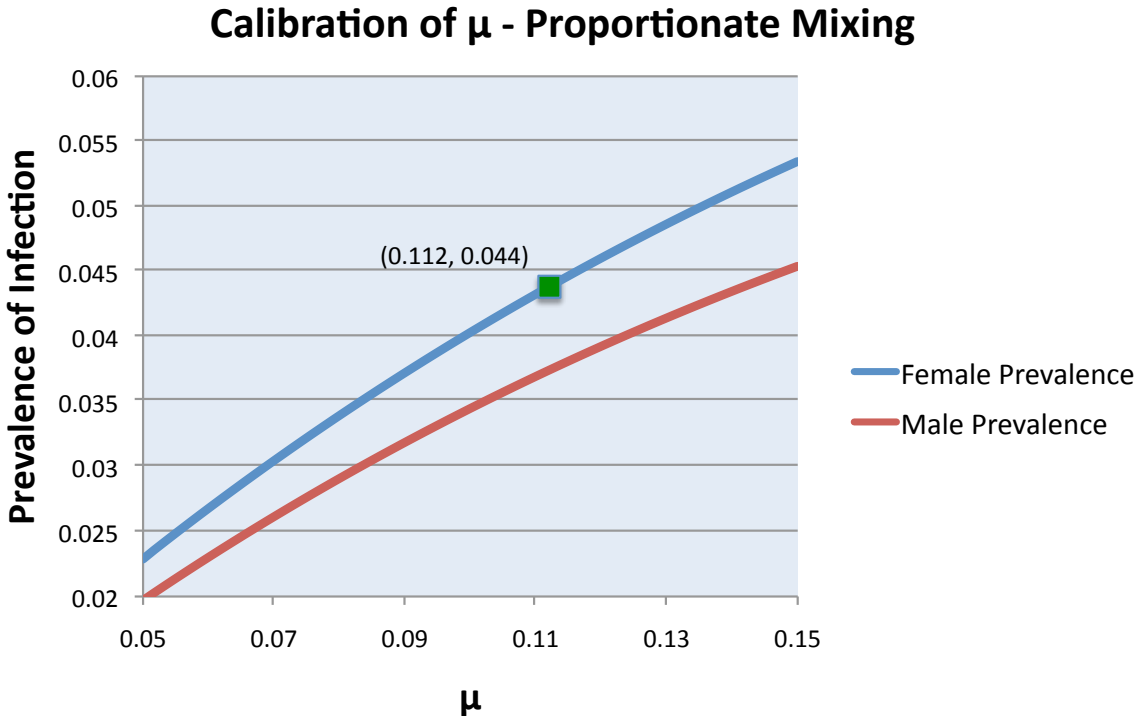


Figure 5.6: Calibration of μ under the assumption of Proportionate Mixing

Finally, under the assumption of proportionate mixing, the prevalence of infection increases approximately linearly as μ increases as shown in Figure 5.6. For females, the prevalence of infection is known to be 0.044, which is predicted by the model when μ is equal to 0.112, as shown in the labelled data point in Figure 5.6. This value corresponds to an average duration of 11.1 years, or $\mu = \frac{1}{46} + \frac{1}{11.1} = 0.112$. As discussed in section 4.2.3, the prevalence of HPV infection in males is assumed to equal that of females, thus we use the associated estimate for μ.

From these investigative plots, a more mathematically appropriate value for μ was found for the model. Under the assumptions of disassortative and proportionate mixing, the prevalence of HPV infection in the Irish population is accurately simulated

when μ is approximately 0.114. This equates to an average duration of about 11 years in the sexually active population, Hughes *et al.* (2002) estimated this value at 15 years for a risk structured model in a general population.

In the following section, the model is simulated under the three sexual mixing assumptions using the previously defined parameter values from Tables 5.1 and 5.2 and the new estimate for $\mu = 0.114$.

5.7 Numerical Analysis of a Heterogeneous Model with Various Sexual Mixing Patterns

The model was simulated three times under each of the sexual mixing scenarios described in the previous sections using MATLAB. For each of the three simulations the prevalence of infection, i, was plotted against time in years. The total prevalence for both genders was plotted, and was further stratified by risk class. All simulations resulted in the population reaching a stable endemic equilibrium after approximately 100 years. The model is assumed to be at an endemic equilibrium when the prevalence rate remains unchanged to six decimal places for a period of 20 years (Garnett and Anderson 1996). The prevalence of infection at the stable endemic equilibrium is representative of the prevalence of HPV 16 and 18 in the Irish population in recent years, that is, before the introduction of a vaccination programme against these two strains. The validity of these results is constrained by the model assumptions and limitations outlined in section 5.4. Figure 5.7 shows the results of these three model simulations.

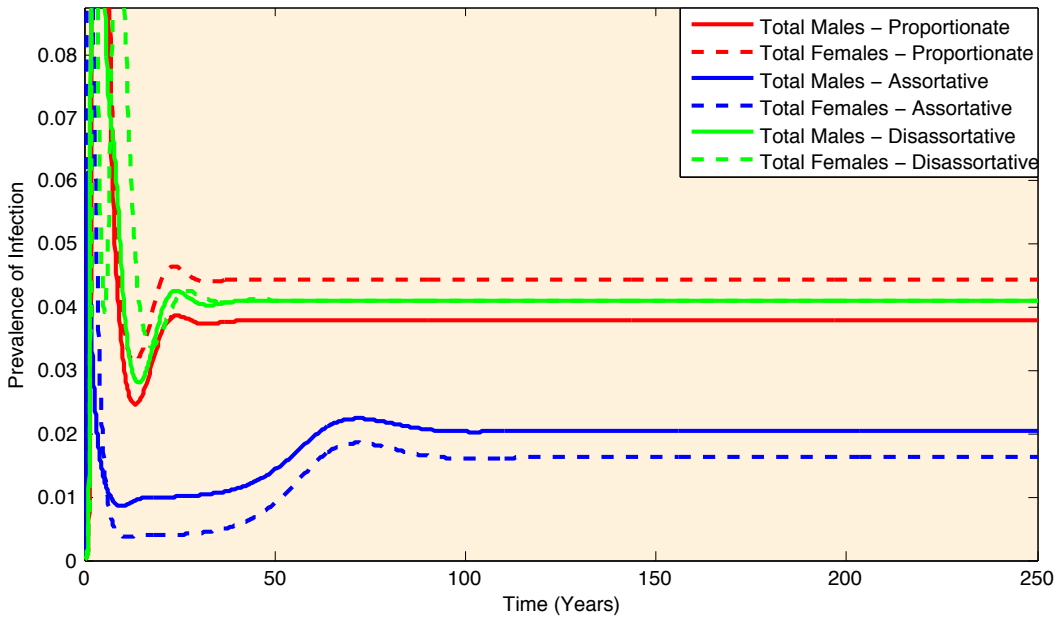


Figure 5.7: Prevalence Plots under Three Sexual Mixing Scenarios

Table 5.7 shows the predicted prevalence of infection at the endemic equilibrium for males and females under the three mixing scenarios, as presented in Figure 5.7.

Mixing Pattern	Prevalence in Males	Prevalence in Females
Assortative	0.021	0.016
Disassortative	0.041	0.041
Proportionate	0.04	0.044

Table 5.7: Prevalence in Males and Females under various Mixing Scenarios

As was the case in the preliminary simulation of the model presented in Table 5.3, the assumption of fully assortative mixing largely underestimates the prevalence of infection in the population, with an average population prevalence of about 0.019, or

1.9%.

In the case of fully disassortative mixing, male and female prevalence is 4.1% a proportion of 0.041, which is relatively close to the known prevalence of 4.4% in the Irish female population, given the level of parameter uncertainty associated with a HPV model, as shown in the parameter sensitivity analysis in section 5.5.

The assumption of proportionate mixing provides the most accurate simulation of HPV infection in Ireland. Prevalence is at 4.4% in the female population, which matches the known prevalence rate. This rate is slightly lower in the male population, at 4%, which is to be expected since the probability of transmission per-partnership is higher from males to females than from females to males, that is, $\beta_f > \beta_m$.

Therefore, these three model simulations show that disassortative mixing and proportionate mixing provide accurate estimates of the prevalence of infection in Ireland. However, this does not assure confidence in the validity of these assumptions. Further analysis of the dynamics of the model under these assumptions is necessary before any conclusions can be drawn on the most appropriate mixing scenario to represent the Irish sexually active population.

Figure 5.7 clearly displays the variation in population prevalence of HPV infection under the three mixing scenarios, assortative, disassortative and proportionate. However, greater insight into the effects that these mixing assumptions have on male and female prevalence can be achieved by comparing prevalence at the risk sub-group level. For example, Figure 5.7 shows that under the assumption of disassortative mixing, prevalence is almost equal for males and females. What this figure cannot demonstrate is whether high risk individuals contribute equally to this prevalence rate in both genders. The following section shows results taken from the same three simulations that produced Figure 5.7. They show greater detail of the dynamics between risk groups and gender and allow for in depth analysis of the mixing scenarios

and how they affect prevalence in each sub-group of the population.

5.7.1 Effects of Mixing Patterns on Sub-group Prevalence

The following graphs show the proportion of infected individuals in the high and low risk groups under the three different mixing scenarios. In each case, adding together the prevalence in the high and low risk groups gives the total prevalence of infection for each gender, as presented in Figure 5.7 and Table 5.7.

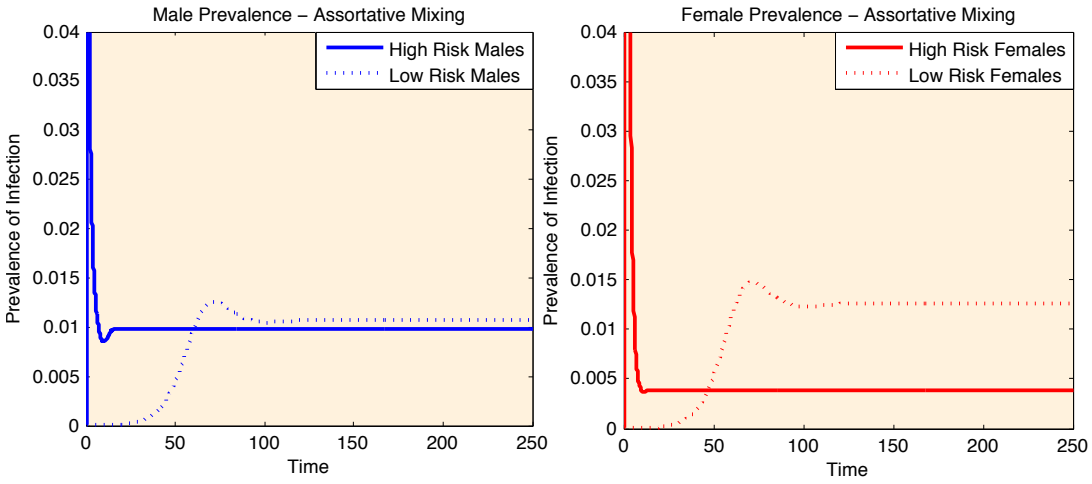


Figure 5.8: Risk group prevalence, with Assortative Mixing

As discussed in section 5.5, assortative mixing in the model results in an underestimate of endemic prevalence. Analysis of the disease dynamics at the sub-group level, as shown in Figure 5.8, shows that although prevalence of infection is slightly

higher in the high risk group than for the other two forms of sexual mixing, as shown in Figures 5.9 and 5.10, infection levels are low in the low risk group. This is because low risk individuals are only being infected by other low risk individuals, and since the partner change rate is low amongst all individuals available for partnerships, infection is unable to reach a high endemic equilibrium in the population. This leads to the question of whether fully assortative mixing, whereby individuals choose partners in the same risk group as themselves, is likely to occur in the Irish population?

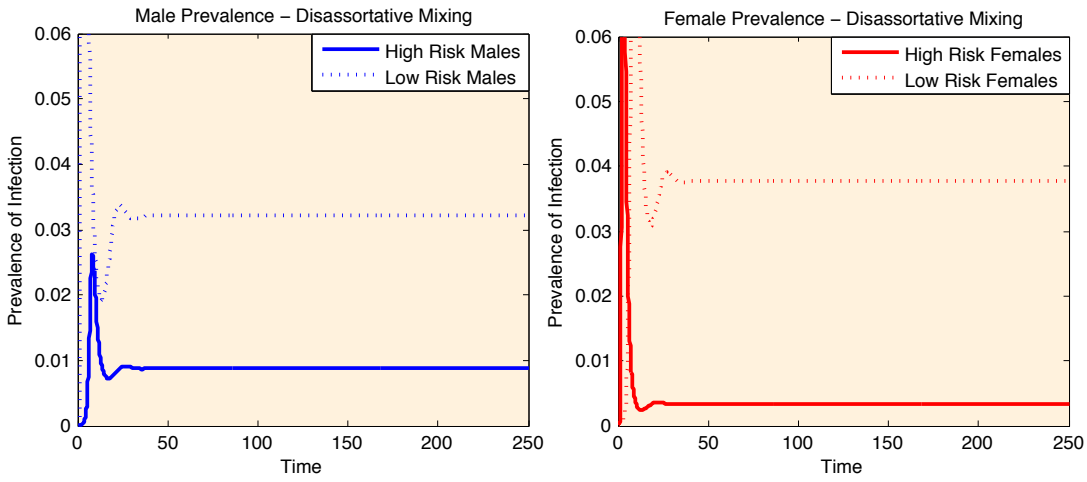


Figure 5.9: Risk group prevalence, with Disassortative Mixing

In the case of disassortative mixing, where individuals choose partners in the opposite risk group to themselves, prevalence in the high risk population is similar to that seen in the assortative mixing case, but the total prevalence in the population is considerably higher as infection reaches a higher endemic equilibrium in the low risk group. This shows the importance of the high risk group in the population. Although the proportion of the population in the high risk group is small (5.4% of females and 14.3% of males), they have a large influence on endemic prevalence. When low risk individuals mix exclusively with other low risk individuals, endemic prevalence is low, as was shown in Figure 5.7. But, when low risk individuals mix exclusively with high

risk individuals, endemic prevalence in the low risk group is more than doubled (3.2% in Males and 3.8% in females).

It is also worth noting that although total prevalence in the disassortative mixing model is equal for both genders, when the dynamics are analysed at the sub-group level as in Figure 5.9, it is clear that there is greater diversity in the female population than in males. Infection is more prevalent in low risk females than in low risk males, but the opposite is true of the high risk group, even though the average annual partner change rate is equal for both genders, that is $c_{Hm} = c_{Hf} = 6$ and $c_{Lm} = c_{Lf} = 1.1$. This disparity in results is directly related to the distribution of males and females into high and low risk groups. Females have a larger proportion of low-risk individuals than the males and this directly affects the prevalence of infection in the each gender. Re-simulating this scenario and allowing $n_{fL} = n_{mL}$ and $n_{fH} = n_{mH}$ results in both genders having almost the same prevalence, with females having a marginally higher prevalence in both sub-groups as a result of the transmission probability being higher from males to females, than from females to males, that is, $\beta_f > \beta_m$.

The final simulation assumes proportionate mixing in the population. As in the previous two simulations, prevalence in the high risk group is $\leq 1\%$. The overall distribution of infection is quite similar to that of the disassortative case. Again, infection is more prevalent in low risk females than in low risk males, but the opposite is true of the high risk group. This is a result of the WAIFW matrices for proportionate mixing, calculated in section 5.3, equation (5.1). From these matrices, low risk partnerships form a larger percentage of the total partnerships in the female population due to the fact that a higher proportion of females practise low risk behaviour than in the male population. This relation, where low risk prevalence is higher in females than in males, holds true when the scenario is re-simulated under the condition where $\beta_f = \beta_m$ and $n_{fL} = n_{mL}$, $n_{fH} = n_{mH}$. This shows that the influencing factor is the terms in the WAIFW matrices.

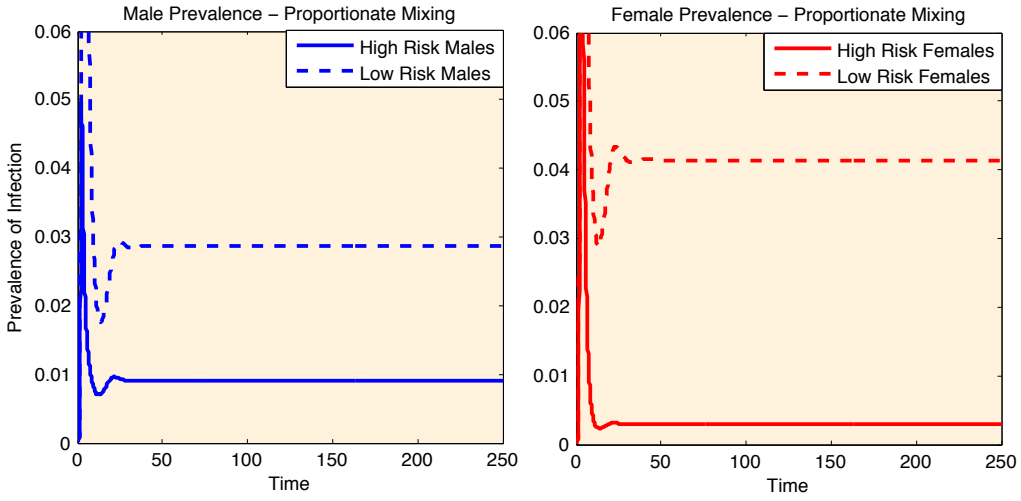


Figure 5.10: Risk group prevalence, with Proportionate Mixing

Tables 5.8 and 5.9 show the exact values of the prevalence of infection shown by the sub-group plots Figures 5.8, 5.9 and 5.10. These exploratory plots highlight the disproportionate role played by the risk groups in males and females and investigate the influence of model parameters such as ρ_{kij} and n_{ki} on sub-group prevalence. Within-group sexual mixing, as shown in Figure 5.8 results in a low endemic prevalence of infection in the low-risk proportion of the population. The following section shows that within-group mixing also equates to a high R_0.

Mixing Pattern	Low Risk	High Risk
Assortative	0.011	0.01
Disassortative	0.032	0.009
Proportionate	0.033	0.007

Table 5.8: Endemic Prevalence in High and Low Risk Males under various Mixing Scenarios

Mixing Pattern	Low Risk	High Risk
Assortative	0.013	0.004
Disassortative	0.038	0.003
Proportionate	0.041	0.002

Table 5.9: Endemic Prevalence in High and Low Risk Females under various Mixing Scenarios

5.8 Basic Reproductive Number

As discussed above, the existence and activities of the "core group", or the high risk group, contributes a disproportionate amount to the persistence of infection in the population. (Hethcote and Yorke 1984). Since model (5.2) is stratified by gender and risk group, the definition of the basic reproductive number, R_0, must be extended to take account of the the different activity levels in the population subgroups (Garnett and Anderson, 1993). As defined previously, the basic reproductive number is the average number of infections caused by an infective individual in an entirely susceptible population. When we assume that sexual mixing is not random, then the number of secondary infections caused by an individual is dependent on the activity group they are in. Therefore, the basic reproductive number must be calculated in terms of the average basic reproductive numbers for each subgroup in the population (Vynnycky and White 2010). The following equations lead to the calculations of R_0 for males and for females under the three previously defined sexual mixing scenarios, assortative mixing, disassortative mixing and proportionate mixing.

5.8.1 Basic Reproductive Number for the General Case

In the general case for a heterosexually mixing population, the basic reproductive number for each subgroup of the population is given by the following equation (Garnett and Anderson, 1993)

$$R_{(0)ij} = \frac{\beta_i \; c_i \; \rho_{ij}}{\gamma + \mu} \tag{5.3}$$

where,

β_i is the probability of the transmission of infection per partnership,

c_i is the average annual partner change rate for group i,

$\frac{1}{\gamma+\mu}$ is the average duration of an infection,

and ρ_{ij} is the proportion of partnerships made by group j that are with individuals from group i, and is given by the WAIFW matrix previously defined in section 5.3, where

$$\rho_{ij} = \begin{array}{c} \\ H \\ L \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{pmatrix} \rho_{HH} & \rho_{HL} \\ \rho_{LH} & \rho_{LL} \end{pmatrix} \end{array}$$

Given that the population is stratified by gender and risk group, two Next Generation Matrices have been calculated, one for each gender. The Next Generation Matrix is a summary of the number of infections caused by an infectious individual in the population. The general form of these matrices is as follows:

$$Male \begin{matrix} \\ H \\ L \end{matrix} \begin{matrix} \begin{matrix} H & \quad L \end{matrix} \\ \begin{pmatrix} R_{(0)HH} & R_{(0)HL} \\ R_{(0)LH} & R_{(0)LL} \end{pmatrix} \end{matrix} \quad , \quad Female \begin{matrix} \\ H \\ L \end{matrix} \begin{matrix} \begin{matrix} H & \quad L \end{matrix} \\ \begin{pmatrix} R_{(0)HH} & R_{(0)HL} \\ R_{(0)LH} & R_{(0)LL} \end{pmatrix} \end{matrix} \tag{5.4}$$

A value for the Basic Reproductive Number for the male population can be found from the Next Generation Matrix by solving for the eigenvalue which is subtracted from the diagonal of the matrix and is set as R_0 (Vynnycky and White, 2010). See definitions for this process in section 3.5.1. The eigenvalues can be found by solving the characteristic equation which is found from the determinant of the matrix M, equation (5.4a)
We have,

$$Male \begin{matrix} \\ H \\ L \end{matrix} \begin{matrix} \begin{matrix} H & \quad L \end{matrix} \\ \begin{pmatrix} R_{(0)HH} - R_0 & R_{(0)HL} \\ R_{(0)LH} & R_{(0)LL} - R_0 \end{pmatrix} \end{matrix} \tag{5.4a}$$

The determinant of this matrix is given by solving for R_0 in the following equation. The basic reproductive number is taken to be the largest value which satisfies the following equation (Vynnycky and White, 2010). The characteristic equation is given by:

$$det|M - \lambda I| = 0$$

where λ represents the eigenvalues of matrix M and is equal to R_0, and I is the identity matrix. The characteristic equation for matrix M is:

$$(R_{(0)HH} - R_0)(R_{(0)LL} - R_0) - R_{LH}R_{HL} = 0 \tag{5.4b}$$

Parameter	Estimate
β_m	0.7
β_f	0.8
μ	0.114
γ	0.58
c_L	1.1
c_H	6

Table 5.10: Parameter values for the Basic Reproductive Number

This equation reduces to a quadratic equation of the form $ax^2 + bx + c = 0$

In the next section, the values of R_0 from equation (5.4b) for each sexual mixing scenario are calculated using the technique described above.

5.8.2 Basic Reproductive Number for the Male Population, Assuming Assortative Mixing

In the case of assortative mixing, the term ρ_{ij} will be given by the previously defined WAIFW matrix for males and females as defined in section 5.3,

$$Assortative \begin{array}{c} \\ H \\ L \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}$$

Given the above definitions and setting all other parameters as in Table 5.10 below, the basic reproductive numbers for the male subgroup of the population are:

$$R_{(0)HH} = \frac{\beta_m \, c_H \, \rho_{HH}}{\gamma + \mu} = \frac{0.7 \times 6 \times 1}{0.58 + 0.114} = 6.05$$
$$R_{(0)HL} = \frac{\beta_m \, c_L \, \rho_{HL}}{\gamma + \mu} = \frac{0.7 \times 1.1 \times 0}{0.58 + 0.114} = 0$$
$$R_{(0)LH} = \frac{\beta_m \, c_H \, \rho_{LH}}{\gamma + \mu} = \frac{0.7 \times 6 \times 0}{0.58 + 0.114} = 0$$
$$R_{(0)LL} = \frac{\beta_m \, c_L \, \rho_{LL}}{\gamma + \mu} = \frac{0.7 \times 1.1 \times 1}{0.58 + 0.114} = 1.11$$

The Next Generation Matrix for the male population follows the format of matrix (5.4). Under the assumption of assortative mixing the Next Generation Matrix for the male population is given by:

$$\begin{array}{cc} & \begin{array}{cc} H & L \end{array} \\ Male \begin{array}{c} H \\ L \end{array} & \begin{pmatrix} 6.05 & 0 \\ 0 & 1.11 \end{pmatrix} \end{array}$$

A value for R_0 in the male population can be found from the determinant of the following matrix:

$$\begin{array}{cc} & \begin{array}{cc} H & L \end{array} \\ Male \begin{array}{c} H \\ L \end{array} & \begin{pmatrix} 6.05 - R_0 & 0 \\ 0 & 1.11 - R_0 \end{pmatrix} \end{array}$$

The determinant of this matrix is given below, and its eigenvalue is our parameter of interest, R_0.

$$(6.05 - R_0)(1.11 - R_0) - 0 = 0$$

Multiplying out the brackets gives the following quadratic equation:

$$R_0^2 - 7.16R_0 + 6.72 = 0$$

Solving for R_0,

$$R_0 = \frac{7.16 \pm \sqrt{7.16^2 - 4(1)(6.72)}}{2(1)}$$

The basic reproductive number for the males under the assumption of assortative mixing is

$$R_0 = 6.05$$

Under the assumption of assortative mixing, R_0 for the female subgroup of the population was calculated as 6.92, details of these calculations can be found in Appendix C.

Details of the calculations for R_0 in the female population, and all subsequent calculations for both genders under the assumptions of disassortative mixing and proportionate mixing are given in Appendix C.

5.8.3 R_0 Assuming Disassortative Mixing

In the case of disassortative mixing, the term ρ_{ij} will be given by the previously defined WAIFW matrix for this case as defined in section 5.3,

$$Disassortative \begin{array}{c} \\ H \\ L \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

For the male subgroup of the population, R_0 was calculated as 2.59, details of these calculations can be found in Appendix C.

Under the assumption of disassortative mixing, R_0 for the female subgroup of the population was calculated as 2.96, details of these calculations can be found in Appendix C.

5.8.4 R_0 Assuming Proportionate Mixing

In the case of proportionate mixing, the term ρ_{ij} for the male and female subgroups of the population will be given by the previously defined WAIFW matrices as defined in section 5.3,

$$Male \begin{array}{c} \\ H \\ L \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{pmatrix} 0.48 & 0.48 \\ 0.52 & 0.52 \end{pmatrix} \end{array} \quad , \quad Female \begin{array}{c} \\ H \\ L \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{pmatrix} 0.24 & 0.24 \\ 0.76 & 0.76 \end{pmatrix} \end{array}$$

Under the assumption of proportionate mixing, R_0 for the male subgroup of the population was calculated as 3.48, details of these calculations can be found in Appendix C.

For the female subgroup of the population, R_0 was calculated as 2.62, details of these calculations can be found in Appendix C.

Mixing Scenario	R_0 for Males	R_0 for Females
Assortative	6.05	6.92
Disassortative	2.59	2.96
Proportionate	3.48	2.62

Table 5.11: Basic Reproductive Number under various Mixing Scenarios

Table 5.11 shows a summary of the calculations described in this section. Comparing these values allows us to evaluate the effects of sexual mixing on the spread of infection, as R_0 is a measure of how rapidly an infection will invade a population. Increasing the degree of within-group mixing generally results in a higher R_0 value and a lower endemic prevalence (Vynnycky & White 2010). This effect is clear from the results in Table 5.11. As was shown in Table 5.7, under the defined set of parameter values, the prevalence of HPV infection in females under the assumption of assortative mixing was 0.016, with an R_0 value of 6.92. Under the same set of parameters, when within-group mixing is reduced to zero for the dissassortative mixing scenario, prevalence rises to 0.041 with an R_0 value of 2.96. Similarly, under the assumption of proportionate mixing which has some within-group mixing prevalence of infection is 0.044 in females with an R_0 value of 2.62. The equilibrium prevalence of infection is higher in scenarios with less within-group sexual mixing because more transmission has to occur in the low activity group to maintain a given R_0 (Vynnycky & White

2010). Because prevalence is lower in the low risk group, within-group mixing in the low-risk group results in a greater transmission of disease, since contacts are not being 'wasted' on high-risk individuals who are already infected. Therefore prevalence in the low-risk group increases.

As discussed in section 5.6, the assumption of assortative mixing underestimates the known prevalence of infection in the population. Table 5.11 shows that under the assumptions of disassortative mixing and proportionate mixing, R_0 for males and females is approximately equal to three, that is, a typical infected individual will infect three individuals per year if they are introduced into an entirely susceptible population. Successfully controlling HPV in a population is dependent on R_0. Knowing the basic reproductive number for HPV infection is useful for shaping vaccination policies. A high R_0 value implies that infection will rise rapidly in the population, since each infected individuals infects a large number of susceptibles.

5.9 Discussion

The results of the three simulations presented here (sections 5.8.2 to 5.8.4) provide valuable insight into the dynamics of the sexually transmitted infection, HPV. In general, the pattern of mixing has a great effect on population prevalence (Garnett & Anderson 1996). Assortative mixing usually increases the spread of infection in the high risk group and results in a lower endemic prevalence compared with proportionate or disassortative mixing (Garnett & Anderson 1996), as was demonstrated in the simulations above.

This effect is the influence of the commonly referred to "core group" dynamics pioneered by Hethcote and Yorke (1984) whereby infection persists in a population

due to the high partner change rate of a small proportion of the population.

5.10 Conclusion

The simulations presented in this chapter have provided many insights into the general dynamics of HPV infection in Ireland. However, the next logical step is to further develop this natural history model and investigate the effects of vaccination on the prevalence of infection.

This chapter explored heterogeneity in sexual behaviour and its effect on HPV prevalence. This heterogeneity can be taken a step further by introducing a third risk class which could create an intermediate risk group and a new high-risk group who would have a higher partner change rate, and be made up of a lower proportion of the population. Although this would provide insight into the prevalence of infection in specific sub-groups of the population, it would add very little to our knowledge of overall level of infection in the population already explored in this chapter. A paper published by Garnett and Anderson (1996) which explores the dynamics of sexual behaviour on sexually transmitted diseases showed that increasing the number of risk groups in a general STI model from three to four groups had a minor effect on population prevalence, but retained a similar relation with the mixing parameter in the model. Thus, adding a third risk group to the model presented in this chapter would increase reliability in the model, but is unlikely to provide new insight.

A possible area of development for the model is to stratify the population by age. As previously stated, sexual behaviour in the Irish population is changing. Younger age groups have a higher partner change rate than the older generation of the population (Layte *et al.* 2006). Assuming this trend continues, Ireland could hypothetically see a rise in prevalence of STIs and related diseases such as cervical cancer (if no control

strategy was put in place) in the coming decades.

In terms of thesis objectives, this chapter addressed a number of the aims outlined at the beginning of the thesis. This chapter looked in detail at the effects of sexual mixing patterns on HPV dynamics and used calibration techniques and sensitivity analyses to refine parameter estimates. This chapter showed the development of a suitable mathematical model for HPV in Ireland. The significant parameters contributing to parameter spread were explored. The parameter n was found to have a significant effect on HPV spread. The effects of sexual behavior and mixing patterns were explored and it was found that the assumption of proportionate mixing was the most appropriate assumption for the model population. The Basic Reproductive Number for the natural history model was calculated to be approximately three. This coincides with previous estimates by Elbasha *et al.* (2010) who estimated R_0 for HPV 16 as 2.63 (Range: 2.00-3.49), and HPV 18 as 2.68 (Range: 1.80-4.27). This also provides confidence in the simplifying assumption made here to group HPV 16 and 18, since estimating them separately resulted in similar ranges for R_0. The only remaining objective to be addressed is to explore the effects of a vaccination programme on the steady state endemic prevalence of infection, which is addressed in chapter 6.

Having explored the dynamics of the natural history of HPV infection in the Irish population, we now introduce a control measure in the form of a prophylactic vaccine. The following chapter explores the effects of a vaccination program on the prevalence of HPV 16 and 18 in Ireland.

Chapter 6

Analysis of a HPV Vaccination Model

6.1 Introduction

The model presented in chapter 5 represents a model for HPV 16 and 18 infection in the Irish population with two risk groups and varying levels of sexual activity and patterns of mixing. This is a natural history model for infection, the case before any control strategies have been implemented.

There are currently two licensed HPV vaccines in Ireland, Cervarix and Gardasil, as mentioned in Chapter 1.
In 2010 the Minister for Health Mary Harney added the HPV vaccine Gardasil into the national immunisation programme. This quadrivalent vaccine induces immunity to HPV types 16 and 18 and also the two most prevalent low risk HPV strains HPV 6 and 11. This is a prophylactic vaccine and not designed as a form of therapy. Therefore a programme that targets individuals before they become sexually active will be most effective. In line with this fact, Ireland's HPV vaccination programme currently targets young women in their first year of secondary school, aged 12-13 years. A second control strategy against HPV infection was introduced in 2008, called Cervical

Check. This is a national cervical screening programme targeting females aged 25 - 60, as discussed in chapter 1.

Given that the HPV vaccine has been introduced into the population, an obvious question one might ask is, what effect will the vaccine have on the prevalence of infection in the population over time? This, and many other questions regarding the relationship between vaccine efficacy and disease dynamics can be answered through the analysis of a transmission dynamic model, such as the model presented in section 5.4.

Thus, this chapter presents a differential equation model for HPV infection in the Irish population under various vaccination scenarios. The population is stratified by gender and risk class and assumes proportionate mixing.

6.2 Vaccination Model Development and Analysis

6.2.1 Vaccination Model Assumptions, Equations and Parameter Estimates

Results so far have concentrated on the natural history of HPV 16 and 18 infection in the Irish population. The model presented in section 5.4 can be further developed to investigate the effects of an infection control strategy on the population endemic prevalence.

A vaccination parameter was added to the previously defined model in section 5.4 to evaluate the effect of a prophylactic vaccine on infection prevalence. The value of this parameter is allowed to vary, to represent different levels of vaccination coverage,

and also to investigate the difference between a female-only vaccination programme and a programme vaccinating males and females.

As discussed in section 4.2.3 individuals are born into the model aged 18 years, that is, they commence sexual activity at age 18. It is assumed that a proportion of these individuals are vaccinated before entering the model population. This simulates the case where individuals are vaccinated at the age of 12 or 13, and do not commence sexual activity until the age of 18. This is a simplification of the model and was discussed in section 4.2.3.

Chapter 5 investigated the effects of sexual mixing on the endemic prevalence of infection. Following detailed analysis of the various population subgroups under the three mixing scenarios, assortative, disassortative and proportionate mixing, it was concluded that the assumption of proportionate mixing yielded the most reliable results. Therefore, in this chapter, the effects of a national vaccination programme is investigated under the assumption of proportionate mixing in the sexually active population.

The vaccination parameter is defined as:

v: A proportion of individuals are vaccinated at birth, before they enter the sexually active population. The vaccine is assumed to be fully effective and provide lifelong immunity to infection with HPV 16 and/or 18. The longest followup study on vaccination against HPV 16 shows lasting antibody persistence and protection from HPV persistent infection 9 years post vaccination (Jit *et al.* 2011).

The model equations are:

$$\begin{aligned}
\frac{ds_{ki}}{dt} &= \mu n_{ki}(1 - v_k) - s_{ki}\lambda_{ki} - \mu s_{ki}, \\
\frac{di_{ki}}{dt} &= s_{ki}\lambda_{ki} - \gamma i_{ki} - \mu i_{ki}, \\
\frac{dr_{ki}}{dt} &= \gamma i_{ki} + \mu n_{ki} v_k - \mu r_{ki},
\end{aligned} \tag{6.0}$$

where $\lambda_{ki} = \beta_k c_{ki} \sum_{j=1}^{n} (\rho_{kij})(\frac{i_{k'j}}{n_{k'j}})$

As previously defined in chapter 4, section 4.2.1, the model parameters are as follows:

μ: The birth/removal rate. This parameter consists of two components, the natural death rate in the population and the rate at which individuals leave the sexually active population. μ is assumed to be constant in the population and universal for all individuals. The birth rate is set equal to the removal rate to maintain a constant population size. This parameter μ is not gender specific.

γ: The recovery rate from HPV infection. The average rate of recovery is calculated as 1/duration of infection, where time is measured in years. For simplicity, the duration of infection is assumed to be equal for both genders.

β: The probability of transmission of HPV 16 and/or 18 per partnership. The term β_f represents the probability of transmission per partnership from males to females, while β_m is the probability of transmission per partnership from females to males.

c: The average number of sexual partners per year.

ρ: Patterns of sexual mixing between the strata are described by the mixing matrix, ρ_{kij}, i and j represent the "chosen" and "choosing" individuals respectively.
λ: The force of infection, the rate at which susceptibles become infected.

As in chapter 5, the model is stratified by gender, which is denoted by the subscript k, which can be either f or m, for females and males respectively. The model is also stratified according to risk group, based on an individual's average annual partner change rate. There are two risk groups, given by the subscripts L for low risk individuals, and H for high risk individuals.
All parameter definitions remain as previously defined in section 4.2.3 with the relevant subscripts for different gender and risk group combinations. The full set of equations for this system are as follows:

Female, Low Risk Model

$$\begin{aligned}
\frac{ds_{fL}}{dt} &= \mu n_{fL}(1-v) - s_{fL}\lambda_{fL} - \mu s_{fL}, \\
\frac{di_{fL}}{dt} &= s_{fL}\lambda_{fL} - \gamma i_{fL} - \mu i_{fL}, \\
\frac{dr_{fL}}{dt} &= \gamma i_{fL} + \mu v n_{fL} - \mu r_{fL},
\end{aligned}$$

where $\lambda_{fL} = \beta_f \; c_{fL} \; (\rho_{LL} \; \frac{i_{mL}}{n_{mL}} + \rho_{LH} \; \frac{i_{mH}}{n_{mH}})$

Female, High Risk Model

$$\begin{aligned}
\frac{ds_{fH}}{dt} &= \mu n_{fH}(1-v) - s_{fH}\lambda_{fH} - \mu s_{fH}, \\
\frac{di_{fH}}{dt} &= s_{fH}\lambda_{fH} - \gamma i_{fH} - \mu i_{fH}, \\
\frac{dr_{fH}}{dt} &= \gamma i_{fH} + \mu v n_{fH} - \mu r_{fH},
\end{aligned}$$

where $\lambda_{fH} = \beta_f \; c_{fH} \; (\rho_{HL} \; \frac{i_{mL}}{n_{mL}} + \rho_{HH} \; \frac{i_{mH}}{n_{mH}})$

Male, Low Risk Model

$$\frac{ds_{mL}}{dt} = \mu n_{mL}(1-v) - s_{mL}\lambda_{mL} - \mu s_{mL},$$
$$\frac{di_{mL}}{dt} = s_{mL}\lambda_{mL} - \gamma i_{mL} - \mu i_{mL},$$
$$\frac{dr_{mL}}{dt} = \gamma i_{mL} + \mu v n_{mL} - \mu r_{mL},$$

where $\lambda_{mL} = \beta_m \; c_{mL} \; (\rho_{LL} \; \frac{i_{fL}}{n_{fL}} + \rho_{LH} \; \frac{i_{fH}}{n_{fH}})$

Male, High Risk Model

$$\frac{ds_{mH}}{dt} = \mu n_{mH}(1-v) - s_{mH}\lambda_{mH} - \mu s_{mH},$$
$$\frac{di_{mH}}{dt} = s_{mH}\lambda_{mH} - \gamma i_{mH} - \mu i_{mH},$$
$$\frac{dr_{mH}}{dt} = \gamma i_{mH} + \mu v n_{mH} - \mu r_{mH},$$

where $\lambda_{mH} = \beta_m \; c_{mH} \; (\rho_{HL} \; \frac{i_{fL}}{n_{fL}} + \rho_{HH} \; \frac{i_{fH}}{n_{fH}})$

Parameter values are as outlined in Table 5.10.

The model was simulated using Berkeley Madonna software. In the initial phase of the simulation, the vaccination parameter, v, was set equal to 0, this simulated the natural history case as in chapter 5. The model reached the endemic equilibrium after 100 years. This value matched the equilibrium point reached in chapter 5 , which is

to be expected since, when $v = 0$, the model presented in equation (6.0) reduces to the exact equations used in model 5.2. As in chapter 5, the endemic equilibrium has been reached when the prevalence of infection remains constant correct to six decimal places for a period of 20 years (Garnett & Anderson 1996).

Following this initial phase, once the endemic equilibrium was reached, the vaccination parameter was given a non-zero value. The value of v represents the proportion of births vaccinated at each time step. For example, the annual birth rate into the model per annum is 0.114, as defined in chapter 5. This is the proportion of the model population that begins sexual activity each year. If $v = 0.6$, we are saying that 60% of the population entering the sexually active population are immunised each year. So, every year, more and more of the total population will be immunised as people are vaccinated before they become sexually active and remain immune until they leave the model through death or cessation of sexual activity.

6.2.2 Vaccination Scenario Analysis

The vaccination parameter was introduced with a time delay of 100 years to allow the population to reach the natural endemic equilibrium. After approximately 80 years, the model reaches a new endemic equilibrium.

The value for v was allowed to vary to investigate the effects of vaccination for various levels of population coverage. Various vaccination proportions were tested for two different vaccination scenarios, one where the vaccine was given to a proportion v of both genders, and the second scenario was a female-only vaccine, which is the current case in the Irish population.

The results of the various vaccination scenarios are summarised in Table 6.2 below. Figures 6.1 - 6.4 illustrate some of the vaccination scenarios reported in Table 6.2. In each of these four figures, the x-axis represents time in years. The first 100 years of

each simulation is the natural history case, as analysed in chapter 5. The population is allowed to reach a stable endemic equilibrium before the vaccine is introduced at year 100. The model is simulated under the chosen vaccination scenario and after approximately 80 years a new stable endemic equilibrium is reached. This represents the prevalence of infection in the population following vaccination.

	Male		Female	
	Endemic Prevalence	**Percentage Reduction**	Endemic Prevalence	**Percentage Reduction**
No Vaccine - Natural History	3.79%	**-**	4.44%	**-**
Vaccinate 50% of 12-13 year old males and females annually	0.7%	**-82%**	0.9%	**-80%**
Vaccinate 60% of 12-13 year old males and females annually	0.3%	**-92%**	0.3%	**-93%**
Vaccinate 70% of 12-13 year old males and females annually	$4.66e^{-9}$ %	**-100%**	$5.9e^{-9}$ %	**-100%**
Vaccinate 80% of 12-13 year old males and females annually	$1.40e^{-30}$ %	**-100%**	$1.78e^{-30}$ %	**-100%**
Vaccinate 50% of 12-13 year old females annually	2.3%	**-39%**	1.8%	**-60%**
Vaccinate 60% of 12-13 year old females annually	1.9%	**-50%**	1.3%	**-71%**
Vaccinate 70% of 12-13 year old females annually	1.4%	**-63%**	0.9%	**-80%**
Vaccinate 80% of 12-13 year old females annually	0.7%	**-82%**	0.4%	**-91%**
Vaccinate 90% of 12-13 year old females annually	$2.57e^{-5}$ %	**-100%**	$1.03e^{-5}$ %	**100%**

Table 6.1: Vaccination scenarios and associated reduction in prevalence

Each percentage vaccinated value represents the new endemic equilibrium for prevalence of infection, reached by the model after approximately 80 years of the relevant vaccination scenario, for example, a female only vaccination programme that vaccinates 60% of 12-13 year old females annually. The "No Vaccine" results represent prevalence of infection in the natural history model. These results show that vaccinating both genders causes a greater, more rapid reduction in prevalence under all levels of population coverage when compared to vaccinating females alone.

As Table 6.2 shows, eradication of HPV 16 and 18 can be achieved by vaccinating at least 70% of males and females. For a female only vaccine, eradication of these HPV strains in males and females is achievable by vaccinating 90% of females. This is the most important result in the table, since it represents the current vaccination strategy in Ireland, with the assumption that the vaccine is 100% effective and provides lifelong immunity to infection with HPV 16 and 18.

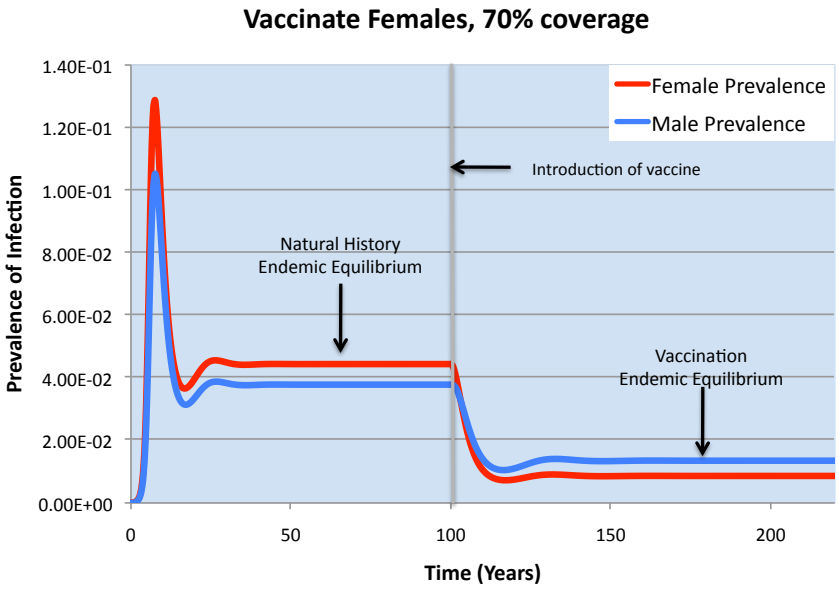


Figure 6.1: Endemic prevalence of infection for a female only vaccine at 70% coverage

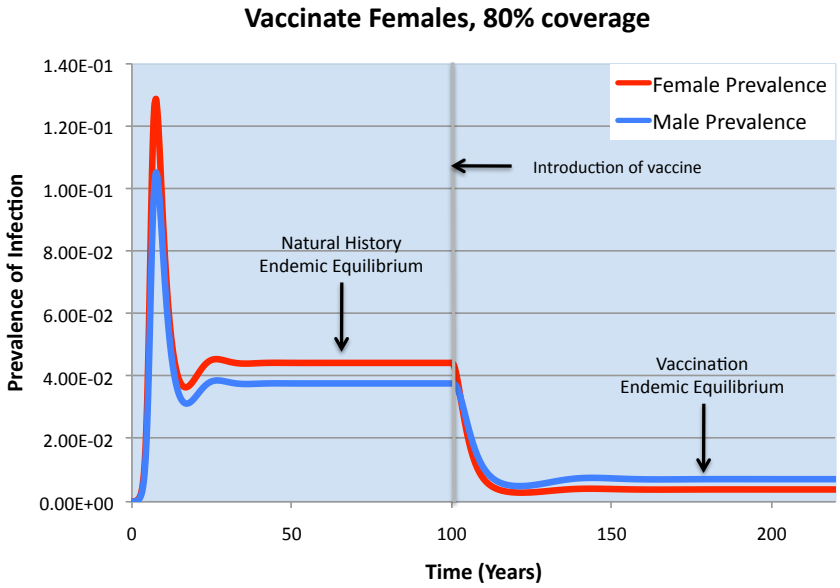


Figure 6.2: Endemic prevalence of infection for a female only vaccine at 80% coverage

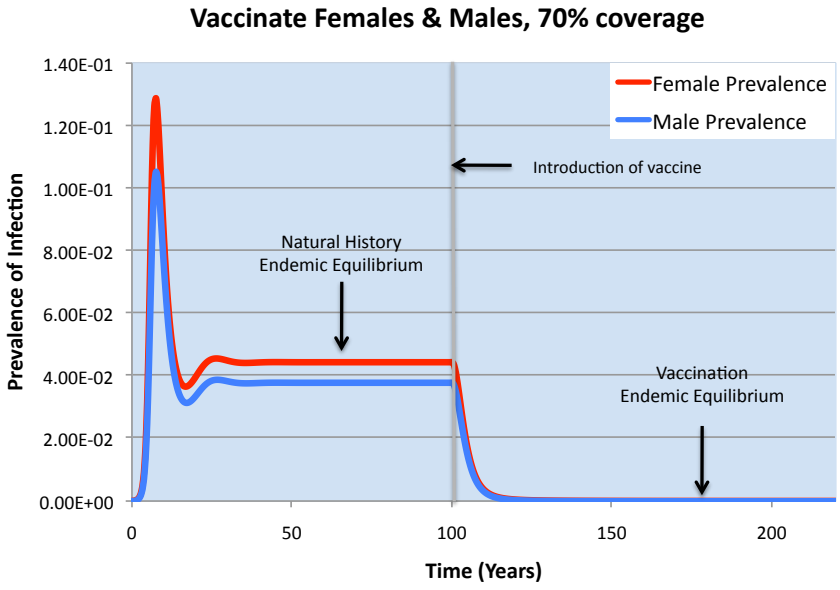


Figure 6.3: Endemic prevalence of infection for a female and male vaccine at 70% coverage

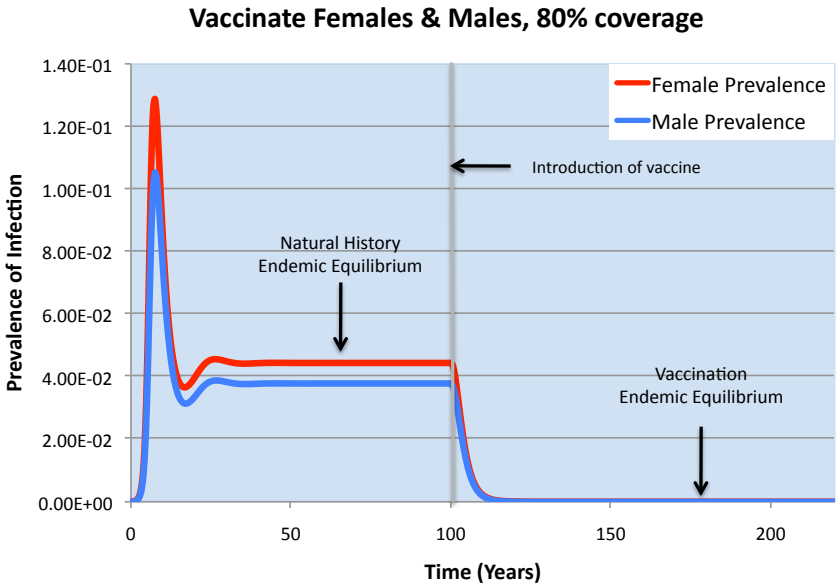


Figure 6.4: Endemic prevalence of infection for a female and male vaccine at 80% coverage

The concluding point from these results is that, under the current vaccination programme in Ireland, at least 90% of the target female population need to be vaccinated annually for a period of approximately 80 years in order to eradicate HPV 16 and 18 in the total sexually active population. Although the results in Table 6.2 are important and act as a predictive guide for policy makers as to what results can be expected in the future under each vaccination scenario, a more urgent question is, what results can be expected in the shorter term, that is, before the population reaches a new endemic equilibrium.

The following plot, Figure 6.5, shows the expected annual prevalence for males and females under the assumption of proportionate mixing, for a female only vaccine targeting 80% of females aged 12-13 each year with a catch up programme which targets females under the age of 18 for the first three years of the immunisation schedule. This means that from the time of the initialisation of the programme, 80% of females entering the sexually active population at age 18, will be vaccinated against infection

with HPV 16 and 18. The vaccine is introduced at year 0. These proposed parameters closely emulate those proposed for the Irish population. Ireland's HPV vaccination programme targets young women aged 12-13 and aims to achieve an uptake of more than 80% for this subgroup of the population (Health Service Executive 2012b). This programme was well received with an uptake rate of 82% in its first year, with 97% of cases receiving the full three doses. (Health Service Executive 2012a). The plot shows 10 years pre-vaccine. This scenario is the natural history case, and results match those presented in chapter 5. The vaccine is implemented at year 0, and its effects are evident from year 1.

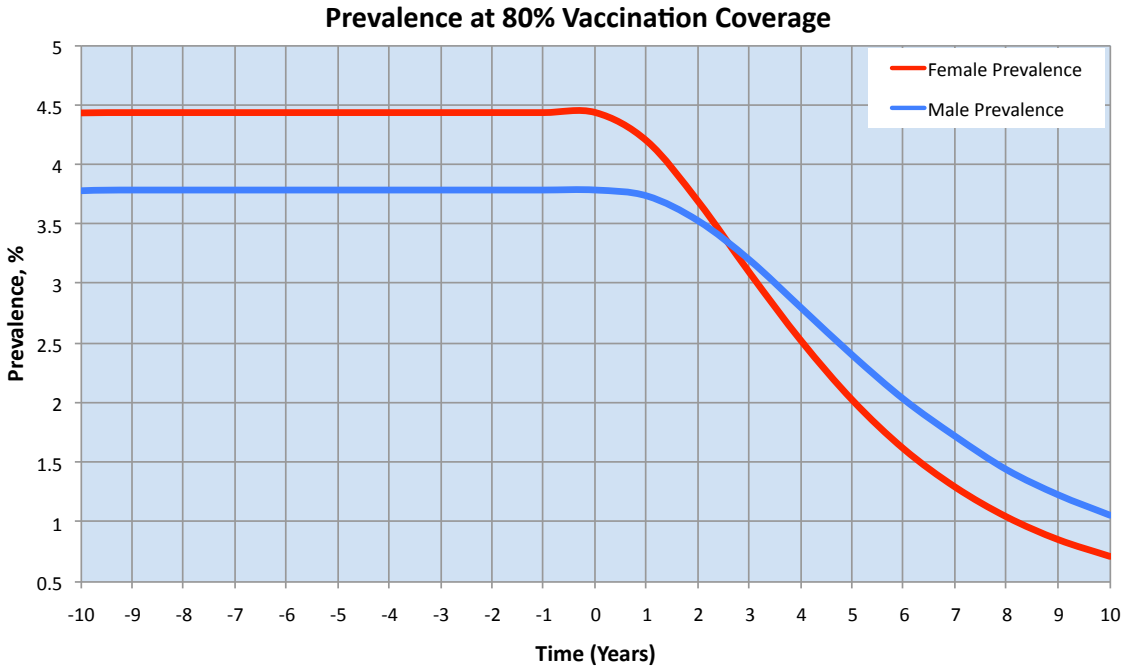


Figure 6.5: Annual prevalence of infection assuming 80% vaccination coverage

The plot shows that although prevalence is higher in the female population prior to the introduction of the vaccine, females see a sharper decline in population prevalence per annum. This is to be expected since 80% of births into the population each

year have induced immunity from infection as a result of vaccination, and the decline in male prevalence is dependent on the accumulation of "wasted contacts", whereby males form partnerships that do not result in the spread of infection since their female partners are already immune.

Vaccine (Years)	Male		Female	
	Infected	**Reduction**	Infected	**Reduction**
0	3.79%	-	4.44%	-
1	3.74%	**-1.3%**	4.20%	**-5.4%**
2	3.53%	**-6.9%**	3.69%	**-17%**
3	3.20%	**-16%**	3.09%	**-30%**
4	2.80%	**-26%**	2.52%	**-43%**
5	2.40%	**-37%**	2.02%	**-55%**
6	2.03%	**-46%**	1.61%	**-64%**
7	1.72%	**-55%**	1.29%	**-71%**
8	1.45%	**-62%**	1.04%	**-77%**
9	1.23%	**-68%**	0.85%	**-81%**
10	1.06%	**-72%**	0.71%	**-84%**

Table 6.2: Annual reduction in prevalence for a female only vaccine at 80% coverage

Table 6.3 shows the exact data presented in Figure 6.5, the annual reduction in prevalence following the implementation of a vaccination programme. As shown in Figure 6.2 the endemic equilibrium for the prevalence of infection of 0.4% for females and 0.7% for males is not reached until approximately 80 years after the introduction of this vaccination programme. The endemic equilibrium is the final rate of prevalence of infection under the defined conditions and it will remain constant assuming all previously defined parameters remain the same. Table 6.3 shows the effect this vaccine has on population prevalence in the short term. We can see that after 10

years, the prevalence of infection in males under these vaccination conditions will be reduced by 72% from the natural history endemic prevalence which was 3.79% of the male sexually active population. This reduction in prevalence continues beyond the 10 years shown in Figure 6.5 as the population moves towards the endemic equilibrium, shown in Figure 6.2.
A similar result is achievable in the female population. After 10 years, this vaccination programme reduces female prevalence by 84%. At the endemic equilibrium the prevalence of infection is reduced by 91% of the natural history prevalence rate.

In terms of the Irish population, the exact number of cases of infection in the female population can be estimated using census data (Central Statistics Office 2006). In 2006, the total number of females in the Irish population aged between 18 and 64 was 1,324,056. This is the same age range as simulated in the model presented in equation (6.0). If we ignore the effects of migration and emigration, then this proportion of the population will be expected to increase to 1,418,927 in the year 2022, which is 10 years after the introduction of the HPV vaccine. With the inclusion of a catch-up programme for this vaccine targeting females up to the age of 18 in 2011, and under the assumption that the vaccine achieves an 80% coverage for each year of the vaccination programme, then as shown in Table 6.3, the prevalence of infection in females is expected to decrease from 4.4% to 0.71%. This equated to a reduction in cases from 58,258 in 2006 to approximately 10,074 in the year 2022, 10 years after the introduction of the vaccine. This population prevalence will continue to decrease towards the endemic equilibrium, as shown in Figure 6.2 and reported in Table 6.2, to a percentage of 0.4%, or approximately 5676 cases. These results are constrained by the model assumptions on which they are based, such as the assumption that the vaccine is fully effective and immunity is lifelong, or a booster vaccination is administered before immunity begins to wane, all individuals begin sexual activity at the age of 18, and sexual mixing is proportionate.

6.3 Reduction in Cervical Cancer Cases

The primary reason for investigating the dynamics of HPV and the HPV vaccination programme is to infer the relative reduction in cervical cancer cases that can be expected from a such a programme. As mentioned in chapter 1, HPV is the cause of cervical cancer, and two strains HPV 16 and 18 are responsible for approximately 70% of cervical cancer cases annually.

Given that this model does not predict the future of cervical cancer, and the likelihood is that the number of cases seen today is not an accurate representation of the incidence expected in the coming decades, perhaps a more appropriate method for assessing the impact of the HPV vaccine on cervical cancer cases would be to report a predicted percentage reduction in cases, as opposed to the number of lives saved, or number of cases prevented.

Thus, we can say, as Table 6.2 shows, a female vaccine targeting 80% of young women annually will lead to a 91% reduction in female HPV prevalence over time. Given that the HPV strains described by this model account for 70% of cervical cancer cases in Ireland, we can infer that this vaccination programme will lead to an approximate reduction of $91\% \times 0.70 = 64\%$ in cases of cervical cancer in Ireland. It is important to note again, that this figure is based on the model reaching an endemic prevalence of infection after approximately 80 years and is not age specific. It also assumes that immunisation against HPV 16 and 18 has no effect on any other strains of HPV. Thus it does not reduce the prevalence of other oncogenic strains, and conversely it does not force the evolution of other oncogenic strains which would have the potential to increase cervical cancer cases.

6.4 Conclusion

This chapter looked at the further development of the natural history model presented in chapter 5. A vaccination parameter v was introduced, which conferred immunity to a varying proportion of individuals entering the susceptible class each year. This represents the current vaccination strategy in Ireland where young girls are being vaccinated before they become sexually active. Vaccination scenarios combining various population coverages with a female only vaccine or a male and female vaccine were simulated to investigate the effects of an immunisation programme on the prevalence of infection. Results from these simulations show that vaccinating males and females results in a greater decrease in population prevalence than vaccinating females alone. Table 6.1 shows that eradication of these two cancer causing strains could be achieved by vaccinating males and females before they become sexually active at a coverage of 70%. This is clearly much more effective in terms of reducing infection prevalence when compared to vaccinating females alone which eradicates infection at a coverage of 80%. However, these figures give no indication of the cost-effectiveness of these scenarios. Applying a cost-effectiveness analysis to this model data would help to determine whether it would be more cost-effective to vaccinate both genders at a rate 0f 70%, or females only at a rate of 80%. A contributing factor which would need to be considered would be the cost associated with disease treatment for both genders.

Based on the government's target coverage of 80% for a female only vaccine, the model presented in this chapter predicts that after 10 years of the vaccination programme, prevalence of HPV 16 and 18 will be reduced by 84% in females and 72% in males. After approximately 80 years of this vaccination programme, the prevalence of infection will reach an endemic equilibrium which will see female prevalence being reduced by 91% from the natural history rate of prevalence, and males by 82%, giving a new endemic prevalence of 0.4% in females and 0.7% in males.

Chapter 7

Conclusions

7.1 Introduction

Using classical mathematical modelling techniques a two-sex, risk structured ODE transmission dynamic model for HPV 16 and 18 was developed for a vaccinated Irish population. This chapter highlights the main findings of this thesis and reflects on the initial aims of the project as outlined in the preliminary summary chapter and at the beginning of each chapter and how these aims were achieved.

This project applied classical mathematical modelling techniques to evaluate a vaccination programme for the sexually transmitted infection, HPV. The model population followed the classic SIR structure whereby individuals were divided into one of three groups based on their disease status. Individuals were either susceptibles who were capable of contracting infection, infectives who were infected and could transmit the disease to others, or recovered individuals who are immune from infection as a result of conferred immunity, vaccination, or death. The population was further divided by gender and sexual behaviour risk group, based on the average number of sexual partners an individual has per year. The model was simulated and analysed under three sexual mixing scenarios with varying levels of within-group mixing.

Chapter 1 explored the history of the Human Papillomavirus (HPV) and the mech-

anisms by which the virus affects humans. Comparisons were made between Ireland and the rest of Europe and the world in terms of infection control strategies such as vaccination. This chapter also looked at the landmark research that has shaped our knowledge of HPV and cervical cancer.

Chapter 2 studied classical mathematical modelling theory. The deterministic ODE model for endemic disease was chosen as the model structure for this thesis and was subsequently presented and defined. The threshold estimates of the model were discussed and the solution paths for the equilibrium points of the system were illustrated using phase plane portraits.

In chapter 3 a simple Ordinary Differential Equation (ODE) model for endemic disease was introduced. The key parameters and model assumptions were defined. These are: μ, γ, β, c and v. These parameter definitions carry through to the more complex models presented in subsequent chapters. Model equations were solved analytically to find solutions to the steady state or equilibrium points of the model. A stability analysis was carried out on the model solutions to evaluate the dynamics of infection. This chapter took an initial step to achieving the first objective outlined in the thesis summary which was to study the classical techniques of mathematical modelling and develop a suitable ODE model for HPV in Ireland. A suitable ODE model for HPV in Ireland was developed in subsequent chapters.

Chapter 4 introduced the first numerical simulation of the thesis. A simple homogeneous model for the natural history of HPV infection was defined and estimates for each parameter were taken from published data. The model was simulated in Berkeley Madonna and a calibration of the model was used to estimate the probability of the transmission of infection per partnership, given by the parameter β. The calibration of β showed that this homogeneous model was over-simplified. To add a degree of complexity to the model, heterogeneous sexual mixing was added. The population was divided into two risk groups based on an individual's average

annual partner change rate, c. The significant effects of β on the spread of HPV and numerical estimates of model parameters were investigated. The effects of sexual risk groups within the population were explored through model simulations. This model does not consider different sexual mixing patterns, this is addressed in chapter 5.

Chapter 5 continued the development of the natural history endemic model introduced in chapter 4. The model was initially developed to simulate the natural history of infection in the population and explore the population dynamics associated with the contributing factors for disease such as gender, sexual behaviour and patterns of sexual mixing. Three WAIFW (Who Acquires Infection From Whom) matrices were introduced to represent three sexual mixing scenarios: assortative, disassortative and proportionate mixing. A sensitivity analysis of the model parameters showed that the system was most sensitive to changes in the parameter μ. This model was subsequently calibrated around μ and re-simulated using the new parameter set under the three sexual mixing scenarios. These simulations revealed that the assumption of assortative mixing underestimated the known prevalence of infection, while the other two sexual mixing scenarios with varying degrees of between-group mixing closely estimated the known prevalence of 4.4% in the female population. The epidemiologically significant parameter R_0, the basic reproductive number, was calculated for this model under the three sexual mixing scenarios: assortative, disassortative and proportionate. Detailed analysis of the effects of sexual mixing on sub-group prevalence in the population was carried out. It was found that increasing within-group mixing resulted in a higher R_0 but lower overall endemic prevalence than a sexual mixing scenario that assumed a degree of between-group mixing. This analysis and the results of the sensitivity analysis also showed that the proportion of the population in each risk group, given by n, has a greater impact on the prevalence of HPV infection than the probability of the transmission of infection β. Therefore, the sexual behaviour of a population is an important catalyst in the transmission of HPV. This chapter concluded that the assumption of proportionate mixing was the most appropriate of the three mixing scenarios for the Irish population and this assumption was carried

into chapter 6 which re-introduced the vaccination parameter, v, from chapter 3.

Given that a HPV vaccination programme has now been introduced in the Irish population, the structured natural history model (5.2) was further developed to include a vaccination parameter and subsequently the effects of various vaccination strategies were explored to investigate their effect on the prevalence of infection in the population. The vaccine was assumed to provide life-long immunity to infection with HPV 16 and 18 and was administered to young people before they became sexually active. The effects of the vaccination programme on the prevalence of infection was evaluated for two separate scenarios: a female-only vaccine, and a vaccine targeted to males and females. Simulation of these scenarios at varying levels of population coverage allowed for the evaluation of the varying effects of a female only vaccine versus a male and female vaccine and the relative reduction in HPV prevalence that could be achieved by these vaccines at various levels of population coverage. The annual reduction in prevalence of HPV for a female-only vaccine was presented in Figure 6.5 and Table 6.3. After 10 years of vaccinating females at a coverage of 80%, the prevalence of HPV 16 and 18 will be reduced by 72% in males and 84% in females from the current natural history prevalence as simulated in chapter 5. Under the same vaccination assumptions, based on Irish census data (Central Statistics Office 2006) the prevalence of HPV 16 and 18 in Irish women will be reduced from approximately 58,258 in 2006, to approximately 10,074 in the year 2022. This chapter also provided an estimate for the reduction in cervical cancer cases. Vaccinating females at 80% population coverage could lead to a reduction of 64% of cervical cancer cases. This figure is based on the model reaching an endemic prevalence of infection after approximately 80 years and is not age specific.

The mathematical models presented in this thesis are the first ODE models for HPV 16 and 18 infection in the Irish population. In contrast to previously published models on HPV dynamics (Usher *et al.* 2008, Jit *et al.* 2008) these models focussed specifically on HPV infection and does not stratify infection by severity. The simplic-

ity of this assumption allowed for the in-depth analysis of the effects of behavioural parameters on HPV prevalence. Analysis of the effects of various sexual mixing scenarios on the sub-group prevalence of infection demonstrated the impact of sexual mixing and the proportion of the population in each risk group on HPV prevalence. These results show that the rate of infection is high across all sub-groups of the population and is not concentrated in the high risk group, as is the case for other sexually transmitted infections, such as gonorrhea (Hethcote & Yorke 1984). The structure of the models presented in this thesis are easily adaptable, meaning the models can be modified as new data becomes available on the factors that influence HPV infection.

Data for these models were taken from two primary sources. Data on sexual behaviour patterns in the Irish population were taken from the ISSHR (Layte *et al.* 2006) and HPV prevalence data were taken from the ARTISTIC trial (Kitchener *et al.* 2006). These models assume the population has a constant population, that is, births are equal to deaths. Individuals enter the sexually active population at age 18 into either the low or high risk group and remain in that risk group until they leave the model at age 64. The models assume exclusively heterosexual mixing.The population is 50% females and 50% males. HPV is a non-fatal disease. Deaths in the model are not HPV-related. The natural history model for HPV, presented in chapter 5 assumes that no control strategies are in place in the population. Therefore, there is no cervical screening programme or vaccination programme. The vaccination programme introduced in chapter 6 assumes that a proportion of individuals are vaccinated at birth, before they enter the sexually active population. The vaccine is assumed to be fully effective and provide lifelong immunity to infection with HPV 16 and/or 18.

The natural history model from chapter 5 demonstrated the significant effect that the parameter n, the proportion of the population in each risk group, has on HPV spread. This result satisfies the second thesis objective which was to explore the epidemiologically significant factors contributing to the spread of HPV and estimate their numerical values. The assumption of proportionate mixing was found to be the

most appropriate mixing scenario to fit the model population, in line with the third objectives. The basic reproductive number for the natural history case was calculated as approximately three, which satisfies the fifth thesis objective. Chapter 6 explored the effects of a vaccination programme on the steady state endemic prevalence of infection, the final objective of this thesis. Results from these simulations showed that under the current vaccination scenario in Ireland which aims to achieve an 80% vaccination coverage in young women, we can expect the prevalence of HPV to be reduced by 84% in 10 years of the vaccination programme and will continue to decrease to 91% as the prevalence approaches a new endemic equilibrium. This is a significant result, since a significant reduction in HPV prevalence will in turn result in a significant decrease in cervical cancer cases.

The next section highlights some aspects of the project which could be developed in future work.

7.2 Further Work

As with any mathematical model for infectious disease, epidemiological results are subject to variability based on the assumptions made when constructing the model. The models presented in this thesis could be further developed by stratifying the population by age. Data shows that HPV incidence peaks in females aged 20-29 (Kitchener *et al.* 2006), and Irish data shows that cervical cancer incidence peaks between the ages of 40-45 (National Cancer Registry 2010). Hence, there is a considerable time delay between the initial contraction of HPV and progression to cervical cancer. Also, data from the ISSHR study shows that patterns of sexual behaviour in Ireland are changing with time (Layte *et al.* 2006). HPV incidence is on the rise in the younger population as a result of their increased levels of sexual activity when compared to the same age group in previous decades. Logically, a rise in HPV incidence may lead to a rise in cervical cancer incidence in the coming decades for this

subgroup. Patterns in sexual behaviour also vary with time across social, economic and geographical groups, (Layte *et al.* 2006; p174), which means that stratifying the population according to these factors may provide further insight. As discussed in section 4.2.3, these models assume that all individuals are sexually inactive until age 18. As data on the prevalence of HPV in people under the age of 18 becomes available, this sector of the population could be analysed and added to the current model to explore the effects of the commencement of sexual activity before the age of 18.

The model could also be improved by adding parameters to represent the annual proportion of females who have a hysterectomy, which removes them from the susceptible class for contracting HPV infection. As previously stated, HPV 16 and 18 account for approximately 70% of cervical cancer cases, which means that a further 30% of cases are caused by strains that are not contained in the current vaccines. A model which simulated these strains in addition to HPV 16 and 18, and explores the possible synergistic or antagonistic effects of the vaccine on the prevalence of these strains would be beneficial to policy makers. Also, the inclusion of the two most prominent low-risk strains, HPV 6 and 11 which are known to cause between 90% and 100% of genital warts cases would provide further insight into the effects of the currently licenced vaccine, Gardasil, on HPV prevalence in Ireland. Adding in the effects of cervical screening would also add accuracy to the model. The models presented here focus on sexual behaviour as a risk factor for contracting HPV infection. A model which considered persistent HPV infection and progression to cervical cancer would need to consider other known risk factors for the progression of disease such as smoking and long-term use of the oral contraceptive pill (Castellsagué & Munoz 2003).

HPV is the known cause of cervical cancer. The specific mechanisms by which this virus causes cervical neoplasia and the risk factors associated with developing cancer are currently being investigated by researchers around the world. The application of this knowledge using mathematical models is a crucial asset to policy makers in the

fight against cervical cancer deaths.

Appendix A

Berkeley Madonna RK4 method

Berkeley Madonna is a differential equation solver that uses the Runge-Kutta 4 method as its default solver for systems of differential equations. The RK4 method is described in Appendix B. This software was used to plot solutions of the model. Other features of this software that were used include:

Parameter sliders to explore the effects of various parameter values on the model solution.
Parameter plots were used to plot a variable as a function of a parameter.
Sensitivity Analyses plotted the partial derivative of a variable with respect to a parameter.

Appendix B

MATLAB ode45 solver method

MATLAB is a computational software that supports both numeric and symbolic modelling approaches for solving Ordinary Differential Equations (ODEs). MATLAB provides a number of solvers for approximating solutions to initial value problems of ordinary differential equations. All solvers solve systems of equations in the form $y' = f(t, y)$. The ode45 solver used throughout this thesis was the ode45 solver. It is based on an explicit Runge-Kutta (4,5) formula, the Dormand-Prince pair. It is a one-step solver in computing $y(t_n)$, it needs only the solution at the immediately preceding time point, $y(t_n - 1)$. This Runge-Kutta works as follows:
Given the initial value problem $y' = f(t, y), \quad y(t_n) = y_n$
We compute,

$$
\begin{aligned}
k_1 &= hf(t_n, y_n) \\
k_2 &= hf(t_n + \frac{h}{2}, y_n + \frac{k_1}{2}) \\
k_3 &= hf(t_n + \frac{h}{2}, y_n + \frac{k_2}{2}) \\
k_4 &= hf(t_n + h, y_n + k_3) \\
y_{n+1} &= y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\
t_{n+1} &= t_n + h
\end{aligned}
$$

Thus, y_{n+1} is determined by the present value y_n plus the weighted average of

k_1, k_2, k_3, k_4 where k_i is the product of the size of the interval h and an estimated slope. The fourth-order method has an error per step of the order of h^5, while the total accumulated error has order h^4.

Appendix C

Calculations for R_0 under various gender and sexual mixing scenarios

A1. R_0 for the Female Population, Assuming Assortative Mixing

Females:

$$R_{(0)HH} = \frac{\beta_f \; c_H \; \rho_{HH}}{\gamma + \mu} = \frac{0.8 \times 6 \times 1}{0.58 + 0.114} = 6.92$$
$$R_{(0)HL} = \frac{\beta_f \; c_L \; \rho_{HL}}{\gamma + \mu} = \frac{0.8 \times 1.1 \times 0}{0.58 + 0.114} = 0$$
$$R_{(0)LH} = \frac{\beta_f \; c_H \; \rho_{LH}}{\gamma + \mu} = \frac{0.8 \times 6 \times 0}{0.58 + 0.114} = 0$$
$$R_{(0)LL} = \frac{\beta_f \; c_L \; \rho_{LL}}{\gamma + \mu} = \frac{0.8 \times 1.1 \times 1}{0.58 + 0.114} = 1.27$$

The Next Generation Matrix for the female population, under the assumption of assortative mixing is given by:

$$Female \begin{array}{c} \\ H \\ L \end{array} \begin{array}{c} \begin{array}{cc} H & \quad L \end{array} \\ \begin{pmatrix} 6.92 & 0 \\ 0 & 1.27 \end{pmatrix} \end{array}$$

A value for R_0 in the female population can be found from the determinant of the following matrix:

$$
\begin{array}{cc} & \begin{array}{cc} H & \quad L \end{array} \\ Male \begin{array}{c} H \\ L \end{array} & \begin{pmatrix} 6.92 - R_0 & 0 \\ 0 & 1.27 - R_0 \end{pmatrix} \end{array}
$$

The determinant of this matrix is given below, and its eigenvalue is our parameter of interest, R_0.

$$(6.92 - R_0)(1.27 - R_0) - 0 = 0$$

Multiplying out the brackets gives the following quadratic equation:

$$R_0^2 - 8.19R_0 + 8.79 = 0$$

Solving for R_0,

$$R_0 = \frac{8.19 \pm \sqrt{8.19^2 - 4(1)(8.79)}}{2(1)}$$

$$R_0 = 6.92 \; or \; 1.27$$

A2. R_0 for the Male Population, Assuming Disassortative Mixing

Males:

$$R_{(0)HH} = \frac{\beta_m \, c_H \, \rho_{HH}}{\gamma + \mu} = \frac{0.7 \times 6 \times 0}{0.58 + 0.114} = 0$$
$$R_{(0)HL} = \frac{\beta_m \, c_L \, \rho_{HL}}{\gamma + \mu} = \frac{0.7 \times 1.1 \times 1}{0.58 + 0.114} = 1.11$$
$$R_{(0)LH} = \frac{\beta_m \, c_H \, \rho_{LH}}{\gamma + \mu} = \frac{0.7 \times 6 \times 1}{0.58 + 0.114} = 6.05$$
$$R_{(0)LL} = \frac{\beta_m \, c_L \, \rho_{LL}}{\gamma + \mu} = \frac{0.7 \times 1.1 \times 0}{0.58 + 0.114} = 0$$

The Next Generation Matrix for the male population, under the assumption of assortative mixing is given by:

$$
\begin{array}{cc} & \begin{array}{cc} H & L \end{array} \\ Male \begin{array}{c} H \\ L \end{array} & \begin{pmatrix} 0 & 6.05 \\ 1.11 & 0 \end{pmatrix} \end{array}
$$

A value for R_0 in the male population can be found from the determinant of the following matrix:

$$
\begin{array}{cc} & \begin{array}{cc} H & L \end{array} \\ Male \begin{array}{c} H \\ L \end{array} & \begin{pmatrix} 0 - R_0 & 6.05 \\ 1.11 & 0 - R_0 \end{pmatrix} \end{array}
$$

The determinant of this matrix is given below, and its eigenvalue is our parameter of interest, R_0.

$$(0 - R_0)(0 - R_0) - (6.05)(1.11) = 0$$

Multiplying out the brackets gives the following equation:

$$R_0^2 - 6.72 = 0$$

Solving for R_0,

$$R_0 = \pm\sqrt{6.72}$$

$$R_0 = 2.59$$

A3. R_0 for the Female Population, Assuming Disassortative Mixing

Females:

$$R_{(0)HH} = \frac{\beta_f \; c_H \; \rho_{HH}}{\gamma + \mu} = \frac{0.8 \times 6 \times 0}{0.58 + 0.114} = 0$$

$$R_{(0)HL} = \frac{\beta_f \; c_L \; \rho_{HL}}{\gamma + \mu} = \frac{0.8 \times 1.1 \times 1}{0.58 + 0.114} = 1.27$$

$$R_{(0)LH} = \frac{\beta_f \; c_H \; \rho_{LH}}{\gamma + \mu} = \frac{0.8 \times 6 \times 1}{0.58 + 0.114} = 6.92$$

$$R_{(0)LL} = \frac{\beta_f \; c_L \; \rho_{LL}}{\gamma + \mu} = \frac{0.8 \times 1.1 \times 0}{0.58 + 0.114} = 0$$

The Next Generation Matrix for the female population, under the assumption of Disassortative mixing is given by:

$$\begin{array}{cc} & \begin{array}{cc} H & L \end{array} \\ Male \begin{array}{c} H \\ L \end{array} & \begin{pmatrix} 0 & 6.92 \\ 1.27 & 0 \end{pmatrix} \end{array}$$

A value for R_0 in the female population can be found from the determinant of the following matrix:

$$\begin{array}{cc} & \begin{array}{cc} H & L \end{array} \\ Male \begin{array}{c} H \\ L \end{array} & \begin{pmatrix} 0 - R_0 & 6.92 \\ 1.27 & 0 - R_0 \end{pmatrix} \end{array}$$

The determinant of this matrix is given below, and its eigenvalue is our parameter of interest, R_0.

$$(0 - R_0)(0 - R_0) - (6.92)(1.27) = 0$$

Multiplying out the brackets gives the following equation:

$$R_0^2 - 8.79 = 0$$

Solving for R_0,

$$R_0 = \pm\sqrt{8.79}$$

$$R_0 = 2.96$$

A4. R_0 for the Male Population, Assuming Proportionate Mixing

Males:

$$R_{(0)HH} = \frac{\beta_m \, c_H \, \rho_{HH}}{\gamma + \mu} = \frac{0.7 \times 6 \times 0.48}{0.58 + 0.114} = 2.90$$
$$R_{(0)HL} = \frac{\beta_m \, c_L \, \rho_{HL}}{\gamma + \mu} = \frac{0.7 \times 1.1 \times 0.48}{0.58 + 0.114} = 0.53$$
$$R_{(0)LH} = \frac{\beta_m \, c_H \, \rho_{LH}}{\gamma + \mu} = \frac{0.7 \times 6 \times 0.52}{0.58 + 0.114} = 3.15$$
$$R_{(0)LL} = \frac{\beta_m \, c_L \, \rho_{LL}}{\gamma + \mu} = \frac{0.7 \times 1.1 \times 0.52}{0.58 + 0.114} = 0.58$$

The Next Generation Matrix for the male population, under the assumption of assortative mixing is given by:

$$\begin{array}{c} \\ Male \end{array} \begin{array}{c} \\ H \\ L \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{pmatrix} 2.90 & 0.53 \\ 3.15 & 0.58 \end{pmatrix} \end{array}$$

A value for R_0 in the male population can be found from the determinant of the following matrix:

$$\begin{array}{c} \\ Male \end{array} \begin{array}{c} \\ H \\ L \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{pmatrix} 2.90 - R_0 & 0.53 \\ 3.15 & 0.58 - R_0 \end{pmatrix} \end{array}$$

The determinant of this matrix is given below, and its eigenvalue is our parameter of interest, R_0.

$$(2.90 - R_0)(0.58 - R_0) - (0.53)(3.15) = 0$$

Multiplying out the brackets gives the following quadratic equation:

$$R_0^2 - 3.48R_0 + 0.01 = 0$$

Solving for R_0,

$$R_0 = \frac{3.48 \pm \sqrt{3.48^2 - 4(1)(0.01)}}{2(1)}$$
$$R_0 = 3.48 \; or \; 0.01$$

A5. R_0 for the Female Population, Assuming Proportionate Mixing

Females:

$$R_{(0)HH} = \frac{\beta_f \; c_H \; \rho_{HH}}{\gamma + \mu} = \frac{0.8 \times 6 \times 0.24}{0.58 + 0.114} = 1.66$$
$$R_{(0)HL} = \frac{\beta_f \; c_L \; \rho_{HL}}{\gamma + \mu} = \frac{0.8 \times 1.1 \times 0.24}{0.58 + 0.114} = 0.30$$
$$R_{(0)LH} = \frac{\beta_f \; c_H \; \rho_{LH}}{\gamma + \mu} = \frac{0.8 \times 6 \times 0.76}{0.58 + 0.114} = 5.26$$
$$R_{(0)LL} = \frac{\beta_f \; c_L \; \rho_{LL}}{\gamma + \mu} = \frac{0.8 \times 1.1 \times 0.76}{0.58 + 0.114} = 0.96$$

The Next Generation Matrix for the female population, under the assumption of Disassortative mixing is given by:

$$\begin{array}{c} \\ Male \end{array} \begin{array}{c} \\ H \\ L \end{array} \begin{array}{c} \begin{array}{cc} H & L \end{array} \\ \begin{pmatrix} 1.66 & 0.30 \\ 5.26 & 0.96 \end{pmatrix} \end{array}$$

A value for R_0 in the female population can be found from the determinant of the following matrix:

$$\begin{array}{c} \\ Male \end{array} \begin{array}{c} \\ H \\ L \end{array} \begin{array}{c} \begin{array}{cc} H & \quad L \end{array} \\ \begin{pmatrix} 1.66 - R_0 & 0.30 \\ 5.26 & 0.96 - R_0 \end{pmatrix} \end{array}$$

The determinant of this matrix is given below, and its eigenvalue is our parameter of interest, R_0.

$$(1.66 - R_0)(0.96 - R_0) - (0.30)(5.26) = 0$$

Multiplying out the brackets gives the following quadratic equation:

$$R_0^2 - 2.62R_0 + 0.01 = 0$$

Solving for R_0,

$$R_0 = \frac{2.62 \pm \sqrt{2.62^2 - 4(1)(0.01)}}{2(1)}$$

$$R_0 = 2.62 \; or \; 0.01$$

Bibliography

Ahrens W. & Pigeot I. (2004) *Handbook of Epidemiology.* 1^{st} edn. Springer-Verlag, New York, LLC.

American Cancer Society (2011) *Global Cancer Facts & Figures* 2^{nd} edn. Atlanta: American Cancer Society.

Anderson RM. & May RM. (1991) *Infectious Diseases of Humans: Dynamics and Control.* Oxford University Press, Oxford.

Australian Institute of Health and Welfare (2008) *Cervical screening in Australia 2005-2006.* Cancer series no. 41. Cat. no. CAN 36. Canberra: AIHW.

Bailey NTJ. (1975) *The Mathematical Theory of Infectious Diseases* 2^{nd} edn. Hafner, New York.

Barnabas RV., Laukkanen P., Koskela P., Kontula O., Lehtinen M. & Garnett GP. (2007) Epidemiology of HPV 16 and cervical cancer in Finland and the potential impact of vaccination: mathematical modelling analyses. *PLoS medicine* **3**(5) 624.

Behera A. & O'Rourke SF. (2008) The Effect of Corrlated Noise in a Gompertz Tumor Growth Model. *Brazilian Journal of Physics* **38** (2), 272-278.

Blower S. & Bernoulli D. (2004) An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. *Reviews in Medical Virology* **14**(5): 275-88.

Braun M. (1975) Differential Equations and their Applications, *Applied Mathematical Sciences*. vol 15, Springer-Verlag: New York.

Cancer Research UK (2008) *Cervical Cancer: UK Cervical Cancer Statistics. September 2008.* Retrieved from http://info.cancerresearchuk.org/cancerstats/types/cervix/incidence/ on 27 April 2009.

Castellsagué X. (2008) Natural history and epidemiology of HPV infection and cervical cancer. *Gynecologic Oncology* **110**(3) Supplement 2: S4-S7.

Castellsagué X. & Munoz N. (2003) Cofactors in human papillomavirus carcinogenesis - role of parity, oral contraceptives, and tobacco smoking. *Journal of the National Cancer Institute Monographs* **31**: 20-28.

Centers for Disease Control and Prevention (CDC) (2011a) *Sexually Transmitted Diseases (STDs). Genital HPV Infection - Fact Sheet.* Retrieved from http://www.cdc.gov/std/HPV/STDFact-HPV.htm on 31 January 2012.

Centers for Disease Control and Prevention (CDC) (2011b) *Sexually Transmitted Diseases (STDs). HPV and Men - Fact Sheet.* Retrieved from http://www.cdc.gov/std/HPV/STDFact-HPV-and-men.htm on 31 January 2012.

Centers for Disease Control and Prevention (CDC) (2011c) *Vaccine Safety. Reports of Health Concerns Following HPV Vaccination.* Retrieved from http://www.cdc.gov/vaccinesafety/vaccines/hpv/gardasil.html on 31 January 2012.

Centers for Disease Control and Prevention (CDC) (2007) *Human Papillomavirus: HPV Information for Clinicians.* Retrieved from http://www.cdc.gov/std/hpv/hpv-clinicians-brochure.htm on 31 January 2012.

Central Statistics Office (2006) *Census of Population Interactive Tables - Persons, males and females classified by single year of age, 2006.* Retrieved from http://census.cso.ie/Census/TableViewer/tableView.aspx?ReportId=77121 on 5 March 2012.

Dicker R., Coronado F., Koo D. & Gibson Parrish R. (2006) *Principles of Epidemiology in Public Health Practice* 3 edn. CDC 2006.

Doll R & Hill AB. (1950) Smoking and carcinoma of the lung; preliminary report. *British Medical Journal* ii:739-48

Doll R & Hill AB. (1954) The mortality of doctors in relation to their smoking habits. A preliminary report. *British Medical Journal* i: 1451-1455.

Doll R & Hill AB. (1956) Lung cancer and other causes of death in relation to smoking. A second report on the mortality of British doctors. *British Medical Journal* ii:1071-1076.

Elbasha E., Dasbach E. & Guris D (2010) P-828 The basic reproduction number for HPV 6, 11, 16, and 18 infections. *26th International Papillomavirus Conference & Clinical and Public Health Workshops.* Montreal Canada

Elbasha EH. & Galvani AP. (2005) Vaccination against multiple HPV types. *Mathematical Biosciences* **197**(1), 88-117.

Encyclopdia Britannica (2009) *Virus.* Retrieved from Encyclopdia Britannica Online: http://www.britannica.com/EBchecked/topic/630244/virus on 31 August 2009.

European Cervical Cancer Association (ECCA) (2009) *HPV Vaccination Across Europe.* Retrieved from http://www.ecca.info/fileadmin/user_upload/HPV_Vaccination/ECCA_HPV_Vaccination _April_2009.pdf on 20 February 2012.

Ferlay J., Shin HR., Bray F., Forman D., Mathers CD., Parkin D. (2010) *GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC Cancer- Base No.10* Lyon, France: International Agency for Research on Cancer. Retrieved from http://globocan.iarc.fr on 14 March 2012.

Fraumeni JF., Lloyd JW., Smith EM. & Wagoner JK. (1969) Cancer mortality among nuns: role of marital status in aetiology of neoplastic disease in women. *Journal of the National Cancer Institute* **42**: 455-468.

Gagnon F (1950) Contribution to the study of the aetiology and prevention of cancer of the cervix of the uterus. *American Journal of Obstetrics and Gynecology* **60**: 516-522.

Garnett GP. (2002) The geographical and temporal evolution of sexually transmitted disease epidemics. *Sexually Transmitted Infections* **78** Supplement 1 i14 - i19.

Garnett GP. & Anderson RM (1996) Sexually Transmitted Diseases and Sexual Behavior: Insights from Mathematical Models. *The Journal of Infectious Diseases* **174** Supplement 2l :S150-61.

Garnett GP. & Anderson RM (1993) Contact Tracing and the Estimation of Sexual Mixing Patterns: The Epidemiology of Gonococcal Infections. *Sexually Transmitted Diseases* **20**(4):181-91.

Garnett GP., Kim JJ., French K. & Goldie SJ. (2006) Chapter 21: Modelling the

impact of HPV vaccines on cervical cancer and screening programmes. *Vaccine* **24** S3/178 - S3/186.

Goldie SJ, Grima D, Kohli M, Wright TC., Weinstein M. & Franco E. (2003) A comprehensive natural history model of human papillomavirus (HPV) infection and cervical cancer: potential impact of and HPV 16/18 vaccine. *International Journal of Cancer* **106**:896-904.

Goldie SJ, Kohli M, Grima D, Weinstein MC., Wright TC., Bosch FX. & Franco E. (2004) Projected clinical benefits and cost-effectiveness of a human papllomavirus 16/18 vaccine. *Journal of the National Cancer Institute* **96**:604-15.

Greer CE., Wheeler CM., Ladner MB., Beutner K., Coyne MY., Liang H., Langenberg A., Benedict Yen TS. & Ralston R (1995) Human papillomavirus (HPV) type distribution and serological response to HPV type 6 virus-like particles in patients with genital warts. *Journal of Clinical Microbiology* **33**(8): 2058-63.

Griffiths M. (1991) *'Nuns, virgins, and spinsters'. Rigoni-Stern and cervical cancer revisited*. British Journal of Obstetrics and Gynaecology **98**(8): 797-802.

Health Protection Agency (2012) *Annual HPV vaccine coverage in England in 2010/2011. Routine programme for school year 8 females (12 to 13 years old)* Department of Health. Retrieved from https://www.wp.dh.gov.uk/immunisation/files/2012/03/120319_HPV_UptakeReport2010-11_acc.pdf on 27 March 2012.

Health Service Executive (2012a) *HPV Cervical Cancer Vaccination Programme* Health Service Executive. Retrieved from http://www.hse.ie/eng/services/News/newsarchive/2012archive/jan2012/hpvcervicalcancervaccination.html on 17 July 2012.

Health Service Executive (2012b) *Human Papillomavirus Programme* Health Service

Executive. Retrieved from http://www.immunisation.ie/en/HPVProgramme/ on 5 March 2012.

Health Service Executive (2011) *HUMAN PAPILLOMAVIRUS VACCINE Frequently Asked Questions for Health Professionals.* Produced by the National Immunisation Office in collaboration with the Health Protection Surveillance Centre. Version 8 - October 2011. Retrieved from http://www.immunisation.ie/en/Downloads/HPV/PDFFile_16845_en.pdf on 27 March 2012.

Health Service Executive (2009) *Epi-Insight. Disease Surveillance Report of HPSC, Ireland.* **10**(4) Retrieved from http://www.hpsc.ie/hpsc/EPI-Insight/Volume102009/File,3543,en.pdf on 15 March 2012.

Hethcote, HW. (1989) Three basic epidemiological models. Applied Mathematical Ecology (Levin SA., Hallam TG. & Gross LJ. eds.), pp. 119-144. Berlin: Springer-Verlag

Hethcote HW. (2000) The Mathematics of Infectious Diseases. *Society for Industrial and Applied Mathematics* **42**(4), 599-653.

Hethcote HW. & Yorke JA. (1984) Gonorrhea Transmission Dynamics and Control. *Lecture notes in biomathematics.* **56**, Springer-Verlag, Berlin.

Ho GY., Bierman R., Beardsley L., Chang CH. & Burk RD. (1998) Natural history of cervicovaginal papillomavirus infection as measured by repeated DNA testing in adolescent and young women. *The New England Journal of Medicine* **338**(7): 423-428.

Hughes JP., Garnett GP. & Koutsky L. (2002) The theoretical population-level impact of a prophylactic human papilloma virus vaccine. *Epidemiology* **13**(6) 631-639

Insinga RP., Perez G., Wheeler CM. Koutsky LA., Garland SM., Leodolter S, Joura EA., Ferris DG., Steben M., Brown DR., Elbasha EH., Paavonen J. & Haupt RM. (2010) Incidence, Duration, and Reappearance of Type-Specific Cervical Human Papillomavirus Infections in Young Women. *Cancer Epidemiology, Biomarkers and Prevention.* **19**: 1585-1594.

Jit, M., Chapman R., Hughes O. & Choi, Y.H. (2011) Comparing bivalent and quadrivalent human papillomavirus vaccines: economic evaluation based on transmission model. *British Medical Journal* **343**:d5775 doi: 10.1136/bmj.d5775.

Jit M., Choi YH. & Edmunds WJ. (2008) Economic evaluation of human papillomavirus vaccination in the United Kingdom. *British Medical Journal* **337**: (a769) (doi:10.1136/bmj.a769)

Keegan H, Ryan F, Malkin A, Griffin M, Lambkin H. (2007) Human papillomavirus prevalence and genotypes in an opportunistically screened Irish female population. *British Journal of Biomedical Science.* **64**: 18-22.

Kermack WO. & McKendrick AG. (1927) A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society London* **115**(772), 700-721.

Kitchener H, Almonte M, Wheeler P, Desai M., Gilham C., Bailey A., Sargent A. & Peto J. (2006) HPV testing in routine cervical screening: cross sectional data from the ARTISTIC trial. *British Journal of Cancer.* **95**: 56-61.

Koshiol JE., Laurent SA. & Pimenta JM. (2004) Rate and predictors of new genital warts claims and genital warts related healthcare utilization among privately insured patients in the United States. *Sexually Transmitted Diseases.* **31**: 748-752.

Kulasingam SL. & Myers ER. (2003) Potential health and economic impact of adding a human papillomavirus vaccine to screening programs. *Jama* **290**(6) 781.

Layte R, McGee H, Quail A, Rundle K., Cousins G., Donnelly C., Mulcahy F. & Conroy R. (2006) *The Irish Study of Sexual Health and Relationships.* Published by the Crisis Pregnancy Agency and the Dept. of Health and Children. October 2006.

Lilienfeld DE. (2007) Celebration: William Farr (18071883) an appreciation on the 200th anniversary of his birth. *International Journal of Epidemiology* **36**: 985987.

Lo CF. (2009) Stochastic Nonlinear Gompertz Model of Tumour Growth. *Proceedings of the World Congress on Engineering 2009* **II** July 1-3, London, UK.

Macey R., Oster G., & Zahnley T. (2000) *Berkeley Madonna User's Guide* Version 8, University of California. Retrieved from http://www.berkeleymadonna.com on 20 March 2012.

Mader S. (1998) *Biology.* 6^{th} edn. The McGraw-Hill Companies, Inc. USA.

McIntyre P. (2005) *Finding the viral link: the story of Harald zur Hausen.* Retrieved from http://www.cancerworld.org/CancerWorld/getStaticModFile.aspx?id=717 On 24th June 2009.

Meyers LA. (2007) Contact Network Epidemiology: Bond Percolation Applied to Infectious Disease Prediction and Control. *Bulletin of the American Mathematical Society* **44**, 1 Jan, 63-86.

Moody CA. & Laimins LA. (2010) Human papillomavirus oncoproteins: pathways to transformation. *Nature Reviews Cancer* **10**(8):550-60.

Murray JD. (2002) *Mathematical Biology I, An Introduction* 3^{rd} edn. Springer, Berlin.

Nahmias AJ., Josey WE., Naib ZM., Luce CF., & Guest BA. (1970) Antibodies to Herpesvirus hominis types 1 and 2 in humans. II. Women with cervical cancer. *American Journal of Epidemiology* **91** (6): 547-52.

National Cancer Registry 2010. *Cancer Registry Data* Retrieved from http://www.ncri.ie/data.cgi/choose-methods.php on 17 February 2012.

Page RM., Cole GE. & Timmreck TE. (1995) *Basic Epidemiological Methods and Biostatistics: A Practical Guide Book*. Jones and Bartlett Publishers, Boston, MA.

Porta, M. (2008) *Dictionary of epidemiology.* 5^{th} edn. New York: Oxford University Press. p. 81.

Rous P. (1911) Transmission of a malignant new growth by means of a cell-free filtrate (Brief Report). *JAMA* **56**: 198

Rous P. & Beard JW. (1935) The progression to carcinoma of virus-induced rabbit papillomas (Shope) *Journal of Experimental Medicine*. **62**: 523-554.

Ruska E., (1986) *The Development of the Electron Microscope and of Electron Microscopy. Nobel Lectures. Physics 1981-1990* (Frngsmyr T. & Ekspng G. eds), World Scientific Publishing Co., Singapore.

Sanders GD. & Taira AV. (2003) Cost effectiveness of a potential vaccine for human papillomavirus. *Emerging Infectious Diseases* **9**(1) 37-48.

Shope RE. (1933) Infectious papillomatosis of rabbits. *Journal of Experimental Medicine* **58**: 607-627.

Shope RE. (1931) Swine Influenza 1. Experimental Transmission and Pathology. *Journal of Experimental Medicine* **54**(3): 349-359.

Skrabanek P. (2000) *False Premise, False Promise: Selected Writings of Petr Skrabanek* (p 118-128).Whithorn: Tarragon Press

Snow J., (1855) *On the Mode of Communication of Cholera. London.* Retrieved from http://www.ph.ucla.edu/epi/snow/html on 28 July 2009.

Stannard LM. (1995) *Virus Ultra Structure Electron Micrograph Images.* Retrieved from http://web.uct.ac.za/depts/mmi/stannard/linda.html on 3 September 2009.

Taira AV., Neukermans, CP. & Sanders GD. (2004) Evaluating human papillomavirus vaccination programs. *Emerging Infectious Diseases* **10**(11), 1915-23

Thoms H. (1943) Diagnosis of Uterine Cancer by the Vaginal Smear. Reviewed by Herbert Thoms. *Yale Journal of Biology and Medicine* **15**(6): 924.

Towne JE. (1955) Carcinoma of the cervix in nulliparous and celibate women. *American Journal of Obstetrics and Gynecology* **69**: 606-613.

Usher C., Tilson L., Olsen J., Jepsen M., Walsh C., Barry M. (2008) Cost-effectiveness of human papillomavirus vaccine in reducing the risk of cervical cancer in Ireland due to HPV types 16 and 18 using a transmission dynamic model. *Vaccine* **26**(44), 5654-5661.

Vynnycky E., White RG. (2010) *An introduction to infectious disease modelling.* Oxford, New York: Oxford University Press.

WHO/ICO Information Centre on HPV and Cervical Cancer (HPV Information Centre) (2007). *Summary report on HPV and cervical cancer statistics in Spain.* Retrieved from www.who.int/hpvcentre on 24 April 2009.

Woodman CBJ., Collins SI. & Young LS. (2007). The natural history of cervical HPV infection: unresolved issues. *Nature Reviews - Cancer* **7**(1): 11-22.

World Health Organisation (2009) *Human Papillomaviruses.* Retrieved from http://www.who.int/vaccine_research/diseases/viral_cancers/en/index3.html. on 2 July 2009.

World Health Organisation (2007) *Human Papillomavirus and HPV Vaccines. Technical Information for Policy Makers and Health Professionals.* Switzerland: WHO.