COST-EFFECTIVENESS OF PRESSURE ULCER PREVENTION AND CALIBRATION ANALYSIS OF DECISION-ANALYTIC MODELS

by

Ba’ Dinh Pham

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy, Department of Health Policy Management and Evaluation, University of Toronto

© Copyright by Ba’ Dinh Pham, 2015
Cost-Effectiveness of Pressure Ulcer Prevention and Calibration Analysis of Decision-Analytic Models

Ba’ Dinh Pham

Doctor of Philosophy

Institute of Health Policy Management and Evaluation

University of Toronto

2015

Abstract

Pressure ulcers are common in the very ill, the elderly, and immobile or neurologically compromised individuals. In Ontario, the estimated prevalence is approximately 15% among hospital patients. Affected patients experience pain and reduced quality of life. The burden of pressure ulcers in Canada is estimated to exceed $1 billion annually, or approximately one-eighth the total health care cost of cardiovascular diseases in Canada.

Current guidelines recommend early prevention to reduce the burden of pressure ulcers but the economic evidence supporting these recommendations is limited. This thesis reports on studies evaluating the cost effectiveness of strategies to prevent pressure ulcers in targeted high-risk patients. In order to conduct the cost-effectiveness analysis, it was necessary to resolve calibration issues in the decision-analytic model that simulates the natural history of pressure ulcers. Model calibration is the practice of comparing projected data from a decision model with corresponding observed data in order to determine unknown model parameters.
A literature review was conducted to provide an inventory of methods for the calibration of decision models and categorize these methods according to their methodological approaches. The first study describes the development of a two-stage random search algorithm for model calibration. The second study reports on the cost-effectiveness analysis of strategies to prevent pressure ulcers in elderly patients seeking help in the emergency departments. The third study describes the cost-effectiveness analysis of strategies to prevent pressure ulcers originating intra-operatively in surgical patients.

Results from the first study show that the proposed two-stage random search algorithm is suitable for routine calibration of moderately complex decision models. Results from the cost-effectiveness studies show that the evaluated prevention strategies are more cost-effective than current practice. The economic evidence supports the implementation of these early prevention strategies to reduce the burden of pressure ulcers in high-risk patients.
Acknowledgments

I’d like to express my love and gratitude to my wife, Joanne Quan, who listened with remarkable patience and support as I worked through the twists and turns of my work and study. Were it not for her love, patience and care of our family, this work would not have been possible. My debt to her is immeasurable, as is my love. I wish to extend my gratitude and love to my daughters Megan and Nicole. They kept me company while I did my research and were a source of love and comfort while I studied.

I’d like to thank Dr. Murray Krahn for his guidance and support during my study. I am grateful for his mentorship and the opportunity to work along his side at the Toronto Health Economics and Technology Assessment Collaborative. I’d like to thank each committee member for their guidance and insights to my research. I’d like to thank Dr. George Tomlinson for his patience when it took me years to work through my research. I’d like to thank Dr. Paul Grootendorst for his encouragement, support and positive energy each and every time I approach him for help. I’d like to thank Dr. Karen Kuntz my external examiner for her careful review and comments to my thesis, and Dr. Allan Detsky my internal examiner for his questions during my defense.

I am eternally grateful for my mother who took care of me and my many brothers and sisters all by herself, alone and in very difficult time. I would forever remember the time I spent with her while writing part of this thesis. I’d like to thank my father for tutoring me when I was little. When he was alive, he expected me to excel and I am grateful for his expectation despite my forever regret about his long absence. I’d like to thank my brothers and sisters for their love and support. In particular, I’d like to thank them for their help with my departure from Vietnam and the opportunity to rebuild my life in Canada.

I’d like to thank Dr. David Moher who mentored me in my new life. I’d like to thank my friend Dr. Andrea Tricco who inspired me to my PhD. I’d like to thank my friends and colleagues at GlaxoSmithKline Canada and the Toronto Health Economics and Technology Assessment for their support and encouragement while I worked and studied.
Publications and manuscripts

Publications and manuscripts from this thesis


Submitted manuscripts


Other publication related to this thesis

# Table of Contents

Acknowledgments ........................................................................................................ iv
Publications and manuscripts .................................................................................. v
Table of Contents ....................................................................................................... vi
List of Tables .............................................................................................................. x
List of Figures ........................................................................................................... xi
Chapter 1 Introduction and Overview ..................................................................... 1
  1.1 Introduction ......................................................................................................... 1
  1.2 Overview and organization of the thesis ......................................................... 2
Chapter 2 Prevention of Pressure Ulcers ................................................................. 4
  2 Prevention of pressure ulcers ............................................................................. 4
    2.1 Pressure ulcer .................................................................................................. 4
    2.2 At-risk individuals .......................................................................................... 7
    2.3 Pressure ulcer classification .......................................................................... 8
    2.4 Prevalence and incidence .............................................................................. 10
    2.5 Health-related quality of life .......................................................................... 12
    2.6 Current practice ............................................................................................. 13
    2.7 Health care costs of pressure ulcers ............................................................. 15
    2.8 Cost-effectiveness of preventive interventions ............................................ 16
    2.9 Prevention policies ......................................................................................... 18
    2.10 Research questions and rationale ............................................................... 19
Chapter 3 Decision Modeling And Cost-Effectiveness Analysis ......................... 21
  3 Decision-analytic modeling and cost-effectiveness analysis .......................... 21
    3.1 Abstract ............................................................................................................ 21
# Table of Contents

3.2 Background .......................................................................................................................... 21
3.3 Decision analytic modeling ................................................................................................. 21
3.4 Cost-effectiveness analysis ................................................................................................. 23
3.5 Calibration analysis of decision models .............................................................................. 26
3.6 Rationale for a literature review of calibration methods ....................................................... 27

Chapter 4 Calibration Analysis of Decision-Analytic Models: A Literature Review ............. 29
4 Calibration analysis of decision models: A literature review .............................................. 29
   4.1 Abstract .............................................................................................................................. 29
   4.2 Introduction ...................................................................................................................... 30
   4.3 Data sources ..................................................................................................................... 31
   4.4 The calibration process .................................................................................................... 31
   4.5 Bayesian approach ......................................................................................................... 35
   4.6 Guided search approach ................................................................................................. 42
   4.7 Random search approach ............................................................................................... 46
   4.8 Grid search approach ...................................................................................................... 51
   4.9 Manual search ................................................................................................................. 53
   4.10 Discussion ..................................................................................................................... 53
   4.11 Methodological question and rationale ........................................................................... 57

Chapter 5 Calibration Analysis of Decision Analytic Models ............................................ 71
5 Calibration analysis of decision analytic models ................................................................. 71
   5.1 Abstract ........................................................................................................................... 71
   5.2 Introduction ...................................................................................................................... 72
   5.3 Methods .......................................................................................................................... 72
   5.4 Results ............................................................................................................................. 88
   5.5 Discussion ....................................................................................................................... 91
List of Tables

Table 1. Prevalence and incidence of pressure ulcers in Canada ......................................... 11
Table 2. Estimated treatment costs attributable to pressure ulcer care .................................. 16
Table 3. Calibration with Bayesian approach ........................................................................ 59
Table 4. Calibration with guided search approach ................................................................. 62
Table 5. Calibration with random search approach ............................................................... 64
Table 6. Calibration with grid search approach .................................................................... 68
Table 7. Suggested methodological issues relevant to the conduct of model calibration ......... 70
Table 8. Sampling designs for the illustration with the two-stage random search approach ....... 80
Table 9. Inputs to the pressure ulcer decision model ............................................................. 97
Table 10. Results of the two-stage random search approach to model calibration ................. 99
Table 11. Results of the illustrated cost-effectiveness analysis .............................................. 101
Table 12. Transition rate estimates (ranges) obtained via calibration analysis ...................... 112
Table 13. An illustration of the value of information calculation .......................................... 116
Table 14. Input data to the cost-effectiveness analysis of early prevention ............................ 131
Table 15. Results of the cost-effectiveness analysis of early prevention ............................... 135
Table 16. Randomized controlled trials of pressure-redistribution overlays ....................... 146
Table 17. Input data to the cost-effectiveness analysis of intra-operative prevention ............. 147
Table 18. Results of the cost-effectiveness analysis of intra-operative prevention ............... 150
List of Figures

Figure 1. Conceptual framework of pressure ulcer risk.................................................. 6

Figure 2. An illustration of Approximate Bayesian Computation.................................. 41

Figure 3. An illustrated Latin hypercube design............................................................. 46

Figure 4. Structure of a pressure ulcer decision model.................................................... 73

Figure 5. The two-stage random search algorithm .......................................................... 78

Figure 6. An illustration of Latin hypercube sampling design ......................................... 83

Figure 7. Ranges and weighted mean estimates of transition parameters ....................... 95

Figure 8. Probability density distributions of calibrated transition parameters ............... 96

Figure 9. Structure of a pressure ulcer decision model .................................................... 107

Figure 10. Observed and projected prevalence of stage-specific pressure ulcers ............ 111

Figure 11. Sensitivity analysis of the cost-effectiveness of early prevention .................... 130

Figure 12. Sensitivity analysis of the cost-effectiveness of intra-operative prevention ....... 145
Chapter 1
Introduction and Overview

1.1 Introduction

Pressure ulcers are common in the very ill, the elderly, and immobile or neurologically compromised individuals. In Ontario, the estimated prevalence ranges from 13% to 17% among hospital patients, approximately 9% among long-term care residents, and 2% among patients receiving home care in the community. Affected patients experience pain from pressure ulcers and their treatment. Also, pressure ulcers significantly and negatively affect the physical, social, psychological, and financial aspects of their life.

Hospital costs attributable to pressure ulcers in Canada are estimated to exceed $1 billion annually, or approximately one-eighth the estimated total health care cost of cardiovascular diseases in Canada. To reduce the incidence and burden of pressure ulcers, practice guidelines recommend risk assessment and individualized care plans for high-risk patients. In particular, the use of pressure-redistribution mattresses has been shown to significantly reduce the risk of developing pressure ulcers in hospital patients. Preliminary results also indicate that although pressure-redistribution mattresses are more expensive than standard hospital mattresses, the health improvement and the treatment saving from averted pressure ulcers more than offsets the additional costs. This is likely the case if prevention is targeted at high risk patients.

Among at-risk groups, hospital patients are at high risk of developing pressure ulcers. In particular, certain patients are at high risk for a well defined duration of exposure, making them ideal for targeted interventions.

This thesis reports on studies evaluating the cost effectiveness of strategies to prevent pressure ulcers in targeted high-risk populations. In order to conduct the cost-effectiveness analysis, it was necessary to first conduct a literature review and a methodological study to resolve calibration issues in the decision-analytic model that simulates the natural history of pressure ulcers. Model calibration can be defined as the practice of comparing projected data from a decision model with corresponding observed data in order to estimate unknown model parameters. The calibrated model was then used to synthesize data from multiple sources and
project outcomes relevant to the cost-effectiveness analysis of interventions to prevent pressure ulcers.

In total, this thesis comprises three studies. The first study describes the development of a random search algorithm for model calibration. The second study reports on the cost-effectiveness analysis of strategies to prevent pressure ulcers in elderly patients seeking help in the emergency departments. The third study describes the cost-effectiveness analysis of strategies to prevent pressure ulcers originating intra-operatively in surgical patients.

1.2 Overview and organization of the thesis

Chapter 2 provides an introduction to pressure ulcers and their burden on patients and health systems. It reviews the clinical and economic evidence of preventive interventions and their target populations. The chapter ends with an introduction to the research questions and the rationale for the cost-effectiveness analysis of a strategy to prevent pressure ulcers in elderly patients seeking care in emergency departments (chapter 7) and the rationale for the cost-effectiveness of a strategy to prevent intra-operative pressure ulcers in surgical patients (chapter 8).

Chapter 3 provides an overview of decision analytic modeling and cost-effectiveness analysis in the context of pressure ulcer prevention. To simulate the prognosis of pressure ulcers, it is necessary to calibrate the decision model to observed data in order to estimate unknown model parameters. The chapter ends with the rationale for the conduct of a literature review of methods for the calibration of decision models (chapter 4).

Chapter 4 is a literature review of calibration analysis of decision-analytic models. It reports on an inventory of existing methods for model calibration. The fundamental concepts underlying the calibration methods as well as the specific details of their applications are summarized. Perusing findings from this review, the chapter identifies the need to develop calibration methods that are suitable for routine calibration of decision models for cost-effectiveness analysis (chapter 5).

Chapter 5 is a study manuscript entitled, “Calibration analysis of disease history models: a two-stage random search algorithm.” The study describes a random search algorithm for the
calibration of decision models, and assesses its feasibility. The algorithm structures the search in stages and provides operational details for each stage.

Chapter 6 describes methods for the cost-effectiveness analyses of early prevention, including the development of a decision model to simulate the prognosis of pressure ulcers in high-risk populations. It also discusses the use of the two-stage random search algorithm developed in chapter 5 to calibrate the decision model to different high-risk populations.

Chapter 7 is a study manuscript entitled, “Support surfaces for early prevention of pressure ulcers among elderly patients admitted through emergency departments: A cost-effectiveness analysis.” The study evaluates the cost-effectiveness of pressure-redistribution foam mattresses on emergency department (ED) stretchers and ED beds for early prevention of pressure ulcers in elderly admitted ED patients.

Chapter 8 is a study manuscript entitled, “Support surfaces for intra-operative prevention of pressure ulcers in surgical patients: A cost-effectiveness analysis.” The study evaluates the cost-effectiveness of a prevention program with pressure-redistribution overlays on operating tables to prevent pressure ulcers originated intra-operatively in surgical patients.

Chapter 9 provides summary of study findings, discusses their implications to research and practice, and a reflection on the work in this thesis.
Chapter 2
Prevention of Pressure Ulcers

2 Prevention of pressure ulcers

2.1 Pressure ulcer

2.1.1 Brief history

A pressure ulcer is a localized injury to the skin or underlying tissue usually over a bony prominence, as a result of a force exerted perpendicular to the tissue – pressure – or a force exerted parallel to the tissue – shearing.\(^1\) Given their first description in Egyptian mummies, mankind has been dealing with pressure ulcers for a long time.\(^2\) In 1593, the surgeon Fabricius Hildanus described for the first time in the Netherlands the clinical characteristics of pressure sores.\(^3\) He identified as causes external natural factors and internal supernatural factors, as well as the interruption in the supply of “pneuma”, blood and nutrients. In 1722, the French surgeon de la Motte noticed that mechanical pressure and incontinence played a part in the development of pressure sores.\(^4\) Since the research performed by Groth in 1942,\(^5\) Husain in 1953,\(^6\) and Kosiak in 1959 and 1961,\(^7,\,8\) the importance of continuous or intermittent pressure as a cause of pressure sores has been generally accepted.\(^9\) In 1958, Reichel suggested that shearing is also a contributing risk factor,\(^10\) but the importance of shearing was not acknowledged until the 1970s.\(^9\)

In 1970, Lowthian introduced a conceptual scheme, a theoretical basis for pressure ulcer risk assessment, including factors for general physical condition, mental state, activity, mobility and incontinence.\(^11\) Using this conceptual scheme, he explored possible relationships between risk assessment and pressure ulcer development.\(^11\) Improving insights into the pathophysiology and risk factors led to the acceptance of a new concept. The tissue tolerance concept consists of a series of risk factors which are known to influence the pressure ulcer risk of an individual but operate independently of the intensity and duration of pressure or shearing.\(^9\)

2.1.2 Conceptual framework

Figure 1 displays a conceptual framework for pressure ulcer risk.\(^9,\,12,\,13\) As a causal factor, the existence of pressure, alone or in combination with shearing, is needed.\(^9,\,12\) Pressure is the primary factor, since without pressure, there can be no shear. Shearing forces cannot be
considered separately from pressure, but they increase the harm of a given pressure because stagnation of blood flow through kinking blood vessels would occur at a relatively low pressure intensity.

Both the intensity and the duration of pressure seem to play a role in the development of pressure ulcers. The intensity of pressure is determined to a large extent by the hardness of the support surfaces. A pressure induced by body weight at bony prominences higher than the capillary pressure would slow down the flow in the capillaries and lymph nodes, resulting in insufficient supply of oxygen and nutrients and insufficient evacuation of metabolic waste. Capillary pressure typically ranges between 32 and 47 mmHg; higher values may be associated with increasing risk of pressure ulcer development.

The duration of pressure is influenced by the level to which a patient perceives pain (due to oxygen shortage at tissue level), and by the timing to which they are able to react. Patients with arthritis, multiple sclerosis, spinal cord injury, pain, and reduced consciousness are less able to change position, because they are insufficiently aware of the pain signals, incapacitated, or restricted in their movement to reduce the associated pain. According to most practice guidelines, an immobile patient should be repositioned every two hours to prevent continuous pressure on bony prominences and assist in decreasing the risk of adverse physiological responses.

High pressure or shearing forces applied for a sufficiently long time will increase pressure ulcer risk (Figure 1). How high the pressure must be and how long it must be maintained to cause damage depends on an individual’s tissue tolerance. This may vary between subjects, depending on factors that change the capacity of the tissue to redistribute pressure – the tissue tolerance for pressure – and factors that influence oxygen distribution within the tissue and the oxygen need of the tissue – the tissue tolerance for oxygen.

The tissue tolerance for pressure decreases with increasing age. The elderly have less subcutaneous tissue and decreased muscle tone. The slow genesis of skin cells in the elderly may also contribute to tissue intolerance for pressure. In the elderly, dehydration due to insufficient liquid intake decreases skin elasticity and increases the capacity for deformation of the tissue, thus increasing the risk of tissue damage.
As long as the oxygen supply to the tissue matches its needs, pressure ulcers do not occur. If oxygen supply decreases or oxygen needs increases, the risk of pressure sores increases. The oxygen requirement of cells can increase with fever; both increased and decreased temperature differences within plus or minus 1.0 degree F can be used to indicate reactive hyperemia or a stage-1 pressure ulcer, but a tissue integrity problem may still exist despite the absence of a temperature difference. On that account, an association was reported for heightened body temperature and the development of pressure sores. Externally applied heat increases the metabolic rate of the tissue, raises the need of tissues for oxygen and nutrients, and consequently, increases pressure ulcer risk.

According to the conceptual framework in Figure 1, interventions to prevent pressure ulcer can be directed at pressure, at shearing or at tissue tolerance. Interventions that influence only tissue tolerance can reduce the risk of pressure ulcers. It is however questionable whether they can prevent pressure ulcers in high-risk patients. The most efficient measures directly influence the causes of pressure ulcers. For example, special mattresses that redistribute the intensity of the interface pressure between the support surface and bony prominences decrease tissue-
interface pressure and therefore can prevent the development of pressure ulcers in high-risk patients.29

2.2 At-risk individuals

Pressure ulcers are common in the very ill, the elderly, and immobile or neurologically compromised individuals.30,31 There are many possible contributing factors to pressure ulcer development; but the significance of all these factors has yet to be elucidated.1,32

2.2.1 Intrinsic risk factors

An individual's potential to develop pressure ulcers may be influenced by the following intrinsic risk factors: reduced mobility or immobility, sensory impairment, acute illness, reduced level of consciousness, extremes of age, previous history of pressure damage, severe chronic or terminal illness, and malnutrition.33 Other physiologic risk factors include diabetes mellitus, peripheral vascular disease, cerebrovascular disease, sepsis, and hypotension.32 These physiologic risk factors may place the patients at risk of developing pressure ulcers because they impair the microcirculatory system.32 Microcirculation is controlled in part by sympathetic vasoconstrictor impulses from the brain and secretions from localized endothelial cells. Patients whose neural and endothelial control of blood flow are impaired during an illness may be more susceptible to ischemic organ damage such as pressure ulcers.32

2.2.2 Extrinsic risk factors

Extrinsic risk factors include pressure, shearing, friction, and excessive moisture on the skin.33 Pressure is a perpendicular force that results in the compression of tissues between a bony prominence and a support surface (e.g., a bed or chair).32 At-risk bony prominences include the occiput, shoulders, scapulae, spine, elbows, iliac crest, greater trochanters, ischial tuberosities, knees, sacrum, coccyx, malleoli, lateral edges of the feet, and the heels.34

Shearing forces are produced when adjacent surfaces slide across one another. For example, shear is exerted on the body when the head of the bed is elevated, especially over 30 degrees. In this position, the skin and superficial fascia remain fixed against the bed linens, while the deep fascia and skeleton slide down. The result of these movements can cause stretching, pulling, and change to the angle of blood vessels, resulting in tissue ischemia.34
Friction created by repeated movements of the skin over support surfaces may result in loss of the superficial skin. The outer layers of skin may become macerated, denuded or broken through prolonged exposure to moisture, particularly moisture related to urinary and fecal incontinence, wound drainage, and perspiration.\textsuperscript{34}

2.2.3 Risk assessment tools

Due to the large number of risk factors for pressure ulcer development, risk assessment tools are widely used, most notably the Norton and Braden scales.\textsuperscript{19} The Norton scale includes five broad clinical categories: physical condition, mental state, activity, mobility, and incontinence. Each category is rated on an ordinal scale ranging from 1 to 4. The total score ranges from 5 to 20, and a score $\leq 16$ indicates high risk.\textsuperscript{19} The Braden scale includes six broad clinical categories: sensory perception, moisture, activity, mobility, nutrition, and friction and shear. Each of the first five categories is evaluated using an ordinal scale ranging from 1 to 4, and the friction and shear category is rated from 1 to 3. The total score ranges from 6 to 23, and a score $\leq 18$ indicates high risk.\textsuperscript{32} It has been reported that at the recommended cut-off values to identify individuals at high risk of developing pressure ulcer, both scales tend to over predict the number of high-risk patients.\textsuperscript{32}

The effectiveness of risk assessment scales is uncertain.\textsuperscript{35} A large trial randomized hospital patients to be assessed with two less-commonly-used scales, the Waterlow scale ($n=410$ patients), the Ramstadius scale ($n=411$), or by clinical judgment ($n=410$). There were no differences among the randomized groups with respect to pre-specified processes of care, including the use of special mattresses, documentation of an explicit pressure care plan, referral to skin care nurses or referral to dieticians. Overall, similar incidence of hospital-acquired pressure ulcers was reported: 6.8\% with clinical judgment, 7.5\% with Waterlow, and 5.4\% with Ramstadius ($p=0.44$). It was concluded that nursing time used to administer these scales might be better spent on daily skin inspection, and on assessing and managing specific risk factors.\textsuperscript{36}

2.3 Pressure ulcer classification

A pressure ulcer classification describes a series of numbered grades or stages, each determining a different degree of tissue damage.\textsuperscript{37}
2.3.1 Common classification systems

Many classification systems have been used over time,\textsuperscript{38, 39} contributing to difficulties in interpreting results across studies.\textsuperscript{40}

According to the joint working group between the U.S. National Pressure Ulcer Advisory Panel and the European Pressure Ulcer Advisory Panel, pressure ulcers are classified as stage 1 (a persistent skin redness), stage 2 (a loss of partial skin thickness that appears as an abrasion, blister, or shallow crater), stage 3 (a loss of full skin thickness, presented as a deep crater), and stage 4 (a loss of full skin thickness, exposing muscle or bone).\textsuperscript{1} The classification system by the U.S. National Pressure Ulcer Advisory Panel includes deep tissue injury (a pressure-related injury to subcutaneous tissues under intact skin) that may herald the subsequent development of a stage 3 or 4 pressure ulcer even with optimal management.\textsuperscript{39}

According to the U.S. National Pressure Ulcer Advisory Panel, a pressure ulcer that presents with an eschar (i.e., significant necrotic tissue obscuring the base of the wound) cannot be staged until the necrotic tissue has been removed.\textsuperscript{37, 39} Partial-thickness skin losses secondary to moisture or incontinence may be described as stage 2 pressure ulcers.\textsuperscript{34} During the healing of a stage 3 or 4 pressure ulcer, the tissue lost is not regenerated, but filled with granulation tissue and scar. It is therefore physiologically incorrect to use the staging system in reverse order to indicate healing.\textsuperscript{41}

2.3.2 Reliability of common classification systems

The inter-rater reliability of pressure ulcer classification systems, most notably the European Pressure Ulcer Advisory Panel and the U.S. National Pressure Ulcer Advisory Panel, has recently been examined in a systematic review.\textsuperscript{42} The systematic review reports that a meaningful comparison between the classification systems was impossible because of the heterogeneity in the conduct and reporting of included studies. In conclusion, the review suggests that there is not enough evidence to recommend a specific pressure ulcer classification system for use in daily practice.
2.4 Prevalence and incidence

2.4.1 Definitions

The method used most commonly to indicate prevalence is ‘point prevalence’. The point prevalence of pressure ulcer is the proportion of the population with pressure ulcer at a particular point in time. It is calculated by dividing the number of individuals with pressure ulcer at a particular point in time by the total number of individuals in the population. "Period prevalence" is the proportion of a population that has pressure ulcer at some time during a given period (e.g., "monthly prevalence"), and includes individuals who already have pressure ulcer at the start of the study period as well as those who acquire it during that period.

Among those who are at risk of developing pressure ulcer, incidence is defined as the proportion of those who develop pressure ulcer in a particular time period. Facility-acquired incidence is usually expressed as the proportion of patients who did not have a pressure ulcer on admission who subsequently developed one or more pressure ulcer during their length of stay.

Reported prevalence and incidence estimates can vary substantially across studies due to different definitions, data sources, characteristics of the study population, care settings, pressure ulcer classifications (e.g., whether deep tissue injury was included), and existing prevention protocols.

2.4.2 Prevalence and incidence of pressure ulcers in Canada

There are no national or provincial surveys of pressure ulcer prevalence and incidence in Canada. A systematic review of Canadian prevalence studies identified 7 published studies. It also included 33 unpublished studies that were provided by members of the Canadian Association for Wound Care and manufacturers of support surfaces that are designed to prevent pressure ulcers. The pooled prevalence estimate was 25% (range: 5-35%) in study participants from acute care hospitals, 30% (6-53%) in study participants from non-acute care settings (i.e., sub-acute care, chronic care, complex continuing care, long-term care and nursing homes), and 15% (13-24%) in study participants from community care.

Using selected data sources that are deemed to be relatively representative of populations across Canadian health care settings, Table 1 displays relatively lower prevalence and incidence
estimates. Estimates for acute care hospitals were derived from a study of 12,787 inpatients during a 1-day annual census conducted in an acute care hospital in Ottawa from 1994 to 2008.\textsuperscript{45} In the study, a full head-to-toe skin and risk assessment was conducted for every patient in the hospital between 6:00am and 6:00pm, and all patients admitted during the 12 hours of the survey.

Estimates for long-term care were derived from a cohort consisting of all residents (n=18,325 residents from a total of approximately 72,000 residents) from 89 long-term care facilities in Ontario (a total of 613 homes) that contributed data to the implementation phase of the Resident Assessment Instrument - Minimum Data Set (RAI-MDS) from May 2004 to November 2007. Residents were included if they had a full assessment (e.g., physical functioning, cognition, nutritional status, skin health, and incontinence) and at least one reassessment (e.g., every three months or when health status changed) in the RAI-MDS.\textsuperscript{46}

Table 1. Prevalence and incidence of pressure ulcers in Canada

<table>
<thead>
<tr>
<th></th>
<th># facilities; # pts$</th>
<th>Time period</th>
<th>Prevalence</th>
<th>Incidence</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute care facilities</td>
<td>1; 12,787</td>
<td>1994-2008</td>
<td>13% - 17%</td>
<td>7% - 11%*</td>
<td>45</td>
</tr>
<tr>
<td>Long-term care facilities</td>
<td>89; 18,325</td>
<td>2004-2007</td>
<td>9%</td>
<td>2.6% over 3 mths</td>
<td>47</td>
</tr>
<tr>
<td>Home care clients\textsuperscript{†}</td>
<td>12 CCACs; 77,381</td>
<td>2005-2008</td>
<td>2%</td>
<td>1.1% over 6 mths</td>
<td>RAI-MDS\textsuperscript{‡}</td>
</tr>
</tbody>
</table>

Notes: CCAC: Community Care Access Centres. RAI-MDS: Resident Assessment Instrument – Minimum Data Set.

$Patients or participants. $Incidence was defined as the number of patients with a pressure ulcer on the day of the annual census of all patients who were ulcer free on the day of admission. \textsuperscript{†}Long-stay clients only, including clients who were on services with Community Care Access Centres for ≥ 60 days. \textsuperscript{‡}Data were provided to the Toronto Health Economics and Technology Assessment Collaborative from the Resident Assessment Instrument – Minimum Data Set, Home Care version (personal communication, Dr. Jeff Poss, University of Waterloo, December 5, 2009).

Estimates for home care were derived from all clients (n=77,381) who received home care services for ≥60 days with all 14 Community Care Access Centres in Ontario that contributed data to the RAI-MDS – Home Care version from April 1, 2005 – March 31, 2008. Clients were assessed upon admission and every six months or when their health status changed. The average follow-up duration in this dataset was 52 weeks.
In Canada, pressure ulcers are common across a variety of care settings, with estimated prevalence ranging from 13%-17% among patients in acute care hospitals, to 9% among long-term care residents, and 2% among long-stay home care clients (Table 1). From 7% to 11% of patients acquire pressure ulcers during their stay in hospitals. The estimated incidence is 2.6% over 3 months among long-term care residents, and 1.1% over 6 months among clients who receive home care services for ≥60 days. The accuracy of prevalence and incidence estimates, however, is closely related to the identification and classification of pressure ulcers, and the quality of the collecting and recording data. In particular, skin inspections are not always thoroughly conducted upon a patient’s admission to a health care facility, and there is generally a lack of documentation of pre-admission skin damage.

2.5 Health-related quality of life

2.5.1 Pain

Pressure ulcers can cause patients considerable pain, discomfort, and suffering. Particularly painful times are dressing changes and other wound care. According to the results of a systematic review that includes studies of patient-reported pain associated with pressure ulcers (n=10 studies, 108 patients), patients described their associated pain using the following common sensory words: tender (68% of 107 patients), sharp (61%), throbbing (54%), aching (54%), and hot burning (42%). The associated pain can be debilitating, can impede physical and social activities, and can compromise psychological well-being. Patients have to manage their pain and its impact on daily life. Self-management behaviors are mediated by individual factors such as mood, coping, and motivation.

2.5.2 Health-related quality of life

A systematic review identified 31 studies with a total of 2,463 participants who described the impact of pressure ulcers on their health-related quality of life. The review included 10 qualitative and 21 quantitative studies. It used a method to synthesize results and findings from these qualitative and quantitative studies. It reported that 11 health-related quality of life themes were identified through the synthesis. The themes were physical impact, social impact, psychological effect, pressure ulcer symptoms, general health, and other impacts of pressure ulcers: healthcare professional–client relationships, need for versus effect of interventions,
impact on others, financial impact, perceived etiology, and need for knowledge. In particular, eleven studies reported pressure ulcers as having a significantly impact on physical aspects of patients’ health-related quality of life, including imposing physical restrictions, lifestyle changes, and the need for environmental adaptation. Pain associated with pressure ulcers was their primary concern, but other symptoms and treatments also contributed to an overall reduction in their quality of life. Participants were also concerned that they had been a burden to others, causing additional anxiety and worry.49

2.6 Current practice

2.6.1 Prevention

Prevention practice guidelines recommend risk assessment and individualized care plan for high-risk patients with daily skin assessment, repositioning patients frequently (e.g., every 2 hours), using pressure-redistribution support surfaces, improving skin health (especially protecting skin from excessive moisture and incontinence), and improving nutrition (with nutritional assessment and a nutritional support plan if necessary).50

The prevention of pressure ulcers is a multidisciplinary responsibility.51 Few health care facilities, however, have a multidisciplinary skin and wound care team and program in place.52 Bundles, or targeted systematic interventions (e.g., risk assessment, individualized care plan for high-risk patients), have been suggested to be effective in reducing the incidence of pressure ulcers.52, 53

A systematic review identified 56 randomized controlled trials evaluating the safety and efficacy of preventive interventions.54 The methodological quality of the included trials was variable. Strategies that effectively addressed impaired mobility included pressure-redistribution foam mattresses, mattress overlays on operating tables, and specialized foam and specialized sheepskin overlays. Although turning or repositioning is widely recommended, there is insufficient evidence to recommend specific turning regimens for patients with impaired mobility. In patients with nutritional impairments, dietary supplements may be beneficial. The incremental benefit of specific topical agents over simple moisturizers for patients with impaired skin health is unclear. Accordingly, there is a need for well-designed randomized controlled
trials that follow established criteria for trial conduct and reporting to evaluate the cost-effectiveness of preventive interventions.

2.6.2 Treatment

The pressure ulcer treatment guidelines by the Registered Nurses' Association of Ontario recommend wound dressings (to maintain a moist wound environment), pain management, optimal nutrition support, adjunctive therapies (for wound healing), and weekly monitoring of treatment effects.\textsuperscript{55} The guidelines also recommend choosing the support surface which best fits with the overall care plan for the patient, considering the goals of treatment, client bed mobility, transfers, caregiver impacts, and ease of use.

A systematic review identified 103 randomized controlled trials evaluating the effectiveness of pressure ulcer treatments.\textsuperscript{56} Included trials were assessed by a validated quality assessment regarding adequate allocation sequence generation, concealed treatment allocation, adequate participant blinding, adequate outcome assessor blinding, comparable treatment groups, and intention-to-treat analysis. Only 15\% of the included trials were deemed to be of reasonable quality. Product manufacturers provided funding for approximately half of the trials, and about 80\% of the trials did not provide sufficient information about authors’ potential financial conflicts of interest.

According to the systematic review, there was no clear evidence favoring one support surface over another (n=12 trials).\textsuperscript{56} No trials compared a specialized support surface with a standard mattress and repositioning. Among 7 trials evaluating nutritional supplements, only 1 relatively higher quality trial with nursing home residents reported that protein supplementation significantly improved mean wound healing scores. Among 54 trials evaluating absorbent wound dressings, only 1 trial reported that compared with dextranomer paste, calcium alginate dressings significantly reduced wound surface area. No other dressing was superior to alternatives. Among 9 trials evaluating biological agents, several trials reported benefits with different topical growth factors. However, the incremental benefit of these biological agents over less expensive standard wound care remains uncertain. No clear benefit was identified in 21 trials evaluating adjunctive therapies including electric current, ultrasound, light therapy, and vacuum therapy.
2.7 Health care costs of pressure ulcers

2.7.1 Costs attributable to pressure ulcer care

Estimating the health care costs attributable to pressure ulcer care requires the disaggregation of the resources consumed in managing the primary illnesses (and comorbidity conditions) from the resources used to care for the associated pressure ulcers.\(^{57}\) Table 2 displays estimated costs attributable to pressure ulcer care; original costs reported in other currencies were converted, adjusted for purchasing power parity for health expenditures and pro-rated to 2009.\(^{58}\) From a risk pool of approximately 1.9 million US hospital patients, Zhan et al. 2003 used multivariate matching to select controls for each of the 41,440 cases with pressure ulcers.\(^{59}\) They then compared cases and controls on different outcome measures. According to their results, hospital-acquired pressure ulcers were associated with an estimated absolute excess length of stay of 4 days, and an estimated absolute excess mortality rate of approximately 7%. The average hospital cost attributable to pressure ulcer care was approximately $13,500 per patient.\(^{59}\) Bennett et al. 2004 used a micro-costing approach to derive stage-specific costs of pressure ulcer care in the UK. They directly enumerated and cost every input used in pressure ulcer treatment according to guideline recommendations.\(^{60}\) The treatment cost estimate ranged from approximately $8,500 to approximately $20,000 across classification stages of pressure ulcer (Table 2).

2.7.2 Hospital costs attributable to pressure ulcer care in Canada

Every year, there are approximately 2.8 million admissions to acute care hospitals in Canada.\(^{61}\) Even at a low incidence estimate of pressure ulcers in acute care patients (e.g., 7%, Table 1) and a low attributable cost (e.g., $9,000, Table 2), hospital costs attributable to hospital-acquired pressure ulcers in Canada could exceed $1.7 billion, or approximately one-eighth the total health care cost of cardiovascular diseases in Canada.\(^{62}\)
Table 2. Estimated treatment costs attributable to pressure ulcer care

<table>
<thead>
<tr>
<th>Source</th>
<th>Method</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Stage 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bennett et al. 2004; UK $^{60}$</td>
<td>Micro-costing of pressure ulcer care according to protocol of care for pressure ulcers in the UK</td>
<td>$8,587</td>
<td>$14,099</td>
<td>$20,341</td>
</tr>
<tr>
<td>Zhan et al. 2003; USA $^{59}$</td>
<td>Multivariate matching to control for confounding factors using 2000 NIS data*</td>
<td>$13,542</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Costs are in 2009 Canadian dollar. Estimates were converted from other currencies and adjusted for purchasing power parity for health expenditures and pro-rated to 2009. $^{63}$ Hospital costs, 2000 Healthcare Cost and Utilization Project Nationwide Inpatient Sample (NIS) developed by the Agency for Healthcare Research and Quality. $^{59}$

2.7.3 Costs of prevention and treatment

In one study reporting both average costs for treatment and prevention, 276 patients were randomly selected from a pool of patients admitted to a community hospital and followed over three months. $^{64}$ The study participants comprised 40 patients with pressure ulcers and 106 patients assessed to be at risk for pressure ulcer development. In patients with pressure ulcers, the authors estimated the average treatment costs of staff-time and products associated with topical therapies (e.g., creams and ointment), dressings (e.g., films, hydrocolloids and gauze), and incontinence management. In at-risk patients, they estimated the average cost of using pressure-redistribution support surfaces or devices for the prevention of pressure ulcers. They reported that the average treatment cost was 2.5 times the average prevention cost.

2.8 Cost-effectiveness of preventive interventions

A UK pressure ulcer practice guideline included a systematic review of three cost effectiveness studies of prevention strategies that were based upon input data from randomized controlled trials. $^{65}$ Russell et al. 2003 compared the cost effectiveness of pressure-redistribution foam mattresses to standard hospital mattresses in elderly patients. $^{66}$ Pressure-redistribution mattresses typically mold around the shape of the patient’s body to distribute the patient’s weight at bony prominences over a large area to reduce interface pressure, and thus reduce the risk of pressure ulcer. The authors reported that pressure-redistribution foam mattresses were cost effective; the
probability that the incremental hospital cost per pressure ulcer averted was £100 was 0.95. Gebhardt et al. 1996 compared the cost effectiveness of alternating-pressure mattresses and pressure-redistribution foam mattresses in acutely ill patients. Alternating-pressure mattresses have air-filled channels that alternately fill and empty to keep bearing weight off bony prominences of at-risk patients. The authors reported that alternating air-pressure mattresses were less expensive and more effective than pressure-redistribution foam mattresses. Inman et al. 1993 compared air-suspension beds and standard intensive care beds in intensive care patients. They concluded air-suspension beds were less expensive and more effective than standard intensive care beds. The systematic reviewers assessed the quality of the included studies using the Drummond checklist. They reported that although the included studies are based on trial evidence, the study quality is low. In particular, no studies presented disaggregated data on resource utilization and none explored the impact of input uncertainty.

In order to explore the cost-effectiveness of pressure-redistribution foam mattresses for patients at different risk levels, the economic evaluation team involved in the development of the UK practice guideline constructed a cost effectiveness model. The results indicated that although pressure-redistribution mattresses are more expensive than standard hospital mattresses, the treatment saving from averted pressure ulcers more than offsets the incremental costs. Also, it is likely that pressure-redistribution mattresses are cost effective for all groups of at-risk patients.

We extended the literature search by the UK National Institute for Clinical Excellence guideline from May 2002 to May 2011, resulting in the identification of five additional economic evaluation studies (including three studies from our group that are discussed in subsequent chapters of this thesis). Iglesia et al. 2006 conducted a cost-effectiveness analysis from the perspective of the UK health and social services using data from a large pragmatic randomized controlled trial evaluating alternating-pressure mattresses and alternating-pressure overlays for the prevention of pressure ulcers. They reported that alternating-pressure mattresses were more likely to be cost effective and were more acceptable to patients than alternating pressure overlays.

Padula et al. 2011 evaluated the cost-effectiveness of a comprehensive prevention program (including risk assessment, bed and chair support surfaces, nutritional supplements, repositioning, and moisture and incontinence management) versus standard practice in hospital
The authors concluded that prevention had close to a 100% chance of improving health and saving costs from a societal perspective. Both the clinical and economic evidence support the use of pressure-redistribution mattresses for the prevention of pressure ulcers among at-risk patients.

2.9 Prevention policies

2.9.1 Prevention policy

In the United States, there is an emerging belief that hospitals should not be reimbursed for serious complications that should never occur in a safe hospital. Pressure ulcers have been designated as one of the “never events”, medical errors that are identifiable and reasonably preventable through the use of evidence-based guidelines. Since 2008, the Centers for Medicare and Medicaid Services no longer reimburse hospitals for additional care for hospital-acquired pressure ulcers. The Centers herald this move as an effort to align financial incentives with patient safety and quality of care.

2.9.2 Prevention practice in Ontario

Like the United States, Ontario has an aging population. By 2026, when many of the baby boom generation will be retired, the proportion of elderly will increase from the current 13.2% to 21.2%. Elderly patients are at risk of developing pressure ulcers because of immobility, poor nutritional status, impaired mental status, and incontinence. The percentage of hospital patients reporting all these deficits increased over 15 years, from 9.1% in 1994 to 17.2% in 2008.

In Ontario, it is challenging to prevent pressure ulcers, especially in acute care hospitals. A high proportion (approximately 38% as of 2010) of hospitals reported a financial deficit. To cope with a high demand for hospital services from an aging population, hospitals strive to discharge patients as early as they can. As a result, discharged patients with hospital-acquired pressure ulcers are managed in the community. This arrangement of transferring care for pressure ulcers from acute to community care does not incentivize hospitals to improve prevention (personal communication, Ms. Laura Teague, the Ontario Wound Care Interest Group, the Registered Nurses Association of Ontario, May 7, 2009). This arrangement also subjects pressure ulcer care to variation in the quality of care transition, including problems with care coordination and
Consequently, the lack of emphasis on prevention in acute care, coupled with an aging population, could lead to a gradual increase in the burden of pressure ulcers over time.

2.10 Research questions and rationale

In summary, pressure ulcers are common in the very ill, the elderly, immobile patients, and neurologically compromised patients. They can cause considerable pain and significantly impair the quality of life of affected patients. Pressure ulcers represent an important issue for health care providers and policy decision makers because they are common and are associated with high treatment cost.

Prevention practice guidelines recommend risk assessment and individualized care plans for high-risk patients with daily skin assessment, frequent repositioning, using pressure-redistribution mattresses, and improving skin health and nutrition. In particular, the use of pressure-redistribution mattresses has been shown to significantly reduce the risk of developing pressure ulcers in acute care patients. Preliminary results also indicate that although pressure-redistribution mattresses are more expensive than standard hospital mattresses, the health improvement and the treatment saving from averted pressure ulcers more than offsets the additional costs. But this may vary for different groups of at-risk patients.

Among at-risk groups, hospital patients are at high risk of developing pressure ulcers. In particular, certain patients are at high risk for a well-defined duration of exposure, making them especially ideal for targeted interventions.

We undertake cost-effectiveness analysis to address the following questions.

Study question 1

What is the cost-effectiveness of early prevention of pressure ulcers among elderly admitted ED patients?

Rationale

Elderly patients are at high risk of developing pressure ulcers because of immobility, poor nutritional status, impaired mental status, and incontinence. Elderly patients account for approximately 30% of ED visits. They typically spend hours in the EDs and could be at
increasing risk of developing pressure ulcers as they lie for considerable time on unyielding diagnostic equipment surfaces, stretchers and standard hospital mattresses in the EDs.\textsuperscript{57} Approximately 6\% of elderly admitted ED patients acquired pressure ulcers within 48 hours of admission, with higher incidence among patients with severe immobility, malnutrition, and incontinence.\textsuperscript{77} However, only half of the patients assessed to be at high risk of developing pressure ulcers are supported by pressure-redistribution mattresses early in their hospital admission.\textsuperscript{83}

**Study question 2**

What is the cost-effectiveness of intra-operative prevention of pressure ulcers in patients undergoing prolonged surgery?

**Rationale**

Patients undergoing prolonged surgical procedures are immobilized for long periods and are therefore at risk of developing pressure ulcers.\textsuperscript{84} In patients undergoing prolonged surgical procedures, the sub-dermal tissue under bony areas is under very high body-induced pressure and sustained tissue injury may occur with immobility lasting more than one hour. The incidence of developing pressure ulcer varies from 2\% to 10\% among surgical patients, with higher incidence among patients with operation lasting >4 hours and patients undergoing cardiac surgery. However, the use of pressure-redistribution support surfaces to prevent pressure ulcers during surgery is low and based more on facility-related factors than on patient risk.\textsuperscript{85}

In summary, elderly admitted ED patients and patients undergoing prolonged surgery are at high risk of developing pressure ulcers. However, the clinical and economic evidence in support of early prevention among these high-risk populations is limited. The questions regarding the cost-effectiveness of early prevention in the emergency and surgery departments are important since at the current time, it is unclear as to whether additional effort and resources should be allocated to pressure ulcer prevention in these settings. The answers to these questions would inform clinical practice and policy decisions regarding early prevention that targets these high-risk populations.
Chapter 3
Decision Modeling And Cost-Effectiveness Analysis

3 Decision-analytic modeling and cost-effectiveness analysis

3.1 Abstract
This chapter provides an overview of decision analytic modeling and cost-effectiveness analysis in the context of pressure ulcer prevention. Decision analytic modeling provides an explicit, quantitative and systematic approach to evaluate alternative interventions. Cost-effectiveness analysis can be defined as the comparison of alternative interventions in terms of their costs and consequences. The development of a decision model for pressure ulcer prevention involves the estimation of uncertain transition probabilities that characterize the worsening or healing of a pressure ulcer. Model calibration involves the estimation of uncertain model parameters, so that modeled projections are consistent with corresponding observed data. Methodological issues related to the calibration of a decision model for early prevention of pressure ulcers are highlighted, and a literature review of calibration methods is suggested.

3.2 Background
The prevention of pressure ulcers requires multidisciplinary efforts and dedicated resources that are generally not readily available. An important question in the allocation of resources to prevention is how much health improvement can be gained with the effort and resources spent on any particular preventive interventions. This question can be addressed by cost-effectiveness analysis, the analytic method that quantifies the trade-off between the costs and health effects associated with alternative interventions. In particular, cost-effectiveness analysis that is based upon a decision modeling approach can be used to evaluate the cost-effectiveness of multiple interventions. This chapter provides an overview of decision analytic modeling and cost-effectiveness analysis in the context of pressure ulcer prevention.

3.3 Decision analytic modeling
Decision analytic modeling provides an explicit, quantitative and systematic approach to evaluate alternative interventions. A decision model uses mathematic relationship to define a
series of possible consequences that would flow from a set of alternative interventions under
evaluation. Based upon model inputs, the likelihood of each consequence is expressed in terms
of probabilities, and each consequence has a cost and a health outcome. This makes it possible to
calculate the expected cost and expected health outcome of each intervention. The expected cost
(or outcome) is the sum of the costs (or outcomes) of each consequence weighted by the
probability of that consequence.

A key purpose of decision modeling is to allow for individual patient variability, parameter
uncertainty and heterogeneity associated with all decisions. Individual patient variability
represents the uncertainty in patient-level outcomes. This uncertainty is entirely due to chance;
it’s also known as stochastic uncertainty or first-order uncertainty. In decision models each
chance node contributes to this uncertainty. For example, at a specific chance node a patient has
an estimated probability of 5% of getting a pressure ulcer. This stochastic uncertainty reflects the
uncertainty related to the actual outcome— a patient may or may not fall within the 5% of patients
who develop a pressure ulcer – which should be distinguished from the uncertainty around the
5% estimate because the limited sample of patients in which the probability estimate was
obtained. The latter uncertainty is known as parameter uncertainty, or second-order uncertainty.

A decision model can also be structured to incorporate parameters that quantify the heterogeneity
among patients because variation in the underlying factors leads to variation in the consequences.
In instances, incorporating heterogeneity into the model is important to inform policy decisions
affecting individuals with different risk profiles.

There are alternate ways to structure a model for the same decision problem. This results in
another source of uncertainty, namely structural uncertainty. A key requirement for decision
modeling is to cascade uncertainty from different sources in the model into a quantifiable
decision uncertainty. For example, decision uncertainty can be characterized by the probability
that a given decision with the supporting data from the model is the correct one.

Decision models have been widely used to evaluate the cost-effectiveness of screening
programs, therapeutic interventions, and diagnostic procedures. They can be used to
simultaneously compare the cost-effectiveness of multiple interventions. For example, different
interventions can be used to reduce the risk of developing pressure ulcers, including those that are directed at pressure, at shearing or at tissue tolerance. No clinical studies can simultaneously evaluate these interventions. However, a decision model has been used to compare the cost-effectiveness of current prevention practice in nursing homes with four quality improvement strategies that target different groups of varying risk for pressure ulcer development.\textsuperscript{95} The preventive strategies include support surfaces, nutritional supplements, skin emollients, and incontinence care. According to the study results, the first intervention is cost-effective but the cost-effectiveness of the remaining interventions is uncertain.

Decision modeling can also be used to quantify different sources of uncertainty, including uncertainty in the disease prognosis under investigation. For example, newly developed pressure ulcers with limited tissue damage may or may not be detected. Over time, the associated tissue damage may increase such that a pressure ulcer can attain a higher classification stage than the stage when it was first developed. On average, higher-stage pressure ulcers are associated with longer median healing time, higher treatment costs and lower quality of life. In a recent cost-effectiveness analysis, a decision model was constructed to characterize the prognosis of pressure ulcers among high-risk hospital patients.\textsuperscript{72} The analysis evaluated a bundle of preventive services, including risk assessment, support surfaces, nutritional supplements, frequent repositioning, and managing of moisture and incontinence. According to the study results, the preventive bundle is cost-effective as it improves health outcomes and reduces hospital cost.

### 3.4 Cost-effectiveness analysis

Cost-effectiveness analysis can be defined as the comparison of alternative interventions in terms of their costs and consequences.\textsuperscript{86} There are trial-based and model-based cost-effectiveness analyses. The former generally evaluate the effectiveness and cost-effectiveness of specific interventions in a well-defined patient population using individual patient data from randomized controlled trials. For example, a pilot randomized controlled trial of negative pressure wound therapy for severe pressure ulcers was conducted to evaluate the feasibility of patient recruitment and data collection, and to provide inputs into a model-based cost-effectiveness analysis that was conducted alongside the pilot trial.\textsuperscript{96} A decision modeling approach to cost-effectiveness analysis is suitable when all relevant interventions are to be considered using existing evidence.\textsuperscript{86}
With this approach, all consequences of the interventions can be identified, measured, and valued; and all sources of uncertainty that exist about these consequences can be quantified. This approach is particularly useful when there are needs to link intermediate to final endpoints, to extrapolate intervention effects beyond the observed duration of clinical studies, and to make results applicable to the decision-making context.

One aim of cost-effectiveness analysis is to identify interventions that lead to improvement in population health. With cost-effectiveness analysis, we ask the question of how much health improvement can be gained for each dollar spent, compared to some alternative use of the resources. Specifically, the trade-off between the cost and health effect of an intervention $A$, compared to its alternative $B$, is quantified using the incremental cost-effectiveness ratio (ICER), defined as the incremental cost of the intervention $A$ (relative to $B$) divided by the corresponding incremental effectiveness. Specifically,

$$ICER = \frac{Expected\ Cost_A - Expected\ Cost_B}{Expected\ Effect_A - Expected\ Effect_B}$$

To calculate the ICER of an intervention, a specific measure of effectiveness is needed. For example, the number of pressure ulcers prevented is suitable for evaluating the cost-effectiveness of preventive interventions. However, some effectiveness measure that allows for the broad comparison of different interventions is also needed. One commonly used measure is the quality-adjusted life year (QALY). The QALY simultaneously captures the potential gains from reduced morbidity (quality gains) and reduced mortality (quantity gains) associated with the consequences of an intervention. The QALY is derived by weighting the length of time affected through intervention by the health-related quality of life measures of the resulting health status. This is calculated over a time period in which the costs or outcomes of the range of alternative interventions may be expected to differ.

The quality weights in the QALY calculation must be based on preferences and must be measured on an interval scale with proper anchors for death (e.g., typically represented by 0) and perfect health (i.e., 1). Note that all QALYs are not the same. Weights may be based on standard gamble utility measurements, time trade-off value measurements, visual analogue scale
value measurements, estimates by physicians and researchers, or preference-weighted systems like the Health Utilities Index or the EuroQol-5D.86

The QALY concept is controversial. Because the QALY reflects the added years of life and the improved quality of life, it could be argued that the QALY is a reasonable measure of health gain. However, there are reasons to suggest that QALY may not adequately reflect social value.99 First, given the blunt nature of some of the instruments used to assess changes in quality of life, it is possible that these instruments may not capture all aspects of life that individuals care about. Second, each gain in QALYs is treated as being equally valuable, no matter whether the gain arises mainly from life extension or improved quality of life. In addition, QALYs are valued the same no matter who receives them. However, some surveys suggest that improving the health of an individual with a very serious health condition may be valued more highly by the general public than improving the health of someone who is already reasonably healthy.100 Despite these limitations, QALY is still frequently used in cost-effectiveness analysis although there is a gradual trend towards rethinking the notion of value in health care.99

Decision problems addressed by cost-effectiveness analysis typically concern the best choice among alternative interventions.86 Ideally, the opportunity cost of re-allocating resources to fund a selected intervention must be identified.88 This opportunity cost is the QALYs forgone by reducing provision of currently funded interventions, services or programs from which any resources are re-allocated for the purpose of funding the intervention under consideration. Identifying the marginal programs that would be displaced (i.e. the least cost-effective program of those currently funded) and quantifying their costs and associated QALYs determines the opportunity cost to fund the selected intervention.88 A selected intervention should be paid for if the associated QALYs gained exceed the QALYs forgone due to displacement of the marginal program(s).

In practice, the marginal programs that would be displaced are unknown, and any new interventions must be evaluated regarding whether they are worth doing (or cost-effective) on the basis of some externally defined value for money threshold λ (e.g., $50K per QALY).88 There is no consensus for threshold values although some are used more often than others. Often, the
threshold values are selected from published sources (which are arbitrary), implied values from previously funded programs, or cost per QALY league tables.\textsuperscript{86}

In practice, the information on the opportunity cost of funding a particular intervention is generally unknown. As such, a simplified decision rule is generally used to determine its cost-effectiveness. Specifically, if an intervention $A$ in comparison to some alternative intervention $B$ is associated with a reduction in the expected cost and an increase in the expected QALY, then it is cost-effective. If $A$ increases the expected cost as well as the expected QALY, $A$ may still be cost-effective given the right trade-off between cost and QALY. In this instance, the ICER of $A$ can be compared to a threshold $\lambda$ to inform as to whether $A$ is more cost-effective than $B$.\textsuperscript{88} The intervention $A$ is considered cost effective if its ICER fulfills the inequality $\text{ICER} \leq \lambda$, or $\frac{\Delta C}{\Delta Q} < \lambda$, where $\Delta C$ and $\Delta Q$ representing the incremental cost and the incremental QALY, respectively. This inequality is often re-arranged to calculate either the net health benefit ($\Delta Q - \Delta C/\lambda$) or the net monetary benefit ($\Delta Q^*\lambda - \Delta C$), with a positive net benefit indicating that $A$ is more cost-effective than $B$.

### 3.5 Calibration analysis of decision models

Decision modeling can be used to simulate a disease prognosis under investigation and to project costs and health consequences for cost-effectiveness analysis of alternative interventions.\textsuperscript{88} For example, a decision model can be used to simulate the natural history of pressure ulcer, including the following prognosis,

- the development of a new pressure ulcer;
- the potential for worsening of the associated tissue damage, indicating the transition of the newly developed pressure ulcer to other classification stages (e.g., from stage 1 to stage 2, stage 2 to 3, stage 3 to 4);
- the potential for healing of a stage-specific pressure ulcer; and
- the possibility of local and systemic infection associated with an open wound.

In the model, stage-specific transitions are uncertain because serial staging assessment of open wounds typically requires debridement of eschar that is not done unless clinically indicated.\textsuperscript{32}
Estimates of these transitions are necessary for the development of the decision model and quantify the downstream consequences of prevention in the cost-effectiveness analysis.\textsuperscript{70} Model calibration involves the determination of uncertain model parameters, so that modeled projections are consistent with corresponding observed data.\textsuperscript{101} It is typically effected by searching a collection of random points in a high-dimensional space for good-fit parameter values or set of values such that model outputs reproduce the corresponding observed data. In the above example, the distribution of stage-specific pressure ulcers can be obtained from a survey of prevalent pressure ulcers among hospital patients. Information from this observed stage-specific distribution can be used to estimate the stage-specific transition rates in the decision model. One simple approach to model calibration is to conduct a random search across some plausible ranges of the transition probabilities and select the plausible values of the probabilities so that the projected distribution of the stage-specific pressure ulcers from the decision model reproduces the observed stage-specific distribution.\textsuperscript{102}

In the context of pressure ulcer prevention, a challenge in the conduct of cost-effectiveness analyses of preventive interventions is the calibration of the decision model to different at-risk populations in order to estimate transition probabilities. The development, deterioration and healing of pressure ulcers involves many patient-specific risk factors, suggesting that transition probabilities may vary by population. The estimated transition probabilities are necessary to characterize the prognosis of pressure ulcer in these populations. There is a need for calibration methods that are suitable for routine calibration of the decision model to different at-risk populations.

### 3.6 Rationale for a literature review of calibration methods

According to recent literature reviews, calibration is widely used in the development of cancer simulation and infectious disease models, but infrequently used in model-based studies of cardiovascular diseases.\textsuperscript{101, 103, 104} A recent proposal recommends a stepwise approach to the conduct of model calibration and presents a theoretical discussion on the options available at each step of the calibration process.\textsuperscript{102, 105} The proposal suggests that further research is required to systematically investigate the engineering and operations research literature for calibration methods that could be used in cost-effectiveness analysis. In addition, there is still limited evidence to support good practice in the calibration of decision models.\textsuperscript{101, 102, 105, 106} In order to
address this gap, it is necessary to conduct a literature review of methods for the calibration of decision models for cost-effectiveness analysis of alternative interventions (chapter 4).
Chapter 4
Calibration Analysis of Decision-Analytic Models: A Literature Review

4 Calibration analysis of decision models: A literature review

4.1 Abstract

Guidelines for good practice in decision modeling recommend improving the consistency and transparency of decision models. Model calibration involves the estimation of uncertain model parameters, so that the aggregate outputs from the model are consistent with observed data.

This chapter reports on a literature review of parameter search strategies for model calibration. The search strategies are identified from decision modeling studies with detailed reporting of calibration. Concepts underlying the search strategies as well as details of their applications were summarized. A series of synopses on the search strategies was compiled. These synopses were stratified by broad categories, including Bayesian, guided search, random search and grid search approaches. Common and complimentary aspects of the strategies across their broad approaches are identified.

Results from the review support: i) the proposed stepwise approach to calibration since the technical considerations involved in the different steps are different; ii) obtaining not only the best estimates for model parameters but also their uncertainty distributions; iii) the propagation from uncertainty distributions of parameters to subsequent analyses; and iv) considering model calibration as an evidence synthesis exercise and if feasible, carrying out calibration within the Bayesian modeling framework for decision analysis.

Methodological issues documented in this chapter require further study, as they are relevant to the uptake of calibration methods for model development, revising and updating. In particular, a study was proposed regarding the development of a calibration approach that is suitable for routine calibration of decision models for cost-effectiveness analysis of alternative interventions.
4.2 Introduction

Decision-analytic models play an increasingly important role in the economic evaluation of health technologies. They simulate the course of the disease under study and provide a flexible framework for synthesizing data from a variety of sources, for linking intermediate to final outcomes, and for extrapolating beyond the horizon of clinical trials. Models are based on evolving scientific knowledge of disease, simplifying assumptions and inputs with different levels of uncertainty. Recent guidelines recommend best practice in conceptualizing models, dealing with uncertainty, and model validation. Guidelines for the conduct of decision modeling studies highlight the importance of ensuring consistency between model predictions and the best available evidence. Model calibration involves the estimation of uncertain model parameters, so that the aggregate outputs from the model are consistent with observed data.

Calibration is an essential step in the development of most cancer simulation models, microsimulation models of individual event histories, discrete event simulation models of systems with constrained resources, and dynamic transmission models of infectious diseases. There has been recent interest in model calibration, including new methods, practice surveys and recent proposals for the conduct and reporting of model calibration. For example, calibration is frequently used in the development of cancer simulation models, but the process of model calibration is infrequently described. When model calibration is described, the methods used are from amongst a variety of techniques, are not consistently applied, and have many details omitted.

An early use of calibration analysis is dated back to the late 1980’s in the development of a coronary heart disease model. Calibration analysis of computer codes that simulate physically based systems has been researched and conducted since the late 1970’s in environmental science, engineering and computer science, and mathematics and statistics. However, cross-disciplinary perspectives on the fundamental concepts of calibration analysis are rarely discussed.

This chapter reports on a literature review of methods used to calibrate decision-analytic models. The next section outlines the data sources used in the review. The calibration process is described in the next section. Subsequent sections describe a series of synopses on the methods
for model calibration, stratified by broad categories. Next, limitations of this literature review are described. We then discuss the advantages and disadvantages of calibration approaches.

4.3 Data sources

We conducted a restricted literature search of MEDLINE (1946 to November Week 1 2013) and EconLit (1987 to November Week 1 2013) to identify potentially relevant studies (Table A-1, Appendix 1). A study is deemed to be relevant if it

i) describes a decision-analytic model,

ii) explicitly reports on the use of model calibration in the development of the decision model,

iii) reports details of the calibration process, and

iv) is an English-language study.

Key words for the literature search include [“model” adjacent to (“decision”, “screening”, “cost-effectiveness”, “disease”, “natural history”, “pathway” OR “simulation”)] and [“calibration”].

The MEDLINE search yielded 683 potentially relevant citations and the EconLit search yielded 110 citations (Appendix 1).

Citation screening was conducted by a single reviewer. This reviewer also (i) conducted the full-text review of potentially relevant studies identified from the citation screening and (ii) screened the reference lists of relevant studies to identify other potentially relevant studies, including studies describing the methods and techniques for parameter search strategies.

The literature search yielded 16 relevant studies that are included in the literature review. The included studies describe 12 cancer simulation models,105, 116, 117, 124, 126-133 2 infectious disease models,134, 135 and 1 coronary heart disease model.136

4.4 The calibration process

We abstracted data pertaining to each step of a stepwise approach using items from the reporting checklist proposed by Stout et al. 2009,101 and a seven-step approach to the conduct of calibration proposed by Vanni et al. 2011.101,102

Tables 3 to 6 describe the calibration process with different approaches to model calibration. The included studies cover a broad spectrum of unique parameter search strategies, including
Bayesian approaches to calibration (Table 3), guided search approaches (Table 4), random search approaches (Table 5) and other search approaches (Table 6).

- The Bayesian approaches include Bayesian evidence synthesis methods that incorporate model calibration (2 studies), Bayesian Markov Chain Monte Carlo (MCMC) methods for the estimation of calibration parameters (2 studies), and the Approximate Bayesian Computation (ABC) method (1 study).

- The guided search approaches include the generalized reduced gradient search strategy (1 study), a simulated annealing search strategy and a genetic algorithm search strategy (1 study) and the Nelder-Mead downhill simplex search strategy (1 study).

- The random search approaches include a three-stage random search strategy with simple random sampling (1 study), a two-stage random search strategy with simple random sampling (1 study), a two-stage random search strategy with stratified random sampling (1 study), an approximate Bayesian approach based upon a random search strategy (1 study) and single-stage random search strategy with Latin-hypercube sampling (1 study).

- The grid search approaches were described in 3 studies (Table 6).

- The manual search approach was used in conjunction with other search strategies (described below).

**Step 1. Which parameters should be varied in the calibration process?**

Calibration was commonly used to estimate uncertain (unobservable) model parameters by only allowing these parameters to vary in the calibration process (Tables 3-6). This was the case for all the cancer simulation models in which model parameters associated with unobservable portions of the disease process were calibrated against observed data that quantify the consequences of the disease (e.g., colorectal, lung, breast, cervical and gastric cancer).\(^{116, 124, 127, 129, 132}\) Calibration was also used to establish initial conditions in a dynamic transmission model of live attenuated HIV vaccines,\(^{134}\) to estimate uncertain age-specific incidence and case-fatality rates in the lowest and highest age groups for a coronary heart disease model,\(^{136}\) and to determine model parameters associated with influential assumptions regarding the uncertain prognoses among untreated and treated HIV infected women in a microsimulation model of HIV and AIDS.\(^{135}\)
Even when parameters have been estimated directly, these parameters may have different levels of uncertainty, leading to the possibility that all natural history and other relevant parameters in the model (unobservable and observable) could be allowed to vary in the calibration process.\textsuperscript{102} This was the case with the calibration of some models in Tables 3-6.\textsuperscript{105, 116, 117, 126, 128}

**Step 2. Which calibration targets should be used?**

Across the first two steps in Tables 3-6, most calibration targets were selected by the modelers to inform the calibration parameters as these targets were deemed to be functionally related to the parameters.\textsuperscript{117} All the selected studies used multiple targets for calibration. Examples of high-quality targets include outcomes from a randomized trial of HPV testing in primary cervical screening; a meta-analysis of randomized control trials;\textsuperscript{105} a multicenter, prospective, observational HIV study;\textsuperscript{135} and a large longitudinal study of the natural history of HPV infection and cervical neoplasia.\textsuperscript{131} In the last study, target outcomes (e.g., mean duration of HPV infection and associated hazard ratios) were derived by analyzing simulated individual patient data according to the analysis of the original study data.\textsuperscript{131}

Cohort and cross-sectional data were often used together to inform different parameters in the simulated disease process (Tables 3-6).\textsuperscript{116, 124, 127, 128, 130-132} In particular, data from cancer screening programs, disease registries, and vital statistics were used to ensure that modeled projections reflect local disease conditions.\textsuperscript{116, 124, 127, 130}

**Step 3. What measure of goodness-of-fit (GOF) should be used?**

A GOF measure quantifies how close the model predictions are to the target data.\textsuperscript{102} It has been suggested that there is no consensus on the most appropriate GOF measure for model calibration.\textsuperscript{102} Also, different measures are used because there are no theoretical justifications to select one measure over the other.\textsuperscript{137}

In Tables 3-6, different measures were used for the parameter search strategies, ranging from simple measures with less data requirements to measures that require more data for their derivation. These include visual inspection,\textsuperscript{134, 135} acceptance windows,\textsuperscript{126, 129, 136} measures of differences,\textsuperscript{128, 135, 138} sum-of-squared errors,\textsuperscript{127} chi-squared GOF measures,\textsuperscript{105} and likelihood statistics.\textsuperscript{105, 130, 131} When models were calibrated to multiple targets, composite GOF measures
were derived by summing across individual targets with equal or differential weights.\textsuperscript{105, 127, 131, 132}

One study reports that the chi-squared GOF measure differentiated between the good-fit parameter sets identified from the search to a far greater degree than the likelihood-based GOF measure (Tables 3-6; see also the related synopsis).\textsuperscript{105} This study and its companion study describe how to derive different GOF measures.\textsuperscript{102, 105}

\textit{Step 4. What parameter search approach should be used?}

Different parameter search strategies and Bayesian methods have been proposed for obtaining solutions to the calibration problem (Table 3). Bayesian methods calibrate model parameters against target data to derive their posterior distributions and, at the same time, propagate parameter uncertainty to model outputs.\textsuperscript{133} The parameter search strategies that are classified as the guided search approach iteratively learn from the currently selected set of parameter values and move towards a better fitting set (Table 4).\textsuperscript{102} The search strategies that are classified as the random search approach select parameter sets at random throughout the search and evaluate each and every set to identify an ensemble of GOF parameter sets (Table 5).\textsuperscript{102} In the grid search approach, the search takes place across the different possible combinations of parameter values in the parameter search space (Table 6). Manual search was less well defined in the studies that reported its use in conjunction with other search strategies (Table 6).\textsuperscript{128, 129}

\textit{Step 5. What determines acceptable good-fit parameter sets?}

GOF parameter sets are solutions to the calibration problem. They were obtained in different forms from the broad approaches (Tables 3-6). With Bayesian methods, posterior distributions of model parameters (and functions of model parameters) were obtained. The guided search strategies generally reported a set of local optima (including probably the global optimum) and associated good-fit parameter sets.\textsuperscript{127, 128} Random search strategies (as well as grid search strategies) generated an ensemble of good-fit parameter sets.

\textit{Step 6. What determines the termination of the calibration process?}

Bayesian methods were terminated upon the convergence of the MCMC simulation that allows for sampling from the posterior distributions of model parameters (Table 3). Different guided search strategies used different termination rules, depending on the search algorithms (Table 4).
As an example, in the calibration of a lung cancer model with a simulated annealing search algorithm, each iteration was limited to a maximum of 1000 iterations. Random search strategies (as well as the grid search strategies) were conducted with predefined large numbers of model evaluations (Table 5). Generally, there were no justifications on how the predefined numbers were determined.

**Step 7. How should the model calibration results and economic parameters be integrated?**

This step is omitted from Tables 3-6 as it was only described in one study (see section 4.6 and the synopsis of the Karnon et al. study).

**Step 8. How should the calibrated model be validated?**

Across studies, the assessment of internal and external consistency could take place during or after calibration (Tables 3-6). In the calibration of a population-based breast cancer model, no solutions were obtained after the first-stage search, leading to an important change in the model structure (see synopsis). In the calibration of an illustrated HPV model with a Bayesian evidence synthesis method, projected prevalence could not be simultaneously reconciled with observed prevalence from a trial and a national screening program. This led to the introduction of an extra model parameter (and an additional assumption) in the revised model.

After calibration, all studies assessed the internal consistency between modeled projections and corresponding targets (Tables 3-6). There were examples of model validation with the use of data that were not used in the calibration process. A calibrated coronary heart disease model was prospectively validated when additional data became available and the model was updated (including calibration) with the additional data. A few studies corroborated results from the calibrated models with results from similar models. One study assessed the external consistency of the calibrated model by comparing modeled projections with observed data not used in the calibration.

**4.5 Bayesian approach**

Bayesian modeling generally starts with defining prior distributions for the model parameter \( \theta \). Once defined, the priors can be updated via Bayes’ theorem using information in the likelihood function of the data, given the model. This gives the posterior distribution of the parameter \( \theta \).
that characterizes the uncertainty around the true values of the parameter and allows for uncertainty propagation to model outputs.\textsuperscript{139}

MCMC methods are commonly used in Bayesian modeling.\textsuperscript{140} They are a class of algorithms that draws random values of the model parameter $\theta$ from an approximate distribution and then corrects those draws to better approximate the posterior distribution. The samples are drawn sequentially, with the distribution of the sample draws depending on the last value drawn; hence the draws form a Markov chain. With each step of the simulation, the approximate distribution is improving, in the sense of converging to the posterior distribution.

**Bayesian evidence synthesis**

Jackson et al. describe a Bayesian evidence synthesis (BES) method that incorporates model calibration, with illustration using a Markov model of human papillomavirus (HPV) infection (Table 3).\textsuperscript{117} With this approach, model calibration is an integral part of a BES model in which the broad term ‘‘model’’ consists of not only a mathematical representation of the disease process, but also the network of distributional and functional relationships that connects the parameters of that process with observed data and prior information.

Jackson et al. used a series of examples to illustrate the BES model. This selected example illustrates part of the technical details. The diagnosed prevalence of HPV-16 infection detected by a HPV test at age $t$, $p^D(t)$, is expressed as

$$p^D(t) = p^H(t) + p^{FP}(t)(1 - p^H(t)),$$

where $p^H(t)$ denotes the probability of being infected with the HPV-16 virus (assuming 100% sensitivity), and $p^{FP}(t)$ the probability of a false-positive of HPV-16 detection among non-infected individuals. The true HPV-16 prevalence $p^H(t)$ is itself assumed to be a non-linear function of age $t$ (over the age range from 10 to 70 years) with three unknown parameters. Upon obtaining the posterior distributions of these parameters, this function is used to extrapolate prevalence observed in women aged 20 to 64 years from an HPV testing trial to younger (or older) ages $t$.

The BES model is structured as a graphical network. The network provides a basis for an iterative MCMC simulation that generates samples from posterior distributions of model parameters and their functions. In the above example, the diagnosed prevalence of HPV-16
infection \( p^D(t) \) is a node in the network. The likelihood of observing the number of HPV-positive participants (given the number of all participants in the HPV testing trial) depends (in part) on \( p^D(t) \). Also depending in part on \( p^D(t) \) is the likelihood of observing the number of cytological abnormalities from participants in a national cervical screening program. The prior distribution of \( p^D(t) \) depends on its parent, the true HPV-16 prevalence \( p^H(t) \). Given current values of all other nodes in the network, a value of the parameter \( p^D(t) \) is sampled from its full-conditional distribution, which is the product of the prior distribution of \( p^D(t) \) and the likelihood of the observed data. Through the likelihood, any data directly or indirectly generated by the node \( p^D(t) \) contribute to the posterior distribution of \( p^D(t) \).

Upon convergence of the MCMC simulation, posterior distributions of model parameters and their functions are used in model checking, verifying potential conflicts from different data sources, and assessing sensitivity of model outputs to data inclusion and prior distributions.

Welton et al. illustrate how calibration can be done as part of a BES method using a 3-state Markov model (Table 3).\(^{141}\) The method uses the Kolmogorov’s forward equation to express transition probabilities (that are generally estimated using directly observed data) as a function of underlying (unobserved) transition rates in the model. The transition rates are estimated when observations are available on all transitions and exact time at risk in each Markov state (fully observed data), and observations are available on initial and final states after a fixed interval of time but not on the sequence of transitions (partially observed data). The authors show how to calibrate the Markov model to observed data from all included studies or from one specific study. The former assumes that the calibrated model is used for a target population that is exchangeable with the populations in all studies, with inherent heterogeneity. The latter considers the possibility that one particular study best represents the target population under consideration.

**Bayesian MCMC methods for model calibration**

Rutter et al. propose a Bayesian calibration method based on two common MCMC algorithms and illustrate the method using a microsimulation model for colorectal cancer (CRC).\(^{124}\) Mathematical equations are used to characterize four components of the adenocarcinoma process, including adenoma risk, adenoma growth, transition from adenoma to preclinical
cancer, and transitions from pre-clinical to clinical cancer. Each equation contains unknown parameters in which their posterior distributions are derived from the Metropolis-Hastings (MH) algorithm embedded in a Gibbs sampler algorithm.\cite{140}

With a Gibbs sampler, the model parameter $\theta$ is partitioned into $d$ components such that $\theta = (\theta_1, ..., \theta_d)$.\cite{140} In the CRC model, the parameter $\theta$ is partitioned according to individual parameters involved in the four mathematical equations. The Gibbs sampler goes through $d$ steps in each iteration $t$ in which an ordering of the $d$ components of $\theta$ is chosen. Each component $\theta_j$ is updated conditional on the latest values of other components. Because the conditional distribution of each component given others is unknown in the microsimulation CRC model, the authors use the MH algorithm for each step of the Gibbs sampler.

The MH algorithm assumes that one can compute the value of a function $f(\theta)$ that is proportional to the posterior density $p(\theta \mid data)$.\cite{140} The second ingredient of the algorithm is a jumping function $J(. \mid .)$ that is used to obtain a proposal for the new value of $\theta^t$ given the previous value of $\theta^{t-1}$. A usual choice is to let $J_t(\theta^t \mid \theta^{t-1})$ be a normal distribution centered at the previous point $\theta^{t-1}$ in the parameter space, so that points closer to $\theta^{t-1}$ are more likely to be visited next – making the sequence of samples into a random walk. In successive iterations, the algorithm considers jumping to a point $\theta^t$ from the previous point $\theta^{t-1}$ according to a set of acceptance and rejection rules. A jump is accepted if it increases the posterior density. If the jump decreases the posterior density, then the jump is accepted with a probability equal to the ratio of posterior densities $r = \frac{f(\theta^t \mid data)}{f(\theta^{t-1} \mid data)}$. The MH algorithm always accepts steps that increase the posterior probability but sometimes accepts downward steps to generate samples for the posterior distribution.

At each step $t$ of the MH algorithm, direct calculation of the posterior density $p(\theta^t \mid data)$ is not possible because of the complexity in the CRC model.\cite{124} For example, when calibrating to observed cancer incidence, the associated probability of incident cancer depends on the full CRC model. Within the MH algorithm, the authors propose an imbedded simulation of $m$ steps with the CRC model to obtain the likelihood of the projected incident cancer. The posterior
probability $p(\theta^t \mid data)$ is then derived from the likelihood of the projected data and the prior distribution of the parameter $\theta^t$.

The authors show that their proposed MCMC approach converges to the posterior distribution of the parameter $\theta$ and use a simulation sub-study to demonstrate the performance of the proposed MCMC approach. They calibrate the CRC model to determine unknown transitions of the adenocarcinoma process. The identifiability of each transition parameter is assessed by comparing the prior $p(\theta)$ and posterior $p(\theta \mid data)$ distributions of the parameter visually and by using an overlap statistic, defined as $\int \min(p(\theta), p(\theta \mid data))d\theta$. This overlapping measure ranges from 0 to 1, with values closer to 1 corresponding to greater similarity between the prior and posterior distributions, suggesting a lack of identifiability of the parameter.

Whyte et al. propose a similar Bayesian MCMC method for the calibration of a population-based state-transition CRC model using screening and epidemiological data from the U.K. (Table 3). The MH algorithm is used to estimate transition parameters of the natural history of CRC and diagnostic characteristics the fetal occult blood screening test. The observed data $p(data \mid \theta)$ is assumed to be normally distributed and centered on the predicted values from the model. This distribution is used to derive the acceptance and rejection rules in the MH algorithm. With this design of the MCMC algorithm, the authors suggest that the algorithm is simple to code (in Visual Basic) and run (overnight on a standard desktop PC).

**Approximate Bayesian Computation (ABC)**

Seigneurin et al. use a microsimulation model to replicate standardized incidence rates of breast cancer (BC) in a region of France from 1991 to 2006 (Table 3). The aim is to quantify the magnitude of overdiagnosis from non-progressive disease detected by screening mammography, adjusting for lead time bias, secular trend in the underlying BC risk, and opportunistic screening. The proportion of non-progressive BC cancers is considered as an unknown parameter in the model and estimated via an ABC analysis.

Figure 2 displays a hypothetical relationship between some accepted values of a parameter $\theta$ and their projected summary statistics $S_i$’s (e.g., annual standardized incidence rates of invasive cancer and carcinoma in situ) that are close to the observed summary statistic $S_{obs}$ within a tolerance window $\pm \epsilon$. An acceptable value of $\theta_i$ is corrected to $\theta^*_i$ by using the regression line
between $\theta$ and the summary statistic $S$. The fitted regression line is estimated by weighting the accepted values of the parameter $\theta$ by some distance measure between $S_i$ and $S_{obs}$. Under certain conditions, the set of corrected parameter values $\theta_i^c$ forms a sample from the posterior distribution.\(^{143}\)

On a technical note, ABC analysis has its roots in rejection sampling and has been suggested to be suitable for solving complex problems with mathematically or computationally intractable likelihood.\(^{143}\) Although the method arose in population genetics, ABC is increasingly used in other fields, including epidemiology, systems biology, ecology, and agent-based modeling.\(^{142, 144, 145}\)

**Other potentially relevant studies**

There are other potentially relevant studies that are not included in Table 3. The Bayesian approaches that are used to calibrate the CRC models are also briefly discussed in subsequent publications.\(^{116, 124}\) Bogaards et al. use a BES model to estimate disease transition and sexual contact parameters in a transmission model of high-risk HPV infection.\(^{146}\) Vergel et al. use a BES and decision-analytic model in a cost-effectiveness analysis of primary angioplasty.\(^{147}\) Other studies outline the principles of BES applied to decision-analytic modeling.\(^{148-150}\)
Figure 2. An illustration of Approximate Bayesian Computation


Notes: Linear regression is conducted within the tolerance window ±ε in order to adjust the good-fit value $\theta_i$ to $\theta_i^*$ so that projected data from the decision model are more consistent with observed data (see text).
4.6 Guided search approach

As discussed previously, a guided search strategy iteratively learns from the currently selected set of parameter values and moves towards a better fitting set.\textsuperscript{102} A random search strategy selects parameter sets at random throughout the search.

**Generalized reduced gradient (GRG) search**

*Karnon* et al. illustrate a stepwise approach to calibration by comparing a random search strategy with a guided search strategy in the calibration of a Markov model simulating the prognosis of early breast cancer. The model was used for a cost-effectiveness analysis of adjuvant therapy with tamoxifen after primary surgery in women with early breast cancer (Table 4).\textsuperscript{105} The model is calibrated to survival data in order to estimate relapse rates, progression and post-relapse mortality rates due to BC and other causes. Chi-squared and log likelihood GOF measures are used to compare projected and observed overall survival data at each of the four different time points after surgery. A composite GOF measure is derived by summing the time-specific GOF measures with equal weights.

The breast cancer model is built in Microsoft® Excel, leading in part to the decision by Karnon et al. to select the GRG search algorithm (available as tool in an Excel add-in package).\textsuperscript{102, 105} For example, let the function $f(\theta)$ be the chi-squared GOF measure discussed above. The search makes use of the gradient of the function $f(\theta)$, which is an $n$-parameter vector of partial derivatives $\nabla f(\theta) = (\frac{\partial f(\theta)}{\partial \theta_1}, ..., \frac{\partial f(\theta)}{\partial \theta_n})$.\textsuperscript{151} Locally, the gradient vector $\nabla f(\theta)$ points in the direction of the greatest rate of increase of $f(\theta)$. Hence $-\nabla f(\theta)$ points locally in the direction of the steepest descent of $f(\theta)$. Gradient descent search is based on the observation that if one goes from a point $\theta^{(0)}$ in the direction of the steepest descent direction $-\nabla f(\theta^{(0)})$ by some *sufficiently* small step to a point $\theta^{(1)}$, then $f(\theta^{(0)}) \geq f(\theta^{(1)})$. The gradient search continues the descent from $\theta^{(1)}$ and repeatedly descends from successive points along a steepest descent path until it reaches a minimum point.

In the calibration, the GRG search is set to start at multiple starting points (that are systematically selected to cover the parameter search space) in order to identify a set of multiple local minima (and a global minimum).\textsuperscript{102, 105} Further start points are selected until no further
improvements in the composite GOF scores are identified. The search identifies a final ensemble of good-fit parameter sets according to two alternative convergence criteria (Table 4).

To integrate the calibration results with the cost-effectiveness analysis, a probabilistic sensitivity analysis is conducted in which the final ensemble of good-fit parameter sets is sampled according to their probability weights (i.e., this is the 7th step in the proposed seven-step approach to calibration). The weight assigned to each parameter set is estimated as the reciprocal of its composite GOF measure, divided by the sum of the reciprocals across all acceptable parameter sets (Table 4). The reciprocal is used because smaller absolute values of the GOF measure represent better fitting parameter sets, thus requiring higher probability weights.

Compared to no calibration, the results show that the calibrated models consistently provide higher mean estimates of projected life years gained with adjuvant therapy (and with reduced model uncertainty (Table 4). The chi-squared GOF measure differentiated between the accuracy of different parameter sets to a far greater degree than the likelihood GOF measure, in the sense of better consistency between observed and projected data. Compared with the random search strategy, the guided search strategy also produces higher mean estimates of projected life years gained.

**A simulated annealing algorithm**

*Kong* et al. calibrate a microsimulation lung cancer policy model to a series of primary targets (i.e., cancer incidence by lung cancer cell type) and secondary targets (i.e., age-specific mortality rates of non-smokers) in order to estimate coefficient parameters \( \theta \) of logistic regression equations that predict the development of adenocarcinoma, and lung cancer with large, small or squamous cells (Table 4). Given a model parameter \( \theta \), the global GOF function \( f(\theta) \) is considered as a multi-dimensional surface with peaks (that are defined by poor fitting parameter sets), and valleys (by good-fit parameter sets). A simulated annealing algorithm and a genetic algorithm are used to locate a set of global minima of the function \( f(\theta) \).

A simulated annealing algorithm operates on a cooling schedule that includes a temperature function \( T(t) \) at time \( t \), and a temperature decrement strategy \( \Delta T(t) \) that allows for fine-tuning the search. At high temperature \( T(t) \), the algorithm allows more jumps and jumps that likely
reduce the values of the function $f(\theta)$. The algorithm also permits uphill moves with increasing values of $f(\theta)$ to allow chances of jumping out of local minima and potentially falling into a more promising downhill path to a set of global minima.\textsuperscript{153} At low temperature $T(t)$, the search allows less jumps and only jumps with large decrease in the function $f(\theta)$. The faster the temperature decrement $\Delta T(t)$, the higher the probability of being trapped in a local minimum. A slow decrement however causes the heuristic to be unacceptably slow. In the study, the search is initiated at 10 random locations, with each search limited to 1000 iterations (additional details in Table 4).

The acceptance-rejection rule that permits jumps in simulated annealing is adopted from the Metropolis algorithm.\textsuperscript{154} If the temperature $T(t)$ is held constant over time $t$, the sequence of moves from the simulated annealing heuristic forms a homogeneous Markov chain. Under broad conditions and as $T(t) \downarrow 0$, it can be shown that the sequence of moves converges to a set of global minima of the GOF function $f(\theta)$.\textsuperscript{153}

A genetic algorithm

Kong et al. also use a genetic algorithm in the above study (Table 4).\textsuperscript{127} A genetic algorithm requires a genetic representation of the multi-dimensional search space, and a fitness function to evaluate solutions of the search.\textsuperscript{155} To start, the initial generation consists of a population of candidate parameter sets. Parameter sets are evaluated according to their GOF scores. Each GOF score is used to define the fitness of a candidate relative to the fitness of other candidates in the generation. The candidate with the largest lack-of-fit value has a zero probability of reproduction. Using a one-point crossover operation, the encoded genetic history of two candidate parameter sets are combined to produce new parameter sets of the next generation. Each newly generated parameter set is also subjected to random changes of individual parameters, with new values chosen within 10% of the values of the parent parameter sets. The algorithm is run until a total of 1000 parameter sets have been tested (Table 4).

Kong et al. show that the simulated annealing algorithm is slightly better than the genetic algorithm for calibrating the model in the study (Table 4).\textsuperscript{127} Subsequently, the authors used simulated annealing but not genetic algorithms in other studies.\textsuperscript{156, 157}
Nelder-Mead downhill simplex search

Taylor et al. calibrate a Markov model of cervical cancer using different search strategies, including the Nelder-Mead downhill simplex search strategy, a random search strategy (with 100,000 input parameter sets), and a manual search strategy (by an analyst who uses the best-fitting parameter sets identified by the random search as a starting point). The numbers of calibration parameters and other model inputs are large (Table 4).

The Nelder-Mead downhill simplex is a function minimization algorithm that only requires function evaluation. The algorithm makes almost no assumption about the function being minimized. A simplex is a geometrical figure consisting, in N dimensions, of N+1 points (or vertices) and their interconnecting line segments. In two dimensions the simplex is a triangle. In three dimensions it is a tetrahedron. The downhill simplex method must be initialized with N+1 points (e.g., candidate parameter sets). To locate a local minimum, the search takes a series of steps, including reflection, expansion, contraction, and reduction of the simplex. Given a current simplex, the next step just moves the vertex of the simplex where the lack-of-fit value is largest (‘highest point’) through the opposite face of the simplex to a lower point to search for potential areas of the parameter space that might better fit the data. The movement of the simplex resembles an amoeba searching for a ‘valley floor’. In order to locate multiple local minima, the search is usually conducted a few times with different simplexes.

Taylor et al. show that the Nelder-Mead downhill simplex method obtained the best-fit parameter set with a mean deviation between observed and projected targets of 7% over 1640 iterations. The corresponding deviation was 39% with a restricted random search method (which was aborted after 35 days of run time), and 10% with the manual search involving 100 iterations of fine tuning over 4 days of a data analyst’ time.

Other potentially relevant studies

There are other potentially relevant studies that are not included in Table 4. Jit et al. use a hybrid approach of a simulated annealing algorithm (as a global search) and the generalized reduced gradient algorithm (as a local search) to estimate progression rates for HPV infection from epidemiological data. Hur et al. use simulated annealing to calibrate a population-based
Penny et al. calibrate a stochastic simulation model of malaria to 61 field scenarios from sub-Saharan Africa using a genetic algorithm.\textsuperscript{161}

4.7 Random search approach

A random search strategy selects parameter sets at random throughout the search. Random search approaches to calibration described below are conducted in stages with different random sampling techniques, including simple random sampling, stratified random sampling and Latin hypercube sampling (Table 5). With simple random sampling, random values for the set of $d$ parameters $\theta = (\theta_1, \ldots, \theta_d)$ are independently sampled from each of the input distributions of $\theta_1, \ldots, \theta_d$.\textsuperscript{162} With stratified random sampling, a set of $n$ values of a parameter $\theta_j$ is generally obtained by dividing the plausible range of the parameter into $n$ strata; then a single point is randomly selected from each stratum.\textsuperscript{162} Latin hypercube sampling generates a collection of points (or a design) in a multidimensional space that spread points evenly over the range of each input separately.\textsuperscript{162} Figure 3 illustrates a Latin hypercube sampling design with 5 points. The points are selected to form a Latin square such that in each vertical or horizontal grid of cells, each interval is selected exactly once.

Figure 3. An illustrated Latin hypercube design

Three-stage random search with simple random sampling

Salomon et al. develop a state-transition model of hepatitis C virus (HCV) infection. Health states in the model are defined according to serologic infection status with HCV or hepatitis B
virus (HBV), and clinical liver disease status (Table 5).\textsuperscript{130} The authors use calibration to gain insights into uncertain transition parameters from infections to advanced liver disease. The random search strategy is conducted in 3 stages. In the first stage, the search is conducted with 50,000 random parameter sets drawing from independent uniform distributions spanning the plausible ranges of transition parameters to identify the best 90 good-fit parameter sets: 30 each for HCV, HBV prevalence, and HCC mortality. In the second stage, the search is conducted with 27,000 new parameter sets (30 HCV x 30 HBV x 30 HCC) to identify the best 50 good-fit parameter sets. Each of these sets is then perturbed to generate a local sample of 1000 parameter sets by sampling from independent, uniform distributions spanning 90-110 percent of its parameter estimates. In the third stage, the search is conducted with 50,000 of these newly generated parameter sets to identify the best 50 good-fit sets.

**Two-stage random search with simple random sampling**

Kim et al. use a two-stage random search approach to calibrate a microsimulation model of cervical cancer (Table 5).\textsuperscript{131} The first stage explores the influence of individual parameters and establishes plausible ranges of all input parameters by calibrating the model to primary data from a longitudinal study of HPV infection and cervical neoplasia. The second stage is intended to vary all input parameters over the identified plausible ranges to calibrate the model to epidemiologic data from multiple sources.

In the first stage, the authors simulate from the distributions of baseline characteristics of the longitudinal study cohort (e.g., enrollment age, HPV status) to generate a dataset of 100,000 women that is a replication of the cohort study dataset of over 2400 participants (Table 5). The same statistical procedures that were used to analyze the study data are used with the simulated data. Parameter values are systematically varied over a broad range to explore the independent influence of these parameters on model outputs and to visually improve the GOF of model outputs to calibration targets. The age patterns of probabilities governing the transition between health states are assumed to be fixed but the magnitude of these probabilities varies; as such, the authors apply multipliers to the parameters and vary the multipliers over a broad range. Plausible ranges for all parameters are defined by parameter values in which model projections fall within or nearing the 95 percent confidence intervals of the corresponding calibration targets.
In the second stage, the authors sample over the ranges of multipliers to conduct multiple simulations of the model. Specifically, one value for each multiplier is randomly selected from a uniform distribution over the identified range. In total, simulations are conducted with 555,000 sampled parameter sets. Model outcomes using each parameter set are compared with multiple epidemiologic targets (Table 5).

With respect to GOF measures, likelihood functions are specified for each calibration target (e.g., prevalence of HPV), assuming that each target follows an independent normal distribution. An overall GOF score is computed as negative two times the sum of the log-likelihood scores for each target. In terms of acceptable parameter sets, the best-fitting parameter set is identified as the one with the lowest overall GOF score. The likelihood ratio test is then used to identify a subset of good-fitting parameter sets, comprising those parameter sets that do not produce a significantly worse fit than the best-fitting parameter set at a significance level of 5 percent. In sensitivity analysis, the authors considered alternative scoring algorithms for assessing GOF and alternative thresholds for acceptable parameter sets.

**Two-stage random search with stratified random sampling**

Fryback et al. develop the Wisconsin breast cancer epidemiology simulation model to replicate breast cancer incidence and mortality for the U.S. population from 1975 to 2000 (Table 5). The model is a discrete-event, stochastic simulation model. Model components are governed by a large number of fixed inputs and 29 parameters, including natural history parameters and breast cancer detection probabilities. The calibration targets are five age-adjusted curves summarizing incidence rates for in situ, localized, regional, and distant breast cancers, and breast cancer mortality rates in the U.S. over time.

Biologically plausible ranges are set for each model parameter (Table 5). A vector of parameter values is drawn at random by drawing each parameter from an independent uniform distribution over a discrete partition of its range so that all ordinal constraints on detection rates and probabilities are satisfied. Each projected incidence curve is assigned a score by using acceptance envelopes around the corresponding observed incidence curve. The envelopes are specified heuristically to encompass variation around the incidence rates that may naturally be expected given the size of the simulated population. Simulation output is scored by counting the number of time points across the four incidence curves at which the simulation-based incidence
rate falls outside of the envelopes; the best possible score is zero, and the worst possible score is 104 (= 4 stages × 26 time points from 1975 to 2000). Empirically the authors determine a minimally acceptable parameter set to have a score ≤10, and exceptionally good parameter set to have a score ≤5.

The calibration is carried out in two stages (Table 5). In the first stage, the authors assume all onset breast cancers progress to be invasive breast cancer, and eventually lethal, if undetected and untreated. Under this assumption they do extensively sampling with 475,000 random parameter sets but cannot obtain any acceptable solutions; the best solution is with a score of 21. In the second stage, they introduce three more parameters that characterized the natural regression of a fraction of breast tumors with limited malignant potential (LMP). Allowing the LMP fraction to be >0, but ≤10%, improves the best score to 15 (from 289,000 candidate parameter sets), and all solutions with scores near 15 has LMP fractions at 9-10%. Ad hoc sampling indicates that much better solutions are available with LMP fraction at 30-50%, so the authors focus sampling for LMP fraction in this range, still allowing other parameters to vary throughout their plausible ranges. With these added parameters, they sample 30,188 parameter sets and found 91 parameter sets with scores ≤5 and 363 parameter sets with scores ≤10. The best scoring model suggested the potential regression of LMP tumors, although no such depletion in the occult pool of localized invasive cancers has been observed in the Surveillance, Epidemiology, and End Results data within the interval from 1975 to 2000.

The authors report issues related to non-identifiability of parameters in their model calibration. The results indicate that good-fit parameter sets tend to fall in four relatively compact clusters in the parameter space, connected by thin bridges of good-fit sets. These clusters seem to be distinguished by a faster versus slower mean for the tumor growth rate distribution, plus some compensatory changes in the growth rate distribution variance and the lag between tumor onset and average tumor incidence. The authors suggest that the biological implications of these clusters should be explored.

**A random search strategy with simple random sampling**

Berry et al. describe a Bayesian approach to model the impact of treatment and screening on breast cancer (BC) mortality (Table 5). A microsimulation model is developed to simulate various BC-related events for each woman in a simulated cohort of 2,000,000 US women from
1975 to 2000. The model accounts for the stage shift among tumors that are detected via screening mammograms, as well as accounts for the possibility of an additional survival benefit due to screening beyond stage shift. The model is calibrated to cancer-specific mortality incidence in order to estimate the reduction in the risk of BC mortality associated with tamoxifen, chemotherapy and mammography screening in the U.S. from 1990 to 2000. Separate and independent prior distributions are assumed for parameters pertaining to the reduction in the risk of BC-related mortality due to adjuvant tamoxifen, adjuvant chemotherapy, and screening effect beyond stage shift.

For each of the 80,000 simulated cohorts, a different parameter set is sampled from the prior distributions. Posterior distributions of model parameters are derived using the “rejection method”. This method is used because the likelihood function involved in the posterior estimation is complicated. Specifically, acceptance windows are pre-defined around BC mortality incidence from 1975 to 2000 (±2.5 per 100,000) and around the estimated slopes in incidence (±0.17 per year) over three 5-year periods from 1985 to 2000. Parameter sets are accepted to be from the posterior distribution if they fall within all the acceptance windows.

**Single-stage random search with Latin-hypercube sampling**

*Blower et al* develop a dynamic transmission model of live attenuated HIV (LAHV) vaccines to predict the trade-off between vaccine efficacy and safety with respect to the burden of HIV infection and AIDS in Zimbabwe and Thailand (Table 5). To establish the initial model conditions, the authors used Latin hypercube sampling to generate random samples from probability distributions assumed for input parameters of the differential equations that govern the temporal dynamics of individuals in five groups, including susceptible individuals, unvaccinated individuals infected with wild-type HIV, individuals uninfected with the wild-type strain but infected with the vaccine strain, individuals dually infected with the vaccine and the wild-type strain, and individuals with AIDS. The model is calibrated using annual AIDS mortality rates in Zimbabwe and Thailand.

The calibrated model at initial conditions (without vaccination) is used to conduct an uncertainty analysis of the endemic equilibria to predict whether mass vaccination with LAHV vaccines could eradicate wild-type HIV in Zimbabwe and Thailand. Latin hypercube sampling is used to
generate 1,000 different parameter sets that define the characteristics of different mass vaccination campaigns. The results suggest that vaccination with any LAHV that causes more than 5% of vaccinated individuals to progress to AIDS in 25 years would lead to perversity (i.e., increase the annual AIDS death rate) in Thailand (where transmission is low); these same vaccines would lead to decreases in the annual AIDS death rate in Zimbabwe (where transmission is high).

**Other potentially relevant studies**

Random search approaches have been used in other studies (that are not included in Table 5).\(^{105,164-169}\) Most notable is a detailed report on the development and calibration of an HPV and cervical cancer model for analyses of screening and vaccination programs.\(^{164}\)

### 4.8 Grid search approach

With a grid search approach, the parameter search takes place across the different possible combinations of parameter values in the parameter space. Weinstein et al. develop a coronary heart disease (CHD) model to project mortality, morbidity and cost of CHD for the United States population (Table 6).\(^{136}\) Substantial uncertainty exists for some model parameters, such as CHD incidence (particularly in the oldest age range) and case-fatality rates associated with myocardial infarction (in the oldest and lowest age ranges). The authors vary these parameters over their plausible ranges, simulate all combinations within this hypercube of the parameter space, and select combinations that lead to better agreement between modeled projections and corresponding national vital statistics and hospital discharge data from 1980. Parameter sets are selected if model projections are within 10% for all targeted outputs, and within 5% for most of the targeted outputs.\(^{107}\) The calibrated model is used to evaluate the population aging effect on CHD burden in 2010. The model is validated using national data from 1986 and re-calibrated accordingly.\(^{170}\)

Yeh et al. develop a Markov model for gastric cancer (Table 6).\(^{132}\) The model is calibrated to specific targets for four subgroups of men and women in China (24 precancerous and 16 cancerous targets) and Columbia (16 and 12). Model calibration is to estimate progression and regression parameters across health states, defined on the basis of normal gastric mucosa, chronic non-atrophic gastritis, gastric atrophy, intestinal metaplasia, dysplasia, and gastric
cancer. Disease progression and regression parameters are varied simultaneously according to a multidimensional grid that includes 65,000 unique candidate parameter sets. Projected targets associated with each parameter set are compared with observed targets. Good-fit parameter sets are selected using a procedure similar to that described above in the study by Kim et al. The model is calibrated for each of the four subgroups. Distributions of good-fit parameters are assessed across subgroups to gain insights into disease progression and regression.

Rydzak et al. describe a series of exercises related to the adaptation of a validated microsimulation model of HIV and AIDS to the Women’s Interagency HIV Study (WIHS) cohort (Table 6). Calibration targets include 24-month survival curves stratified by treatment status (untreated HIV-infected women and women treated with highly active antiretroviral therapy - HAART) and baseline CD4 cell count. The calibration is conducted in three steps. First, parameters relevant to the natural history of HIV infection in untreated women are systematically varied. Projected survival curves are visually compared with observed survival curves. Secondly, for HIV-infected women receiving HAART, parameters pertaining to heterogeneity of treatment response, treatment adherence and loss to follow-up are varied. Projected survival curves are visually compared to observed survival curves. Parameter sets with low absolute differences between projected and observed survival rates (at 6, 12, 18 and 24 months) are considered to be more consistent with the observed data. Thirdly, projected outputs (median time on treatment and median time to switching HAART regimen) from the 50 best-fit parameter sets are compared with corresponding estimates derived from a distinct subset of data from the WIHS study that is not used in the adaptation.

Findings from this calibration exercise indicate that the most influential model assumption in untreated women is to be related to chronic HIV-associated mortality following an opportunistic infection. The most influential assumption in treated women is to be related to the clinical effectiveness of HAART and the ability of HAART to prevent HIV complications independent of virologic suppression. Information from the identified good-fitting parameter sets suggest reductions in the clinical effectiveness of 1st and 2nd line HAART and improvements in 3rd and 4th line regimens.

Other relevant studies
The grid search approach is also used in the calibration of other HIV models.\textsuperscript{171, 172}

4.9 Manual search

The studies by Fryback et al. and Taylor et al. described above use a combination of manual and random searches to identify solutions to the calibration problem.\textsuperscript{128, 129} It has been suggested that automated techniques alone may not be able to obtain solutions to the calibration of complex physics-based simulation models; it is a combination of the knowledge of the process under simulation and the techniques that yield solutions to the calibration problem.\textsuperscript{122}

4.10 Discussion

We review existing parameter search strategies that are used to calibrate decision-analytic models. The strategies and methods are identified from selected studies of decision models that report details of the calibration process. We abstract details pertaining to each step of the calibration process to provide context for the review. The context of the decision model is also described in the review of the search strategy. We summarize the fundamental concepts underlying the strategies as well as the specific details of their applications. Based upon their fundamentals, the strategies are broadly categorized into Bayesian, guided search, random search and grid search approaches. Common and complimentary aspects of the strategies across their broad approaches are identified.

One of the limitations of our review is that it is not fully systematic and there are further studies and methods that we were not able to cover. The studies we selected must explicitly describe the calibration process as part of the model description. This is not the case for other potentially relevant studies that use other terms to describe model calibration, such as ‘model fitting’ or ‘model identification’.\textsuperscript{101} In particular, we did not search for studies that did not explicitly state that calibration was used, but did provide references regarding model calibration.\textsuperscript{166}

Table 7 summarizes a number of suggested methodological issues that emerge from this review. These issues are outlined below; we subsequently discuss how each calibration approach deals with these issues.
- Posterior distribution - the approach yields the posterior distribution of model parameters or an approximation to the posterior distribution.
- Uncertainty propagation - there is a provision to propagate uncertainty in model parameters to the relevant model outputs (e.g., life years).
- Data conflicts - there is a provision to identify and resolve the conflicts between modeled projections and multiple data points.
- Non-identifiability - there is a provision to identify model parameters that are not identifiable during the calibration such that there is a large overlapping between the prior and posterior distributions.
- Accessibility - model calibration should not be restricted to applications that use less accessible modeling techniques.

**Bayesian approach** - Approaching model calibration from a Bayesian modeling perspective is advantageous because there are established methods and practice to address these methodological issues. Within the Bayesian approach category, at least one of the five included studies addresses these issues, including obtaining posterior distributions of model parameters (5 studies), simultaneously propagating of parameter uncertainty to model outputs (5), identifying and resolving conflicts between modeled projections and multiple data points, non-identifiable parameters that are non-identifiable with large overlap between their prior and posterior distributions. Bayesian methods to calibration are however more complex than other approaches, as illustrated in the technical details pertaining to the specification of the Bayesian evidence synthesis network in one study, and the design of the Metropolis-Hastings algorithm embedded in a Gibbs sampler algorithm in another study. We therefore suggest that Bayesian methods are relatively less accessible than other calibration approaches.

**Guided search strategies** - We review guided search strategies using search algorithms that are well validated and widely used in other fields of studies (e.g., engineering and optimization). These algorithms are efficient in identifying the set of local optima (and the global optimum) associated with solutions of a calibration problem, often with small number of model evaluations. Most notably, the general algorithm for simulated annealing is an adaptation of the Metropolis-Hasting algorithm that is widely used in Bayesian modeling for obtaining posterior distributions of model parameters.
A simulated annealing algorithm has been shown to be slightly more efficient than a genetic algorithm in the calibration of a lung cancer policy model.\textsuperscript{127} The guided search strategy with the generalized reduced gradient search algorithm has been shown to be more accurate than a random search strategy, in the sense of projected outputs from the calibrated model are more consistent with the observed data.\textsuperscript{105} A proposed hybrid approach combining a simulated annealing algorithm (as a global search) with the generalized reduced gradient search (as a local search) has been used to calibrate an HPV infection and cervical cancer model.\textsuperscript{159} The Nelder-Mead downhill simplex algorithm has been shown to be efficient compared to a random search approach in the calibration of a HPV model.\textsuperscript{128}

Two of the included studies with a guided search strategy did not report information regarding the uncertainty associated with the best-fit estimates.\textsuperscript{127, 128} In principle, the confidence region of a locally best-fit estimate can be approximated based upon the estimated covariance matrix of the parameters of interest.\textsuperscript{174} However, confidence regions were not reported in any of the included studies with a guided search strategy. One study reports the propagation of uncertainty in the good-fit parameter sets to a cost-effectiveness evaluation through the use of probabilistic sensitivity analysis.\textsuperscript{105} None of the included studies with a guided search strategy addresses methodological issues related to approximate posterior distributions, data conflicts and non-identifiability (Table 7).

\textit{Random search approach} – The random search strategies we review are generally conducted in stages to allow for using results from the current stage of the search to improve the search in the next stage. The strategies use different random sampling schemes, including simple random sampling, stratified random sampling and Latin hypercube sampling. Varying the stages and the sampling schemes allows for considerable flexibility to deal with a number of reported scenarios during model calibration, including implausible model structure that was in conflict with observed data;\textsuperscript{129} stage-specific targets that are very different (e.g., individual patient data and aggregate summary statistics);\textsuperscript{131} or broad search versus locally search.\textsuperscript{130} The ensemble of good-fit parameter sets obtained from a random search strategy is used to demonstrate the uncertainty associated with the calibrated estimates.\textsuperscript{126, 129-131}
Random search strategies have been used to calibrate complex decision models, but in high dimensional space, the search could be very time consuming. The searches we review used very large and arbitrary numbers of candidate parameter sets to provide adequate coverage of the parameter search space. Despite this, it is unclear whether the ensemble of good-fit parameter sets identified from the search would contain a reasonable estimate to the best-fitted value and would adequately characterize the associated uncertainty. Results from one study suggest that an approximate Bayesian computational analysis can be conducted on an ensemble of good-fit parameter sets to obtain approximate posterior distributions of model parameters.

The included studies using a random search strategy we review discuss methodological issues related to approximate posterior distributions, uncertainty propagation, data conflicts, and non-identifiability (Table 7).

**Grid search approach**- We did not attempt to assess these issues with the grid search approach as there is not enough information from the included studies.

As illustrated by the abstracted data in Tables 3-6, the technical considerations are different for each of the different steps in the stepwise approach to calibration proposed by Vanni et al and Stout et al. These considerations however are interdependent. Variation in the steps of the proposed stepwise approach is possible. For example, in the validation of computational engineering simulation models, a target for validation is structured to be consisted of field data, the definition of a distance measure between the observed and projected data, and a predefined set of conditions for which the projected data are supposed to match with the field data. This combines a few steps of the proposed stepwise approach into a single step that defines a calibration target.

In some studies included in our literature review, the uncertainty from the distributions of the calibrated parameters is propagated to subsequent cost-effectiveness analysis. As suggested elsewhere, the integration between the uncertainty distributions into the probabilistic sensitivity analysis of a cost-effectiveness analysis reduces model uncertainty and improves the relevance of model-based cost-effectiveness analysis.
Results from this literature review also support considering model calibration as an evidence synthesis exercise and if feasible, within the Bayesian modeling framework for decision analysis. From this perspective, it would be optimal to obtain not only the best estimates for model parameters but also their uncertainty distribution. This issue is particularly relevant for parameter search strategies within the guided search approach since it remains challenging to approximate confidence regions associated with locally optimal solutions to the calibration problem. For strategies within the random search approach, investigations are needed to assess whether the yielded ensemble of good-fit parameter sets could adequately characterize uncertainty distributions of the calibrated parameter estimates.

In summary, we suggest that a better understanding of the fundamental concepts of the search strategies and estimation methods (and their similarities and differences) would facilitate the uptake of model calibration and support studies evaluating existing methods or investigations into new methods for model calibration. It appears that much work is still ahead to communicate and promote better understanding of existing methods, support the development of new methods, and facilitate the uptake of these methods into routine practice for model development. In particular, we are interested in developing methods for routine calibration of decision models for cost-effectiveness analysis of alternative interventions.

4.11 Methodological question and rationale

How to structure a random search approach to calibration in order to obtain the joint estimates of model parameters?

Rationale

Random search approaches have been used to calibrate complex simulation models in computer science, environmental science, engineering, and disease history models. There is a large body of literature on random search for model calibration in other disciplines that is relevant to address the question. The random search approach to calibration is similar to the Monte Carlo approach to model evaluation for cost-effectiveness analysis, including the use of Monte Carlo approach for probabilistic sensitivity analysis and value-of-information analysis. If calibration analysis is to become a routine practice in model development, a random search approach to model calibration would be a natural fit into the
current toolbox of decision modeling. In chapter 5, we set out to explore ways to improve a random search approach that is suitable for routine calibration.
### Table 3. Calibration with Bayesian approach

<table>
<thead>
<tr>
<th>Bayesian approach</th>
<th>Decision model</th>
<th>1. Which parameters were allowed to vary in the calibration process?</th>
<th>2. Which calibration targets were used?</th>
<th>3. What measures of GOF were used?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jackson et al. 2013(^\text{117})</td>
<td>Multiparameter evidence synthesis (MPES). Illustrated HPV-16 Markov model.</td>
<td>Progression and regression probabilities from HPV infection to cervical cancer, cytological screening accuracies, HPV-16 test specificity and parameters of the age-specific function of HPV-16 prevalence.</td>
<td>Age-specific counts (e.g., women with CIN), and prevalence (e.g., CIN, ICC) from the ARTISTIC trial, UK National Health Services cervical cancer screening program, cancer registry, vital statistics, and published literature.</td>
<td>Fullconditional distribution of each node (V) in the MPES structure is defined by the prior distribution of (V) and the likelihood of observing all children of (V).</td>
</tr>
<tr>
<td>Welton et al. 2005(^\text{141})</td>
<td>Illustrated Markov models with up to 3 states</td>
<td>Underlying transition rates between Markov states</td>
<td>Transition probabilities when observations are available on all transitions and exact time at risk in each state (fully observed data), and observations are on initial states and final states after a fixed interval of time but not on the sequence of transitions in between (partially observed data).</td>
<td>Likelihoods of observed data are derived using the Kolmogorov's forward equation that links transition probabilities to underlying transition rates.</td>
</tr>
<tr>
<td>Whyte et al. 2012(^\text{116})</td>
<td>UK population-based Markov state transition model for colorectal cancer (CRC).</td>
<td>Natural history parameters characterizing the adenoma-carcinoma progression and FOBT screening test characteristics.</td>
<td>Incidence of CRC in the absence of screening (2004-2006) and after 2006, screening outcomes (e.g., polyp detection, positive FOBT rates). Data on underlying polyp prevalence from published autopsy studies.</td>
<td>Given the model, the likelihood of the observed data was taken to be normally distributed and centered on the fitted values. The joint posterior distribution factors as a product of Gaussians.</td>
</tr>
<tr>
<td>Rutter et al. 2009(^\text{124})</td>
<td>Method for Bayesian calibration. Microsimulation model for colorectal cancer.</td>
<td>Parameters for equations characterizing adenoma risk, adenoma growth, transition from adenoma to preclinical cancer, and from pre-to clinical cancer.</td>
<td>Prevalence data of adenoma and pre-clinical cancer. Progression data from pre-clinical to clinical cancer. SEER colon and rectal cancer incidence rates from 1975 to 1979.</td>
<td>Data likelihoods were derived from distributions assumed for the calibration targets, e.g. binomial (e.g., CRC prevalence), Poisson (e.g., #s of adenomas), multinomial (e.g., adenoma size).</td>
</tr>
<tr>
<td>Seigneurin et al. 2011(^\text{138})</td>
<td>Microsimulation model of breast cancer to estimate overdiagnosis from non-progressive cancer detected by screening.</td>
<td>Life-time probabilities of invasive cancer and carcinoma in situ across birth cohorts, age at onset, parameters for in situ and invasive phases, progressive and non-progressive tumours, sojourn time, and screening characteristics.</td>
<td>Standardised annual incidence rates of invasive cancers and carcinoma in situ observed from 1991 to 2006 in a region of France.</td>
<td>Euclidean distance between projected and observed values of 16 standardized incidence rates each for invasive cancers and cancer in situ.</td>
</tr>
</tbody>
</table>
### Table 3. Calibration with Bayesian approach (continued)

<table>
<thead>
<tr>
<th>Bayesian approach</th>
<th>4. What parameter search strategy was used?</th>
<th>5. What determined acceptable GOF parameter sets?</th>
<th>6. What determined the termination of the calibration process?</th>
<th>8. How was the model validated?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jackson et al. 2013(^{117})</td>
<td>Markov chain Monte Carlo (MCMC) algorithm to generate samples from posterior distributions.</td>
<td>Sampling from the converged MCMC sequences yields samples from posterior distributions.</td>
<td>Convergence of MCMC simulation.</td>
<td>Checked for internal validation. Investigated data conflicts. Conducted sensitivity analyses to assess relative fit between alternative model specifications and relative influence of priors.</td>
</tr>
<tr>
<td>Welton et al. 2005(^{141})</td>
<td>MCMC methods in WinBUGS with sampling of the underlying transition parameters from their posterior distributions.</td>
<td>Convergence of multiple MCMC chains.</td>
<td>10,000 MCMC simulations after a burn-in period of 10,000 simulations.</td>
<td>Not relevant as the models are used for illustration only.</td>
</tr>
<tr>
<td>Whyte et al. 2012 (^{116})</td>
<td>Metropolis-Hastings algorithm to generate parameter sets from their posterior distributions.</td>
<td>After convergence of the 6 MCMC chains, parameter sets were combined to derive the posterior distributions of model parameters.</td>
<td>Convergence was examined by standard diagnostic techniques for MCMC simulation.</td>
<td>The 99% credible intervals of model predictions were displayed against observed data for CRC incidence, polyp prevalence and screening outcomes. Results from the current calibration were contrasted with the calibration of a previous version of the model.</td>
</tr>
<tr>
<td>Rutter et al. 2009 (^{113})</td>
<td>Metropolis-Hastings algorithm embedded in Gibbs sampler. The MH-Gibbs sampler was used to derive posterior distributions of model parameters.</td>
<td>Samples were obtained from systematically selecting every 25(^{th}) iteration from the last 75,000 iterations from each converged chain.</td>
<td>Convergence was assessed using trace plots and Gelman and Rubin statistics for multiple chains.</td>
<td>Estimated posterior means and 95% predictive intervals of calibration parameters were compared with estimates from the calibration targets. Predicted incidence rates replicated SEER cancer incidence rates.</td>
</tr>
</tbody>
</table>
Table 3. Calibration with Bayesian approach (continued)

<table>
<thead>
<tr>
<th>Bayesian approach</th>
<th>4. What parameter search strategy was used?</th>
<th>5. What determined acceptable GOF parameter sets?</th>
<th>6. What determined the termination of the calibration process?</th>
<th>7. How was the model validated?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seigneurin et al. 2011</td>
<td>Approximate Bayesian computation analysis that bypasses exact likelihood calculation by rejection sampling based on summary statistics.</td>
<td>Model calibration was conducted with 100,000 random parameter sets. The top 500 (0.5%) with the smallest Euclidean distance between the simulated and observed values of standardised incidence rates were retained.</td>
<td>Completion of a pre-defined fixed number of randomly generated parameter sets (e.g., 100,000). Parameter values were corrected to form an approximate sample from the posterior distribution.</td>
<td>Sensitivity analyses were conducted with varying sojourn time, excluding data at the beginning of screening program, varying prior distribution for non-progressive carcinoma in situ, varying termination condition (e.g., 200,000 random sets). Compared estimates of over-diagnosis with those from other studies.</td>
</tr>
</tbody>
</table>

Notes: ICC: invasive cervical cancer. CIN Cervical intraepithelial neoplasia; CRC colorectal cancer; FOBT fecal occult blood test; HPV human papillomavirus; HR HPV high-risk HPV; ICC invasive cervical cancer; MH Metropolis Hastings algorithm; RCT randomized controlled trial; SEER Surveillance, Epidemiology, and End Results Program; WCRS Wisconsin Cancer Reporting System; MPES Multi-parameter evidence synthesis; *† Step 7 was not displayed as it is only relevant to one selected study describing a guided search strategy; this step was described in the synopsis of the study. 105
### Table 4. Calibration with guided search approach

<table>
<thead>
<tr>
<th>Guided search approach</th>
<th>Decision model</th>
<th>1. Which parameters were allowed to vary in the calibration process?</th>
<th>2. Which calibration targets were used?</th>
<th>3. What measures of GOF were used?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karnon et al. 2011 105</td>
<td>Markov model, women with breast cancer treated with tamoxifen for 5 yrs after primary surgery.</td>
<td>Relapse rates, progression and death post-relapse, other-cause mortality, and other clinically relevant parameters.</td>
<td>Target data from a meta-analysis of RCTs: 4 overall survival data points after primary surgery: 0-4 yrs, 5-9, ≥10 for women aged ≥50; 15-20 for all women.</td>
<td>Chi-squared and log likelihood GOF measures for each target. Derived composite GOF measure by summing across targets with equal weights.</td>
</tr>
<tr>
<td>Kong et al. 2009 127</td>
<td>Lung cancer policy micro-simulation model.</td>
<td>Population-specific coefficient parameters of logistic regression equations predicting the development of 3 of the 4 most common types of lung cancer.</td>
<td>Primary targets: cancer incidence by cell type, stage-specific survival rates, stage distribution at diagnosis. Secondary targets: age-specific mortality rates of non-smokers, cancer-specific mortality ratios for current versus never smokers.</td>
<td>GOF measure: sum-of-squared error for each target. Global GOF measure is defined as a weighted sum of GOF measures across primary (weight of 1) and secondary (0.5) targets.</td>
</tr>
<tr>
<td>Taylor et al. 2010 128</td>
<td>Markov model of HPV infection and cervical disease in the US.</td>
<td>A total of 79 parameters were calibrated and a total of 565 parameters were directly or indirectly varied in the calibration.</td>
<td>Thirty-five calibration targets, including age-specific and overall age-adjusted prevalence of CIN1, CIN2 and CIN3, cervical cancer incidence and mortality.</td>
<td>Absolute deviation between predicted and observed target (divided by the observed target). Global deviation was derived with weights of ICC incidence and mortality endpoints of 6- to 3-fold greater than CIN endpoints, respectively.</td>
</tr>
</tbody>
</table>
### Table 4. Calibration with guided search approach (continued)

<table>
<thead>
<tr>
<th>Guided search approach</th>
<th>4. What parameter search strategy was used?</th>
<th>5. What determined acceptable GOF parameter sets?</th>
<th>6. What determined the termination of the calibration process?</th>
<th>8. How was the model validated?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karnon et al. 2011</td>
<td>Guided search with generalized reduced gradient method and random search.</td>
<td>Two acceptance criteria: modeled outputs fulfilled all four 95% CIs of targets and fulfilled any three of the four 90% CIs.</td>
<td>Random search: an iterative stopping rule was used that doubles the sample size of the search after each iteration. Guided search: small changes in the objective function between successive iterations.</td>
<td>Comparing the distributions of discounted life-years produced by probabilistic sensitivity analyses using calibrated and non-calibrated input parameters.</td>
</tr>
<tr>
<td>Kong et al. 2009</td>
<td>Guided search: simulated annealing and genetic algorithm were used in the search for good-fit-parameter sets.</td>
<td>The best-fit parameter set from each algorithm was chosen for the calibrated model.</td>
<td>Multiple starting points; each run limited to 1000 iterations or the returned global GOF value below a predefined stopping value.</td>
<td>Internal validation against target data: e.g., annual incidence of lung cancer, cell type distribution, and stage distribution at diagnosis.</td>
</tr>
<tr>
<td>Taylor et al. 2010</td>
<td>The main search strategy was Nelder-Mead downhill simplex method. Random search was conducted with 100,000 input parameter sets. Manual calibration was conducted by an analyst who used the best-fitting random search parameter set as a starting point.</td>
<td>Not explicitly reported. It appeared that the best-fit parameter set associated with the smallest GOF measures was selected.</td>
<td>The random search of 100,000 random parameter sets was discontinued after 35 days. Manual calibration took over 4 days (e.g., 100 iterations). The downhill simplex method took 13 days.</td>
<td>Comparing of aggregated population endpoints by calibration method.</td>
</tr>
</tbody>
</table>

Notes: CIN Cervical intraepithelial neoplasia; GOF Goodness-of-fit; HPV human papillomavirus; ICC invasive cervical cancer; RCT randomized controlled trial; *+ Step 7 was not displayed as it is only relevant to one selected study describing a guided search strategy; this step was described in the synopsis of the study. 105
### Table 5. Calibration with random search approach

<table>
<thead>
<tr>
<th>Random search approach</th>
<th>Decision model</th>
<th>1. Which parameters were allowed to vary in the calibration process?</th>
<th>2. Which calibration targets were used?</th>
<th>3. What measures of GOF were used?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fryback et al. 2006</td>
<td>Wisconsin Breast Cancer Epidemiology Simulation Model, a discrete event simulation model.</td>
<td>Natural history parameters; breast cancer detection probabilities; baseline cure fractions for in situ, local, regional, and distant stage tumors.</td>
<td>Stage-specific incidence statistics from the Wisconsin Cancer Reporting System (WCRS) and National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER).</td>
<td>Each parameter set was assigned a score by using acceptance envelopes around the observed SEER and WCRS incidence rate curves. Parameter set was scored by counting the number of time points across the four graphs at which the projected incidence rate fell outside of the envelopes.</td>
</tr>
<tr>
<td>Salomon et al. 2002</td>
<td>State transition model of Hepatitis C virus infection in the United States.</td>
<td>Parameters related to HBV and HCV infection (vertical, transfusion-related and community infections), age- and sex-specific transition rates between consecutive fibrosis stages, and cancer-related mortality.</td>
<td>NHANES III data for markers of HCV and HBV infection from serologic examinations. Epidemiologic data for age and sex-specific mortality rates from primary hepatocellular carcinoma (HCC).</td>
<td>GOF was assessed by separate likelihoods for HCV prevalence, HBV prevalence, and HCC mortality. Composite score was calculated as the sum of the individual likelihoods after rescaling each measure by its maximum possible value.</td>
</tr>
<tr>
<td>Kim et al. 2007</td>
<td>Natural history microsimulation model of cervical cancer.</td>
<td>Calibration parameters for progression and regression rates from non-infected to HPV infection, CIN, and cervical cancer.</td>
<td>Targets from the Ludwig-McGill longitudinal study: mean duration of HPV infection, and hazard ratios of cytological abnormalities associated with HPV status at enrolment and time-dependent HPV status, and incident CIN. Targets for a likelihood-based approach to calibration: Ludwig-McGill study targets, age-specific prevalence of HPV, CIN, HPV-type distributions, and age-specific cervical cancer incidence.</td>
<td>Likelihood function for each target, assuming that each followed an independent normal distribution. Overall GOF was computed by summing the log-likelihood measure for each model outcome.</td>
</tr>
<tr>
<td>Decision model</td>
<td>1. Which parameters were allowed to vary in the calibration process?</td>
<td>2. Which calibration targets were used?</td>
<td>3. What measures of GOF were used?</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------------------------------------</td>
<td>---------------------------------------</td>
<td>-----------------------------------</td>
<td></td>
</tr>
<tr>
<td>Berry et al. 2006</td>
<td>Bayesian updating with rejection sampling method. Monte Carlo simulation with sampling from the prior distributions of the calibration parameters. In total, 80,000 populations were simulated and monitored from 1975 to 2000. Parameter sets are accepted if they fall within the annual acceptance window for the incidence curve and the acceptance window for the slopes of the 5-year periods.</td>
<td>It appears that the Monte Carlo simulation was stopped after completion of the fixed number of parameter sets (80,000 parameter sets).</td>
<td>Comparing observed BC mortality rates with average simulated mortality rates from 1975 to 2000.</td>
<td></td>
</tr>
<tr>
<td>Blower et al. 2001</td>
<td>Dynamic transmission model of live attenuated HIV vaccines for Zimbabwe and Thailand Initial model conditions involving mean survival time with AIDS, mean transmission rates of wild-type HIV per sexual partnership, and mean progression rate to AIDS due to wild-type infection, and mean rate of acquiring new partners.</td>
<td>Current annual AIDS death rates in Zimbabwe and Thailand.</td>
<td>For establishing the initial model conditions, projected median estimates for the two countries were inspected to be closely matched to the annual AIDS death rates.</td>
<td></td>
</tr>
</tbody>
</table>
Table 5. Calibration with random search approach (continued)

<table>
<thead>
<tr>
<th>Random search approach</th>
<th>4. What parameter search strategy was used?</th>
<th>5. What determined acceptable GOF parameter sets?</th>
<th>6. What determined the termination of the calibration process?</th>
<th>8. How was the model validated?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fryback et al. 2006 129</td>
<td>Two-stage random search. 1st stage: all onset cancers were assumed to progress to invasive breast cancer. No adequate solutions were found. 2nd stage: three more parameters were introduced to allow low malignant potential (LMP) cancers to regress. Sampling was conducted jointly involving all parameters.</td>
<td>Empirically determined a minimally acceptable parameter set to have a score ≤10; exceptionally good parameter set scored ≤5.</td>
<td>1st stage: no acceptable solutions. 2nd stage: modeling the LMP fraction yielded acceptable parameter sets. Ad hoc sampling was conducted to identify optimal range of the LMP fraction. Random sampling was used to obtain the final ensemble of good-fit parameter sets.</td>
<td>Projected stage-specific incidence rates were compared to observed data in SEER and the WCRS by year from 1975 to 2000. Projected age-adjusted incidence and breast cancer mortality rates were compared to data from WCRS and SEER data.</td>
</tr>
<tr>
<td>Salomon et al. 2002 130</td>
<td>Three-stage random search. 1st stage: identified 90 good-fit sets: 30 each for HCV, HBV prevalence, and HCC mortality. 2nd stage: 27,000 new sets cross the 30x30x30 identified sets to select 50 good-fit sets. 3rd stage: local perturbation of each best 50 with 1000 sets.</td>
<td>Identified the best 50 good-fit sets at the end of the 3rd stage.</td>
<td>Fixed number of samples defined in advance.</td>
<td>Model fits data on seroprevalence of HCV antibody by age and sex. Model fits to US vital statistics on death rates from HCC by age and sex.</td>
</tr>
<tr>
<td>Kim et al. 2007 131</td>
<td>Two-stage random search. 1st stage model was fit to primary data from the Ludwig-McGill study. 2nd stage, model was calibrated to epidemiological data from multiple sources (e.g., meta-analyses and published literature).</td>
<td>Likelihood ratio test was used to identify a good-fitting subset, comprising parameter sets that did not produce a significantly worse fit than the best-fitting parameter set (e.g., top 5, top 50, and top 100 good-fit parameter sets).</td>
<td>Fixed number of random parameter sets. In total, simulations were conducted with 555,000 uniquely sampled candidate parameter sets.</td>
<td>Compared modeled projections with calibration targets from the first and second stages of the search.</td>
</tr>
<tr>
<td>4. What parameter search strategy was used?</td>
<td>5. What determined acceptable GOF parameter sets?</td>
<td>6. What determined the termination of the calibration process?</td>
<td>8. How was the model validated?</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>----------------------------------</td>
<td></td>
</tr>
</tbody>
</table>

Blower et al. 2001 134 LHS method was used to initially fit the model to observed data. LHS was also used to generate the 1000 parameter sets that characterize the live attenuated HIV vaccines in the uncertainty analysis to quantify the trade-off between the efficacy and safety effects of the vaccine. The study did not report the number of parameter sets used in the model fitting of the initial conditions. Not reported. At the initial endemic steady state, each model was shown to replicate the observed annual AIDS death rate.

Notes: AIDS acquired immunodeficiency syndrome; BC breast cancer; CIN Cervical intraepithelial neoplasia; GOF Goodness-of-fit; HAART highly active antiretroviral therapy; HCC hepatocellular carcinoma; HIV human immunodeficiency virus; HPV human papillomavirus; LHS Latin hypercube sampling; LMP low malignant potential; HBV hepatitis B virus; HCV hepatitis C virus; NHANES National Health and Nutrition Examination Survey; SEER Surveillance, Epidemiology, and End Results Program; WCRS Wisconsin Cancer Reporting System.

*† Step 7 was not displayed as it is only relevant to one selected study describing a guided search strategy; this step was described in the synopsis of the study. 105
Table 6. Calibration with grid search approach

<table>
<thead>
<tr>
<th>Grid search approach</th>
<th>Decision model</th>
<th>1. Which parameters were allowed to vary in the calibration process?</th>
<th>2. Which calibration targets were used?</th>
<th>3. What measures of GOF were used?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weinstein et al. 1987</td>
<td>A coronary heart disease (CHD) state-transition model.</td>
<td>CHD incidence, particularly in the oldest age range, MI case-fatality rates in the oldest and lowest age ranges; CABG incidence rates; and CHD event rates among persons with a history of angina or CABG only.</td>
<td>National data on the age-specific number of CHD and non-CHD deaths and the number of MI's.</td>
<td>Modeled projections are within 10% of target values.</td>
</tr>
<tr>
<td>Rydzak et al. 2010</td>
<td>Microsimulation model of HIV and AIDS, adapted from the Cost-Effectiveness of Preventing AIDS Complications model using data from the Women Interagency HIV Study (WIHS).</td>
<td>Parameters related to the assumptions in untreated women included chronic HIV-associated mortality following an opportunistic infection, and in treated women, the clinical effectiveness of HAART and the ability of HAART to prevent HIV complications independent of virologic suppression.</td>
<td>24-month survival curves based on untreated HIV-infected women, according to starting CD4 cell count were used to calibrate the natural history model. 24-month survival curves of women in the WIHS study treated with HAART.</td>
<td>Consistency between model projections and empiric data was assessed by visually comparing the average model outcomes with the means and the 95% CI of the observed survival data. GOF measures were also calculated as the sum of the absolute differences between model projections and corresponding observed mean at 6, 12, 18 and 24 months.</td>
</tr>
<tr>
<td>Yeh et al. 2008</td>
<td>Markov state transition model of gastric carcinogenesis for China and Columbia.</td>
<td>Progression and regression probabilities from normal gastric mucosa, to precancerous lesions and cancer.</td>
<td>Age-specific and country-specific prevalence of gastritis, atrophy, intestinal metaplasia, dysplasia, and age-specific symptomatic gastric cancer incidence.</td>
<td>A GOF score was computed as negative two times the sum of the log-likelihood scores of each calibration target.</td>
</tr>
<tr>
<td>Weinstein et al. 1987</td>
<td>Vary the most uncertain parameters over their plausible ranges, and simulate all the combinations within this hypercube of the parameter space.</td>
<td>Parameter sets are selected if model projections are within 10% for all targeted outputs, and within 5% for most of the targeted outputs.</td>
<td>Completion of all combinations of calibration parameters.</td>
<td>In a subsequent publication, predictions from the model were validated using national data from 1986.</td>
</tr>
</tbody>
</table>
Table 6. Calibration with grid search approach (continued)

<table>
<thead>
<tr>
<th>Grid search approaches</th>
<th>4. What parameter search strategy was used?</th>
<th>5. What determined acceptable GOF parameter sets?</th>
<th>6. What determined the termination of the calibration process?</th>
<th>8. How was the model validated?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rydzak et al. 2010 135</td>
<td>A combination of one-way sensitivity analyses and multi-way sensitivity analyses were used to explore model assumptions. As a result, varying each of the uncertain assumptions individually, in combination, and according to HAART regimen, generated more than 1500 unique combinations of parameters for each CD4 stratum.</td>
<td>Parameter sets were ranked based on their GOF values; those with the lowest values were considered to be more consistent with survival data. For the assessment of model performance, the authors selected the best of the 50 good-fit parameter sets for each CD4 strata.</td>
<td>Completion of model evaluations for the unique combinations of parameters for each CD4 stratum.</td>
<td>Projected time on treatment and time to switch HAART therapy were compared to corresponding times from independent analyses that used a distinct subset of WIHS data not used in the calibration of the natural history model.</td>
</tr>
<tr>
<td>Yeh et al. 2008 132</td>
<td>A multidimensional search of the parameter space defined by the plausible ranges of the model parameters using a grid search and identified 65,000 unique candidate parameter sets.</td>
<td>The best-fitting parameter set with the lowest GOF score was selected. A subset of good-fitting parameter sets, defined as those whose fit was statistically indistinguishable from the GOF score of the best-fitting set was also accepted.</td>
<td>The grid search was conducted with a fixed number of parameter sets.</td>
<td>Assessing model fit to calibration targets on precancerous lesions prevalence and cancer incidence using 50 randomly selected good-fitting parameter sets for men in China and Colombia.</td>
</tr>
</tbody>
</table>

Notes: AIDS acquired immunodeficiency syndrome; CABG coronary arteries bypass graft surgery; CHD coronary heart disease; CI confidence interval; GOF Goodness-of-fit; HAART highly active antiretroviral therapy; HIV human immunodeficiency virus; MI myocardial infarction; WIHS Women Interagency HIV Study.

*† Step 7 was not displayed as it is only relevant to one selected study describing a guided search strategy; this step was described in the synopsis of the study. 105
Table 7. Suggested methodological issues relevant to the conduct of model calibration

<table>
<thead>
<tr>
<th>Approach to calibration</th>
<th>“Posterior” distribution</th>
<th>Uncertainty propagation</th>
<th>Data conflicts</th>
<th>Non-identifiability</th>
<th>Accessibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian approach</td>
<td>🟢</td>
<td>🟢</td>
<td>🟥</td>
<td>🟥</td>
<td>🟥</td>
</tr>
<tr>
<td>Guided search approach</td>
<td>🟥</td>
<td>🟥</td>
<td>🟥</td>
<td>🟥</td>
<td>🟢</td>
</tr>
<tr>
<td>Random search approach</td>
<td>🟥</td>
<td>🟢</td>
<td>🟥</td>
<td>🟥</td>
<td>🟢</td>
</tr>
</tbody>
</table>

Notes: The methodological issues are described as follows: **Posterior distribution**: The approach yields the posterior distribution of model parameters or an approximation to the posterior distribution. **Uncertainty propagation**: There is a provision to propagate uncertainty in model parameters to the relevant model outputs (e.g., life years). **Data conflicts**: There is a provision to identify and resolve the conflicts between modeled projections and multiple data points. **Non-identifiability**: There is a provision to identify model parameters that are not identifiable during the calibration such that there is a large overlapping between the prior and posterior distributions. **Accessibility**: Model calibration should not be restricted to applications that use less accessible modeling techniques. Harvey balls (similar to those used by Consumer Reports) are used to indicate the degree to which a particular approach addresses a particular methodological issue.180
Chapter 5
Calibration Analysis of Decision Analytic Models

5 Calibration analysis of decision analytic models

5.1 Abstract

**Purposes:** Model calibration involves the determination of uncertain parameters, so that model outputs are consistent with observed data. Calibration is typically effected by searching a sampling design (i.e., a collection of random points in a high-dimensional space) for good-fit parameter values or set of values such that model outputs reproduce the corresponding observed data. We propose a two-stage random search algorithm for routine calibration of decision models.

**Methods:** The algorithm structures the calibration in stages and suggests operational details regarding sampling design, sample size, and stopping rules. The first-stage search is conducted with sequential random sampling designs with increasing sample size to obtain stable estimates of the marginal distributions of individual parameters and their rank correlations. Results from the first-stage search were used to update the parameter search space. The second-stage search is conducted with a random sampling design from the updated search space to obtain a final ensemble of good-fit parameter sets and an associated set of probability weights for use in subsequent evaluation of the calibrated model. We evaluated the feasibility of the proposed algorithm with a decision model for a cost-effectiveness analysis.

**Results:** The results suggest that much information obtained via calibration was gained early in the first-stage search, especially with the initial sampling design. This included early detection of non-identifiable parameters and approximation to the best-fit estimates of model parameters. Subsequent iterations of the search were mainly used to characterize parameter uncertainty. It was easier to obtain stable estimates for marginal distributions of model parameters than to obtain stable estimates of the correlation matrix. There were little differences in the results of the cost-effectiveness analysis when the search was conducted with designs from simple or stratified random sampling.

**Conclusion:** The proposed algorithm is feasible for routine calibration of moderately complex decision models. Additional studies are needed to evaluate the impact of the varying options suggested for different steps of the algorithm on the cost-effectiveness results.
5.2 Introduction

Decision models are important tools in developing our understanding of disease prognoses and enabling the evaluation of health technologies.\textsuperscript{181} In our case, we modeled the natural history of pressure ulcers in ill patients and evaluated prevention strategies.\textsuperscript{182} Pressure ulcers are generally preventable but without proper care, could progress through severe stages.\textsuperscript{32} Higher stages are associated with skin breakdown, tissue injury and infections, and high treatment costs.\textsuperscript{32} Stage-specific transition rates are generally unknown because of unobserved deep tissue injury and serial staging assessment of open wounds typically requires debridement of eschar that is not done unless clinically indicated.\textsuperscript{32} Stage-specific prevalence data are however available through prevalence surveys.\textsuperscript{183}

Model calibration involves the determination of uncertain model parameters, so that model outputs are consistent with observed data.\textsuperscript{111} Calibration is typically effected by searching a sampling design (i.e., a collection of random points in a high-dimensional space) for good-fit parameter values or set of values such that model outputs reproduce the corresponding observed data.\textsuperscript{120} We calibrated transition rates of our pressure ulcer decision model to ensure that projected frequencies of stage-specific pressure ulcers replicate observed frequencies. Routine calibration was necessary to reflect the burden of pressure ulcers in nursing homes residents, elderly patients in emergency rooms, and patients undergoing prolonged surgical procedures. The calibrated models were used for cost-effectiveness analyses of setting-specific prevention strategies.\textsuperscript{95, 182, 184} We have been experimenting with different approaches for routine calibration. Below, we report on a two-stage random search algorithm for the calibration of decision models.

5.3 Methods

The two-stage random search algorithm is described and illustrated with the calibration of a decision model for a cost-effectiveness analysis.\textsuperscript{182} The methods section is organized as follows. First, we describe the cost-effectiveness analysis and the decision model. In subsequent sections, we describe the two-stage random search algorithm, the calibration process and the cost-effectiveness analysis.
Notes: PrU: pressure ulcer. Pressure ulcers are classified as stage 1, 2, 3 or 4 as described above. The symbol $P_{01}$ denotes the daily incidence of developing a stage-1 pressure ulcer; $P_{12}$ the daily probability that a pressure ulcer worsens from stage 1 severity to stage 2 severity; $P_{23}$ from a stage 2 to a stage 3; and $P_{34}$ from a stage 3 to a stage 4. The symbol $H_{10}$ denotes the healing rate of a stage-1 pressure ulcer, $H_{20}$ the healing rate of a stage-2 pressure ulcer, and so on.
5.3.1 Cost-effectiveness of a prevention strategy

We conducted a cost-effectiveness analysis of a prevention strategy that calls for replacing standard hospital mattresses on stretchers and beds in emergency departments (EDs) with pressure-redistribution mattresses. The latter mattresses may reduce the risk of developing pressure ulcers during ED wait time by distributing the body weight at bony prominences (e.g., tailbone, heels, elbows) over a large contact area. Pressure-redistribution mattresses are associated with reducing the risk of developing pressure ulcers (a relative risk of 0.78, 95% confidence interval: 0.42, 1.46). The additional cost of replacing each standard hospital mattress was $480 (range: $46, $700). The analysis was conducted from a health system perspective over a time horizon of one year, including the hospital length of stay and a post-discharge recovery time that is twice the average healing time of the most severe hospital-acquired pressure ulcers.

We used a Markov model with a 1-day cycle to simulate health events related to pressure ulcers in elderly patients admitted to inpatient care via EDs. A full description of the decision model is described in the chapter 6 and elsewhere. In brief, Figure 4 displays the model structure, which was implemented in TreeAge Pro 2009 Suite (TreeAge Software, Williamstown, MA). The health states are structured to reflect a validated staging classification (see footnotes in Figure 4) and by wound status, including pressure ulcers in which the underlying risk factors remain and pressure ulcers that are susceptible to healing with improvement in the underlying risk factors. If the risk factors (e.g., immobility) associated with the development of a pressure ulcer remain unchanged, the pressure ulcer may progress to higher severity stage (stage 1 ➔ 2, 2 ➔ 3, 3 ➔ 4) and it may be at risk for local infection (stage 2-4 pressure ulcers) and systemic infection (stage 3-4 pressure ulcers). If the underlying risk factors are reduced (e.g., with improving mobility and overall health), the tissue damage may heal under intact skin (for stage 1 pressure ulcers) or the open wound may start to re-epithelialise until full closure (stage 2-4 pressure ulcers). In the absence of the risk factors, healing pressure ulcers are assumed to complete the healing process over an average duration, and during this time, they are not at risk of worsening tissue damage and infection.
Figure 4 also shows the daily transition rates among health states, including daily incidence of developing pressure ulcers ($P_{01}$); daily progression rates of pressure ulcers from stage 1 to 2 ($P_{12}$), stage 2 to 3 ($P_{23}$) and stage 3 to 4 ($P_{34}$); and daily healing rates of stage 1 ($H_{10}$), stage 2 ($H_{20}$), stage 3 ($H_{30}$), and stage 4 ($H_{40}$) pressure ulcers. Because the transition rates are uncertain, we obtained their estimates via model calibration.

5.3.2 Two-stage random search algorithm

In a random search approach, a set (or vector) of input values to a decision model is generated by drawing each value independently and randomly from the distributions of the model parameters. Using the generated candidate parameter sets, modeled outputs are compared with corresponding observed data to identify an ensemble of good-fit parameter sets for which the outputs are consistent with the corresponding observed data. Different random search approaches have been used to calibrate disease history models. These approaches are generally conducted in stages to allow using results from the current stage of the search to guide the search at the next stage. Varying the stages (e.g., 2, 3 stages) and the sampling schemes (e.g., simple random sampling, stratified random sampling) that generate the random candidate parameter sets allows for considerable flexibility to the search.

Figure 5 displays the proposed two-stage random search algorithm. The objective of the first-stage search is to obtain an ensemble of good-fit parameter sets in which their input values are used to generate stable estimates of the marginal distributions of the model parameters and their correlations. Here, we use a collection of marginal distributions and the correlation matrix to partially characterize the joint distribution of model parameters. While obtaining the joint distribution of the model parameters is desirable, it is generally possible with Bayesian approaches to calibration.

Results from the first-stage search are used to update the parameter search space to increase the likelihood of identifying good-fit parameter sets. The objective of the second-stage search is to sample the updated parameter search space densely enough to obtain a final ensemble of good-fit parameter sets that provides accurate estimates of the marginal distributions and correlation matrix of the model parameters (in the sense of ensuring overall consistency of the model, given the observed data) and to fully characterize parameter uncertainty. Also estimated from the
second-stage search is a set of weights that quantifies the probability that within the final ensemble, some good-fit parameter set is better fit than others. The final ensemble and its accompanying set of probability weights are used in probabilistic sensitivity analysis to allow uncertainty propagation to the cost-effectiveness analysis.

The algorithm calls for an iterative first-stage search, including descriptions of an iterative loop and a stopping rule (Figure 5). Within the loop, estimates from the current iteration are compared with those from the previous iteration to determine as to whether the search objective has been fulfilled. Key elements of the algorithm are described below.

We denote the sets of random input values for the model parameters of interest (e.g., transition rates) as the “experimental design”. The experimental region refers to the sets of input values over which we study model projections. A design point in this region corresponds to a specific set of input values. Thus, a sampling design is a specification of design points in the experimental region at which we evaluate model responses.

5.3.2.1 Initial sample size estimation
The sample size of the initial sampling design is estimated so that the search could identify a number of good-fit points that is ≥10 times the input dimension. To obtain this estimate, we used the informal rule that the number of runs for an effective initial computer experiment should be about 10 times the input dimension. For example, to calibrate a decision model with 8 unknown parameters, the initial sampling design for the random search in an 8-dimensional space would need to identify approximately 80 good-fit points. On average, a random search approach typically yields approximately 1 good-fit point for every 100 random points (or a yield of approximately 1%, ranging from 0% to 5%). Therefore, to use the random search approach in an 8-dimensional search space, the sample size of the initial sampling design to identify approximately 80 good-fit points would need to be approximately 8,000 random points (with each point represents a parameter set of 8 random values that are sampled independently from the parameter distributions), or approximately 1000 random design points for each model input. If one decides to conduct the random search of the 8 unknown model parameters while allowing for the 9 other model inputs to vary during the calibration, the sample
size of the initial sampling design for the random search in a 17-dimensional space should be 17,000.
Figure 5. The two-stage random search algorithm

1. **Initialize the random search**
   - Estimate the sample size of the initial sampling design
   - Generate the initial sampling design

2. **Conduct the search with the current sampling design**
   - Obtain an ensemble of good-fit points
   - Estimate marginal distributions of model parameters
   - Estimate the rank correlations of model parameters

3. **Stopping rule** - Do the estimates converge to stable values?
   - Yes
     - **FIRST-STAGE SEARCH**
     - Obtain the final ensemble of the good-fit parameter sets
   - No
     - **SECOND-STAGE SEARCH**
     - Record the sample size at end of the first-stage search
     - Update the parameter search space
     - Generate the sampling design for the updated space
     - Conduct the search in the updated search space

4. **Generate the next sampling design with double sample size**

---

- **FIRST-STAGE SEARCH**
- **SECOND-STAGE SEARCH**
5.3.2.2 Sample size of consecutive sampling designs

In the iterative loop of the first-stage search, two scenarios can be considered for increasing the sample size of the current sampling design when the sample size of the previous sampling design is inadequate to achieve the search objective. In the first scenario, the former sample size is automatically increased to be twice as large as the latter. In the second scenario, variances of a continuous outcome measure (e.g. life years) projected from the pair of consecutive designs are compared; their ratio determines the number of additional design points. Compared with the first scenario, the second scenario is generally associated with larger bias in the estimate of the variance of the population mean of the outcome measure. It is also more complicated. For simplicity, we propose to keep doubling the sample size of consecutive sampling designs until the first-stage search reaches some stopping criteria.

5.3.2.3 Stopping rules

The stopping rules are based upon the estimates of the marginal cumulative probability density functions of individual model parameters and their correlations. The changes between the estimates from successive iterations of the first-stage search are to be small within some prespecified tolerance values, as specified below.

\[
\text{max}_i |F_i(\theta_i)^{\text{current iteration}} - F_i(\theta_i)^{\text{previous iteration}}| \leq \varepsilon_F \quad (1) \\
\text{max}_i |\Sigma^{\text{current iteration}} - \Sigma^{\text{previous iteration}}| \leq \varepsilon_\Sigma \quad (2)
\]

where \( F_i(\theta_i = a) = P[\theta_i \leq a] \) is the cumulative probability density function of an individual parameter \( \theta_i \) from the model parameter set \( \theta \) (e.g., \( i = 1, \ldots, 8 \) for the illustrated model), and \( \Sigma \) is the correlation matrix of the parameter set \( \theta \).

For each parameter \( \theta_i \), expression (1) quantifies the maximum distance between the cumulative probability density functions of the parameter estimated from any pair of successive iterations. This distance is the Kolmogorov–Smirnov (K-S) statistic. The null hypothesis of the K-S test is that the tested samples of the parameter \( \theta_i \) from successive iterations are drawn from the same
distribution. We select the K-S test because it is simpler to calculate than the more sensitive Anderson-Darling test.\(^{190}\)

The off-diagonal elements of the rank correlation matrix \(\Sigma\) are the pairwise rank correlation estimates from parameters \(\theta_i\) and \(\theta_j\) (\(i \neq j\), e.g., \(i\) or \(j = 1, \ldots, 8\)). A rank correlation quantifies the degree to which large (or small) values of one parameter associate with large or small values of another parameter. Unlike a linear correlation that assumes a linear relationship between two parameters, a rank correlation measures the strength of the association in terms of ranks. It can be used with all types of marginal distributions and any sampling schemes for which correlated inputs could logically be considered.\(^{191}\) This is relevant to model calibration as the calibration process induces correlation among model parameters by constraining modeled projections to replicate the corresponding observed data.\(^{102}\)

### 5.3.2.4 Second-stage search

The parameter search space is updated using results from the first-stage search (Figure 5). This is likely to result in a reduced search space. The second-stage search is conducted using the same sample size that provides adequate coverage of the full parameter search space in the first-stage search. This sample size is likely to be adequate for the reduced search space. Thus, the design in the second-stage search oversamples the reduced search space to obtain a final ensemble of good-fit parameters that fulfills the objective of the second-stage search.

#### Table 8. Sampling designs for the illustration with the two-stage random search approach

<table>
<thead>
<tr>
<th>Sampling scheme</th>
<th>First-stage search</th>
<th>Second-stage search</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRS</td>
<td>Space-filling design</td>
<td>Space-filling design</td>
</tr>
<tr>
<td>LHS</td>
<td>Space-filling design</td>
<td>Space-filling design</td>
</tr>
<tr>
<td>LHS</td>
<td>Space-filling design</td>
<td>Sampling design with correlations</td>
</tr>
</tbody>
</table>

Notes: SRS - simple random sampling. LHS - Latin hypercube sampling.

### 5.3.2.5 Sampling design

Table 8 describes the sampling designs used in the calibration of the decision model. Initially, good-fit points are assumed to be just as likely to occur in one part of the space as another.\(^{122}\) Suitable random sampling schemes are used to generate *space-filling* designs that spread points evenly across the space in order to provide broad coverage of the search space and to maximize
the chance of identifying a representative ensemble of good-fit points. Space-filling designs are used in the first-stage search of the original parameter search space and also in the second-stage search of the updated search space, with one exception (below).

For the calibration, space-filling designs are generated by simple random sampling (SRS) or Latin hypercube sampling (LHS) scheme. With SRS, random values for the model parameter \( \theta = (\theta_1, \ldots, \theta_8) \) are independently sampled from each of the distributions of the parameters \( \theta_1, \ldots, \theta_8 \). In high-dimensional space, the generated space-filling designs from the SRS scheme (especially those with small sample size) may exhibit some clustering and may fail to provide points in portions of the space. Latin hypercube sampling is a stratified SRS scheme that can be used to generate space-filling designs that spread points evenly over the range of each parameter. This is illustrated with an example below.

Figure 6 illustrates two LHS designs, each with 10 design points in the three-dimensional space of the parameters \( (\theta_1, \theta_2, \theta_3) \). The points are projected onto the planes defined by the pairs \( (\theta_1, \theta_2) \) and \( (\theta_1, \theta_3) \), according to their marginal probability density distributions: \( \theta_1 \sim \text{uniform}(0, 2) \), \( \theta_2 \sim \text{beta}(2,5) \), and \( \theta_3 \sim \text{gamma}(2,3) \). With LHS, plausible ranges of these parameters are stratified into equally probable intervals. Within each interval, each point of the design is randomly selected proportionally to the probability density. Pairwise, the points are selected to form a Latin square such that each interval (being selected exactly once in each row) is paired with another interval (being selected exactly once in each column).

LHS can generate space-filling designs with pairwise random pairing of row and column intervals. In Figure 6, this is illustrated by the design points that are displayed as open circles. The points tend to spread out across the square, proportionally to the marginal probability density values.

When the rank correlations of a set of inputs are known, LHS can be used to generate a sampling design taking this into account. In this case, the row and column intervals are paired in a restricted way that replicates the pairwise rank correlation. In Figure 6, this is illustrated with the design points that are displayed as black circles. The points are arranged to induce a high
positive rank correlation between the pair of parameters \((\theta_1, \theta_2)\) and a high negative rank correlation between \((\theta_1, \theta_3)\). In this LHS design, the effect of restricted pairing is to rule out the upper-left and lower-right corners of the plane defined by \((\theta_1, \theta_2)\), and to rule out the lower-left and upper-right corners of the plane defined by \((\theta_1, \theta_3)\). The restricted pairing option in the LHS scheme can be used to generate space-filling designs (ensuring that all pairwise rank correlations are zero) or designs with correlated inputs. In Table 8, one of the second-stage searches was conducted with a LHS design that took into account the estimated rank correlation matrix from the first-stage search.
Figure 6. An illustration of Latin hypercube sampling design

Notes: Latin hypercube sampling designs with a space-filling design (open circles) and a design of correlated inputs (filled circles).
5.3.3 The calibration process

We conducted the model calibration according to the reporting checklist proposed by Stout et al. 2009,\textsuperscript{101} and a seven-step approach proposed by Vanni et al. 2011.\textsuperscript{101, 102}

**Step 1.** Which inputs should be varied in the calibration process?

Table 9 displays model inputs to the decision model and their distributions.\textsuperscript{182} The following distinction was made among the groups of model inputs. The first group of inputs is the set of model parameters comprising the 8 uncertain transition rates (Figure 7). These parameters are context-specific as they vary for different at-risk populations. They however remain constant when the calibrated model is used for the target population.\textsuperscript{120} Another group of model inputs comprises the variable inputs (e.g., hospital length of stay, mean healing time of hospital-acquired pressure ulcers) that affect the prognosis of pressure ulcer (n=9 inputs with italic distributions in Table 9). We conducted the calibration analysis when these variable inputs were fixed at their expected values, and when they were allowed to vary during the calibration (see also the descriptions of the analysis scenarios in the Data Analysis section).

Other model inputs pertaining to costs and health-related quality of life (that are not affecting the pressure ulcer prognosis) were set at their expected values during the calibration (Table 9). Additional model inputs that were assumed to be with fixed values during the calibration process and the cost-effectiveness analysis are reported elsewhere.\textsuperscript{182}

The transition rates across pressure ulcer stages are uncertain; we therefore conducted a literature search (Medline, MESH terms “pressure ulcer” AND (“epidemiology” OR “incidence” OR “prognosis”), January 1980 to May 2012) to inform their plausible ranges (Table 9).\textsuperscript{182} Similar to other calibration studies, we assumed uniform distributions over their plausible ranges.\textsuperscript{129-131}

**Step 2.** Which calibration targets should be used?

The use of multiple targets for calibration is recommended.\textsuperscript{102, 193, 194} There is limited data on the prevalence of pressure ulcers in ED patients.\textsuperscript{57} Specifically, stage-specific prevalence of hospital-acquired pressure ulcers was derived using data from a cross-sectional prevalence survey from three acute care hospitals.\textsuperscript{182} We used survey data for a subgroup of 745 elderly (age $\geq$65) ED-
admitted patients. The observed stage-specific prevalence was not used as input to the decision model. However, it is functionally (non-linearly) related to the transition rates that were to be determined via calibration. Therefore, the observed prevalence was deemed to be an appropriate calibration target.

To match the observed frequencies of pressure ulcers (i.e., none, stage 1-4), the projected frequencies were generated for a simulated cohort of elderly admitted ED patients. The simulated cohort was characterized according to baseline data from the prevalence survey, including a mean age of 70 years, a mean length of hospital stay of 7 days, and an average ED wait time of approximately 15 hours (Table 9). The simulated disease process was sampled at the 7th day after admission to obtain the projected frequencies. We assumed that the projected frequencies would replicate the observed frequencies, although the history model did not simulate the variation in the characteristics of the patient population. Specifically, the observed frequencies were from a cross-sectional survey of patients at varying age (with a mean age of 70 years, Table 9), at varying length of hospital stay (with a mean length of stay of 7 days), and at varying ED wait time (a mean wait time of 15 hours). In other words, one of the sources of discrepancy between the projected and observed frequencies was due to model discrepancy; i.e., the model did not simulate the effects of the cited sources of heterogeneity.

**Step 3.** What measure of goodness-of-fit (GOF) should be used?

Pearson’s goodness-of-fit (GOF) statistic $X^2$ and the log likelihood ratio statistic $G^2$ are commonly used GOF measures for comparing observed frequencies and expected frequencies, assuming a multinomial distribution. Specifically,

$$X^2 = \sum_{k=0}^{4} (Observed\ frequency_k - Projected\ frequency_k)^2 / Projected\ frequency_k$$

$$G^2 = 2 \sum_{k=0}^{4} Observed\ frequency_k \times \log (Observed\ frequency_k / Projected\ frequency_k)$$

with $k = 0$ representing no pressure ulcers, and $k = 1-4$ representing stage 1 to 4 pressure ulcers. Lower values of $X^2$ or $G^2$ indicate better fit between the frequencies (i.e., lack of fit measures). We evaluated the impact of the choice of the GOF measures on the results of the cost-effectiveness analysis.
**Step 4.** What parameter search strategy should be used?
The two-stage random search algorithm described above was used.

**Step 5.** What determines acceptable GOF parameter sets?
A good-fit point identified in the random search had to be associated with transition rates that fulfilled the following acceptance criteria:
- the daily healing rates are decreasing with increasing pressure ulcer stages, and
- the projected stage-specific prevalence was within the sampling variation from the observed prevalence (see step 3).

The latter criterion was defined as the Pearson’s $X^2$ measure or the log likelihood ratio $G^2$ measure $\leq 9.49$, the 95th quantile of the chi-squared distribution (with 4 degrees of freedom, see step 3). We also calibrated with an alternative cut-off value, the 50th quantile of the chi-squared distribution (a GOF measure $\leq 3.36$). The use of the 95th quantile cut-off value is expected to yield a larger ensemble of good-fit points compared to the use of the 50th quantile cut-off value. We evaluated the impact of the threshold selection on the results of the cost-effectiveness analysis.

**Step 6.** What determines the termination of the calibration process?
The first-stage search was initiated with an initial design of 8,000 design points (see the Sample size section above). This stage stopped when the maximum distance between the cumulative probability density functions from two consecutive iterations was deemed to be small (i.e., $\varepsilon_F < 0.1$, expression 1), and the maximum difference in the estimated rank correlation matrices from any two consecutive iterations was deemed to be small (i.e., $\varepsilon_\Sigma < 0.1$, expression 2). With expression (1), a cut-off value of the K-S statistic of 0.1 corresponds to a yield of approximately 400 good-fit points from the first-stage search.

**Step 7.** How should the calibrated model be validated?
Ideally, benchmarking data for the validation of the calibrated model should include only data that are not used in the calibration analysis. Given the limited data available in the illustrated
example, we could only conduct an internal validation by comparing the calibrated transition rates to the corresponding reported rates from published studies that had been used previously to inform their input ranges (step 1).197

**Step 8.** How should the model calibration results and economic inputs be integrated? Probabilistic sensitivity analysis was conducted with Monte Carlo simulation (n=20,000) to evaluate the calibrated model for the cost-effectiveness analysis. 88 The Monte Carlo evaluation sampled all input distributions, including random sampling (with replacement) the final ensemble of good-fit parameter sets with probability weights inversely proportional to the lack-of-fit LOF measure (i.e., Pearson’s $X^2$ or log likelihood ratio $G^2$; these measures were denoted as GOF measures for ease of the presentation in step 3). The probability weight assigned to each good-fit parameter set was estimated as the reciprocal of the LOF measure, divided by the sum of the reciprocals across all good-fit parameter sets in the final ensemble. The reciprocal was used because smaller absolute values of the LOF measure represent better-fit parameter sets and thus require higher probability weights.105

We also used an alternative set of probability weights that was based upon the significance values of the chi-squared distribution (with 4 degrees of freedom) of the LOF measures. Again, smaller absolute values of the LOF measure correspond to larger significance values (indicating that the projected frequencies are consistent with the observed frequencies), thus representing higher probability weights. We evaluated the impact of the choice of probability weights on the results of the cost-effectiveness analysis.

**5.3.4 Data analysis**

The calibration analysis involves varying options along its various steps. We therefore considered the following scenarios to evaluate how these options affected the final results of the cost-effectiveness analysis.

- No-calibration - a Monte Carlo evaluation of the history model was conducted by random sampling from the initial uniform distributions of the 8 transition parameters (Table 9).
• Initial calibration – this scenario was equivalent to the first-stage search with the initial sampling design (n=8,000) involving only the 8 transition parameters (Figure 7).
• 2-S SRS (8) – this scenario was conducted with the two-stage random search approach (2-S) with simple random sampling (SRS) involving only the 8 transition parameters.
• 2-S SRS (8, 9) – this scenario was similar to the 2-S SRS (8) scenario but also involved sampling from the 9 variable inputs (Table 9).
• 2-S LHS (8, 9) – this scenario was similar to the 2-S SRS (8, 9) scenario but LHS was used instead of SRS.

All SRS designs were generated by random sampling from individual distributions of model inputs, assuming no dependences among these inputs. For the first-stage search with LHS, space-filling designs were generated with restricted pairing to reproduce a correlation matrix with all pairwise rank correlations equal to zero.\(^{191}\) For the second-stage search with LHS, sampling designs were generated with restricted pairing to reproduce the estimated rank correlation matrix from the first-stage search. LHS designs were generated using the library “lhs” in the \textit{R} statistical package.\(^{198, 199}\) \textit{R} codes for the restricted pairing option with Latin hypercube design are described in Appendix 2. \textit{R} functions were used for Spearman rank correlation calculations (“cor”), K-S statistic (“ks.test”), and density estimates of the calibrated parameters (“density” with the following options: “wt” is set to be equal to the probability weights associated with the final ensemble of good-fit points, and the smoothing bandwidth “(bw= ‘sj’)” and the adjustment factor “adjust=0.5”).\(^{39}\)

In the cost-effectiveness analysis, results from the probabilistic sensitivity analysis were reported as the probability that the prevention strategy was more cost effective than current practice at a willingness-to-pay value of Canadian $50,000 per quality-adjusted life year (QALY) gained.\(^{200}\)

### 5.4 Results

Overall, this calibration analysis was characterized by a lack of observed data to fully inform the 8 transition parameters, leaving many parameters not identifiable from the analysis.
5.4.1 Calibration analysis results

Table 10 displays the results of the two-stage random search under different calibration scenarios. The first-stage search started with a sampling design with a sample size of 8,000 candidate parameter sets and ended with a sampling design with a sample size of 256,000 sets. The yield of good-fit points was <1% in the first-stage search and ranged from 5% to 8% in the second-stage search. In the first-stage search, it took a sampling design with a larger sample size (256,000) to obtain the stable rank correlation estimates, and a design with a smaller sample size (128,000) to obtain stable estimates of the cumulative probability density functions.

The mean cost associated with pressure ulcer history was estimated to be approximately $7000 (range: $7,038; $7,077) when the disease history model was calibrated allowing only the eight transition parameters to vary during calibration (i.e., the 2-S SRS (8) scenario, Table 10). The mean cost ranged from approximately $4,500 to $14,000 when calibration was conducted allowing both the transition parameters and other model inputs to vary (i.e., the 2-S SRS (8, 9) scenario). This suggests that calibration under the latter scenario seems to better reflect the parameter uncertainty in the model.

Results from the calibration process displayed in Table 10 suggest that there were no differences between SRS and LHS when these sampling schemes were used in the two-stage random search. The final ensemble consisted of approximately 5% of good-fit points with both sampling schemes. Most likely because of the lack of observed data to inform the transition parameters, the estimated rank correlation matrix showed only weak association between a few pairs of parameters (e.g., \( H_{10} \) and \( H_{20} \), \( P_{01} \) and \( H_{10} \); Table A-2, Appendix 2). As a result, the use of LHS design taking into account this correlation structure only slightly improved the yield of good-fit points from approximately 5% to 8% (Table 10).

Figure 7 displays the ranges of each transition parameter under different scenarios: no calibration (i.e., the original range as displayed by the left most vertical line), initial calibration (the second left most red line), 2-S SRS (8) (the blue line), 2-S SRS (8, 9) (the green line) and 2-S LHS (8,9) (the black line). The calibration target only informed a few transition parameters, namely \( P_{01} \) and \( P_{12} \). The weighted mean estimates of transition parameters from the calibration analysis were
reasonably consistent across scenarios, suggesting fairly quick convergence starting with the initial calibration. The information gain is most pronounced with the initial calibration, relative to no calibration. Relative to the 2-S SRS (8, 9) and 2-S LHS (8,9) scenarios, the initial calibration and 2-S SRS (8) scenarios underestimate the parameter uncertainty in the model (Figure 7).

For illustration, Figure 8 displays the estimated probability density distributions of four transition parameters (i.e., \( P_{01} \), \( P_{12} \), \( H_{10} \), and \( H_{20} \)) under three scenarios: no calibration (the rectangle), initial calibration (the green line), and 2-S LHS (8, 9) (the blue line). The calibration results show quick shrinkage of the calibrated probability distributions for \( P_{01} \) and \( P_{12} \). The probability distributions for \( H_{10} \) and \( H_{20} \) exhibit large overlap of the un-calibrated and calibrated distributions. These patterns seemed to emerge early with the initial calibration.

### 5.4.2 Validation of the calibrated model

The projected prevalence from the calibrated model replicates the observed prevalence to within the sampling errors (Figure A, Appendix 2). In a prospective cohort study of elderly admitted ED patients (\( n=3233 \)), the incidence of pressure ulcers in the first two days after ED admission was 6.2%; the calibrated incidence \( P_{01} \) was 5.8%.77 A prospective study of surgical patients (\( n=280 \)) reports that 12 of the 33 stage-1 pressure ulcers evolved to stage 2 pressure ulcers in one day - a daily transition rate of approximately 36.4%; the calibrated transition rate \( P_{12} \) was 31.0%.201 This study also reports that 3 of the 23 stage 2 pressure ulcers evolved to stage 3 in one day - a transition rate of approximately 13%; the calibrated transition rate \( P_{23} \) was 4.8%.

### 5.4.3 Cost-effectiveness results

There were very small differences in the cost-effectiveness results with or without calibration (Table 11). The cost-effectiveness results were insensitive to variation in GOF measures, cut-off values for determining good-fit parameter sets, and choices of probability weights assigned to good-fit parameter sets in the probabilistic sensitivity analysis. Across these sources of variation, the projected mean cost with prevention ranged from $15,634 to $16,219 with calibration, and was $16,421 without calibration. The corresponding mean cost with current practice ranged from $15,683 to $16,271 with calibration and was $16,505 without calibration. The projected mean QALDs were very similar with or without calibration. Prevention was associated with a saving
of approximately $50 with calibration and $80 without calibration. Prevention was unlikely to be harmful as indicated by very small incremental QALDs. Overall, the probability that prevention is cost-effective was approximately 80%.

5.5 Discussion
We propose a two-stage random search algorithm for the calibration of moderately complex disease history models with uncertain (unknown) parameters. The first-stage search is conducted iteratively with sequential random sampling designs with increasing sample size to obtain stable estimates of marginal probability distributions of the model parameters and their rank correlations. Results from the first-stage search are used to update the parameter search space to increase the likelihood of identifying good-fit parameter sets. The second-stage search is conducted oversampling the updated parameter search space to obtain a final ensemble of good-fit parameter sets that defines the marginal distributions and the correlation structure of the model parameters, in the sense of ensuring overall consistency between modeled projections and corresponding observed data. The final ensemble is used in probabilistic sensitivity analysis to allow for uncertainty propagation from the calibrated model to the cost-effectiveness analysis.

We evaluated the feasibility of the proposed algorithm with the calibration of a pressure ulcer history model. Key findings from the results are as follows. First, there were little differences when the search was conducted with SRS or LHS sampling. SRS is however easier to implement. Secondly, it was easier to obtain stable estimates of the marginal cumulative probability density functions of individual parameters than of their correlation matrix. The estimates of interest became relatively stable after 5 iterations, with a final sample size of 256,000 candidate parameter sets. These observations suggest that the proposed approach may be useful for the calibration of moderately complex disease history models.

It is unclear whether the random sampling approach is feasible for the calibration of complex decision models. We cited results from two selected studies to further inform the feasibility of the proposed algorithm. In one study, a random search approach with 100,000 candidate parameter sets could not locate the best-fit parameter sets in the calibration of a Markov model of cervical cancer (after over a month of model evaluations). In another studies, a random search
approach has been used to calibrate a complex cervical cancer model with 1,000,000 candidate parameter sets in which the microsimulation model is evaluated with 100,000 simulated individuals.\textsuperscript{164} Apparently, the different software platforms in which these models are based contribute to the feasibility of the calibration process.

Our results also suggest that much of the information obtained via calibration was gained early in the first-stage search, especially with the initial calibration. This included early detection of parameters that were not identifiable via calibration, and approximation to the best-fit estimates of model parameters. It appears that subsequent iterations of the search were used to characterize the uncertainty around the parameter estimates. The joint distribution of model parameters appeared to be complex, with multiple peaks and valleys (Figure 8). This complexity is likely due to the limited observed data we used for calibration. Also, the complex joint distribution makes it difficult to identify all the solutions of the calibration problem with a random search approach.

Bayesian approaches to the calibration of decision models combine evidence from diverse sources and calibrate model parameters against target data to obtain posterior distributions of model parameters and their functions, and at the same time, propagate uncertainty to model outputs relevant to the decision problem.\textsuperscript{141, 202, 203} A method for the calibration of disease history models through Bayesian multiparameter evidence synthesis (MPES) has been recently proposed.\textsuperscript{117} Other Bayesian approaches to calibration have been proposed for disease history models, including microsimulation and Markov models.\textsuperscript{116, 124} The Bayesian updating process requires the specifications of distributional relationships between the observed data used for calibration, conditioning on the corresponding modeled projections.\textsuperscript{148} Bayesian approaches are technically more complex than random search approaches.\textsuperscript{204} Additional studies are needed to examine whether it is possible to obtain an approximation to the posterior distribution of model parameters with a random search approach.

Guided search approaches to calibration iteratively learn from the currently selected set of parameter values and move towards a better fitting set of parameter values.\textsuperscript{102} These approaches typically utilize search algorithms that are well validated and widely used in other disciplines.
The search algorithms are designed to locate local minima (and the global minimum) on a response surface in high dimensional space. In a study describing the calibration of a cervical cancer model that we discussed above, a random search approach could not locate the best-fit parameter set but a guided search approach with the Nelder-Mead downhill simplex method could. Guided search approaches typically focus on obtaining the best-fit parameter set with much less emphasis on quantifying the associated uncertainty. For example, contour confidence volumes associated with the best-fit parameter set are generally not reported. Additional studies are needed to examine issues related to how random search and guided search approaches characterize and propagate uncertainty in the calibration of decision models.

Current approaches to the calibration of decision analytic models generally assume no discrepancy between modeled projections and the real disease process that the model is supposed to simulate (with one exception discussed below). However, the GOF measures between model projections and corresponding observed data are subjected to different sources of variation. First, there is sampling variation in the observed data (e.g., prevalence estimates obtained from a sample of study participants). Also, there are possible mismatches between conditions under which the observed data are obtained and the simulated conditions under which the projected data are generated (e.g., step 2). Most importantly, the disease history model is a crude approximation of the disease process, typically including a series of structural assumptions, and in instances, a lack of understanding of certain components of the disease process. Calibrated estimates therefore are functionally dependent on these sources of variation, most importantly model discrepancy. Relevant methods for model calibration that account for model discrepancy have been developed and refined in other disciplines for well over a decade, but have not been put to good use in decision modeling. The exception is a recent proposal for managing structural uncertainty in decision models.

Random search approaches have been used to calibrate complex simulation models in engineering and computer science, environmental science, and disease modeling. We have studied previous methodologies and practice in developing this proposed algorithm (see chapter 4). This algorithm is particularly suitable for routine calibration. However, to fully integrate model calibration into routine practice, additional studies are needed.
to address a number of methodological issues, most notably routine calibration approaches that produce approximation to the posterior distributions of model parameters, and early detection of non-identifiable parameters that may allow for additional verification of model assumptions and possible acquisition of additional target data.

5.6 Limitations

In our study, the calibration target we used was with limited information for the determination of the unknown transition parameters. In retrospect, we would have preferred obtaining additional observed data. In principle, calibration should be conducted with multiple sources of observed data, preferably data that are correlated to different subsets of calibration parameters to reduce problems with non-identifiability. With the illustrated cost-effectiveness analysis of pressure ulcer prevention, we could not assess the impact on the cost-effectiveness results of different options for the steps of the calibration process. This was in part due to the relatively low impact of the prevention strategy we evaluated, the short time horizon of the economic evaluation, and the limited observed data we used as targets in the calibration. Additional studies are needed to evaluate the impact of the proposed random search algorithm in other cost-effectiveness models.
Figure 7. Ranges and weighted mean estimates of transition parameters

Notes: Vertical lines: ranges. Small round circles: Weighted mean estimates. Mean estimates were derived with weights inversely proportional to the Pearson’s GOF measures. The following lines are displayed for each transition: the left most black line displays the plausible range before calibration, the red line the final range estimated from the initial calibration (see text), the blue line the final range estimated from the 2-S SRS (8) calibration scenario, the green line the final range estimated from the 2-S SRS (8, 9) calibration scenario, and the right most black line the final range estimated from the 2-S LHS (8, 9) scenario.
Figure 8. Probability density distributions of calibrated transition parameters

Notes: For each transition parameter, the following plots are displayed: initial uniform distribution prior to calibration (black rectangle), estimated distribution under the initial calibration scenario (green line), and estimated distribution under the 2-S LHS (8, 9) scenario.
Table 9. Inputs to the pressure ulcer decision model

<table>
<thead>
<tr>
<th>Prevention Effect Estimates</th>
<th>Estimate</th>
<th>Variation</th>
<th>Distribution</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative risk of developing pressure ulcers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRM versus SM, ED patients, 1 RCT</td>
<td>0.78</td>
<td>0.42, 1.46&lt;sup&gt;CI&lt;/sup&gt;</td>
<td>Log-normal</td>
<td>185</td>
</tr>
<tr>
<td>PRM versus SM, ED and hospital patients, 6 RCTs</td>
<td>0.45</td>
<td>0.27, 0.74&lt;sup&gt;CI&lt;/sup&gt;</td>
<td>Log-normal</td>
<td>29</td>
</tr>
</tbody>
</table>

| Natural History Model of Pressure Ulcers | | | | |
| Daily transition rates during hospitalization | | | | |
| *Daily incidence of stage-1 pressure ulcer* | | | | |
| | Calibrated | 0.02, 0.14<sup>R</sup> | Uniform | |
| | | | | |
| Daily progression rate during hospitalization | | | | |
| *Stage 1 → 2* | Calibrated | 0.13, 0.36<sup>R</sup> | Uniform | |
| *Stage 2 → 3* | Calibrated | 0.01, 0.13<sup>R</sup> | Uniform | |
| *Stage 3 → 4* | Calibrated | 0.01, 0.13<sup>R</sup> | Uniform | |

| Daily healing rate during hospitalization | | | | |
| *Healing stage 1* | Calibrated | 0.05, 0.23<sup>R</sup> | Uniform | |
| *Healing stage 2* | Calibrated | 0.05, 0.14<sup>R</sup> | Uniform | |
| *Healing stage 3* | Calibrated | 0.02, 0.07<sup>R</sup> | Uniform | |
| *Healing stage 4* | Calibrated | 0.01, 0.03<sup>R</sup> | Uniform | |

| Post-discharge mean healing time (week) | | | | |
| *Stage 1, 2, 3, 4 (respectively)* | 4, 13, 18, 22 | Not applicable | Exponentials | 60 |

| Costs | | | | |
| Mattress on ED stretchers | | | | |
| Additional cost of obtaining PRMs for ED stretchers | $480 | $46, $700<sup>R</sup> | Gamma | 182 |
| Additional cost of obtaining PRMs for ED beds | $200 | $30, $3775<sup>R</sup> | Gamma | 182 |

| In-patient costs | | | | |
| Mean cost attributable to pressure ulcer care | | | | |
| Stage 2 | $11,967 | $3,702<sup>SD</sup> | Gamma | 182 |
| Stage 3 | $12,951 | $7,849<sup>SD</sup> | Gamma | 182 |
| Stage 4 | $21,797 | $12,031<sup>SD</sup> | Gamma | 182 |
| Mean hospitalization cost | $6,806 | $10,745<sup>SD</sup> | Gamma | 182 |
| Mean in-patient physician billings | $445 | $728<sup>SD</sup> | Gamma | 182 |

<p>| Post-discharge home care costs | | | | |</p>
<table>
<thead>
<tr>
<th>Weekly cost attributable to pressure ulcer care</th>
<th>Estimate</th>
<th>Variation</th>
<th>Distribution*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 2</td>
<td>$57</td>
<td>$113 $^{SD}$</td>
<td>Gamma</td>
<td>182</td>
</tr>
<tr>
<td>Stage 3</td>
<td>$81</td>
<td>$116 $^{SD}$</td>
<td>Gamma</td>
<td>182</td>
</tr>
<tr>
<td>Stage 4</td>
<td>$105</td>
<td>$119 $^{SD}$</td>
<td>Gamma</td>
<td>182</td>
</tr>
<tr>
<td>Mean weekly cost</td>
<td>$134</td>
<td>$111 $^{SD}$</td>
<td>Gamma</td>
<td>182</td>
</tr>
</tbody>
</table>

**Health-Related Quality-of-Life Weights**

<table>
<thead>
<tr>
<th>Relative health utility decrement - stage 2-4 vs 0-1 (%)</th>
<th>6.10</th>
<th>1.10 $^{SD}$</th>
<th>Beta</th>
<th>208</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute health utility decrement</td>
<td>0.0268</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-patient health utility</td>
<td>0.44</td>
<td>0.32 $^{SD}$</td>
<td>Gamma</td>
<td>209</td>
</tr>
<tr>
<td>Full recovery after 1 year</td>
<td>0.79</td>
<td>0.12 $^{SD}$</td>
<td>Gamma</td>
<td>210</td>
</tr>
</tbody>
</table>

**Hospital Discharge Data**

<table>
<thead>
<tr>
<th>Average time in ED for admitted elderly patients (hour)</th>
<th>15.4</th>
<th>13.7 $^{SD}$</th>
<th>Gamma</th>
<th>182</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated length of stay (day)</td>
<td>6.5</td>
<td>9.7 $^{SD}$</td>
<td>Weibull</td>
<td>182</td>
</tr>
<tr>
<td>Estimated hospital mortality (%)</td>
<td>7.2</td>
<td>0.55 $^{SD}$</td>
<td>Time-specific</td>
<td>182</td>
</tr>
</tbody>
</table>

**Pressure Ulcer Related Complications**

<table>
<thead>
<tr>
<th>Daily incidence of local infection (stage 2-4) (%)</th>
<th>0.14</th>
<th>0.07 $^{SD}$</th>
<th>Beta</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily incidence of sepsis (locally infected stage 3-4) (%)</td>
<td>2.22</td>
<td>0.64 $^{SD}$</td>
<td>Beta</td>
<td>95</td>
</tr>
<tr>
<td>Mortality due to sepsis (%)</td>
<td>12.8</td>
<td>0.7 $^{SD}$</td>
<td>Beta</td>
<td>211</td>
</tr>
</tbody>
</table>

Abbreviations: PRM: pressure-redistribution foam mattress. SM: standard mattress. ED: emergency department. RCT: randomized controlled trial. Notes: *Italic model inputs were allowed to vary during the calibration. *Plausible ranges were pre-determined based upon results from a literature search (see text). $^{CI}$Confidence interval. $^{SD}$Standard deviation. $^{R}$Range.
Table 10. Results of the two-stage random search approach to model calibration

<table>
<thead>
<tr>
<th>2-Stage SRS (8 calibration parameters)</th>
<th>Random samples</th>
<th>8k*</th>
<th>16k</th>
<th>32k</th>
<th>64k</th>
<th>128k</th>
<th>256k</th>
<th>First-stage search</th>
<th>Second-stage search</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.46</td>
<td>0.27</td>
<td>0.14</td>
<td>0.12</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.17</td>
<td>0.16</td>
<td>0.12</td>
<td>0.08</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stopping rules</td>
<td>Max Δ correlation ρ§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max Δ CDF¶</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yield</td>
<td># GF points (%)</td>
<td>44 (0.6)</td>
<td>81 (0.5)</td>
<td>164 (0.5)</td>
<td>329 (0.5)</td>
<td>679 (0.5)</td>
<td>1,351 (0.5)</td>
<td>19,570 (7.6)</td>
<td></td>
</tr>
<tr>
<td>Best-fit point</td>
<td>Min Pearson GOF</td>
<td>0.92</td>
<td>1.01</td>
<td>0.17</td>
<td>0.20</td>
<td>0.46</td>
<td>0.35</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean costs*</td>
<td>$7,038</td>
<td>$7,060</td>
<td>$7,056</td>
<td>$7,065</td>
<td>$7,077</td>
<td>$7,039</td>
<td>$7,068</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean QALDs*</td>
<td>153.36</td>
<td>153.36</td>
<td>153.36</td>
<td>153.36</td>
<td>153.36</td>
<td>153.36</td>
<td>153.36</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2-Stage SRS (8 calibration parameters, 9 model inputs)</th>
<th>Random samples</th>
<th>8k*</th>
<th>16k</th>
<th>32k</th>
<th>64k</th>
<th>128k</th>
<th>256k</th>
<th>First-stage search</th>
<th>Second-stage search</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.28</td>
<td>0.32</td>
<td>0.16</td>
<td>0.15</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.15</td>
<td>0.13</td>
<td>0.15</td>
<td>0.07</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stopping rules</td>
<td>Max Δ correlation ρ§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max Δ CDF¶</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yield</td>
<td># GF points (%)</td>
<td>65 (0.8)</td>
<td>125 (0.8)</td>
<td>221 (0.7)</td>
<td>458 (0.7)</td>
<td>953 (0.7)</td>
<td>1,893 (0.7)</td>
<td>12,925 (5.1)</td>
<td></td>
</tr>
<tr>
<td>Best-fit point</td>
<td>Min Pearson GOF</td>
<td>1.30</td>
<td>1.00</td>
<td>0.59</td>
<td>0.48</td>
<td>0.17</td>
<td>0.08</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean costs*</td>
<td>$4,873</td>
<td>$7,606</td>
<td>$7,644</td>
<td>$14,007</td>
<td>$9,573</td>
<td>$4,510</td>
<td>$7,624</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean QALDs*</td>
<td>153.37</td>
<td>153.36</td>
<td>153.36</td>
<td>153.37</td>
<td>153.37</td>
<td>153.36</td>
<td>153.36</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2-Stage LHS (8 calibration parameters, 9 model inputs)</th>
<th>Random samples</th>
<th>8k*</th>
<th>16k</th>
<th>32k</th>
<th>64k</th>
<th>128k</th>
<th>256k</th>
<th>First-stage search</th>
<th>Second-stage search</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.51</td>
<td>0.49</td>
<td>0.13</td>
<td>0.13</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.24</td>
<td>0.13</td>
<td>0.09</td>
<td>0.07</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stopping rules</td>
<td>Max Δ correlation ρ§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max Δ CDF¶</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yield</td>
<td># GF points (%)</td>
<td>45 (0.6)</td>
<td>108 (0.7)</td>
<td>224 (0.7)</td>
<td>491 (0.8)</td>
<td>928 (0.7)</td>
<td>1,808 (0.7)</td>
<td>13,334 (5.2)</td>
<td>21,126 (8.3)</td>
</tr>
<tr>
<td>Best-fit point</td>
<td>Min Pearson GOF</td>
<td>1.07</td>
<td>0.83</td>
<td>0.56</td>
<td>0.42</td>
<td>0.14</td>
<td>0.16</td>
<td>0.18</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Mean costs*</td>
<td>$7,585</td>
<td>$4,297</td>
<td>$17,450</td>
<td>$3,464</td>
<td>$1,205</td>
<td>$5,893</td>
<td>$24,177</td>
<td>$11,631</td>
</tr>
<tr>
<td></td>
<td>Mean QALDs*</td>
<td>153.37</td>
<td>153.38</td>
<td>153.38</td>
<td>153.38</td>
<td>153.37</td>
<td>153.37</td>
<td>153.36</td>
<td>153.37</td>
</tr>
</tbody>
</table>
Abbreviations: GOF: goodness-of-fit. QALD: quality-adjusted life days.
Notes: All calibration analyses were conducted with the Pearson’s GOF measure and an acceptance cut-off value of 9.49 (the 95\textsuperscript{th} quantile of the chi-squared distribution with 4 degrees of freedom) to identify good-fit parameter sets. The following scenarios are described in the data analysis section:

- Initial calibration.
- 2-S SRS (8).
- 2-S SRS (8, 9).
- 2-S LHS (8,9).
- Maximum difference in the estimates of the rank correlation matrices among the transition rates from consecutive sampling designs (e.g., 128k- and 256k-designs).
- Maximum difference in the estimates of the cumulative probability density functions of transition parameters from consecutive designs.
- The second-stage search was conducted with the Latin hypercube sampling design \textit{without} taking into account the rank correlation estimates among the transition parameters from the first-stage search.
- The second-stage search with the Latin hypercube sampling design was conducted taking into account the rank correlation estimates among the transition parameters from the first-stage search.
- Mean costs and mean QALDs were obtained by evaluating the model at the best-fit points (i.e., cohort evaluation).
Table 11. Results of the illustrated cost-effectiveness analysis

<table>
<thead>
<tr>
<th>Calibration scheme</th>
<th>GOF</th>
<th>Acceptance value</th>
<th>Weight</th>
<th># GOF points</th>
<th>Prevention Mean cost</th>
<th>Mean QALD</th>
<th>Current Practice Mean cost</th>
<th>Mean QALD</th>
<th>Incr. cost</th>
<th>Incr. QALD</th>
<th>% prevention is CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>No calibration†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$16,421</td>
<td>135.64</td>
<td>$16,505</td>
<td>135.63</td>
<td>-$84</td>
<td>0.0016</td>
<td>79.19</td>
</tr>
<tr>
<td>Initial cal.‡</td>
<td>Pearson’s 95%</td>
<td>1/GOF</td>
<td>44</td>
<td></td>
<td>$15,736</td>
<td>135.73</td>
<td>$15,787</td>
<td>135.73</td>
<td>-$51</td>
<td>0.0010</td>
<td>80.95</td>
</tr>
<tr>
<td>Initial cal.‡</td>
<td>Pearson’s 50%</td>
<td>1/GOF</td>
<td>5</td>
<td></td>
<td>$15,897</td>
<td>136.30</td>
<td>$15,951</td>
<td>136.30</td>
<td>-$54</td>
<td>0.0010</td>
<td>81.31</td>
</tr>
<tr>
<td>Initial cal.‡</td>
<td>LLKHR††</td>
<td>95%</td>
<td>39</td>
<td></td>
<td>$15,773</td>
<td>135.57</td>
<td>$15,824</td>
<td>135.57</td>
<td>-$51</td>
<td>0.0010</td>
<td>81.29</td>
</tr>
<tr>
<td>2-S SRS (8)☆</td>
<td>Pearson’s 95%</td>
<td>1/GOF</td>
<td>19,570</td>
<td>$16,076</td>
<td>134.82</td>
<td>$16,128</td>
<td>134.82</td>
<td>-$51</td>
<td>0.0009</td>
<td>81.21</td>
<td></td>
</tr>
<tr>
<td>2-S SRS (8)☆</td>
<td>Pearson’s 50%</td>
<td>1/GOF</td>
<td>2,587</td>
<td>$16,003</td>
<td>136.12</td>
<td>$16,055</td>
<td>136.11</td>
<td>-$52</td>
<td>0.0010</td>
<td>80.94</td>
<td></td>
</tr>
<tr>
<td>2-S SRS (8)☆</td>
<td>LLKHR††</td>
<td>95%</td>
<td>17,782</td>
<td>$15,981</td>
<td>134.92</td>
<td>$16,031</td>
<td>134.92</td>
<td>-$50</td>
<td>0.0009</td>
<td>80.91</td>
<td></td>
</tr>
<tr>
<td>2-S SRS (8, 9)§</td>
<td>Pearson’s 95%</td>
<td>1/GOF</td>
<td>12,925</td>
<td>$16,219</td>
<td>134.12</td>
<td>$16,271</td>
<td>134.12</td>
<td>-$53</td>
<td>0.0009</td>
<td>81.09</td>
<td></td>
</tr>
<tr>
<td>2-S SRS (8, 9)§</td>
<td>Pearson’s 50%</td>
<td>1/GOF</td>
<td>1,974</td>
<td>$15,819</td>
<td>135.27</td>
<td>$15,871</td>
<td>135.27</td>
<td>-$53</td>
<td>0.0010</td>
<td>80.84</td>
<td></td>
</tr>
<tr>
<td>2-S SRS (8, 9)§</td>
<td>LLKHR††</td>
<td>95%</td>
<td>11,715</td>
<td>$15,634</td>
<td>136.46</td>
<td>$15,683</td>
<td>136.46</td>
<td>-$49</td>
<td>0.0010</td>
<td>81.03</td>
<td></td>
</tr>
<tr>
<td>2-S LHS (8, 9)‖</td>
<td>Pearson’s 95%</td>
<td>1/GOF</td>
<td>21,126</td>
<td>$16,073</td>
<td>136.22</td>
<td>$16,124</td>
<td>136.22</td>
<td>-$51</td>
<td>0.0010</td>
<td>80.85</td>
<td></td>
</tr>
<tr>
<td>2-S LHS (8, 9)‖</td>
<td>Pearson’s 95%</td>
<td>p-val ††</td>
<td>21,126</td>
<td>$15,880</td>
<td>135.50</td>
<td>$15,933</td>
<td>135.50</td>
<td>-$53</td>
<td>0.0010</td>
<td>81.16</td>
<td></td>
</tr>
<tr>
<td>2-S LHS (8, 9)‖</td>
<td>Pearson’s 50%</td>
<td>1/GOF</td>
<td>3,256</td>
<td>$15,941</td>
<td>135.18</td>
<td>$15,993</td>
<td>135.18</td>
<td>-$52</td>
<td>0.0010</td>
<td>81.05</td>
<td></td>
</tr>
<tr>
<td>2-S LHS (8, 9)‖</td>
<td>LLKHR††</td>
<td>95%</td>
<td>18,837</td>
<td>$15,873</td>
<td>136.12</td>
<td>$15,921</td>
<td>136.12</td>
<td>-$48</td>
<td>0.0010</td>
<td>80.72</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The cost-effectiveness analysis was conducted via a probabilistic sensitivity analysis (PSA) of the calibrated model (n=20,000).

The probability weights are defined in step 8 of the calibration process. The number of good-fit parameter sets obtained at different stages of the proposed algorithm. Prevention is more cost-effective than current practice at a willingness-to-pay value of $50,000 per quality-adjusted life year gained. The following scenarios are described in the data analysis section: \textdagger No calibration. \textdaggerdbl Initial calibration. \textasteriskcentered SRS (8). \textasteriskcentered SRS (8, 9). \textasteriskcentered LHS (8, 9). \textdaggerdbl The significance levels (i.e., p-values) assuming that Pearson’s GOF measures associated with good-fit parameter sets are from a chi-squared distribution (with 4 degrees of freedom) are used to derive the probability weights for the PSA (see step 8 of the methods section).
Chapter 6
Cost-Effectiveness of Pressure Ulcer Prevention

6 Cost-effectiveness of early prevention - Methods

6.1 Overview
Preliminary results indicate that although early prevention is costly, the health improvement and the treatment saving from averted pressure ulcers more than offset the additional costs, if the prevention is targeted at high-risk patients. In particular, early prevention is relevant to elderly patients admitted to inpatient care through emergency departments, and patients undergoing elective surgery with long surgical duration. This chapter describes a collection of methods that are used in the cost-effectiveness analysis of early prevention in these high-risk populations.

A decision model was used to structure the decision problem of adopting early intervention, and simulated the natural history of pressure ulcer. The model was used to project the downstream consequences of early prevention and to quantify the uncertainty in the decision to adopt early prevention. The latter includes probabilistic sensitivity analysis to estimate the probability that early prevention is cost effective and a value-of-information analysis to estimate of the cost of uncertainty associated with the adoption decision.

6.2 Pressure ulcer decision model

6.2.1 Intended use
The decision model was developed to support multiple cost-effectiveness applications that evaluate pressure ulcer prevention strategies for different settings, including acute care, long-term care, and home care. The underlying economic assumptions are described in subsequent cost-effectiveness analyses (chapters 7 and 8). In brief, these analyses were conducted according to the guidelines by the Canadian Agency for Drugs and Technologies in Health. All costs, expressed in 2009 Canadian dollars, were calculated from the perspective of a single health care payer. Future health outcomes and costs were discounted at 5%.
6.2.2 Rationale
We used a Markov model to simulate health events related to the prognosis of pressure ulcers and to project associated costs and QALYs. In brief, the time horizon of a Markov model is divided into discrete and non-overlapping cycles. Markov models assume that at any cycle, a simulated patient is in one of a set of mutually exclusive health states. All health events are represented as transitions between health states. Each health state is assigned an incremental cost and an incremental QALY. In evaluating the model, each simulated patient begins in an initial health state and over the time horizon, may move between health states in subsequent cycles according to some transition probabilities. The specific sequence of transitions generates the projected cost and QALYs for that patient. At the end of the simulation, the mean QALYs and mean costs are estimated across all simulated patients.

We selected a Markov model with a cycle length of one day over a time horizon of one year because the risk of pressure ulcer is continuous in hospitalized patients, the timing of a pressure ulcer is important for prevention, and patients may experience recurrent pressure ulcers over time and at different bony prominences.

We used a time horizon of one year that spans the hospital length of stay and a period of post-discharge convalescence that allows for more than twice the average healing time of the most severe hospital-acquired pressure ulcers. Longer time horizons were also considered to evaluate the impact of the selected time horizon on the simulation results.

6.2.3 Model structure
The decision model was structured and developed with inputs from an expert panel convened by the Ontario Health Technology Assessment Committee and the Toronto Health Economics and Technology Assessment Collaborative. In the Markov model, the likelihood of health events related to pressure ulcers depends on the effectiveness of existing skin care protocols for the prevention and treatment of pressure ulcers in the hospitals. Simulated patients are exposed to a daily risk of developing pressure ulcer. The severity of pressure ulcers is categorized according to the classification system by the joint collaboration between the U.S. National Pressure Ulcer Advisory Panel and the European Pressure Ulcer Advisory Panel (section 2.3.1), including the following classification stages.
- Stage 1 - a persistent non-blanchable skin redness,
- Stage 2 - a loss of partial skin thickness that appears as an abrasion, blister, or shallow crater,
  Stage 3 - a loss of full skin thickness, presented as a deep crater,
- Stage 4 - a loss of full skin thickness, exposing muscle or bone.¹

In the development of the model structure, we had also considered deep tissue injury as a health state, but ended up excluding this pressure ulcer stage due to a lack of studies describing its manifestation in the targeted populations.⁵⁷, 201, 215

Figure 9 displays the key health states of a pressure ulcer history model for hospital patients. The health states are structured to reflect a validated staging classification (see footnotes in Figure 9) and are stratified by wound status, including pressure ulcers in which the underlying risk factors remain and pressure ulcers that are susceptible to healing with improvement in the underlying risk factors.¹⁸⁶ If the risk factors (e.g., immobility) associated with the development of a pressure ulcer remain unchanged, the pressure ulcer may progress to higher severity stage (stage 1 → 2, 2 → 3, 3 → 4) and it may be at risk for local infection (stage 2-4 pressure ulcers)²¹⁶ and systemic infection (stage 3-4 pressure ulcers).²¹⁷ If the underlying risk factors are reduced (e.g., with improving mobility and overall health), the tissue damage may heal under intact skin (for stage 1 pressure ulcers) or the open wound may start to re-epithelialize until full closure (stage 2-4 pressure ulcers). In the absence of the risk factors, healing pressure ulcers are assumed to complete the healing process over an average duration, and during this time, they are not at risk of worsening tissue damage and infection. We assumed that systemic infection was preceded by local infection. Simulated patients with systemic infection were assumed to be monitored in the hospital, and in rare instances, could die from the infection.²¹¹ At any time, simulated patients could die due to causes unrelated to pressure ulcers. Although not shown in Figure 9, the health states are also stratified by setting (i.e., in hospital or at home post-discharge).

We assumed that all patients who had survived an acute episode of care were discharged home. For simplicity, we did not simulate the scenario in which patients were directly admitted to long-term care facilities from hospitals. Also, we did not simulate the scenario in which patients were subsequently admitted to long-term care facilities after being discharged home. Patients with hospital-acquired pressure ulcers were to be receiving home care for their pressure ulcers. All hospital-acquired pressure ulcers were assumed to heal according to their average stage-specific
healing times, ranging from 11 to 22 weeks. The pressure ulcer history model was constructed in TreeAge Pro 2009 Suite (TreeAge Software, Williamstown, Massachusetts).
Figure 9. Structure of a pressure ulcer decision model

Notes: Reproduced Figure 4. PrU: pressure ulcer. Pressure ulcers are classified as stage 1, 2, 3 or 4 as described above. The symbol $P_{01}$ denotes the daily incidence of developing a stage-1 pressure ulcer; $P_{12}$ the daily probability that a pressure ulcer worsens from stage 1 severity to stage 2 severity; $P_{23}$ from a stage 2 to a stage 3; and $P_{34}$ from a stage 3 to a stage 4. The symbol $H_{10}$ denotes the healing rate of a stage-1 pressure ulcer, $H_{20}$ the healing rate of a stage-2 pressure ulcer, and so on.
6.3 Model inputs

We obtained model input estimates from the Tri-Hospital Pressure Ulcer Prevalence Survey, the Ontario Case Costing Initiative,\textsuperscript{218} a large data set for home care in Ontario,\textsuperscript{219} and additional literature search.

6.3.1 Prevalence of hospital-acquired pressure ulcers

We first obtained data from the Toronto Tri-Hospital Acute Care Pressure Ulcer Prevalence Survey, an annual one-day survey of pressure ulcer prevalence in three acute care hospitals (personal communication, Ms. Laura Teague, Wound Care Program, St. Michael’s Hospital, Toronto, October 14, 2008). The survey uses a standard data collection form to collect patient characteristics and prevalence data from two tertiary care and one community care hospitals (Appendix 3). On the day of surveillance, skin care nurses conducted head-to-toe inspections of patients who had been admitted to the medical and surgical wards of the participating hospitals and who had provided informed consent to the skin inspection. Clinical bedside audit and full chart review were performed for all detected pressure ulcers to determine if they had been hospital-acquired or had been existed upon hospital admission.

Annual prevalence data from the Toronto Tri-Hospital Acute Care Pressure Ulcer Prevalence Survey (2005-2007, n=3307 patients) were pooled to smooth out year-to-year variation. The prevalence of hospital-acquired pressure ulcers was calculated by dividing the number of patients who developed pressure ulcers after admission by the total number of surveyed participants. The prevalence of hospital-acquired pressure ulcers was 14.4\% among hospital patients (n=3307), 19.6\% in elderly admitted ED patients (n=745), and 17.1\% in surgical patients (n=1118).\textsuperscript{182,184}

There are a number of limitations to the survey data. First, the survey participants who consented to the full-body skin inspection could have been at a higher risk of developing pressure ulcer than other patients who declined to participate in the survey. Secondly, documentation of pressure ulcers at the time of hospital admission may be incomplete since other care activities at the time of admission take higher priority than skin and risk assessment for the developing of pressure ulcer.\textsuperscript{57} Thus, there could have been more pressure ulcers detected among the surveyed
participants, and fewer pressure ulcers recorded upon hospital admission. This might result in an exaggeration of the reported prevalence of hospital-acquired pressure ulcers. We verified the potential for this over-estimation below.

We conducted a literature search of Medline (MESH terms: “pressure ulcer” AND “Canada”, January 1990 to May 2012) to identify studies reporting the prevalence of pressure ulcers in Canadian hospitals. We identified three relevant studies. VanDenKerkhof et al. 2011 conducted an annual 1-day census of all inpatients in an acute care hospital in eastern Ontario. The census involved a full head-to-toe skin assessment and risk assessment of every patient in the hospital between 6:00 a.m. and 6:00 p.m. and all patients admitted during the 12 hours of the survey. Pressure ulcer status on admission was captured from the patients’ admission notes. Between 1998 and 2008, the prevalence of hospital-acquired pressure ulcers ranged between 11% and 14%. Woodbury et al. 2005 reported a similar prevalence of 14.3% (95% confidence interval: 9.6%, 19.0%) using data from 213 patients from three acute care hospitals. Hurd et al. 2009 reported wound prevalence from wound audits carried out at 13 acute hospitals in Canada in 2006 and 2007. Funded by device manufacturers for wound prevention, the audits were carried out in each hospital by a team of advanced practice nurses using standard data-collection forms. The mean prevalence of pressure ulcers was 22.9% (range: 15%, 39%). A consistently high proportion (mean 79.3%, range: 68-91%) of pressure ulcers were hospital-acquired, suggesting a prevalence of hospital-acquired pressure ulcer of approximately 18%. These reported prevalence estimates of hospital-acquired pressure ulcers suggest that the prevalence estimates we used in the development of the decision model are reasonably consistent with other external data sources.

6.3.2 Progression and healing rate estimates

Figure 9 displays the daily progression and healing rates of stage-specific pressure ulcers, including daily incidence of developing pressure ulcers ($P_{01}$), progression rates from stage 1 to 2 ($P_{12}$), stage 2 to 3 ($P_{23}$), stage 3 to 4 ($P_{34}$), and healing rates of stage 1 ($H_{10}$), stage 2 ($H_{20}$), stage 3 ($H_{30}$), and stage 4 ($H_{40}$) pressure ulcers. With few exceptions, data regarding the prognosis of pressure ulcer is limited.
Because there are no data to concurrently determine the transition rates, we calibrated the decision model to the observed prevalence of stage-specific pressure ulcers in order to estimate the transition rates. Full details of this calibration analysis are in chapter 5. The methods we used for this calibration analysis were supported by evidence from a literature review of different approaches to calibration of decision models (chapter 4). Figure 10 displays the observed and projected prevalence of stage-specific pressure ulcers for the target populations. Table 12 displays the corresponding estimates of daily progression and healing rates, including their 95% credible ranges derived from the calibration.

6.3.3 Incidence of ED-acquired pressure ulcers
Evaluating preventive interventions for patients in the ED is difficult because few studies examine risk factors and incidence of pressure ulcers originating from the EDs.\textsuperscript{77,222} The decision model used a calibrated daily incidence of developing stage-1 pressure ulcer of 2.95% (Table 12). The projected incidence from the decision model in the first two days following an ED visit was 5.8%. According to results from a prospective cohort study including 3,233 elderly admitted ED patients, the reported cumulative incidence of hospital-acquired pressure ulcers in 2 days after admission was 6.2% (95% confidence interval, 5.4\%–7.1\%).\textsuperscript{77}

We derived the hourly incidence from the calibrated daily incidence of developing pressure ulcers, assuming that pressure ulcers would occur at a constant hourly rate over any day in the hospital length of stay. This assumption was made mainly due to a lack of data to characterize the risk of pressure ulcer over time.
Figure 10. Observed and projected prevalence of stage-specific pressure ulcers

Notes: Prevalence was estimated from 745 elderly admitted ED patients (left panel) and 1,118 surgical patients (right panel). Bars are standard errors of observed prevalence estimates, derived assuming binomial distributions given the sample sizes of the target populations. For the left panel, only prevalence estimates of stage 1-3 pressure ulcers are plotted. The observed prevalence of stage-4 pressure ulcers was 0% and the projected prevalence was 0.04%. Analytical details regarding how the projected prevalence estimates were obtained via calibration analysis are described in chapter 5.
Table 12. Transition rate estimates (ranges) obtained via calibration analysis

<table>
<thead>
<tr>
<th></th>
<th>Elderly admitted ED patients</th>
<th>Surgical patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily incidence of stage-1 pressure ulcer (%)</td>
<td>2.95 (2.69, 3.58)</td>
<td>5.46 (1.67, 13.68)</td>
</tr>
<tr>
<td>Daily progression rate during hospitalization (%)</td>
<td>20.80 (18.66, 21.62)</td>
<td>19.57 (15.20, 23.06)</td>
</tr>
<tr>
<td>Stage 1 → 2</td>
<td>1.11 (0.91, 1.22)</td>
<td>2.23 (2.13, 4.88)</td>
</tr>
<tr>
<td>Stage 3 → 4</td>
<td>0.30 (0.06, 0.55)</td>
<td>4.64 (3.69, 4.78)</td>
</tr>
<tr>
<td>Daily healing rate during hospitalization (%)</td>
<td>2.34 (0.50, 4.97)</td>
<td>9.24 (3.42, 13.75)</td>
</tr>
<tr>
<td>Stage 1</td>
<td>0.97 (0.12, 1.78)</td>
<td>2.23 (1.58, 10.69)</td>
</tr>
<tr>
<td>Stage 2</td>
<td>0.37 (0.05, 0.76)</td>
<td>1.11 (0.43, 5.40)</td>
</tr>
<tr>
<td>Stage 4</td>
<td>0.04 (0.01, 0.07)</td>
<td>0.14 (0.03, 0.80)</td>
</tr>
</tbody>
</table>

Notes: Estimates and ranges were obtained via calibration to ensure that projected stage-specific prevalence from the pressure ulcer policy model reproduced observed prevalence from the Toronto Tri-Hospital Acute Care Pressure Ulcer Prevalence Survey (2005-2007). Full details of the calibration analysis are in chapter 5.

Given the hourly incidence, we estimated the incidence of ED-acquired pressure ulcers over an average ED stay of approximately 15.4 hours. To evaluate the effects and costs of early prevention in the ED setting, the associated relative risk estimate of early prevention was applied to the incidence of ED-acquired pressure ulcers.

6.3.4 Intra-operative incidence of pressure ulcers

In Table 12, the calibrated incidence of stage-1 pressure ulcers on the day of surgery was 5.5% (range: 1.7%-13.7%). According to results from studies identified in the literature review to inform the plausible ranges of the parameters that were determined via the calibration analysis (section 5.3.3, step 1 of the calibration process), the reported incidence estimates vary according to at-risk populations and the duration of immobility due to surgery: 1.7% in patients who underwent open chest cardiac surgery, 2.0% in patients who underwent elective major general, gynecological and vascular surgery, 3.8% in patients who underwent elective cardiac surgery, 5.2% in patients who underwent scheduled surgery lasting > 4 hours from nine
surgical specialties, and 10.3% in patients who underwent scheduled cardiac surgery with extracorporeal circulation.

We derived the hourly incidence from the calibrated incidence of developing pressure ulcers on the day of surgery, assuming that the hourly incidence was constant over the one-day duration of exposure to the risk of developing pressure ulcer. Given the hourly incidence, we then estimated the intra-operative incidence of developing pressure ulcers over an average surgical duration of 4.6 hours. We estimated the intra-operative incidence reduction associated with prevention by applying the relative risk estimate due to prevention to the intra-operative incidence.

Surgical patients are at risk of developing pressure ulcers when their mobility is restricted due to surgery. The duration of restricted mobility starts pre-operatively with anesthesia, encompasses the surgical duration, and does not end post-operatively until arrival in the recovery room. We recognize that the duration of restricted mobility due to surgery is longer than the surgical duration. However, given the lack of reliable data on the former duration, we used the surgical duration in the calculation of the intra-operative incidence. Our calculation is likely to underestimate the intra-operative risk of pressure ulcers; consequently, our cost-effectiveness analysis is likely to bias against the intra-operative prevention strategy.

6.3.5 Hospital costs attributable to pressure ulcers
We requested costing data from a validated data source, the Ontario Case Costing Initiative. The Ontario Case Costing Initiative uses the same case costing standards to collect data from 47 teaching and community hospitals in Ontario. Direct hospital costs are aggregated from ward care costs, pharmacy costs, overhead costs and capital depreciation costs of hospital infrastructure.

We used estimates of home care costs attributable to pressure ulcers after discharge from an additional analysis of data from a published study (personal communication, Dr. Jeff Poss, University of Waterloo, December 16, 2008). A costing dataset previously used to validate the Resource Utilization Groups for Home Care (RUG-III/HC) was used for this estimation. In brief, episodic costs over a 13-week period were aggregated from billing records of individual
home care clients. Direct home care costs included costs for nursing, personal support, dietary services, social work, physical and occupational therapies.

6.3.6 Quality weights and QALYs
We conducted a literature search (Medline, keywords (“EQ-5D”[text word] or “Health Utility” [text word] or “HUI”[text word] and “pressure ulcer”[MESH terms], January 1990 to December 2009) to identify data sources that could be used to derive quality weights (or health utilities) for the cost-effectiveness analysis. The relevant studies we identified used the Health Utilities Index scale to derive preference-based health utilities. The Health Utilities Index-Mark 3 assesses the following attributes: vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain. In general, pain is the most significant consequence of having a pressure ulcer and the induced pain affects every aspect of patients’ lives. Patients also reported that pressure ulcers restrict normal activities of daily living, resulting directly from reduced physical activity due to pressure ulcer care.

We derived the mean quality weight among hospitalized patients from a study of 1,207 hospital patients. A retrospective population-based study of 16,531 long-term care residents in Ontario was the only study we identified that report differential quality weights among those with and without pressure ulcers. Compared to normal or intact skin, stage 2-4 pressure ulcers were associated with on average a 6% relative decrement in quality weights. We derived QALYs by adjusting the time simulated patients spent in different health states of the decision model using the quality weight estimates.

6.4 Cost-effectiveness analysis
6.4.1 Base case analysis
Relative to current practice, we calculated the incremental cost and incremental QALYs associated with early prevention, and if appropriate, the associated ICER. We also expressed the ICERs in terms of net monetary benefits, using a societal willingness-to-pay threshold of $50,000 per QALY. This value is frequently cited, but an agreed-upon valuation does not exist. We therefore conducted sensitivity analysis by varying the threshold up to $100,000.
6.4.2 One-way sensitivity analysis

We conducted one-way sensitivity analyses by varying each of the key model inputs over its plausible range to assess the corresponding variation in the net monetary benefits of early prevention, with positive net benefits indicating that prevention is more cost-effective than current practice.

6.4.3 Probabilistic sensitivity analysis

The net monetary benefit of prevention in the base case analysis was estimated with input uncertainty and variability. We therefore conducted a probabilistic sensitivity analysis to assess the impact of input uncertainty and variability on the estimated net monetary benefits associated with prevention. We independently sampled 20,000 random values from the input distributions and with the random inputs, repeated the cost-effectiveness analysis 20,000 times to derive the probability that early prevention is more cost effective than current practice via the number of samples with positive net monetary benefits. The plot of this probability at different value of the willingness-to-pay threshold forms the cost-effectiveness acceptability curve.

6.4.4 Value-of-information analysis

The clinical evidence evaluating the use of pressure-redistribution mattresses in the EDs is uncertain. Compared with standard hospital mattresses, the estimated relative risk of developing pressure ulcers associated with pressure-redistribution mattresses is not statistically significant, according to a randomized controlled trial evaluating these mattresses in elderly admitted ED patients. We conducted a value-of-information analysis to quantify the expected cost of making the wrong choice between early prevention and current practice. An illustration of the VOI analysis is briefly illustrated below (Table 13).
Table 13. An illustration of the value of information calculation

<table>
<thead>
<tr>
<th>Realization</th>
<th>Net Monetary Benefit ($)</th>
<th>Decision</th>
<th>VOI or Cost of Uncertainty ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>48</td>
<td>Favour P</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>-2</td>
<td>Favour C</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>52</td>
<td>Favour P</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>-3</td>
<td>Favour C</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>-2</td>
<td>Favour C</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>46</td>
<td>Favour P</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>38</td>
<td>Favour P</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>80</td>
<td>Favour P</td>
<td>0</td>
</tr>
<tr>
<td>Expected value</td>
<td>32</td>
<td></td>
<td>0.88</td>
</tr>
</tbody>
</table>

Notes: Net monetary benefit was calculated for prevention versus current practice, with positive values indicating that prevention is more cost effective than current practice. Abbreviation: VOI: value of perfect information. P: prevention. C: current practice.

Table 13 illustrates the projected net monetary benefits of early prevention, given the current information that for example, consists of eight realizations of model inputs. The expected net benefit is $32, and there is a 63% (5/8) chance that early prevention is more cost-effective than current practice (or equivalently, the decision uncertainty is characterized by a 37% false error rate). Given the current information, the decision must be made before we know how the decision uncertainty will be resolved, that is, we must make a decision now based upon the expected net benefit. In this case, the decision is in favor of prevention. With perfect information, assuming to be contained in each of the realizations that represents a future realization of the true model inputs, we know how the uncertainty will resolve, so we could make different decisions for different resolutions of the net benefit. In Table 13, the resolution of all uncertainties with perfect information leads to the same decisions as with current information five times (i.e., no monetary gain with perfect information or equivalently, no monetary loss due to uncertainty), and to the reversed decisions three times (i.e., some gain with perfect information or equivalently, some cost of uncertainty). The expected value of perfect information (or equivalently, the expected cost of uncertainty) of 0.88 is calculated over all the realizations. This analysis can be done including all model input uncertainties to derive the expected value of perfect information (EVPI). It can also be done for uncertainty associated with some model parameters to derive the expected value of perfect information for the parameters (EVPPI).
In the cost-effectiveness analysis of a strategy to prevent pressure ulcers among elderly admitted ED patients (chapter 7), the relative risk of prevention was estimated with a large confidence interval. This was due to the small randomized controlled trial that evaluates the preventive intervention. We therefore derived the expected value of perfect information for this relative risk to facilitate the decision as to whether (or not) to invest in a larger trial to collect additional information on this parameter. We used a 2-level algorithm to conduct the value-of-information analysis. For the second-order simulation to deal with the parameter uncertainty, we obtained 1000 samples from the log-normal distribution of the relative risk estimate. To evaluate the decision model via microsimulation, we conducted the first-order simulation with 3,000 samples from all the remaining input distributions.

6.5 Limitations

There are important limitations to our cost-effectiveness analyses. First, the decision model we used does not take the anatomic sites of pressure ulcers into account although multiple bony prominences are at risk for skin breakdown. For example, among patients who underwent elective surgery with a surgical duration ≥4 hours, 21% of them experienced pressure ulcers within two days post-surgery, and 39% of whom experienced multiple pressure ulcers. In the case of multiple pressure ulcers, our models represent only the one with the highest classification stage. This simplification tends to under-estimate the clinical effectiveness and cost-effectiveness of early prevention since overall risk reduction at multiple sites and stages is not accounted for in the decision model. Therefore our cost-effectiveness results are likely to be conservative, in the sense that the results tend to bias against prevention.

We used a pressure ulcer classification proposed by the joint collaboration between the US National Pressure Ulcer Advisory Panel and the European Pressure Ulcer Advisory Panel that did not include deep tissue injury or un-staged pressure ulcers (e.g., pressure ulcers that are covered with eschar). Depending on the definition, the incidence of deep tissue injury however is small in our target populations and we did not expect these omissions to greatly influence the cost-effectiveness results.
6.6 Discussion

We describe a collection of methods for the cost-effectiveness analysis of early prevention of pressure ulcers in high-risk hospital patients. A decision model was used to structure the decision problem of adopting early intervention, characterize key assumptions in the analysis, simulate the natural history of pressure ulcers, and quantify the associated decision uncertainty. The model was calibrated to reflect the prognosis and burden of pressure ulcers in high-risk patients, perusing the two-stage random search approach to model calibration (chapter 5). The calibrated decision model was used to evaluate the cost-effectiveness of early prevention in elderly admitted ED patients (chapter 7) and elective surgical patients (chapter 8).
Chapter 7 Early Prevention of Pressure Ulcers Among Elderly Patients Admitted through Emergency Departments

7 Early prevention in Emergency Departments

7.1 Abstract

**Background:** Every year, approximately 6.2 million hospital admissions through emergency departments (ED) involve elderly patients who are at risk of developing pressure ulcers. We evaluated the cost-effectiveness of pressure-redistribution foam mattresses on ED stretchers and beds for early prevention of pressure ulcers in elderly admitted ED patients.

**Methods:** Using a Markov model, we evaluated the incremental effectiveness (quality-adjusted life-days) and incremental cost (hospital and home care costs) between early prevention and current practice (with standard hospital mattresses) from a health care payer perspective during a 1-year time horizon.

**Results:** The projected incidence of ED-acquired pressure ulcers was 1.90% with current practice and 1.48% with early prevention, corresponding to a number needed to treat of 238 patients. The average upgrading cost from standard to pressure-redistribution mattresses was $0.30 per patient. Compared with current practice, early prevention was more effective, with 0.0015 quality-adjusted life-days gained, and less costly, with a mean cost saving of $32 per patient. If decision makers are willing to pay $50,000 per quality-adjusted life-year gained, early prevention was cost-effective even for short ED stay (i.e., 1 hour), low hospital-acquired pressure ulcer risk (1% prevalence), and high unit price of pressure-redistribution mattresses ($3,775). Taking input uncertainty into account, early prevention was 81% likely to be cost-effective. Expected value-of-information estimates supported additional randomized controlled trials of pressure-redistribution mattresses to eliminate the remaining decision uncertainty.

**Conclusion:** The economic evidence supports early prevention with pressure-redistribution foam mattresses in the ED. Early prevention is likely to improve health for elderly patients and save hospital costs.
7.2 Background

Pressure ulcers may develop when persisting pressure on bony prominences obstructs healthy capillary flow, leading to tissue necrosis. Elderly patients are at high risk of developing pressure ulcers because of immobility, poor nutritional status, impaired mental status, and incontinence. Annually, elderly patients account for 30% of the 117 million emergency department (ED) visits, resulting in 6.2 million admissions to US hospitals. Elderly admitted patients typically spend hours in the EDs, especially during crowding period. Prior to admission, these patients could be at risk of developing pressure ulcers as they lie for considerable time on unyielding diagnostic equipment surfaces, stretchers and standard hospital mattresses in the EDs.

Baumgarten et al. 2006 prospectively studied 3,233 elderly admitted ED patients. Approximately 6% of these patients acquired pressure ulcers within 48 hours of admission, with higher incidence among patients with severe immobility, malnutrition, and incontinence. Because pressure ulcers that developed early in the hospital stay could account for approximately one third of all hospital-acquired pressure ulcers among elderly patients, prevention may need to start early.

7.2.1 Importance

Most pressure ulcers are considered to be preventable. Clinical approaches to prevention include overall assessment of the patients, active mobilization of patients who are able to walk, and regular repositioning of bedbound patients. In particular, patients at risk of developing a pressure ulcer should not remain on a standard hospital mattress; a pressure-redistribution mattress should be used. Pressure-redistribution mattresses typically mould around the shape of the patient to distribute the patient’s weight at bony prominences over a large area to reduce pressure-induced ischemia.

Rich et al. 2009 studied the use of pressure-redistribution devices early in the hospital stay of 792 elderly admitted ED patients. Only half of the patients assessed to be at high risk of developing pressure ulcers had a preventive device. The results suggest that early prevention is suboptimal even among high-risk patients.
Ayello et al. 2009 examined legal issues in the prevention and treatment of pressure ulcers; they suggested that failure to prevent pressure ulcers increases the risk of litigation. The Centers for Medicare and Medicaid Services has designated pressure ulcers as preventable complications of medical care (“never events”), and no longer reimburse hospitals for the cost of treating hospital-acquired pressure ulcers. The aim of these policy changes is patient safety.

7.2.2 Goal of this investigation
According to recent systematic reviews, pressure-redistribution foam mattresses significantly reduce the incidence of pressure ulcers in hospital patients by approximately 60%. The price of pressure-redistribution mattresses varies considerably, from approximately $350 for foam types to over $4,000 for low air-loss mattresses. The trade-off between the additional costs and health benefits of using pressure-redistribution mattresses in the ED setting has not been evaluated. Commissioned by the Ontario Health Technology Advisory Committee, a panel that makes recommendations to the Ontario Ministry of Health and Long-Term Care regarding the uptake and diffusion of health technologies, we evaluated the cost-effectiveness of pressure-redistribution foam mattresses on ED stretchers/beds for early prevention of pressure ulcers among elderly patients admitted to hospitals through EDs.

7.3 Methods
7.3.1 Study design
We conducted a cost-effectiveness analysis using the pressure ulcer history model described in chapters 5 and 6. We followed guidelines for economic evaluation by the Canadian Agency for Drugs and Technologies in Health. The base case analysis was conducted from a health care payer perspective and sensitivity analysis from a hospital perspective. We included hospital costs and post-discharge costs of home care for hospital-acquired pressure ulcers. Health outcomes included incidence of ED-acquired pressure ulcers and quality-adjusted life years.

We used a time horizon of one year, including the hospital length of stay and a period of post-discharge convalescence, and allowing for more than twice the average healing time of severe pressure ulcers (Table 14). Sensitivity analysis was conducted with a 5-year time horizon to evaluate the effect of early prevention on pressure-ulcer related mortality. Due to the short time
horizon, health outcomes and costs were not discounted. All costs are expressed in 2009 Canadian dollars.63

7.3.2 Setting and participants
For the base case analysis, we simulated a cohort of elderly admitted ED patients, defined as patients who were ≥ 65 years of age, sought emergency care in the EDs and were subsequently admitted to inpatient medical or surgical services of an acute care hospital.

7.3.3 Early prevention strategy
We compared current practice that uses standard mattresses on ED beds (e.g., 5-inch thick mattresses) and on ED stretchers (e.g., 3-inch thick mattresses) with an early prevention strategy which uses pressure-redistribution foam mattresses on ED beds (e.g., 8-inch thick mattresses) and on ED stretchers (e.g., 5-inch thick mattresses). Upon admission to inpatient wards, we assumed that simulated patients were nursed according to current prevention practice, with patients assessed to be at high risk of developing pressure ulcers being placed on pressure-redistribution mattresses.1,21,237

7.3.4 Pre-model data analysis
We obtained input estimates from current systematic reviews of pressure ulcer prevention,29,54 the Tri-Hospital Pressure Ulcer Prevalence Survey, the Ontario Case Costing Initiative,218 a large data set for home care in Ontario,219 and additional literature searches. Table 14 displays all model inputs, including relative risk estimates of pressure-redistribution mattresses, transition rates of the pressure ulcer history model, cost estimates, and health-related quality of life weights. Model inputs related to the pressure ulcer history model are described in chapter 6. For example, the transition rates across the stages of pressure ulcers were estimated via a calibration analysis. We described specific inputs related to the cost-effectiveness analysis below.

7.3.4.1 Effectiveness of pressure-redistribution mattresses
We identified only one randomized controlled trial that involved ED patients from two recent systematic reviews of pressure ulcer prevention.29,24 In terms of quality of the evidence, the trial by Gunningberg et al. 2000 has adequate reporting of randomization and blinded outcome assessment.185 In the trial’s experimental group, 48 participants with a suspected hip fracture were each placed on a 10-cm thick visco-elastic foam mattress immediately upon arrival in the
ED, and when transferred to the ward, were placed on a 7-cm visco-elastic foam overlay on top of a standard hospital mattress. In the control group, 53 participants were placed on a standard trolley mattress (5-cm foam) and then on a standard hospital mattress (10-cm foam), respectively. Over an average follow-up of 12 days, 12 and 17 patients in the experimental and control group developed pressure ulcers, respectively. The relative risk estimate associated with the pressure-redistribution mattresses was 0.78 [95% confidence interval: 0.42, 1.46], a statistically non-significant prevention effect (Table 14). We used this prevention effect estimate in our base case analysis.

We also considered evidence from other settings that supports the effectiveness of pressure-redistribution foam mattresses. According to data from five randomized controlled trials (n=2,016 patients) included in the McInnes et al. 2008 systematic review, pressure-redistribution foam mattresses, compared to standard hospital mattresses, significantly reduced the incidence of pressure ulcers in hospital patients (relative risk estimate of 0.40 [0.21, 0.74], Table 14). Assuming that the prevention effect with pressure-redistribution foam mattresses is similar for patients during ED stay or inpatient care, we derived a pooled relative risk estimate of 0.45 [0.27, 0.74] from all six trials, including data from the Gunningberg et al. 2000 trial (n=2127 patients, Table 14). We used this pooled estimate in a sensitivity analysis.

**7.3.4.2 Incidence of ED-acquired pressure ulcers**

The procedure used to estimate the incidence of ED-acquired pressure ulcers is described as part of the development of the pressure ulcer history model in chapter 6. To evaluate the health benefits and costs of pressure-redistribution mattresses in the ED setting, the associated relative risk estimate was applied to the incidence of ED-acquired pressure ulcers.

**7.3.4.3 Costs of pressure-redistribution foam mattresses**

The average prices and lifespan of pressure-redistribution mattresses for ED stretchers and beds were obtained from a telephone survey with 3 manufacturers in October 2009. The differential cost of upgrading from a standard hospital mattress to a pressure-redistribution mattress was amortized over the average lifespan of the mattresses to derive the per-patient cost of early prevention, assuming that over an average ED stay, patients were supported by pressure-redistribution mattresses (Table 14).
7.3.5 Outcome measures
Outcome measures were described as part of the development of the pressure ulcer history models in chapter 6.

7.3.6 Analytical approaches
Analytical approaches for the cost-effectiveness analysis were described as part of the development of the pressure ulcer history model in chapter 6. The joint distribution of model transition rates derived via calibration in chapter 5 was used in the probabilistic sensitivity analysis to estimate the probability that prevention is more cost effective than current practice. This joint distribution was also used in the value of information analysis to estimate the cost of uncertainty associated with the prevention effect estimates of pressure-redistribution mattresses.

7.4 Results

7.4.1 Clinical effectiveness
Table 15 displays results of the base case analysis. Over an average ED stay of 15.4 hours, the model predicted that the incidence of pressure ulcers that originated in the ED (i.e., ED-acquired pressure ulcers) would be 1.48% (range: 0.73%, 3.38%) with early prevention and 1.90% (1.74%, 2.31%) with current practice, corresponding to a difference of -0.42% (-1.52%, 1.10%). On average, one would need to institute early prevention for 238 elderly admitted ED patients to prevent one ED-acquired pressure ulcer.

7.4.2 Cost-effectiveness analysis - Base case
The average cost of upgrading from standard to pressure-redistribution mattresses on ED beds and stretchers was 30 cents per patient (Table 14). Compared to current practice, early prevention was not associated with any harmful effects, as indicated by a very small health benefit of 0.0015 quality-adjusted life days gained (Table 15). The net benefit of prevention was $32.36, consisting of $0.21 from the health benefit of attaining higher mean quality-adjusted life days, and $32.15 of cost saving per patient.

Annually, approximately 240,000 elderly patients are admitted via EDs to acute care hospitals in Ontario, a province with approximately 13 million people, of whom 14% are elderly.
Targeting these patients for early prevention, the model projected that the strategy would prevent 1,005 ED-acquired pressure ulcer cases and save approximately $7.2 million in hospital costs per year.

### 7.4.3 One-way sensitivity analysis

Figure 11 displays the net benefit of early prevention according to uncertain model inputs. The two top bars of Figure 11 show variation in the net benefit of prevention according to the 95% confidence intervals around the prevention effect estimates. From the Gunningberg et al. 2000 trial that involved ED patients, the relative risk estimate of pressure-redistribution mattresses versus standard mattresses was 0.78 [0.42, 1.46]. Across this confidence interval, the net benefit of prevention ranged from a net loss of $74 to a net gain of $89 per patient, indicating that the cost effectiveness of prevention is uncertain if the Gunningberg relative risk estimate is accepted as the best evidence of effectiveness (top bar). Borrowing strength from the evidence concerning hospital patients, the pooled relative risk estimate was 0.45 [0.27, 0.74], as displayed by the second bar from the top. Across this confidence interval, the net benefit of prevention ranged from $89 to $163 per patient, indicating that prevention is cost effective.

The mean ED stay was 15.4 hours, according to data from the Tri-Hospital survey (Table 14). For ED cases admitted to acute care hospitals in Canada, the median ED stay is 9.5 hours (National Ambulatory Care Reporting System, 2008-2009 data), corresponding to a net benefit of early prevention of $19.31. Because of the low per-patient cost of prevention, prevention was cost effective even for a very short ED stay. For elderly admitted patients with an average ED wait time of 1 hour, the net benefit of early prevention was $4 per patient (Figure 11).

The prevalence of hospital-acquired pressure ulcers is an accessible indicator of pressure ulcer risk among inpatients. We varied the prevalence of hospital-acquired pressure ulcers among the study patients from 20% down to 1%, corresponding to an approximate incidence of ED-acquired pressure ulcers of 0.1%. At this low risk of pressure ulcers, prevention was still cost effective, with a net benefit of approximately $1 per patient (Figure 11).

The additional care for hospital-acquired pressure ulcers is mostly provided by hospitals. Early prevention was cost saving from the hospital perspective. The average cost of upgrading from a
standard mattress to a pressure-redistribution mattress for an ED stretcher was $480 (max: $700). The corresponding upgrading cost for an ED bed was $200 (max: $3,775). The average maximum upgrading cost of $2,238 represents the high price of pressure-redistribution mattresses, the high cost of replacing standard mattresses, or both. At the maximum upgrading cost, the net benefit of prevention was $19 per patient (Figure 11). Also according to Figure 11, the net benefit of prevention changed slightly with the time horizon of the analysis (30 days to 5 years), and within plausible ranges for estimated transition rates in the natural history of pressure ulcers.

7.4.4 Probabilistic sensitivity analysis
Taking into account the joint uncertainty in key model inputs, the probability that early prevention was cost effective was 81%, suggesting that early prevention is likely to be cost effective despite uncertainty in the clinical evidence. This probability did not change when the monetary value of a quality-adjusted life-year was varied up to $100,000 (Table 15).

7.4.5 Value-of-information analysis
The total expected value of perfect information associated with the overall uncertainty in the cost-effectiveness analysis was $9.11 per patient, ranging from $9.07 to $9.14 when the monetary value of a quality-adjusted life-year was varied up to $100,000 (Table 15). The total expected value of perfect information was approximately $2.2 million for the target population of elderly admitted ED patients in Ontario. This means that if implementation of early prevention is delayed to obtain further information, the maximal return that could be provided by doing additional research is $2.2 million per year. On the other hand, implementation of early prevention without further research would result in a saving of $7.2 million per year.

While the current economic evidence seems to be sufficient in support of early prevention, the clinical evidence remains incomplete without resolving the uncertainty in the relative risk estimate of pressure-redistribution mattresses in the ED. For the target population in Ontario, the partial expected value of perfect information associated with uncertainty in the relative risk estimate was approximately $1.4 million per year. It would be cost effective to carry out a larger trial to resolve this uncertainty if the expected value of perfect information exceeded the expected costs of conducting the trial. The expected value of perfect information should be expressed for the target population of current and future patients who will benefit from early
assuming that high-density foam remains the core technology in pressure-redistribution mattresses for the ED in the next 10 years, the expected value of perfect information for current and future patients will be approximately $1.4 million \times 8.6$ (discounted), where 8.6 is the annuity factor for a period of 10 years with an interest rate of 3%. For even larger (national and international) populations, the expected value of perfect information would be much larger, supporting the cost-effectiveness of doing additional randomized controlled trials to evaluate pressure-redistribution mattresses designed for the ED.

7.5 Limitations

Our analysis only captured the health benefits of early prevention among elderly admitted ED patients. We were not able to capture all potentially relevant health benefits. High-risk elderly patients who are not admitted to hospitals, or high-risk younger patients who are admitted, might also benefit. Our base case analysis is therefore conservative.

7.6 Discussion

We synthesized the available clinical and economic evidence for early prevention of pressure ulcers in elderly admitted ED patients. The current clinical evidence is mixed. In the ED, pressure-redistribution mattresses may reduce the incidence of ED-acquired pressure ulcers but the prevention effect derived from a well designed, but small trial is not statistically significant. The magnitude of the prevention effect is, however, consistent with a significant prevention effect observed in hospital patients. The economic evidence, on the other hand, is stronger. Early prevention with pressure-redistribution mattresses is likely to be more effective and less costly than current practice with standard hospital mattresses. The cost saving associated with early prevention is likely for both hospitals and health systems.

Existing practice guidelines call for the use of pressure-redistribution mattresses for patients at risk of developing pressure ulcers. To date, little attention has been given to patient support surfaces in the ED. Tarpey et al. 2000 evaluated support surfaces for patients in transit through an ED; they reported deficiencies in all of them, due to both design and deterioration. Our survey of EDs in Ontario hospitals showed that approximately 12% of ED stretchers and beds are currently equipped with pressure-redistribution mattresses (n=30 randomly selected
departments, 100% response rate). We suspect similarly low utilization rates of pressure-redistribution support surfaces in the ED in other acute care hospitals in developed countries.

The results of our cost-effectiveness analysis were similar to those from cost-effectiveness analyses accompanying RCTs evaluating support surfaces for pressure ulcer prevention in hospital patients. All studies reported qualitatively similar results; the cost saving associated with the reduction or delay of pressure ulcer development due to prevention was larger than the cost of upgrading to pressure-redistribution surfaces.

Given the clinical and economic evidence that is now available, the policy options are widespread adoption of pressure-redistribution mattresses in EDs, more research, or both. The decision to either act now or delay until further evidence is acquired depends largely on the question of effectiveness. If we think that the prevention effect of pressure-redistribution mattresses can be extended to the ED setting from the ward bed setting, both the clinical and economic evidence strongly favor adopting this technology now. If, on the other hand, only evidence gathered in the ED setting is deemed relevant, further study prior to an adoption decision might be justified.

Our view is that because the patients, risk factors and physiology of pressure ulcer development share many similarities across settings, it is legitimate to pool effectiveness evidence gathered from hospitalized inpatients and elderly admitted ED patients. The Ontario Health Technology Advisory Committee has agreed with this view. Partially based upon data reported here, the Ontario Health Technology Advisory Committee has recommended using pressure-redistribution mattresses for all persons accessing emergency room care.

We, however, recognize dissenting views. According to the systematic review by McInnes et al. 2008, the Cochrane reviewers interpreted the same effectiveness evidence we summarized and concluded that patient support surfaces designed for use in the ED have not been adequately evaluated. Our value-of-information analysis suggests that given the prospectively large number of elderly admitted ED patients that could benefit from early prevention, it is cost effective to conduct additional trials of pressure-redistribution mattresses designed particularly for the ED setting.
In summary, the economic evidence supports early prevention with pressure-redistribution foam mattresses in the ED. Early prevention is likely to improve health for elderly patients and save hospital costs. The value-of-information evidence supports additional studies to evaluate the risk and incidence of ED-acquired pressure ulcers and preventive support surfaces designed for the ED.
Figure 11. Sensitivity analysis of the cost-effectiveness of early prevention

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Range</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR of PRM vs SM, ED patients, 1 RCT (0.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RR of PRM vs SM, ED &amp; hospital patients, 6 RCTs (0.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED stay (15.4 hrs)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prevalence of hospital-acquired pressure ulcers (19.6%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost component (hospital and home care costs)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upgrading cost to PRM for ED stretchers/beds ($340)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily incidence of pressure ulcer (2.95%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily healing rate - stage 1 (2.34%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily progression rate - stage 1 -&gt; 2 (20.80%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time horizon (1 year)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily healing rate - stage 2 (0.97%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily progression - stage 2 -&gt; 3 (1.11%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily progression rate - stage 3 -&gt; 4 (0.30%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily incidence of systemic infection (2.22%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily incidence of local infection (0.14%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily healing rate - stage 3 (0.37%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net monetary benefit of early prevention per patient</td>
<td></td>
<td>PRMs are less cost-effective</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PRMs are more cost-effective</td>
</tr>
<tr>
<td></td>
<td>-$100</td>
<td>$100</td>
</tr>
<tr>
<td></td>
<td>-$80</td>
<td>$80</td>
</tr>
<tr>
<td></td>
<td>-$60</td>
<td>$60</td>
</tr>
<tr>
<td></td>
<td>-$40</td>
<td>$40</td>
</tr>
<tr>
<td></td>
<td>-$20</td>
<td>$20</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>$0</td>
</tr>
<tr>
<td></td>
<td>$20</td>
<td>$20</td>
</tr>
<tr>
<td></td>
<td>$40</td>
<td>$40</td>
</tr>
<tr>
<td></td>
<td>$60</td>
<td>$60</td>
</tr>
<tr>
<td></td>
<td>$80</td>
<td>$80</td>
</tr>
<tr>
<td></td>
<td>$100</td>
<td>$100</td>
</tr>
<tr>
<td></td>
<td>$120</td>
<td>$120</td>
</tr>
<tr>
<td></td>
<td>$140</td>
<td>$140</td>
</tr>
</tbody>
</table>

Abbreviations: RR: relative risk estimate. ED: emergency departments. PRMs: pressure-redistribution foam mattresses. SM: standard mattresses. RCT: randomized controlled trial. Notes: The horizontal bars represent variation in the net benefit between pressure-redistribution foam mattresses and standard mattresses. Base case values are displayed in parentheses. Input values are displayed to the left and right of the horizontal bars. The solid line indicates no difference in the net benefit of prevention (zero net benefit).
Table 14. Input data to the cost-effectiveness analysis of early prevention

<table>
<thead>
<tr>
<th>Prevention Effect Estimates</th>
<th>Base case value</th>
<th>Range or SD</th>
<th>Distribution*</th>
<th>Reference #, Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative risk of developing pressure ulcers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRM versus SM, ED patients, 1 RCT</td>
<td>0.78</td>
<td>0.42, 1.46CI</td>
<td>Log-normal</td>
<td>185</td>
</tr>
<tr>
<td>PRM versus SM, ED and hospital patients, 6 RCTs</td>
<td>0.45</td>
<td>0.27, 0.74CI</td>
<td>Log-normal</td>
<td>29*</td>
</tr>
</tbody>
</table>

Natural History Model of Pressure Ulcers

Daily transition rates during hospitalization

<table>
<thead>
<tr>
<th>Event</th>
<th>Base case value</th>
<th>Range or SD</th>
<th>Distribution</th>
<th>Reference #, Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily incidence of stage-1 pressure ulcer (%)</td>
<td>2.95</td>
<td>0.2, 15.37</td>
<td>Joint distribution</td>
<td>Hospital Survey†</td>
</tr>
<tr>
<td>Stage 1 → 2</td>
<td>20.80</td>
<td>18.66, 21.62</td>
<td>Joint distribution</td>
<td>†</td>
</tr>
<tr>
<td>Stage 2 → 3</td>
<td>1.11</td>
<td>0.91, 1.22</td>
<td>Joint distribution</td>
<td>†</td>
</tr>
<tr>
<td>Stage 3 → 4</td>
<td>0.30</td>
<td>0.06, 0.55</td>
<td>Joint distribution</td>
<td>†</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event</th>
<th>Base case value</th>
<th>Range or SD</th>
<th>Distribution</th>
<th>Reference #, Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily healing rate during hospitalization (%)</td>
<td></td>
<td></td>
<td>Joint distribution</td>
<td></td>
</tr>
<tr>
<td>Healing stage 1</td>
<td>2.34</td>
<td>0.50, 4.97</td>
<td>Joint distribution</td>
<td>†</td>
</tr>
<tr>
<td>Healing stage 2</td>
<td>0.97</td>
<td>0.12, 1.78</td>
<td>Joint distribution</td>
<td>†</td>
</tr>
<tr>
<td>Healing stage 3</td>
<td>0.37</td>
<td>0.05, 0.76</td>
<td>Joint distribution</td>
<td>†</td>
</tr>
<tr>
<td>Healing stage 4</td>
<td>0.04</td>
<td>0.01, 0.07</td>
<td>Joint distribution</td>
<td>†</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event</th>
<th>Base case value</th>
<th>Range or SD</th>
<th>Distribution</th>
<th>Reference #, Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-discharge mean healing time (week)</td>
<td></td>
<td></td>
<td>Exponential</td>
<td></td>
</tr>
<tr>
<td>Stage 1, 2, 3, 4 (respectively)</td>
<td>4, 13, 18, 22</td>
<td>Not available</td>
<td>Hospital Survey†</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event</th>
<th>Base case value</th>
<th>Range or SD</th>
<th>Distribution</th>
<th>Reference #, Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed prevalence of hospital-acquired pressure ulcers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pressure ulcers – All stages (%)</td>
<td>19.6</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>9.0</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 2</td>
<td>9.9</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costs</td>
<td>Stage 3</td>
<td>Stage 4</td>
<td>Reference #, Data Source</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------</td>
<td>---------</td>
<td>--------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Costs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mattress on ED stretchers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5” pressure-redistribution mattress for ED stretchers</td>
<td>0.6</td>
<td>Not reported</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3” standard foam mattresses for ED stretchers</td>
<td>0.3</td>
<td>Not reported</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional cost of obtaining PRMs for ED stretchers</td>
<td>0.3</td>
<td>Not reported</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifespan of all support surfaces (year)</td>
<td>0.3</td>
<td>Not reported</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mattress on ED beds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8” pressure-redistribution mattress for ED beds</td>
<td>1:1</td>
<td>Not varied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5” standard foam mattresses for ED beds</td>
<td>1:1</td>
<td>Not varied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional cost of obtaining PRMs for ED beds</td>
<td>1:1</td>
<td>Not varied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifespan of all support surfaces (year)</td>
<td>1:1</td>
<td>Not varied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average time in ED for admitted elderly patients (hour)</td>
<td>15.4</td>
<td>13.7</td>
<td>Gamma</td>
<td></td>
</tr>
<tr>
<td>Average ratio of ED stretchers : ED beds</td>
<td>1:1</td>
<td>Not varied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average additional cost of PRM for ED stretchers/beds</td>
<td>340</td>
<td>38, 2238</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per-patient cost for PRM on ED stretcher and bed</td>
<td>0.30</td>
<td>0.03, 1.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>In-patient costs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean cost attributable to pressure ulcer care</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 2</td>
<td>11,967</td>
<td>3,702</td>
<td>Gamma</td>
<td></td>
</tr>
<tr>
<td>Stage 3</td>
<td>12,951</td>
<td>7,849</td>
<td>Gamma</td>
<td></td>
</tr>
<tr>
<td>Stage 4</td>
<td>21,797</td>
<td>12,031</td>
<td>Gamma</td>
<td></td>
</tr>
<tr>
<td>Mean hospitalization cost</td>
<td>6,806</td>
<td>10,745</td>
<td>Gamma</td>
<td></td>
</tr>
<tr>
<td>Mean in-patient physician billings</td>
<td>445</td>
<td>728</td>
<td>Gamma</td>
<td></td>
</tr>
<tr>
<td>Post-discharge home care costs</td>
<td>Base case value</td>
<td>Range or SD</td>
<td>Distribution*</td>
<td>Reference #, Data Source</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------</td>
<td>-------------</td>
<td>---------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Weekly cost attributable to pressure ulcer care</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 2</td>
<td>$57</td>
<td>$113</td>
<td>Gamma</td>
<td>MDS - HC†</td>
</tr>
<tr>
<td>Stage 3</td>
<td>$81</td>
<td>$116</td>
<td>Gamma</td>
<td></td>
</tr>
<tr>
<td>Stage 4</td>
<td>$105</td>
<td>$119</td>
<td>Gamma</td>
<td></td>
</tr>
<tr>
<td>Mean weekly cost</td>
<td>$134</td>
<td>$111</td>
<td>Gamma</td>
<td></td>
</tr>
</tbody>
</table>

**Health-Related Quality-of-Life Weights**

| Health utility decrement, stage 2-4 vs. stage 0-1 (%) | | | Beta | 208 |
| Absolute health utility decrement | 6.10 | 1.10 | | |
| In-patient health utility | 0.0268 | 0.0048 | | |
| Full recovery after 1 year | 0.44 | 0.32 | Gamma | 209 |

**Hospital Discharge Data**

| Estimated length of stay (day) | 6.5 | 9.7 | Gamma | CIHI-DAD* |
| Estimated hospital mortality (%) | 7.2 | 0.02 | Beta | CIHI-DAD‡ |
| Post-discharge mortality | age-specific | | | Statistics Canada** |

**Pressure Ulcer Related Complications**

| Daily incidence of local infection (stage 2-4) (%) | 0.14 | 0.07 | Beta | 95 |
| Daily incidence of sepsis (locally infected stage 3-4) (%) | 2.22 | 0.64 | Beta | 95 |
| Mortality due to sepsis (%) | 12.8 | 0.7 | Beta | 211 |

Abbreviations: SD: standard deviation. PRM: pressure-redistribution foam mattress. SM: standard hospital mattress. ED: emergency department. RCT: randomized controlled trial. CIHI-DAD: Canadian Institute of Health Information - Discharge Abstract Database. OCCI: Ontario Case Costing Initiative. MDS – HC: Minimum Data Set – Home Care. Notes: *Distributions were used in the probabilistic sensitivity analysis. CI: Confidence interval. †A random-effects model was used to derive a relative risk estimate from the six RCTs. A total of 4,055 correlated
transition rate estimates were identified from the calibration; they reproduced the observed stage-specific prevalence of pressure ulcers from the Tri-Hospital Survey (Appendix 3). \(^1\) Details pertaining to the Tri-Hospital Survey are in section 2.5.1. \(^2\) Survey of three manufacturers/distributers, including Stryker, Hill-Rom, and Waterloo/Gaymar. \(^3\) Ontario Case Costing Initiative data. \(^4\) Costing data were derived from a published study involving individuals receiving home care from the Minimum Data Set – Home Care. \(^5\) Canadian Institute of Health Information – Discharge Abstract Database 2008. \(^6\) Statistics Canada 2008.
Table 15. Results of the cost-effectiveness analysis of early prevention

<table>
<thead>
<tr>
<th>Prevention effect</th>
<th>Pressure-Redistribution Mattress Estimate (Range)*</th>
<th>Standard Mattresses Estimate (Range)*</th>
<th>Difference Estimate (Range)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incidence of ED-acquired pressure ulcers (%)</td>
<td>1.48 (0.73, 3.38)</td>
<td>1.90 (1.74, 2.31)</td>
<td>-0.42 (-1.52, 1.10)</td>
</tr>
<tr>
<td>Cost-effectiveness</td>
<td>Expected value</td>
<td>Expected value</td>
<td>Incremental (95% CI)</td>
</tr>
<tr>
<td>Health care cost†</td>
<td>$7,031</td>
<td>$7,063</td>
<td>-$32 (-$1,099; $23)</td>
</tr>
<tr>
<td>Hospital cost</td>
<td>$6,877</td>
<td>$6,908</td>
<td>-$30 (-$1,086; $21)</td>
</tr>
<tr>
<td>Quality-Adjusted Life Days</td>
<td>153.3648</td>
<td>153.3634</td>
<td>0.0015 (-0.0007, 0.0037)</td>
</tr>
</tbody>
</table>

Probability that pressure-redistribution foam mattresses are cost-effective‡ 81%

Value-of-information

<table>
<thead>
<tr>
<th>Total expected value of perfect information (all uncertainties in model inputs)</th>
<th>Expected value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per patient§</td>
<td>$9.11</td>
</tr>
<tr>
<td>Per population per year*</td>
<td>$2,186,409</td>
</tr>
</tbody>
</table>

Partial expected value of perfect information (uncertainty in the preventive effect estimate)

| Per patient||                                                                 |
|--------------------------------------------------|----------------|
| $5.91                                             |

Per population per year* $1,418,400

Abbreviations: ED: emergency department. CI: confidence interval from the probabilistic sensitivity analysis.
Notes: Costs are reported in 2009 Canadian dollar (adjusting for purchasing power parities, $1 Canadian=$0.89 U.S.).* Estimates and ranges were derived from the good-fit transition rate estimates in the calibration (chapter 5).†Healthcare costs included hospital costs and home care costs post-discharge.‡This probability was relatively consistent when the monetary value of a quality-adjusted life year was varied up to $100,000.§Values ranged from $9.07 to $9.14 when the monetary value was varied.¶Values ranged from $5.86 to $5.95 when the monetary value was varied. The targeted population is approximately 240,000 elderly admitted ED patients to Ontario hospitals per year.
Chapter 8
Prevention of Pressure Ulcers in Surgical Patients: A Cost-Effectiveness Analysis

8 Intra-operative prevention

8.1 Abstract

**Background:** Patients who undergo prolonged surgical procedures are at risk of developing pressure ulcers. Recent systematic reviews suggest that pressure-redistribution overlays on operating tables significantly decrease the associated risk. Little is known about the cost effectiveness of using these overlays in a prevention program for surgical patients.

**Methods:** Using a Markov cohort model, we evaluated the cost effectiveness of an intra-operative prevention strategy with operating table overlays made of dry, viscoelastic polymer from the perspective of a health care payer over a 1-year period. We simulated patients undergoing scheduled surgical procedures lasting $\geq 90$ min in the supine or lithotomy position.

**Results:** Compared with the current practice of using standard mattresses on operating tables, the intra-operative prevention strategy decreased the estimated intra-operative incidence of pressure ulcers by 0.51%, corresponding to a number-needed-to-treat of 196 patients. The average cost of using the operating table overlay was $1.66 per patient. Compared with current practice, this intra-operative prevention strategy would increase slightly the quality-adjusted life days of patients and by decreasing the incidence of pressure ulcers, this strategy would decrease both hospital and home care costs for treating fewer pressure ulcers originated intra-operatively. The cost savings was $46 per patient, which ranged from $13 to $116 by different surgical populations. Intra-operative prevention was 99% likely to be more cost effective than the current practice.

**Conclusion:** In patients who undergo scheduled surgical procedures lasting $\geq 90$ min, this intra-operative prevention strategy could improve patients’ health and save hospital costs. The clinical and economic evidence support the implementation of this prevention strategy in settings where it has yet to become current practice.
8.2 Background

Pressure ulcers develop when persisting pressure on bony prominences obstructs healthy capillary flow, leading to tissue necrosis. They could affect the elderly, immobile and neurological compromised individuals. An estimated 2.5 million pressure ulcers are treated each year in US acute care hospitals. Pressure ulcers can induce pain, impair functional status, and decrease quality-of-life. They tend to increase intensive care duration, hospital length of stay, and treatment costs. The Centers for Medicare and Medicaid Services have designated pressure ulcers as preventable complications of medical care (“never events”), and no longer reimburse hospitals for the cost of treating hospital-acquired pressure ulcers. The aim of these policy changes is patient safety.

Patients undergoing prolonged surgical procedures are immobilized for long periods and are therefore at risk of developing pressure ulcers. The sub-dermal tissue injury under bony prominences likely occurs between the first hour and 4 to 6 hours after sustained loading, according to a cross-synthesis of evidence from clinical, animal and in-vitro studies. Compared to the healthy capillary pressure between 20 to 40 mm Hg, tissue interface pressure (i.e., the force per unit area that acts perpendicularly between the body and the patient support surface) at bony prominences (e.g., scapular, sacral areas and heels) is extremely high (≥120 mm Hg) in patients undergoing prolonged surgical procedures.

Pressure-redistribution overlays for operating tables reduce the tissue interface pressure by either molding around the shape of the patient to distribute the patient’s weight over a larger area (static overlays), or mechanically varying the pressure beneath the patient to reduce the duration of the applied pressure (e.g., alternating pressure overlays). Some pressure-redistribution overlays significantly reduce the post-operative incidence of pressure ulcers in surgical patients, according to recent systematic reviews. The price of these overlays however varies widely, from approximately $800 for static overlays to over $2,000 for alternating pressure overlays. Little is known about the trade-off between the additional costs and clinical benefits of their use in an early prevention program for surgical patients. Commissioned by the Ontario Health Technology Assessment Committee, the current study evaluated the cost-effectiveness of pressure-redistribution overlays for intra-operative prevention of pressure ulcers in surgical patients.
8.3 Methods

8.3.1 Study Design
We conducted the cost-effectiveness analysis using the pressure ulcer history model described in chapter 6. We followed guidelines for economic evaluation by the Canadian Agency for Drugs and Technologies in Health.\textsuperscript{212} We simulated a cohort of patients with an average age of 63 years who underwent scheduled surgical procedures lasting \( \geq 90 \) minutes in the supine or lithotomy position (Table 16).\textsuperscript{224}

The base case analysis was conducted from the health care payer perspective; sensitivity analysis was also conducted from the hospital perspective.\textsuperscript{212} We included hospital costs and post-discharge home care costs for hospital-acquired pressure ulcers. Health outcomes included quality-adjusted life-years and the incidence of pressure ulcers occurring intra-operatively and immediate post-operation. We defined intra-operative risk as the risk associated with the duration of restricted mobility due to surgery and immediately post-operative risk as the cumulative risk in the first two days following surgery.\textsuperscript{201}

A time horizon of one year was used for the analysis, including the hospital length of stay and the period of post-discharge convalescence. Sensitivity analysis was conducted with extended time horizon to evaluate the effect of early prevention on pressure-ulcer related mortality. Due to the short time horizon, health outcomes and costs were not discounted. All costs are expressed in Canadian dollars in 2009.\textsuperscript{63}

8.3.2 Intra-operative prevention strategies
We compared current practice using standard mattresses and padding on operating tables with the intra-operative prevention strategy using dry visco-elastic polymer overlays on operating tables (Action Product Inc.).\textsuperscript{224} Post-operation, we assumed that patients were nursed according to current practice, especially with respect to pressure ulcer prevention.\textsuperscript{19,50}

8.3.3 Input data
We obtained input estimates from current systematic reviews of pressure ulcer prevention,\textsuperscript{29,54} the Tri-Hospital Pressure Ulcer Prevalence Survey, the Ontario Case Costing Initiative,\textsuperscript{218} a large data set for home care in Ontario,\textsuperscript{219} and additional literature searches. The pressure ulcer history
model and related model inputs are described in chapter 6. Unknown transitions across health states related to the severity of pressure ulcers were determined via calibration (chapter 5). Table 16 displays the characteristics and results of randomized controlled trials evaluating pressure-redistribution overlays for operating tables. Table 17 displays model inputs to the cost-effectiveness analysis, including transition rates for the pressure ulcer history model, costs, and quality-of-life weights. We described below specific inputs related to the cost-effectiveness analysis.

8.3.3.1 Clinical effectiveness of pressure-redistribution overlays

From recent systematic reviews, we identified five randomized controlled trials that evaluated three pressure-redistribution overlays for operating tables, including the dry polymer overlay, two foam overlays, and an alternating pressure overlay. The trial results are summarized in Table 16.

The trial by Nixon et al. 1998 evaluates dry polymer overlays for patients who underwent surgical procedures lasting ≥90 minutes in the supine or lithotomy position. In terms of the quality of the evidence, this trial has adequate reporting for randomization and allocation concealment, double-blinding, a-priori sample size determination, and intention-to-treat analysis. According to the trial results, the use of the dry polymer overlays significantly reduced the incidence of pressure ulcers immediately post-operation by 47% (Table 16).

Two different types of foam overlay did not contribute to any reduction in the incidence of post-operative pressure ulcers among surgical patients (Table 16). Alternating pressure overlays (used intra-operatively and post-operatively) were compared with a combination of patient support surfaces on operating tables and ward beds. The use of alternating pressure overlays significantly reduced the risk of pressure ulcers. However, it was not clear whether this prevention effect was due to intra-operative or post-operative pressure reduction, or both. Because of this uncertainty, we considered only the dry polymer overlay in the current cost-effectiveness analysis.

8.3.3.2 Intra-operative incidence of pressure ulcers

The steps we used to estimate intra-operative incidence of pressure ulcers are described as part of the development of the pressure ulcer history model in chapter 6. We estimated the intra-
operative prevention effect with the dry polymer overlays by applying the associated relative risk estimate to the intra-operative incidence.

8.3.3.3 Cost of operating table overlays
The average unit price and lifespan of the dry polymer overlays were obtained from the manufacturer (Table 17). The per-patient cost of the dry polymer overlay was derived by amortizing its unit price over its average life span, assuming that on average, approximately 250 major surgical procedures are performed on an operating table per year.251

8.3.4 Analytical approaches
Analytical approaches for the cost-effectiveness analysis were described as part of the development of the pressure ulcer history model in chapter 6. The joint distribution of model transitions derived via calibration was used in the probabilistic sensitivity analysis to estimate the probability that prevention is more cost effective than current practice (chapter 5).

8.4 Results
8.4.1 Clinical effectiveness
Table 18 displays results of the base case analysis. Over an average surgical duration of 4.6 hours, the projected intra-operative incidence of pressure ulcers was 1.07% for patients with current practice and 0.57% for patients with the intra-operative prevention strategy, corresponding to an absolute reduction of 0.51%. On average, one needs to institute prevention for 196 surgical patients to prevent one pressure ulcer originated intra-operatively.

The projected incidence of stage 2-4 pressure ulcers in the first two days following surgery was 1.06% for patients with current practice and 0.80% for patients with the intra-operative prevention strategy. Note that the intra-operative incidence accounts for stage 1 pressure ulcers that may resolve spontaneously without interventions. The immediate post-operation incidence includes only stage 2-4 pressure ulcers that account for the increase in hospital costs and decrease in quality-of-life of surgical patients.
8.4.2 Cost-effectiveness analysis

8.4.2.1 Base case analysis

For the intra-operative prevention strategy, the average cost of using the pressure-redistribution overlay was $1.66 per patient (Table 17). Compared to current practice, the prevention strategy was more effective as it was associated with higher mean quality-adjusted life days (Table 18). The prevention strategy was less costly as it reduced the need to treat hospital-acquired pressure ulcers that occurred immediately post-operation, with a mean cost saving of $45.66 per patient. The net benefit of prevention was $45.95 per patient, consisting of $0.29 from the health benefit of attaining higher quality-adjusted life days and $45.66 of cost saving from the prevention of costly pressure ulcers.

Annually, approximately 152,000 scheduled major surgical cases are conducted in Ontario, a province with approximately 13 million people. By targeting these surgical cases for intra-operative prevention, the prevention strategy could prevent 760 incident pressure ulcers and save approximately $7 million in health care cost per year.

8.4.2.2 Sensitivity analysis

Figure 12 displays results of the one-way sensitivity analysis, illustrating how the net benefit of the intra-operative prevention strategy varies with variation in model inputs. For patients on standard operating table mattresses, the projected intra-operative incidence ranged from 0.32% to 2.79% over a mean surgical duration of 4.6 hours; the corresponding net benefit of the prevention strategy ranged from $13.35 to $116.19 per patient. We estimated that the plausible range for intra-operative incidence ranged from 0.33% to 2.55% in clinical studies we identified in our literature search.223 224 24 201 225 As such, the intra-operative prevention strategy would be cost effective for a number of surgical populations.

The net benefit of the intra-operative prevention strategy was $14.11 for patients with a surgical duration of 90 minutes and increased with longer surgical duration, suggesting that prevention was cost-effective for surgical duration ≥90 minutes (Figure 12). At the surgical duration of 90 minutes, the projected intra-operative incidence ranged from 0.1% to 0.9%; the corresponding net benefit of prevention ranged from $3.46 to $37.10. The intra-operative prevention strategy remained cost-effective for patients at very low intra-operative risk.
The costs of care for hospital-acquired pressure ulcers are mostly provided by hospitals. Intra-operative prevention was cost effective from the hospital perspective (Figure 12). Intra-operative prevention was cost effective across the 95% confidence interval of the relative risk estimate of the dry polymer overlays, or when the unit price of the overlays was varied up to approximately $2000. The cost-effectiveness results were not sensitive to longer time horizon.

Taking into account the joint uncertainty in key model inputs, intra-operative prevention was 99% likely to be cost-effective relative to current practice when the monetary value of one quality-adjusted life year was varied from $50,000 to $100,000, according to results of the probabilistic sensitivity analysis (Table 18).

8.5 Discussion

According to recent systematic reviews, using the dry polymer overlays on operating tables significantly reduces the immediate post-operative incidence of pressure ulcers in patients undergoing surgical procedures lasting ≥ 90 minutes. We projected the health benefits and costs associated with the observed incidence reduction over a one-year time horizon. According to our results, the intra-operative prevention strategy with the dry polymer overlays improves patients’ health and save hospital costs. Both the clinical and economic evidence strongly supports using the dry polymer overlays for intra-operative prevention of pressure ulcers in surgical patients.

Most practice guidelines recommend the use of pressure-redistribution support surfaces for high risk patients undergoing surgical interventions. In a study of elderly patients who had surgery for hip fracture, the use of pressure-redistribution support surfaces however was low and was based more on facility-related factors than on patient risk. In our survey of surgery departments in Ontario hospitals, from 8% to 20% of operating tables are currently equipped with pressure-redistribution overlays. We suspect there is a large gap between current prevention practice and practice recommendations and a potential for large improvement in the prevention of pressure ulcers among surgical patients.
Our results were similar to those from cost-effectiveness analyses accompanying randomized controlled trials evaluating support surfaces for pressure ulcer prevention in hospital patients. All studies reported qualitatively similar results; the increase in per-patient costs between experimental and control surfaces was minimal compared to the cost saving associated with the reduction or delay of pressure ulcer development.

Among studies examining the risk of pressure ulcers in surgical patients, many have reported outcomes of skin assessment before, on the day of surgery, and post-operation. Other studies have not included findings from perioperative skin assessments, thus limiting our understanding of the surgical pressure ulcer risk. Also, it has been reported that sores attributable to pressure experienced during surgery are observed 2-10 days post-operation, but the average time lag between tissue injury and appearance of characteristic changes at the skin surface is uncertain. According to our analysis, these methodological issues have major implications in the cost-effectiveness evaluation of prevention strategies. Future studies should try to distinguish between intra-operative and post-operative risk of pressure ulcers in surgical patients.

8.6 Limitations

Our analysis has several limitations. All the trials evaluating pressure-redistribution overlays for operating tables included in Table 16 report incidence of pressure ulcers. Due to the uncertain validity of tissue interface pressure as a surrogate outcome for pressure ulcer, we excluded studies using tissue interface pressure as the primary outcome measure. The exclusion of this body of evidence has been suggested as a limitation of our study. However, it could be argued that including only studies measuring incidence of pressure ulcers would improve the validity of our findings.

With the use of advanced laparoscopic surgery, the dry polymer overlays we evaluated here may not be useful for certain surgical positions such as the steep Trendelenburg position for laparoscopic prostatectomy or the semi-sitting position for laparoscopic esophagectomy. Additional preventive measures will need to be considered for these special surgical positions. However, our results are valid for prolonged surgical procedures in more conventional surgical positions.
8.7 Policy implications
The Centers for Medicare and Medicaid Services no longer reimburse US hospitals for additional care of patients with hospital-acquired pressure ulcers.\textsuperscript{74} Placing the financial responsibility for pressure ulcer prevention on the hospitals should intensify the focus on early strategies to reduce pressure ulcer incidence.\textsuperscript{261} Partially based upon the results reported here, the Ontario Health Technology Advisory Committee recommended that a high quality support surface should be used during surgical procedures lasting $\geq 90$ minutes.\textsuperscript{214}

8.8 Conclusions
In patients undergoing scheduled surgical procedures lasting $\geq 90$ minutes, the intra-operative prevention strategy with the dry polymer overlays could improve patients’ health and save hospital costs. The current analysis provides clinical and economic evidence to support the implementation of this prevention strategy in settings where it has yet to become current practice.
Figure 12. Sensitivity analysis of the cost-effectiveness of intra-operative prevention

Abbreviations: I-O: intra-operative. PrU: pressure ulcer. Notes: Net benefit was derived assuming a monetary value of $50,000 per QALY gained. Positive net benefits indicate that prevention is more cost effective than current practice. Black horizontal bars represent variation in the net benefit when model inputs were varied from values displayed to the left and right of the bars (value used in the base case analysis was in parenthesis at the end of the left label). For example, the top horizontal bar displays changes in the net benefit of prevention from $13 to $116 when the intra-operative incidence was varied from 0.32% to 2.79% (base value 1.1%).
### Table 16. Randomized controlled trials of pressure-redistribution overlays

<table>
<thead>
<tr>
<th>1&lt;sup&gt;st&lt;/sup&gt; Author/ year</th>
<th>Surgical population</th>
<th>Age (year)</th>
<th>Surgical duration</th>
<th>Tested pressure-redistribution overlays / Controlled patient support surfaces on operating tables</th>
<th>Timing of reporting incidence</th>
<th># pats w. ulcers / total # pats (% incidence)</th>
<th>Relative risk estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nixon 1998</td>
<td>Elective major general, gynecological or vascular surgery</td>
<td>55-69 (56%) 70+ (44%)</td>
<td>≥ 1.5 hrs</td>
<td>Dry visco-elastic polymer overlays (i.e. dry polymer overlays) ‡ / Standard operating table mattresses</td>
<td>Day 1</td>
<td>22/205 (11%)†</td>
<td>0.53 (0.33, 0.85)</td>
</tr>
<tr>
<td>Feuchtinger 2006†</td>
<td>Scheduled cardiac surgery with extracorporeal circulation</td>
<td>Range:33-92 Mean: 68</td>
<td>≥ 1.5 hrs</td>
<td>4-cm thermo-active visco-elastic foam overlays for op. tables with heating sources / Standard op. tables with the heating sources</td>
<td>Day 1-2</td>
<td>13/85 (15%)</td>
<td>1.38 (0.64, 2.97)</td>
</tr>
<tr>
<td>Schultz 1999§</td>
<td>Scheduled surgery</td>
<td>Range:25-91 Mean: 66</td>
<td>≥ 2 hrs</td>
<td>Foam overlays for operating tables / Drug polymer pads, foam egg crate mattresses, and foam donuts for heels</td>
<td>Day 1-6</td>
<td>55/206 (27%)</td>
<td>1.63 (1.11, 2.38)</td>
</tr>
<tr>
<td>Aronovitch 1999</td>
<td></td>
<td>Scheduled cardiothoracic, urology or vascular surgery</td>
<td>Mean: 64</td>
<td>≥ 3 hrs</td>
<td>Alternating pressure overlays both during and after surgery § / Dry polymer overlays, † PRMs after surgery</td>
<td></td>
<td>Day 1-7</td>
</tr>
<tr>
<td>Russell 2000¶</td>
<td>Scheduled cardiovascular surgery</td>
<td>Mean: 65</td>
<td>≥ 3 hrs</td>
<td>Alternating pressure overlays on op. tables, critical care and ward beds $ / Dry polymer overlays, † PRMs [4] in critical care, ¶ and SMs in ward beds</td>
<td></td>
<td>Day 1-7</td>
<td>2/98 (2%)</td>
</tr>
</tbody>
</table>

Abbreviations: “op. table”: operating table. CI: confidence interval. PRM: pressure-redistribution mattress. SM: standard mattress. †Day 1 indicates the first day following surgery. ‡This trial using a pressure ulcer staging system from 0 (intact skin) to 5, including stage 1 as redness to the skin – blanching occurs, 2a as redness to the skin – non-blanching occurs, and 2b as superficial damage to epidermis. §Action Product Inc. 224 $Alternative pressure overlay (MicroPulse System). 243 ||Pressure Guard II (Span-America Medical System, Inc.) as described in Aronovitch et al. 1999. 252 ¶“Hill-Rom with 6-inch foam overlay” and #“Hill-Rom Century with 4-inch foam overlay” as described in Russell et al. 2000. 243
### Table 17. Input data to the cost-effectiveness analysis of intra-operative prevention

<table>
<thead>
<tr>
<th>Relative risk estimate</th>
<th>Estimate</th>
<th>Variation</th>
<th>Dist’n</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry visco-elastic polymer overlays versus standard mattresses on operating tables</td>
<td>0.53</td>
<td>0.33, 0.85 CI</td>
<td>LN</td>
<td>Nixon et al. 224</td>
</tr>
</tbody>
</table>

| Characteristics of surgical patients (n=1,118) | | | | Tri-Hospital Survey* |
| Mean age (year) | 63.1 | 16.1 SD | | |
| Female (%) | 44.2 | 1.5 SD | | |
| Mean surgical duration (hour) | 4.6 | 3.9 | | |

| Transition rates for the natural history model | | | | Tri-Hospital Survey* |
| Daily incidence rate of stage-1 pressure ulcers (%) | 5.5 | 0.3, 13.7 R | JD† | |
| Daily progression rate (%) | | | | |
| Stage 1 to 2 | 19.6 | 15.2, 23.1 R | JD† | |
| Stage 2 to 3 | 2.2 | 2.1, 4.9 R | JD† | |
| Stage 3 to 4 | 4.6 | 3.7, 4.8 R | JD† | |
| Daily healing rate (%) | | | | |
| Healing stage 1 | 9.2 | 3.4, 13.8 R | JD† | |
| Healing stage 2 | 2.2 | 1.6, 10.7 R | JD† | |
| Healing stage 3 | 1.1 | 0.4, 5.4 R | JD† | |
| Healing stage 4 | 0.1 | 0.03, 0.8 R | JD† | |

| Observed prevalence among surgical patients | | | | Tri-Hospital Survey* |
| Hospital-acquired pressure ulcers – All stages (%) | 17.1 | 1.1 SD | | |
| Stage 1 | 6.4 | 0.7 SD | | |
| Stage 2 | 9.4 | 0.9 SD | | |
| Stage 3 | 0.8 | 0.3 SD | | |
| Stage 4 | 0.5 | 0.2 SD | | |

Post-discharge mean healing time
<table>
<thead>
<tr>
<th>Costs</th>
<th>Estimate</th>
<th>Variation</th>
<th>Dist’n</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry polymer overlays</td>
<td>4, 13, 18, 22</td>
<td>$796, $878 R</td>
<td>G</td>
<td>Bennett et al. 60</td>
</tr>
<tr>
<td>Lifespan (year)</td>
<td>2</td>
<td>fixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated cost per patient</td>
<td>$1.66</td>
<td>$0.88, $1.76 R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of surgical cases per operating table per year</td>
<td>250</td>
<td>250, 500 R</td>
<td></td>
<td>Schultz et al. 251</td>
</tr>
<tr>
<td>Hospital costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean cost attributable to pressure ulcer care</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 2</td>
<td>$11,967</td>
<td>$3,702 SD</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>Stage 3</td>
<td>$12,951</td>
<td>$7,849 SD</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>Stage 4</td>
<td>$21,797</td>
<td>$12,031 SD</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>Mean hospitalization cost</td>
<td>$14,216</td>
<td>$12,245 SD</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>Mean in-patient physician billings</td>
<td>$445</td>
<td>$728 SD</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>Post-discharge home care costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean weekly cost</td>
<td>$105</td>
<td>$119 SD</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>Health-related quality of life weights</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decrement associated with pressure ulcers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative decrement of stage 2-4 versus stage 0-1 (%)</td>
<td>6.10</td>
<td>1.10 SD</td>
<td>B</td>
<td>Thein et al. 262</td>
</tr>
<tr>
<td>Absolute quality-of-life weight decrement</td>
<td>0.0268</td>
<td>0.0048 SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average quality of life weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital patients</td>
<td>0.44</td>
<td>0.32 SD</td>
<td>G</td>
<td>Hays et al. 209</td>
</tr>
<tr>
<td>Patients full recovery at 1 year after hospitalization</td>
<td>0.79</td>
<td>0.12 SD</td>
<td>G</td>
<td>Mittmann et al. 210</td>
</tr>
<tr>
<td>Hospital discharge data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated length of stay (day)</td>
<td>11.1</td>
<td>12.7 SD</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>Estimated hospital mortality (%)</td>
<td>2.2</td>
<td>0.4 SD</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Estimate Category</td>
<td>Estimate</td>
<td>Variation</td>
<td>Distribution</td>
<td>Data Source</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>----------</td>
<td>-----------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Post-discharge mortality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pressure ulcer related complications</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily incidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local infection with stage 2-4 pressure ulcer (%)</td>
<td>0.14</td>
<td>0.07 SD</td>
<td>B</td>
<td>**</td>
</tr>
<tr>
<td>Sepsis with a locally infected stage 3-4 ulcer (%)</td>
<td>2.22</td>
<td>0.64 SD</td>
<td>B</td>
<td>**</td>
</tr>
<tr>
<td>Mortality due to sepsis (%)</td>
<td>17.8</td>
<td>0.8 SD</td>
<td>B</td>
<td>Tourangeau et al.²¹¹</td>
</tr>
</tbody>
</table>

Abbreviations: SD: standard deviation. R: range. CI: confidence interval. Dist.: Distributions used in the decision analytic model to represent uncertainty in model inputs, including log-normal (LN), joint-distribution of transition rates from the calibration procedure (JD), exponential (E), gamma (G), and beta (B).⁸⁷

Notes: Costs are reported in 2009 Canadian dollar ($1 US = $1.19 Canadian, after adjusting for purchasing power parity).⁶³
*Toronto Tri-Hospital Acute Care Pressure Ulcer Prevalence Survey. †The eight daily transition rates were simultaneously determined from the prevalence of hospital-acquired pressure ulcers from the Tri-Hospital Survey (section 2.5.1). ‡Retail unit costs and average lifespan of the dry polymer overlays were obtained from Action Products Inc.® in October 2009. §Data were from the Ontario Case Costing Initiative data from 2002-2007.¶Data from 29,921 episodes of home care services in Ontario.²¹⁹ Data were obtained from the Canadian Institute of Health Information - Discharge Abstract Database from 2002 to 2008. #Data were obtained from Statistics Canada, 2008. **Data were obtained from the Ontario Minimum Data Set – Long Term Care homes in Ontario, including 18,321 residents from 2004 to 2007.⁴⁷
### Table 18. Results of the cost-effectiveness analysis of intra-operative prevention

<table>
<thead>
<tr>
<th>Clinical effectiveness</th>
<th>Overlays Estimate (Range)</th>
<th>Standard mattresses Estimate (Range)</th>
<th>Difference Estimate (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-operative incidence of PrUs (%)</td>
<td>0.57 (0.11, 2.37)</td>
<td>1.07 (0.32, 2.79)</td>
<td>-0.51 (-1.31, 0)</td>
</tr>
<tr>
<td>Immediate post-op. incidence of stage 2-4 PrUs (%)†</td>
<td>0.82 (0.16, 2.43)</td>
<td>1.06 (0.20, 2.63)</td>
<td>-0.26 (-1.10, 0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cost-effectiveness</th>
<th>Mean</th>
<th>Mean</th>
<th>Incremental (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health care cost ($)</td>
<td>14,680</td>
<td>14,726</td>
<td>-46 (-238,-2)</td>
</tr>
<tr>
<td>Hospital cost ($)</td>
<td>14,649</td>
<td>14,693</td>
<td>-44 (-236,-1)</td>
</tr>
<tr>
<td>Life Days</td>
<td>355.9820</td>
<td>355.9820</td>
<td>0</td>
</tr>
<tr>
<td>Quality-Adjusted Life Days</td>
<td>158.3484</td>
<td>158.3463</td>
<td>0.0021 (-0.0010; 0.0134)</td>
</tr>
<tr>
<td>Likelihood that intra-operative prevention is cost-effective‡</td>
<td>99.41%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: “post-op.”: post-operation.

Notes: Costs are reported in 2009 Canadian dollar ($1 US = $1.19 Canadian, after adjusting for purchasing power parity).63

*This is the incidence of pressure ulcers that occur during the surgical duration.
†This is the cumulative incidence of pressure ulcers occurring in the first two days following surgery.
‡The probability that the intra-operative prevention strategy with dry polymer overlays is cost effective relative to standard mattresses on operating tables, when the monetary value for one quality-adjusted life-year was set at $50,000.
Chapter 9
Conclusions and Implications

9 Conclusions

The work of this thesis makes several contributions to the body of evidence in pressure ulcer prevention, and also makes some contributions to the methods for the calibration of decision models for cost-effectiveness analysis.

Study 1 summary

In the study entitled, “Calibration analysis of disease history models: a two-stage random search algorithm,” we propose an algorithm for the calibration of moderately complex decision models with uncertain model parameters. We evaluated the feasibility of the proposed algorithm with the calibration of the decision model for the cost-effectiveness analysis of early prevention of pressure ulcers in elderly admitted ED patients.

We report that there are no differences when the proposed random search algorithm was conducted with simple random sampling or Latin hypercube sampling for the calibration of the pressure ulcer decision model. Our results show that much information that is aimed to be obtained via calibration is gained early in the first-stage search, including early detection of model parameters that are not identifiable via calibration, and approximation to the best-fit estimates of model parameters. It appears that subsequent iterations of the search were used to characterize the uncertainty around the model parameter estimates.

Study 1 implications

We suggest that the proposed algorithm is well suited for routine calibration of moderately complex decision models. However, additional studies are needed to compare the performance of the proposed random search algorithm with other approaches to model calibration, including Bayesian and guided search approaches. In particular, it remains to be verified as to whether the ensemble of good-fit parameter sets identified from the proposed two-stage random search algorithm can be used to characterize the posterior distributions of the model parameters that are to be determined via calibration, and whether early detection of non-identifiable parameters may
allow for the verification of model assumptions and may inform the acquisition of additional data for model development.

**Study 2 summary**

In the study entitled, “Support surfaces for early prevention of pressure ulcers among elderly patients admitted through emergency departments: A cost-effectiveness analysis,” we evaluated the cost-effectiveness of pressure-redistribution foam mattresses on ED stretchers and ED beds for early prevention of pressure ulcers in elderly admitted ED patients.

Due to the uncertainty in the clinical evidence, results of the value-of-information analysis support a randomized controlled trial evaluating the effectiveness and cost-effectiveness of different pressure-redistribution support surfaces for patients in the EDs. Despite this uncertainty, the overall results of the cost-effectiveness analysis show that early prevention is likely to be cost-effective.

**Study 2 implications**

Given the clinical and economic evidence that is now available, the policy options are widespread adoption of early prevention in the EDs, more research, or both. The decision to either act now or delay until further evidence is acquired depends largely on the question of effectiveness. If one thinks that the prevention effect of pressure-redistribution mattresses can be extended to the ED setting from the ward bed setting, both the clinical and economic evidence strongly favor adopting early prevention in this setting now. If, on the other hand, only evidence gathered in the ED setting is deemed relevant, further study prior to an adoption decision might be justified.

Results from this study contribute to policy changes. The Ontario Health Technology Advisory Committee has recommended using pressure-redistribution mattresses for all persons accessing emergency room care. According to the Geriatric Emergency Department Guidelines, the economic evidence supports early prevention in ED patients. An accompanying editorial to the study publication suggests that hospitals that endorse the study findings should consider replacing standard hospital mattresses in the EDs with pressure-redistribution mattresses. It also suggests that creating an environment for emergency care that works for the frail and vulnerable also benefits the greater ED population.
Study 3 summary


We found that compared to current practice with standard mattresses on operating tables, the intra-operative prevention strategy with pressure-redistribution overlays was cost-effective. We conclude that the clinical and economic evidence support the implementation of the intra-operative prevention strategy in settings where it has yet to become current practice.

Study 3 implications

The study findings contribute to policy changes. The Ontario Health Technology Advisory Committee recommended that a high quality support surface should be used during surgical procedures lasting \( \geq 90 \) minutes.\(^{263}\) An international clinical review of acute nursing care for older adults with hip fracture recommends that care organizations should use pressure-redistribution support surfaces for high risk patients on nursing units, in the operating theaters, and in the emergency departments.\(^ {265}\) The Risk Assessment and Prevention of Pressure Ulcers Guidelines by the Registered Nurses Association of Ontario recommend the implementation of intra-operative pressure management devices for surgical procedures lasting \( \geq 90 \) minutes.\(^ {50}\) This study was abstracted for dissemination by the Centre for Reviews and Dissemination of the National Health Services in the United Kingdom.\(^ {266}\)

Overall reflections

With an aging population in Ontario or elsewhere, pressure ulcers will remain an important public health issue for providers and policy decision makers.\(^ {45, 30, 31, 54}\) We expect that the pressure ulcer decision model developed in this thesis will facilitate future evaluation of evidence-based interventions for the prevention and treatment of pressure ulcer.

Decision-analytic models are increasingly used for economic evaluation of health technologies. Some of these models are used to simulate complex disease processes, including those with unobservable pathways. In these instances, model calibration can be a useful tool for the determination of uncertain model parameters related to an unobserved pathway, as illustrated by
the work in this thesis. In particular, the two-stage random search algorithm proposed in this thesis is suitable for routine calibration of moderately complex decision models.
References


63. OECD. Organization for Economic Co-Operation and Development - Prices and Purchasing Power Parities (PPP). 2010. Available at http://www.oecd.org/department/0,3355,en_2649_34347_1_1_1_1_1_00.html. April 2010.


Appendices

Appendix 1. Calibration Analysis of Decision Models - Literature search (chapter 4)

Database(s): Ovid MEDLINE(R) 1946 to November Week 1 2013, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations November 15, 2013, Embase 1980 to 2013 Week 46

**Table A-1: Search Strategy:**

<table>
<thead>
<tr>
<th>#</th>
<th>Searches</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>((model* adj2 (&quot;decision&quot; or &quot;screening&quot; or &quot;cost effectiveness&quot; or &quot;disease&quot; or &quot;natural history&quot; or &quot;simulation&quot; or &quot;pathway&quot;)) and calibration).mp. [mp=ti, ab, ot, nm, hw, kf, ps, rs, ui, an, sh, tn, dm, mf, dv, kw]</td>
<td>983</td>
</tr>
<tr>
<td>2</td>
<td>remove duplicates from 1</td>
<td>702</td>
</tr>
<tr>
<td>3</td>
<td>limit 2 to english language</td>
<td>683</td>
</tr>
</tbody>
</table>

Database: EconLit 1987 to November Week 1 2013
model* NEAR/2 ("decision" OR "screening" OR "cost effectiveness" OR "disease" OR "natural history" OR "simulation" OR "pathway") AND calibration.

Results: 110
Appendix 2. Calibration analysis of decision models (chapter 5)

Figure A. Observed and projected prevalence of pressure ulcers

Abbreviation: PrU: pressure ulcer. Notes: The black columns display the observed prevalence of no pressure ulcers and stage 1-3 pressure ulcers from a study cohort of 745 elderly admitted ED patients. The grey columns display the corresponding projected prevalence from the calibrated model at the best-fit values in the 2-S LHS (8, 9) scenario. Each error bar on top of the columns represents the estimated standard error of the proportion from a binomial distribution.
### Table A-2. Estimated rank correlations of model parameters at the end of the first-stage search

<table>
<thead>
<tr>
<th></th>
<th>$P_{01}$</th>
<th>$P_{12}$</th>
<th>$P_{23}$</th>
<th>$P_{34}$</th>
<th>$H_{10}$</th>
<th>$H_{20}$</th>
<th>$H_{30}$</th>
<th>$H_{40}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{01}$</td>
<td>1.00</td>
<td>-0.10</td>
<td>0.02</td>
<td>-0.04</td>
<td>0.33</td>
<td>0.14</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>$P_{12}$</td>
<td>1.00</td>
<td>0.07</td>
<td>0.02</td>
<td>0.26</td>
<td>0.17</td>
<td>0.02</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>$P_{23}$</td>
<td>1.00</td>
<td>0.01</td>
<td>-0.04</td>
<td>0.10</td>
<td>0.03</td>
<td>-0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_{34}$</td>
<td>1.00</td>
<td>0.45</td>
<td>0.05</td>
<td>0.12</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_{10}$</td>
<td>1.00</td>
<td>0.34</td>
<td>0.31</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_{20}$</td>
<td>1.00</td>
<td>0.34</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_{30}$</td>
<td>1.00</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_{40}$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Illustrated R codes for generating Latin hypercube sampling designs with correlated inputs

```r
n=10  # size of the Latin hypercube samples

library(MASS); # attached library for multivariate normal distributions
cor.m=matrix(c(1.0, 0.8, -0.8, 0.8, 1.0, -0.4, -0.8, -0.4, 1.0), 3, 3) # correlation matrix
norm1=mvrnorm(n,mu=c(0,0,0),Sigma=cor.m) # normal variates with known correlations

library(lhs) # attached library for LHS
lhs1=randomLHS(n,3) # generate LHS on standard uniform distributions

#LHS with restricted pairing - Sort the normal variates by their first column. The rank ordering of the second column determines the rank correlation, so we can use that to order the columns of the LHS samples to have the same rank correlation as the normal variates.
#sort the normal variates in order of the first column
norm1=norm1[order(norm1[,1]),]

#sort the first column of lhs from smallest to largest
lhs1=lhs1[order(lhs1[,1]),]

#Order the second and higher columns of the LHS samples the same way as the second and higher columns of the normal variates
lhs1[,2]=sort(lhs1[,2])[rank(norm1[,2])]
lhs1[,3]=sort(lhs1[,3])[rank(norm1[,3])]

# Transform to other distributions using quantile functions
lhs2[,1]=qunif(lhs1[,1],min=0,max=2)
lhs2[,2]=qbeta(lhs1[,2],shape1=2,shape2=5)
lhs2[,3]=qgamma(lhs1[,3],shape=2,rate=3)

# check the rank correlation
cor(lhs2,method="spearman")
```
Appendix 3. Pressure Ulcer Prevalence Survey (chapter 2)

### Patient Information
- **Date of Admission:**

### Pressure Ulcer Screen

#### Ulcer Identification Table

<table>
<thead>
<tr>
<th>Ulcer #</th>
<th>Location Code</th>
<th>Stage (select one box)</th>
<th>Length (cm)</th>
<th>Width (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Left</td>
<td>DTT 1 2 3 4</td>
<td>unstageable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Left</td>
<td>DTT 1 2 3 4</td>
<td>unstageable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Left</td>
<td>DTT 1 2 3 4</td>
<td>unstageable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Left</td>
<td>DTT 1 2 3 4</td>
<td>unstageable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Left</td>
<td>DTT 1 2 3 4</td>
<td>unstageable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Left</td>
<td>DTT 1 2 3 4</td>
<td>unstageable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Left</td>
<td>DTT 1 2 3 4</td>
<td>unstageable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Left</td>
<td>DTT 1 2 3 4</td>
<td>unstageable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Left</td>
<td>DTT 1 2 3 4</td>
<td>unstageable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Left</td>
<td>DTT 1 2 3 4</td>
<td>unstageable</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Ulcer Code Table
- Occiput (OCC)
- Ear (EAR)
- Shoulder (SHO)
- Scapula (SCA)
- Elbow (ELB)
- Spine (SPI)
- Sacrum/Coccyx (SAC)
- Buttocks (BUT)
- Hip (HIP)
- Thigh (THI)
- Knee (KNE)
- Ankle (ANK)
- Heel (HEE)
- Foot (FOO)
- Other (OTH)

If more than 10 ulcers are present, please complete an additional form and fill in same subject ID.

#### Management of Pressure:
- Current Mattress (shade one only):
  - Pressure Redistribution
  - Non Pressure Redistribution
- Type of Mattress (shade one only):
  - Air
  - Fluid/Gel
  - Foam
  - Other
- Is mattress motorized?: Yes

#### Management of:
- Moisture:
  - Yes
  - No
  - N/I
- Friction/Shear:
  - Yes
  - No
  - N/I
- Nutrition:
  - Yes
  - No
  - N/I

---

*Pressure Ulcer Form: Last Updated: September 2007. Version 0.0.*
### Admitting Diagnosis

- [ ] Cardiovascular
- [ ] Musculoskeletal
- [ ] Skin/Wound
- [ ] Gastrointestinal
- [ ] Neuro
- [ ] Trauma
- [ ] Genito/Urinary
- [ ] Oncology
- [ ] Other
- [ ] Infectious Disease
- [ ] Respiratory

### Location of the Patient over the Last 48 Hours

- [ ] No Change
- [ ] Home
- [ ] Long-Term Care
- [ ] Another Hospital
- [ ] Homeless
- [ ] OR
- [ ] Diagnostic Test
- [ ] Critical Care Unit
- [ ] ER
- [ ] Rehab

### Number of Hours Spent in the ER Prior to Admission

- [ ] N/A

### Has the Patient Been Admitted to a Critical Care Unit During This Admission?

- [ ] Yes
- [ ] No

### Has the Patient Had Surgery During This Admission?

- [ ] Yes
- [ ] No

If yes, how many?

Document the length of time of each surgery:

- [ ] OR 1 hour
- [ ] OR 2 hours
- [ ] OR 3 hours
- [ ] OR 4 hours
- [ ] 5 or more ORs time in hours (add remaining OR times)

### Contributing Factors

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>U/D</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemoglobin in last 4 weeks?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If yes, is most recent Hg less than 100 g/L?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albumin in last 4 weeks?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If yes, is most recent Albumin less than 35 g/L?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPO x3 days within last 7 days?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parenteral/Enteral Feeding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urinary Incontinence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fecal Incontinence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Involuntary Weight Loss</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diarrhea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensory Impairment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diaphoresis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peripheral Arterial Disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anasarca (Total Body Edema)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immobility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute Chronic Pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contracted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is pain interfering with ADLs?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Critical Care Only:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fragile Skin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuromuscular Blockers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immunosuppressive Drugs/ Cancer Therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sedation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemodynamic Instability (Critical Care only)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Skin Assessment

- [ ] Yes
- [ ] No

Were pressure ulcer(s) documented on admission?

- [ ] Yes
- [ ] No

### Pressure Ulcer Risk Assessment Tool Documented within 24 Hours of Admission?

- [ ] Yes
- [ ] No

If yes, admission score

Date of most current pressure ulcer risk assessment score within the last 7 days:

- [ ] DD
- [ ] MMM
- [ ] YY

Current Score:

---

*Pressure Ulcer P and I: 14 September 2007, v. 3*
Copyright Acknowledgements

We would like to thank and acknowledge the publishers and co-authors of all the publications listed below for their kind permission to reproduce material from those original publications.


