A Multi-Criteria Approach to the Evaluation of Food Safety Interventions

A thesis

submitted in partial fulfilment

of the requirements for

the Degree of Doctor of Philosophy

in Management Science

at the University of Canterbury

by

Alexander Hiram Dunn

Department of Management, Marketing and Entrepreneurship

University of Canterbury

2014

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisors, Dr Terri Green, Dr Rob Lake, Dr Don McNickle and Dr Cheryl Brunton for their ongoing support throughout the duration of this research. This thesis would not have been possible without the huge amount of guidance and expertise offered by this supervisory team. Special mention has to go to Dr Terri Green for being a thorough, challenging and extremely helpful Senior Supervisor.

Thank you to the University of Canterbury for providing me with resources I needed to undertake this research, including a scholarship, place to work and a part-time job during my time as a full-time doctoral student. Many thanks also to the staff and fellow postgraduate students of the Department of Management, Marketing and Entrepreneurship for their friendship, advice and help.

Special thanks to Dr Rob Lake and The Institute of Environmental Science and Research Christchurch Science Centre for providing me with access to data, expertise and a place to work following the 2011 Christchurch earthquakes.

Thanks to Dr Beverly Horn from the ESR Christchurch Science Centre for providing her expert opinion during the generation of input data for the MCDA case studies, and for providing her expertise and advice a number of times throughout this research.

Thanks also to Nicola King from the ESR Christchurch Science Centre and to Dr Andrea Menclova from the University of Canterbury for providing their advice and expertise during the early stages of this thesis.

I would like to thank the many individuals from various stakeholder groups who participated in the qualitative research component of this thesis. Many of these individuals went beyond the call of duty when providing their time, expertise and/or data.

Special thanks to my family for their massive support and love always.

Huge thanks to my girlfriend Maddy, who was always there during the highs and lows.

Deputy Vice-Chancellor's Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

'Cost-Effectiveness of Interventions To Control Campylobacter in the New Zealand Meat Food Supply' : Publication is included in Appendix D. Specific cost ($) and effectiveness (DALYs) data is referenced and utilised in Chapter 8 and 9 of this thesis. This cost and effectiveness data has been extracted from this publication.

Please detail the nature and extent (%) of contribution by the candidate:
My contribution to this paper was approximately 20%, and included the researching of cost components and the generation of overall cost estimates. Additionally, I contributed to the writing of the paper and the production of the graphical material.

Certification by Co-authors:
If there is more than one co-author then a single co-author can sign on behalf of all
The undersigned certifys that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name:	Signature:	Date:
Rob Lake	[signature]	16/9/14

Executive Summary

New Zealand faces a range of food safety hazards. Microbial hazards alone were estimated to cause over 2,000 years of lost healthy life in 2011 (Cressey, 2012) and $62m in medical costs and lost productivity in 2009 (Gadiel & Abelson, 2010).

Chemical hazards are thought to be well managed through existing controls (Vannoort & Thomson, 2009) whereas microbial hazards are considered harder to control, primarily due to their ability to reproduce along the food production chain. Microbial hazards are thought to cause the majority of acute foodborne gastroenteritis.

This research reviewed food safety literature and official documentation, and conducted 55 interviews, mostly with food safety experts from different stakeholder groups, to examine the food safety decision-making environment in New Zealand. This research explores the concept of the 'stakeholder' in the context of food safety decision-making and proposes an inclusive 'stakeholder' definition as any group which is able to affect, or be affected by, the decision-making process. Utilising this definition, and guided by interviews, New Zealand stakeholders in food safety decision-making were identified and classified as follows:

- Regulators
- Public health authorities
- Food safety scientists/academics
- Consumers
- Māori
- Food Businesses (further classified as):
    - Farmers
    - Processors
    - Food retailers

- Exporters

Interviews with stakeholders from these groups[1] highlighted twelve criteria as being relevant to multiple groups during food safety intervention evaluation:

- Effectiveness
- Financial cost
- Market Access
- Consumer Perceptions
- Ease of Implementation
- Quality or Suitability
- Quality of Science
- Equity of Costs
- Equity of Benefits
- Workplace Safety
- Cultural Impact
- Animal Welfare

There are a number of different ways to measure or assess performance on these criteria. Some are able to be quantitatively measured, while others may require the use of value judgements. This thesis used the Disability-Adjusted Life Year (DALY) metric for quantifying effectiveness during the testing of different MCDA models.

[1] Although identified as key stakeholders in this research, consumers were not interviewed for a number of reasons outlined in Section 5.2.

This thesis reviews the MCDA process and the food safety specific MCDA literature. There are different ways of conducting MCDA. In particular, there are a large number of models available for the aggregation phase; the process of converting model inputs, in the form of criteria scores and weights, into model recommendations. This thesis has described and reviewed the main classes of model.

The literature review and interview process guided the construction and testing of three classes of MCDA model; the Weighted Sum, Analytic Hierarchy Process (AHP) and PROMETHEE models. These models were selected due to their having different characteristics and degrees of complexity, as well as their popularity in the food safety and Health Technology Assessment (HTA) literature. Models were tested on the problem of selecting the most appropriate intervention to address the historic *Campylobacter* in poultry problem in New Zealand during the mid-2000s. Experimentation was conducted on these models to explore how different configurations utilise data and produce model outputs. This experimentation included:

- Varying the format of input data
- Exploring the effects of including/excluding criteria
- Methods for sensitivity analysis
- Exploring how data inputs and outputs can be elicited and presented using visual tools
- Creating and using hybrid MCDA models

The results of this testing are a key output of this thesis and provide insight into how such models might be used in food safety decision-making. The conclusions reached throughout this research phase can be classified into one of two broad groups:

- Those relating to MCDA as a holistic process/methodology for decision-making
- Those relating to the specific models and mathematical procedures for generating numerical inputs and outputs

This thesis demonstrates that food-safety decision-making is a true multi-criteria, multi-stakeholder problem. The different stakeholders in food-safety decision-making do not always agree on the value and importance of the attributes used to evaluate competing intervention schemes. MCDA is well suited to cope with such complexity as it provides a structured methodology for the systematic and explicit identification, recording and aggregation of qualitative and quantitative information, gathered from a number of different sources, with the output able to serve as a basis for decision-making.

The MCDA models studied in this thesis range from models that are simple and quick to construct and use, to more time consuming models with sophisticated algorithms. The type of model used for MCDA, the way these models are configured and the way inputs are generated or elicited can have a significant impact on the results of an analysis. This thesis has identified a number of key methodological considerations for those looking to employ one of the many available MCDA models. These considerations include:

- Whether a model can accommodate the type and format of input data
- The desired degree of compensation between criteria (i.e. full, partial or no compensation)
- Whether the goal of an analysis is the identification of a ‘best’ option(s), or the facilitation of discussion, and communication of data
- The degree of transparency required from a model and whether an easily understood audit trail is desired/required
- The desired output of a model (e.g. complete or partial ranking).

This thesis has also identified a number of practical considerations when selecting which model to use in food safety decision-making. These include:

- The amount of time and energy required of stakeholders in the generation of data inputs (elicitation burden)
- The degree of training required for participants

- How data inputs are to be elicited and aggregated in different group decision-making environments
- The availability of MCDA software for assisting an analysis

Considering the above points will assist users in selecting a suitable MCDA model that meets their requirements and constraints.

This thesis provides original and practical knowledge to assist groups or individuals looking to employ MCDA in the context of food-safety intervention decision-making. This research could also serve as a guide for those looking to evaluate a different selection of MCDA models.

List of Abbreviations

AGI – Acute Gastrointestinal Illness

AHP – Analytic Hierarchy Process

ALARA – As Low As Reasonably Achievable

ALOP – Appropriate Level of Protection

ALR – Acceptable Level of Risk

BoD – Burden of Disease

CAC – Codex Alimentarius Commission

CBA – Cost-Benefit Analysis

CCP – Critical Control Point

CEA – Cost-Effectiveness Analysis

CER – Cost-Effectiveness Ratio

CFU – Colony Forming Unit

COI – Cost of Illness

COP – Code of Practice

CUA – Cost-Utility Analysis

DALY- Disability-Adjusted Life Year

EFSA – European Food Safety Authority

ESR – Institute of Environmental Science and Research Limited

FAO – Food and Agriculture Organization

FDA – Food and Drug Administration

FSANZ – Food Standards Australia New Zealand

GAP – Good Agricultural Practice

GBD – Global Burden of Disease

GMP – Good Manufacturing Practice

GOP – Good Operating Practice

HACCP – Hazard Analysis and Critical Control Point

HALY – Health-Adjusted Life Year

HPP – High pressure processing

HTA – Health Technology Assessment

MAUT – Multi-Attribute Utility Theory

MCAP – Multi-Criteria Aggregation Procedure

MCDA – Multi-Criteria Decision Analysis

MIA – Meat Industry Association

MoH – Ministry of Health

MPI – Ministry for Primary Industries

MRL – Maximum Residue Limit

NIWA – National Institute for Water and Atmospheric Research

OMAR – Overseas Market Access Requirement

PHARMAC – New Zealand Pharmaceutical Management Agency

PHU – Public Health Unit

PIANZ – Poultry Industry Association New Zealand

PROMETHEE – Preference Ranking Organisation Method for Enrichment Evaluation

QALY – Quality-Adjusted Life Year

QRA – Quantitative Risk Assessment

RMF – Risk Management Framework

SeaFIC – Seafood Industry Council

SPS – Sanitary and Phytosanitary

SSM – Soft Systems Methodology

STEC – Shiga-Toxic Escherichia Coli

TBT – Technical Barriers to Trade

UN – United Nations

VS – Verification Services

WSM – Weighted Sum Model

WTO – World Trade Organization

Chapter 1: Introduction

Foodborne illness In New Zealand is estimated as being responsible for approximately 2000 years of lost healthy life (Cressey, 2012) and an additional $62m in medical costs and lost productivity annually (Gadiel & Abelson, 2010). This represents a considerable burden on the health and productivity of New Zealanders. At a time of increasing health costs it is essential that money is allocated to health interventions in an efficient, transparent and evidence driven manner. Food safety regulators have the important task of deciding which food safety risks need addressing. There are existing frameworks in place for completing this task (Cressey, 2012; NZFSA, 2006). Once it has been established that a food safety risk requires attention there is the equally important task of deciding how this risk can be best managed. The multiple stakeholders involved in the food production chain will have different priorities and criteria when it comes to deciding how to best manage a food safety issue. It is this task of evaluating and selecting appropriate risk management options that is the focus of this thesis.

The practical assessment of interventions is no easy task as there are multiple criteria to consider, and these can be measured using a variety of metrics and techniques. This thesis explores the decision-making process and poses the primary research question:

Research Question 1: How can food safety interventions be systematically evaluated using a framework which incorporates both quantitative data and stakeholder input, and which provides a basis for decision-making?

Historically, the evaluation of regulatory options has centred on the economic analysis tools of Cost-Benefit Analysis (CBA) and Cost-Effectiveness Analysis (CEA). CBA involves assessing interventions on only one criterion, net monetary costs; assigning a dollar value to every impact. The major methodological difficulty with using a cost-benefit framework is in assigning monetary values to health outcomes such as mortality.

Cost-Effectiveness Analysis (CEA) overcomes this problem by allowing the measurement of health outcomes using a non-monetary metric such as Disability-Adjusted-Life-Years[2] (DALYs), but lacks the ability to incorporate other relevant impacts such as those related to consumer perceptions, equity, safety and market access. A natural alternative to these methods is Multi-Criteria Decision Analysis (MCDA), which allows for the systematic integration of multiple quantitative and qualitative criteria, without the need for monetisation.

New Zealand is regarded as having a robust decision-making process for food safety risk-management. This process is co-ordinated by the Ministry for Primary Industries (MPI) and is outlined in their Food Safety Risk Management Framework (RMF) document (NZFSA, 2010b). Through the RMF MPI follows a comprehensive process for identifying and managing food safety risks in foods produced and consumed in New Zealand. Food safety interventions are usually evaluated against a range of criteria.

MCDA usually involves the synthesis of various measurements (or estimates) into an overall score, which can be used to rank alternative intervention schemes. It is a technique that is receiving increasing attention as an appropriate technique for use in food safety decision-making (Anders & Schmidt, 2011; Bartolini & Viaggi, 2010; Caswell, 2008; Fazil et al., 2008; Henson et al., 2007; Henson & Masakure, 2011; König et al., 2010; Mangen et al., 2010; Ragona & Mazzocchi, 2008b). This is primarily due to its transparency and ability to systematically integrate a variety of non-financial impacts, measured on different scales. This research examines the potential for the widespread application of MCDA to New Zealand food safety intervention evaluation.

The body of this thesis consists of eight chapters, each detailing an important phase in my research process. Chapter 2 begins by providing a synopsis of foodborne hazards in New Zealand, with the goal of addressing research question two:

---

[2] A Disability-Adjusted-Life-Year, or DALY, essentially represents one year of healthy life lost through illness, injury or death. It is a metric which combines both the morbidity and mortality associated with a health condition. The DALY is a popular metric in Cost-Effectiveness Analysis (or Cost-Utility Analysis). Section 6.3.2 contains a more thorough description of the DALY metric and how it is used in Cost-Effectiveness/Cost-Utility Analysis.

Research Question 2: What are the characteristics of the different types of food safety hazard faced by new Zealanders?

The main types of food safety hazard are explained; microbial, chemical and physical hazards. The burden of disease associated with microbial hazards has been most well characterised. The most common foodborne illnesses are summarised and the major differences in the characteristics of the wide range of foodborne hazards that contribute to New Zealand's burden of foodborne disease are highlighted, thus emphasising the complexity of reducing this burden.

Chapter 3 provides a comprehensive review of the food safety regulatory environment in New Zealand in order to address research question three:

Research Question 3: How is food safety legislation formulated and applied in New Zealand, and what tools do regulators have at their disposal to ensure the food produced and consumed in New Zealand is adequately safe?

A number of key concepts in contemporary food safety regulation need to be understood in order to generate a complete picture of how decision-making is conducted in New Zealand. Chapter 3 details these fundamental concepts that underpin the current regulatory model, including HACCP, Good Operating Practice (GOP), Appropriate Level of Protection (ALOP) and the Precautionary Principle. Chapter 3 also details the nature of New Zealand's interaction with key international food safety organisations such as the Codex Alimentarius Commission (Codex) and Food Standards Australia New Zealand (FSANZ), and outlines the key food safety treaties to which New Zealand is a signatory. Chapter 3 examines the MPI Risk Management Framework (RMF), the official document detailing the existing process for evaluating and selecting food safety interventions. The widespread adoption of a new decision-support tool (such as MCDA) in New Zealand would require its successful integration with existing risk management processes, especially the RMF. Understanding the current regulatory environment, as well as the nuances of any potential decision-support tools like MCDA, is essential for assessing the compatibility of the two.

Chapter 4 describes the general MCDA methodology in detail and is dedicated to addressing research questions four and five:

Research Question 4: What is Multi-Criteria Decision Analysis and how can it improve decision-making?

Research Question 5: What are the different types of MCDA model and what are their key differences?

There is often confusion regarding the definition of MCDA, with some viewing MCDA simply as a mathematical procedure for converting stakeholder preferences into model outputs, usually in the form of an intervention ranking. This synthesis of results is only a part of MCDA, which actually encompasses an entire decision-making process, from the initial conceptualisation of the problem to the final analysis and validation of model outputs. Each step in MCDA is important, with some involving the mathematical representation and aggregation of stakeholder preferences and intervention performances. The most common MCDA models are also described in Chapter 4.

The evaluation and selection of food safety interventions is a multidisciplinary activity and has the potential to affect, or be affected by, a wide range of stakeholder groups. As such, the decision-making process could benefit from a participative and inclusive approach to ensure that the full spectrum of views are incorporated and the range of potential impacts associated with different intervention schemes enumerated and considered in decision-making. This can help maximise stakeholder uptake or acceptance of the process/solution and identify factors which may affect the successful implementation of the option(s) under consideration. Chapter 5 examines the multi-stakeholder nature of food safety decision-making to address research question 6:

Research Question 6: Who are the key stakeholders in food safety intervention decision-making and how can they be classified according to their role in food safety intervention decision-making?

Qualitative research methodologies including a document review and interviews with key stakeholders are used in Chapter 5 to explore the concept of the stakeholder before settling on an appropriate 'stakeholder' definition for use throughout this thesis. Chapter 5 details the results of the research process used to identify and classify stakeholders into groups. Fifty-five interviews were conducted with stakeholders from these groups as part of the process to elicit stakeholder knowledge and preferences regarding different aspects of intervention decision-making, which was necessary to guide the later modules of research.

MCDA requires that the data inputs used in the modelling phases reflect the preferences of stakeholders (or decision-makers). A key task in MCDA is determining the criteria to include and how performance is to be measured on each. Chapter 6 and 7 explore this issue by addressing research question seven:

Research Question 7: What criteria are relevant to key New Zealand stakeholders during the evaluation of food safety interventions, and how can performance on these criteria be measured or quantified?

Chapters 6 and 7 generate a list of criteria which could potentially be used to evaluate the overall performance of new food safety interventions. Additionally, these chapters investigate how these criteria could be practically measured or quantified. Chapter 6 reviews existing food safety literature while Chapter 7 uses interviews with stakeholders to identify additional criteria and assess the relevance to New Zealand stakeholders of the criteria from the literature. Chapter 7 proposes a set of criteria relevant to New Zealand stakeholders for later use in a MCDA case study.

This thesis uses information gathered during interviews and the literature review process to guide a hypothetical MCDA. The analysis evaluates interventions to address the historic *Campylobacter* in poultry problem in New Zealand during the mid-2000s. Three potentially suitable MCDA models are constructed and tested on this case study using test data, with the intention of addressing research questions eight through ten:

Research Question 8: How can MCDA be utilised in the evaluation of specific food safety interventions?

Research Question 9: How can decision-makers select the most appropriate MCDA model for their problem?

Research question 10: What practical considerations should decision-makers keep in mind when employing MCDA in the context of food safety intervention evaluation?

Chapter 8 documents the construction and testing of these models. The models are compared and evaluated for suitable use in New Zealand food safety decision-making. Experimentation with their formats is conducted to help identify specific advantages and disadvantages of each. These are highlighted and explained through the use of examples in the context of the *Campylobacter* in poultry case study.

Chapter 9 summarises the main conclusions reached in this thesis. The main differences between MCDA and other decision-support methodologies such as Cost-Effectiveness Analysis, in the context of food safety decision-making, are highlighted. Implications of this research for food safety decision-making are discussed, and the strengths and limitations of the research presented in this thesis are also outlined. Suggestions for future research are also provided.

This research utilised a variety of methodologies to address the various research questions posed in this thesis. These methodologies included literature reviews, interviews and the construction, testing and analysis of specific MCDA models. Ethics approval was not required for this research. Thesis chapters are organised as a series of research modules, with their own methodologies, Rather than produce a single, comprehensive 'Methodology' chapter to describe the range of methodologies used throughout the entire thesis, methodology sections are included in the relevant chapters. This structure helps provide context, and assists in guiding the reader through each module of research.

Chapter 2: Foodborne Illness in New Zealand

2.1 Introduction

Foodborne illness is responsible for a considerable burden on the health of New Zealanders. A substantial amount of time and money is focused towards the prevention and management of foodborne illness by food safety authorities and food businesses. In New Zealand there are a wide variety of hazards which have the potential to cause foodborne illness. Significant effort is expended on monitoring chemical residues in food and performing risk management activities, primarily by the Ministry for Primary Industries (MPI), who have reduced the health risks associated with chemical contamination of food to a very low level (MPI, 2014a; Vannoort & Thomson 2009). The health burden associated with microbial hazards such as foodborne bacteria and viruses has been estimated, and these hazards are the subject of risk management activities and target setting (Lim et al., 2012). It is widely accepted that the majority of acute foodborne illness in New Zealand is of a biological origin. This chapter provides some background information regarding the surveillance activities carried out by regulators, and the hazards they monitor.

2.2 Chemical Hazards

The burden of disease associated with chemical hazards is thought to be low (MPI, 2014a; Vannoort & Thomson, 2009). This is largely due to the stringent regulatory activities conducted by MPI. MPI formulate maximum allowable limits for chemical hazards in food and conduct monitoring activities to ensure that producers are complying with them (Chapter 3 further details this regulatory process, including the setting of maximum limits). This section outlines the main types of foodborne chemical hazard and provides examples of those hazards most relevant to New Zealand.

Pesticides

Pesticides have the potential to remain in or on food products post-harvest and can be ingested by consumers. Ingesting high doses of pesticides can cause acute poisoning, while long-term exposure can also lead to health issues. In New Zealand the use of pesticides in food produced and consumed is tightly controlled, and the risk to consumers from exposure to pesticides is thought to be negligible (Vannoort & Thomson, 2009).

Heavy Metals

In the context of food safety, heavy metals refer to any metal or metalloid[3] that is of concern to the health of consumers or animals. Heavy metals can be introduced into the food chain through environmental pollution such as leaded petrols and paints, industrial effluents and leaching of metal ions from the soil into lakes and rivers (Oxford, 2008), which are subsequently ingested by crops and/or livestock.

Oysters are thought to be the most significant dietary source of the heavy metal cadmium, although the majority of New Zealanders do not eat shellfish (Vannoort & Thomson, 2009). Lead is a heavy metal with a *"ubiquitous environmental presence"*, albeit at very low levels, and is found in a range of foods. Mercury is a heavy metal most often associated with fish and shellfish consumption (Vannoort & Thomson, 2009).

Hormones

Hormonal growth promotants (HGPs) are administered to animals in many countries. In New Zealand their use is strictly controlled (MPI 2014e). The acceptable use of HGPs is an internationally debated topic and has led to major disagreements between large international trading partners (MercoPress, 2008). The health effects of HGP are also debated (MercoPress, 2008).

---

[3] A metalloid is any chemical element having properties in between those of metals and non-metals. From: http://en.wikipedia.org/wiki/Metalloid

2.3 Microbial Hazards

The primary objective of the food safety wing of MPI (formerly the New Zealand Food Safety Authority) is to protect New Zealanders and overseas consumers of New Zealand food products from potentially injurious foodborne hazards. High quality surveillance of foodborne disease, especially domestic outbreaks, is an essential component of MPI's Risk Management Framework (RMF) (Lim et al., 2012; NZFSA, 2010b). Regulators commission the collection of a significant amount of data regarding foodborne illness in the New Zealand population. This information is used extensively by a variety of scientific organisations to profile diseases and research strategies to reduce their burden of illness.

The notifiable disease schedule is a list of diseases that, if encountered (or suspected by) a medical health practitioner, are required to be reported to a public health official (MOH, 2015). This allows for appropriate intervention and, if necessary, more intense surveillance in order to determine the source of the disease and to reduce the risk of additional cases. Where there is reasonable suspicion from a medical health practitioner that a patient is suffering from a notifiable disease, a laboratory sample is usually obtained from the patient, with laboratory analysis ultimately confirming or refuting the suspected diagnosis. Disease reporting to public health officials usually occurs directly from laboratories through a rapid messaging system known as LabSurv (ESR, 2014b). In addition to specific enteric diseases, certain categories of acute gastroenteritis must be reported immediately, even if a causative agent has not yet been established. Such categories include, two or more cases sharing a suspected common source, and single cases in areas of higher-risk (e.g. food production, teaching institutions).

Notifications are collected in a central database named EpiSurv (ESR, 2014a). EpiSurv is a web-based application available for use by staff at each of the twenty public health units (PHUs) within New Zealand. The service is maintained by the Institute of Environmental Science and Research (ESR), who are also responsible for aggregation, analysis and reporting of notifiable disease statistics on behalf of the Ministry of Health (MoH).

In New Zealand there are ten notifiable diseases for which foodborne transmission is important:

- Campylobacteriosis
- Cryptosporidiosis
- Acute Gastroenteritis thought to be from a common source
- Giardiasis
- Hepatitis A
- Listeriosis
- Salmonellosis
- Shigellosis
- STEC infection
- Yersiniosis

A brief summary of the disease characteristics and rates for each of these diseases is presented in the following section. There are usually many possible transmission routes for these diseases and estimating the proportion of illness attributable to foodborne sources is a difficult task. With the exception of campylobacteriosis, the foodborne proportions of illness referenced in this research are the results of expert elicitation.

Campylobacteriosis

Campylobacteriosis is an infection caused by the *Campylobacter* bacterium, most commonly *Campylobacter jejuni*. Campylobacteriosis is estimated as being significantly underreported, and responsible for more recorded cases of foodborne illness than any other bacterial agent (Gadiel & Abelson, 2010).

Campylobacter is most commonly found in the digestive tract of birds and animals. Illness resulting from *Campylobacter* infection is most commonly associated with the consumption of undercooked meat products, with poultry being a major contributor to the disease burden. Over half of the estimated 3,568 foodborne cases of Campylobacteriosis in 2011 were estimated as being attributable to poultry consumption (Lim et al., 2012).

Campylobacteriosis can result in significant and sometimes life-threatening sequelae[4]. Such sequelae can include inflammation of the liver and kidneys; reactive arthritis and serious paralytic conditions (ESR, 2007).

Reported incidence rates of campylobacteriosis rose steadily after it became a notifiable disease in 1980, and in the mid 2000s New Zealand had one of the highest rates of campylobacteriosis in the developed world (Gadiel & Abelson, 2010). This number peaked in 2006 at 383.5 cases per 100,000 population, or 15,873 total cases (Gadiel & Abelson, 2010).

At the height of the campylobacteriosis epidemic in the mid 2000s the New Zealand Food Safety Authority (now part of MPI), in conjunction with the poultry industry, implemented a *Campylobacter* Risk Management Strategy (NZFSA, 2010a). The primary goal of the strategy was to reduce the foodborne burden of campylobacteriosis by 50% by 2013 (NZFSA, 2010a). The strategy included the setting of performance targets to reduce bacterial numbers on poultry carcases during primary production, as well as a public education campaign regarding proper handing and preparation of raw poultry meat. The poultry industry as a whole responded to the epidemic by implementing specific interventions during primary processing. Such interventions included the upgrade of evisceration machines and spray washes, more intensive monitoring of carcass bacterial numbers and the occasional use of chemical dip tanks (Lake et al., 2013). This *Campylobacter* Risk Management Strategy has been highly successful at reducing the foodborne burden of campylobacteriosis, with reported rates more than halving over the 5 year period since its inception (MPI, 2013b; Sears, 2011).

[4] Sequelae are secondary medical conditions resulting from an initial disease. In the context of foodborne illness, sequelae often refer to health issues following an initial bout of ‘food poisoning’.

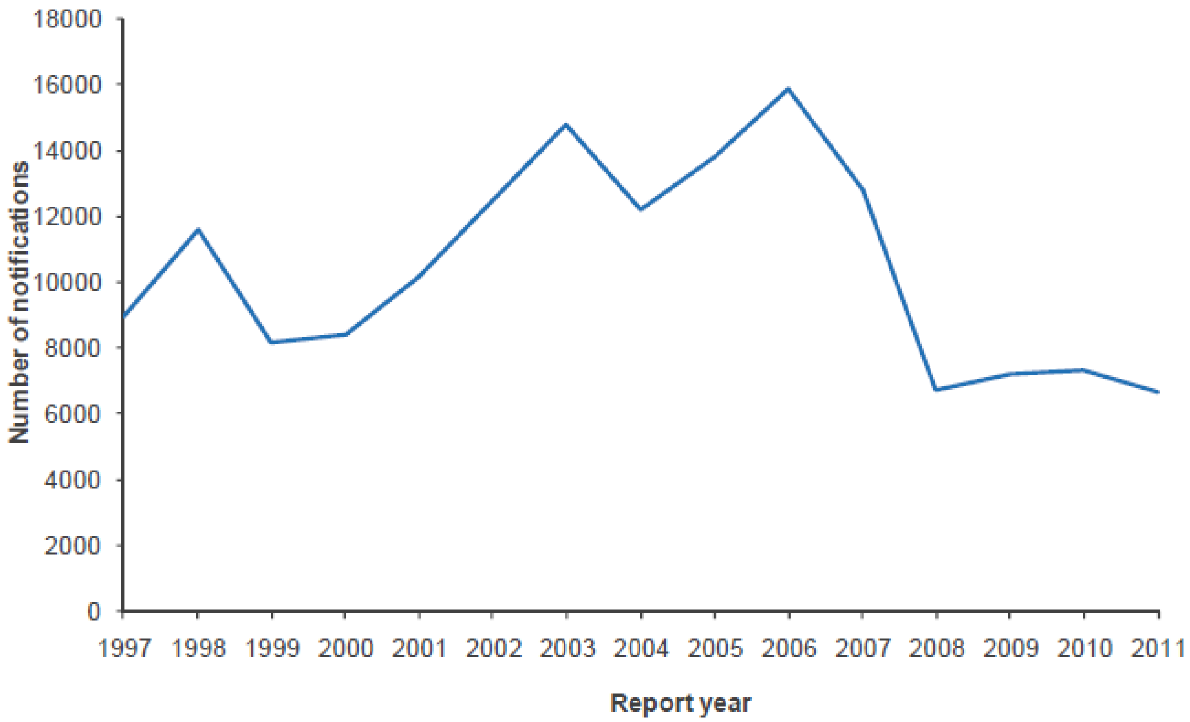


Figure 1: Campylobacteriosis notifications by year, 1997-2011. Graph from (Lim et al., 2012).

Cryptosporidiosis

Cryptosporidiosis is a parasitic infection caused by the *Cryptosporidium* parasite. It affects the digestive tract of humans in a (usually) short-term and self-limiting infection. Cryptosporidiosis is generally considered more of a waterborne rather than foodborne illness. However, infection can occur following the consumption of food which has been washed in, or come into contact with, contaminated water.

Annual reported illness peaked at 1,208 cases in 2001 and has fluctuated between 610 and 954 annual notifications over the last decade, following no discernible pattern. In 2011 there were 610 notifications, representing an incidence rate of 13.8 cases per 100,000 population (Lim et al., 2012).

Giardiasis

Giardiasis is caused by the *Giardia* parasite, which inhabits the digestive systems of a wide variety of animals. Infection is often associated with the consumption of contaminated food or water, but person-to-person transmission is also possible.

There has been a steady increase in reported cases of giardiasis since 2006, and in 2011 there were 1,935 reported cases, representing an incidence of 43.9 cases per 100,000 population (Lim et al., 2012).

Hepatitis A

Hepatitis A is a food and waterborne virus associated with poor sanitation practices, especially poor hand washing after visiting the bathroom. The virus cannot multiply in food or water but it is extremely resistant to environmental conditions (e.g. heat). Shellfish are commonly implicated in Hepatitis A outbreaks due to their tendency to concentrate virus particles from contaminated growing waters in their tissues as a result of filter feeding.

Over the last 15 years there has been an overall downwards trend in reported cases of Hepatitis A (see Figure 2). In 2011 there were 26 Hepatitis A notifications (0.6 per 100,000 population**).**

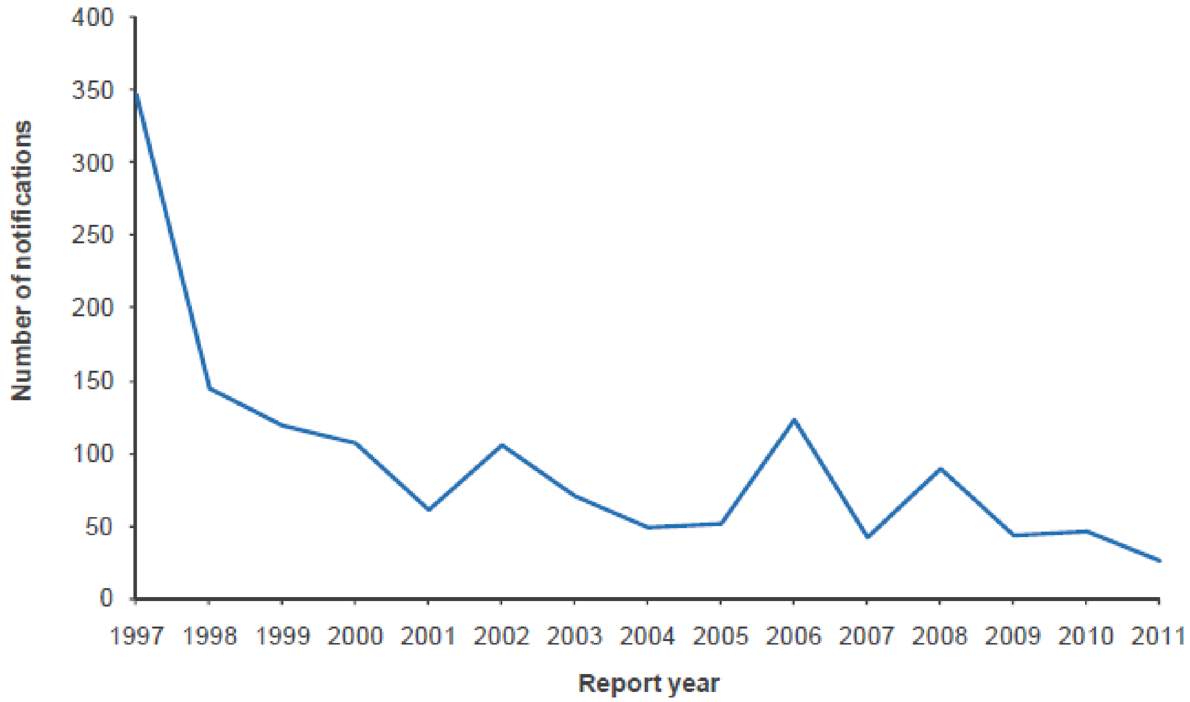


Figure 2: Hepatitis A notifications by year. 1997-2011. Graph from (Lim et al., 2012). Increased notifications in 2002, 2006 and 2008 correspond to significant outbreaks in each of these years (Lim et al., 2012).

Listeriosis

Listeriosis is an illness caused by the bacterium *Listeria monocytogenes.* The illness can remain confined to the digestive system where it manifests itself as typical gastroenteritis, known as non-invasive listeriosis. Alternatively, invasive listeriosis involves the spread of the infection to the brain, where it can cause severe illness and death. The mortality rate for invasive listeriosis is far higher than that of most other foodborne diseases. Despite its relatively low incidence rate, it has been estimated that 28% of all deaths caused by foodborne pathogens can be attributed to invasive listeriosis (Lim et al., 2012). Many of these deaths are perinatal.

MPI formulated a specific *Listeria* risk management strategy in 2008 with the primary goal of ensuring no increase in foodborne listeriosis cases over the 2008-2013 period (MPI, 2013d). The strategy has been successful in achieving this goal. Listeriosis rates have remained relatively constant at around 15-20 notified cases per year over this time (Lim et al., 2012).

Salmonellosis

The exposure pathways for *Salmonella* bacteria have not been as clearly established as those for *Campylobacter*, although consumption of food from retail outlets and poultry products is often reported (Lim et al., 2012; Lim et al., 2011; Lim et al., 2010).

Salmonellosis notifications peaked in 2001 at 2,417 cases before declining until around 2004. In 2008 MPI decided that salmonellosis required its own risk management strategy. The main goal of this strategy was to reduce the annual incidence of foodborne salmonellosis by 30% over the 2008-2012 period (MPI, 2013h). This goal has been achieved, primarily through the application of additional control measures at various points in the food production chain (MPI, 2013h). In 2011 there were 1056 cases of salmonellosis, representing an incidence rate of 24 per 100,000 people (Lim et al., 2012).

Shigellosis

Shigellosis is an illness caused by *Shigella* bacteria. Poor sanitation is a risk factor for *Shigella* transmission and, because of this, uncooked foods that require significant hand contact during preparation are commonly implicated with shigellosis outbreaks (USDHHS, 2013). Shigellosis notifications declined over the second half of last decade, with 101 notifications (2.3 per 100,000 population) in 2011 (Lim et al., 2012).

Norovirus

Norovirus is estimated to be the single greatest non-bacterial cause of foodborne gastroenteritis in New Zealand (Gadiel & Abelson, 2010). Norovirus is also easily transmitted and the virus is commonly implicated in large outbreaks in close-quarter settings such as rest homes, schools and cruise ships (Gadiel & Abelson, 2010). Because of this it is very difficult to estimate the fraction of norovirus cases due to contaminated food rather than person-to-person transmission.

Due to its relatively low severity and self-limiting nature it is thought that the vast majority of norovirus cases go unreported. It is estimated that for every notified case of food-attributable norovirus there are up to 1000 additional food-related cases in the general population (Gadiel & Abelson, 2010). In 2011 there were 72 notifications of norovirus infection, 40% of these cases were thought to be associated with food and, of these foodborne cases, 40% were estimated to be due to consumption of contaminated shellfish (Lim et al., 2012).

Yersiniosis

Yersiniosis is an illness associated with consumption of certain strains of bacteria from the *Yersinia* genus. Causing symptoms similar to those of classical gastroenteritis, yersiniosis is generally self-limiting. Yersiniosis became a notifiable disease in 1996 and, since then, annual recorded cases have remained fairly stable at between 9 and 12 cases per 100,000 (Lim et al., 2012).

Shiga Toxic producing Escherichia Coli (STEC)

Compared to diseases such as campylobacteriosis, STEC is infrequently reported in New Zealand. Cases of STEC infection often require hospitalisation (Gilbert et al., 2006). Meat and meat products like ground beef are commonly implicated in STEC outbreaks (Gadiel & Abelson, 2010; Lake et al., 2002). STEC notifications have been steadily increasing since 1997 (see Figure 3). In 2011 there were 154 STEC cases reported (3.5 per 100,000) with between a quarter and a half of these cases estimated to be of foodborne origin (Lim et al., 2012).

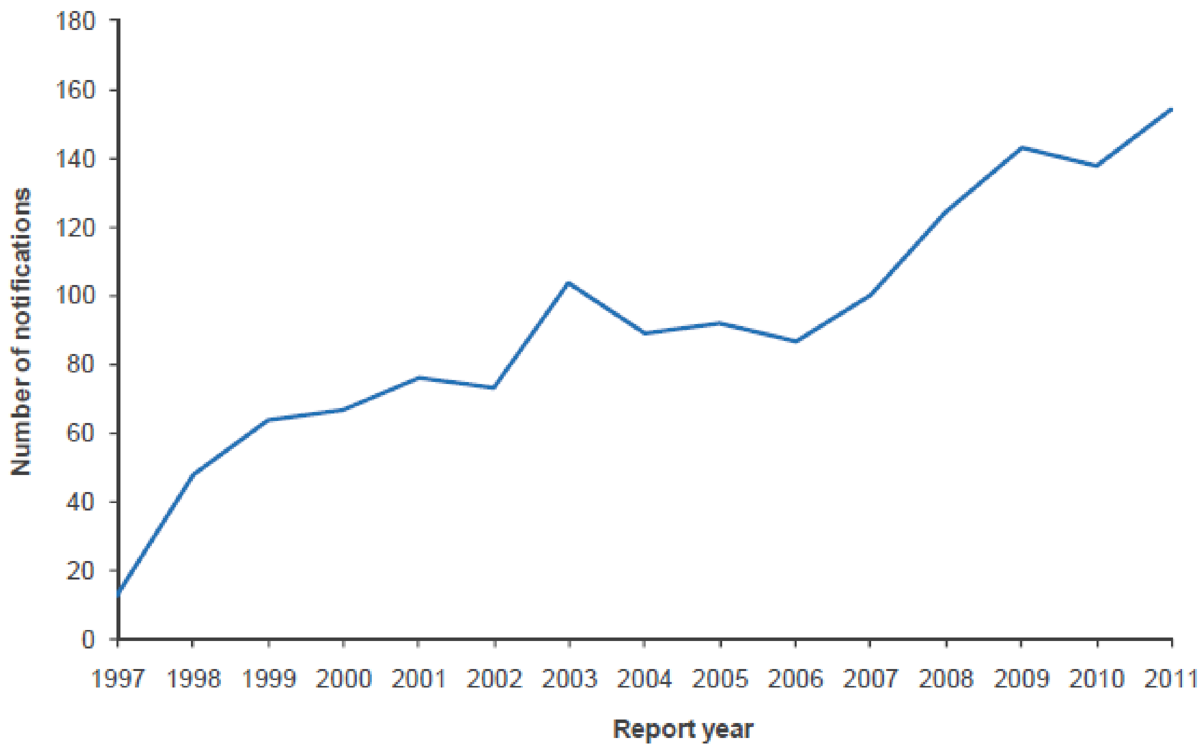


Figure 3: STEC infection notifications by year, 1997-2011. From (Lim et al., 2012).

2.3.1 Microbial Toxins

Some pathogenic organisms can become a food safety hazard through their production of toxic metabolites. Some Bacteria excrete exotoxins, which can be subsequently ingested by consumers even when no bacteria remain. In contrast, endotoxins remain part of the bacteria, making the bacteria harmful due to the presence of such toxins. *Botulinum* neurotoxins are the causative agents of botulism, a serious and sometimes deadly foodborne illness (ESR, 2010). *Shigatoxigenic Escherichia coli* (STEC) are another group of bacteria that can produce very harmful toxins, and are heavily monitored (Gilbert et al., 2006; Lake et al., 2002). Mycotoxins are secondary metabolites of some species of fungi. The term 'Mycotoxin' is usually reserved for the toxic by-products of the fungi that actively colonise crops. Nuts and grains are two food groups routinely monitored for Mycotoxins (Ragona et al., 2011).

2.4 The AGI Pyramid

Although most enteric diseases are notifiable and should be reported to public health authorities under New Zealand's notifiable disease surveillance system, a significant amount of underreporting occurs. Acute Gastrointestinal Illness (AGI) is the clinical term given to the rapid onset of the symptoms resembling 'food poisoning' (usually vomiting and/or diarrhoea). Only a fraction of those who experience AGI visit a medical practitioner, with an even smaller fraction receiving a conclusive diagnosis with regards to the pathogen responsible. However, it is useful from a public health perspective to produce estimates of the overall New Zealand disease burden associated with enteric illness.

When a person in the community becomes ill with suspected AGI there is generally a five step process that is followed before a notification is made to the EpiSurv database. Lake et al., have used community survey and national surveillance data (from 2006) to estimate that for every case of AGI reported to the surveillance database there are approximately 222 cases occurring in the community (Lake, Adlam, et al., 2010).

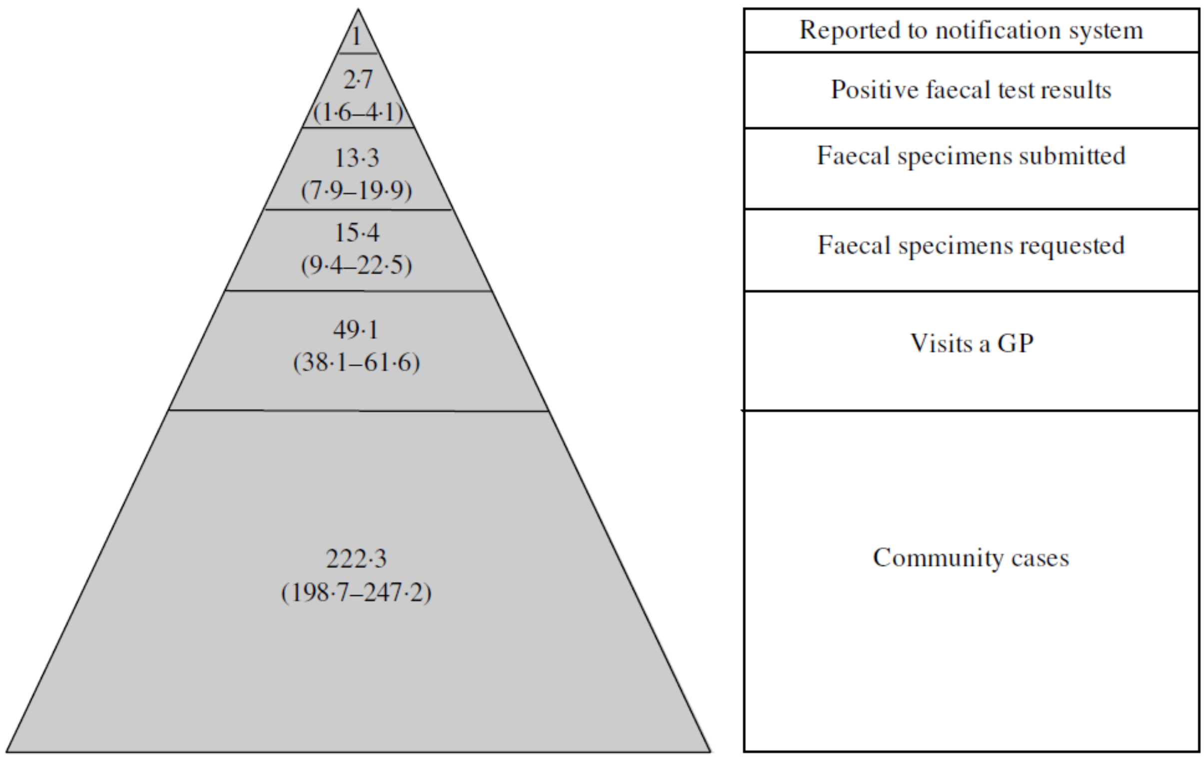


Figure 44: The AGI reporting pyramid showing ratios of cases in the community, general practice and clinical laboratory levels relative to notifiable diseases, 2006, (mean, 5th and 95th percentiles) from (Lake, Adlam, et al., 2010).

Multipliers for individual diseases will differ from the pyramid in Figure 4, which presents aggregated data. The reporting rate is likely to be much higher for more serious diseases such as invasive listeriosis. Estimating multipliers is a difficult and resource intensive task and many multipliers used in New Zealand are from other countries like Britain and the Netherlands.

2.5 Burden of Disease

Regulators and health planners are interested in quantifying the impacts of foodborne illness so they can prioritise the food safety interventions that will lead to the greatest reduction in the burden of foodborne disease. The burden of illness for a selection of New Zealand foodborne diseases is calculated periodically by utilising the Disability-Adjusted Life Year (DALY) and Cost of Illness (COI) approaches. A DALY essentially represents one year of healthy life lost through illness, injury or premature death, whilst the COI approach estimates the value of the direct and indirect financial costs associated with an illness. (A more thorough discussion of these approaches is provided in Section 6.3.2). Following a similar approach to that of Haagsma and Havelaar (2006), Cressey and Lake have estimated the food attributable DALY burden associated with selected New Zealand microbial pathogens (Cressey, 2004), with the most recent update being for 2011 (Cressey, 2012). The results are summarised below.

Disease	Food attributable DALYs (2011)
Norovirus	873
Campylobacteriosis	587
STEC infection	200
Listeriosis	160
Salmonellosis	67
Yersiniosis	62
Total	**1949**

Table 1: 2011 Food attributable DALY impact of various microbial diseases in New Zealand (Cressey, 2012)

DALYs are designed to capture the pain, suffering and overall loss of quality life associated with a disease. There is some disagreement regarding whether such outcomes can, or should, be assigned a dollar value instead, so that the entire burden of disease is expressed in financial terms. This is the approach that Gadiel and Abelson have taken to express the annual burden of foodborne disease based on 2009 data (Gadiel & Abelson, 2010). They calculated the direct and indirect financial costs of medical treatment in a similar way to that of Lake et al (Lake, Cressey, et al., 2010). However, additional to these costs, Gadiel and Abelson have also calculated what they define as *"Residual lifestyle costs"*, which capture the *"welfare gain in avoiding any risk of pain and suffering (including self-care situations) as well as in avoiding the risk of premature death"*. In the studies by Lake and Cressey et al, it could be argued that this last cost component (pain, suffering and loss of life) is captured by the DALY metric, whereas in the Gadiel and Abelson study this cost is monetised and combined with the other, more tangible, costs. Gadiel and Abelson estimate the same burden of disease at $NZ161m (Gadiel & Abelson, 2010), almost double the Cost-of-illness estimates produced by Lake et al (Lake, Cressey, et al., 2010). This is due to the different methodologies used; in particular, the choice by Gadiel and Abelson to monetise pain, suffering and loss of life.

2.6 Conclusion

Irrespective of the methodology used for its quantification, foodborne disease is responsible for a considerable health and financial burden on New Zealand society. Food safety authorities expend much effort monitoring and controlling chemical and biological hazards. Chemical hazards are easier to control and monitor, and New Zealand does a good job at minimising the chemical burden of foodborne disease. The majority of acute foodborne illness in New Zealand is thought to be due to pathogenic hazards, which are much harder to control.

Diseases dominated by AGI as an outcome contribute significantly to the overall disease burden due to the sheer number of annual cases. Norovirus infection and campylobacteriosis are examples of this type of disease. Other diseases such as listeriosis have a low incidence but contribute significantly to the burden of disease, due to their severity and sometimes severe on-going health problems and/or death. Comparing the burden of different foodborne diseases by looking at their incidence rates does not necessarily allow for a clear ranking of diseases according to their aggregate impact on society, because incidence rates do not take into account the duration and severity of the different illnesses. This is the main reason why disease burdens are expressed in DALYs and/or financial costs. Although different, and with their own methodological issues, the DALY and Cost of Illness approaches utilised by Lake et al, and the Willingness to Pay approach utilised by Abelson and Gadiel both allow for the meaningful comparison of diseases with different disease characteristics. Provided adequate data is available, these methodologies allow for the comparison of foodborne diseases with other diseases (e.g. Cancer), medical conditions (e.g. disabilities) or health outcomes (e.g. death from road accidents).

Intensive monitoring and surveillance of foodborne illness allows food safety authorities to build a picture of the effects on society of the different diseases. These activities, combined with additional research into developing reliable risk management options for addressing the different diseases, provides a useful starting point upon which to base food safety decision-making, so that limited resources can be directed towards making the greatest impact on the burden of disease in the most appropriate way. Minimising the burden of foodborne disease is a primary objective for most food safety interventions. The degree to which an intervention can reduce the incidence and burden associated with a disease are two criteria that will typically be considered during intervention decision-making. However, there may be other criteria which play an important role in this type of decision also. Regulators in particular may wish to enumerate and consider these criteria during the food safety decision-making process.

Chapter 3: Food Safety Regulation and Decision-making in New Zealand

3.1 Introduction

The successful implementation of new tools to support New Zealand food safety decision-making relies on their compatibility with the existing regulatory framework. This chapter describes the current New Zealand food safety regulatory and decision-making environment.

3.2 Chapter Methodology

The primary objective of this chapter is to provide an overview of food safety regulation in New Zealand. Specifically, the intent is to describe how food safety policy is formulated and enforced, and to detail the specific roles of, and interactions between, the different organisations responsible for the safety of food produced and imported in New Zealand. This chapter also describes the tools that are commonly utilised by regulators to manage different food safety hazards.

The key Research Question addressed in Chapter 3 is Research Question 3:

Research Question 3: How is food safety legislation formulated and applied in New Zealand, and what tools do regulators have at their disposal to ensure the food produced and consumed in New Zealand is adequately safe?

The information presented in this chapter has been sourced in a two-step process consisting of an initial document review followed by key informant interviews with stakeholders. These stakeholders included regulators, food businesses, public health experts and scientists and academics.

The majority of the information presented in this chapter was sourced from official documents released by organisations involved in food regulation and standard setting, especially the Ministry for Primary Industries (MPI), Codex Alimentarius Commission (Codex) and Food Standards Australia New Zealand (FSANZ). The websites of these organisations contain large databases of documents, ranging from broad descriptions of overarching policies to highly technical descriptions of standards, guidelines and enforcement policies. These databases were searched during June and July 2013 for relevant material, with no restrictions on publication date. Documents were selected based on their ability to inform Research Question 3.

Purposive sampling was used to locate and interview key personnel involved in food safety regulation to supplement the information gathered during the document review phase. An initial brainstorm session identified a small number of key informants in industry and government. These individuals were contacted and asked to participate in this research and to provide recommendations of other key informants who might provide useful information.

3.3 Evolution of Risk Management and the Food Safety Regulatory Model in NZ

In the mid 1980's New Zealand underwent a comprehensive set of structural reforms. These reforms included changes to the way domestic food safety policy was administered; resulting in a shift towards a user-pays approach to food safety regulation and a progressive reduction in the government's involvement in food production and sale.

Before these reforms the government had a more prescriptive role in ensuring food safety, by delivering instructions to industry on what must be done and, in most cases, how it must be done. From the perspective of the government, food businesses had little incentive to truly understand the hazards and risks associated with their products or production processes (Flynn, 1999). The responsibility of ensuring food safety was almost solely the governments, conducted through the manual end product inspection of foods. The government had significant hands-on involvement in the day-to-day food safety management of businesses.

These reforms signalled a move towards 'cost recovery' and 'user pays' policies across the food sectors, with the ultimate goal of improving efficiency while forcing the beneficiaries from goods or services to pay the real costs. An example of this cost recovery in action can be seen in the government's changes to product inspection services for New Zealand agricultural producers in the late 1980's, which resulted in producers essentially paying the full cost of product inspection (NZFSA, 2008b). During this time a change to the way food research was conducted also occurred with, for example, large changes to the way the Meat Industry Research Institute of New Zealand (MIRINZ) was funded. The government heavily reduced its contributions to meat research, from originally funding research in equal partnership with industry, to offering only conditional grants based on *"contracts dealing with issues that were non-appropriable by commercial interests"* (Wright, 2013).

During this time an increasing demand from consumers for safe food and responsible producers resulted in a growing emphasis on sophisticated risk management and quality control procedures by businesses. Advances in risk management made during the previous half century had resulted in some robust and sophisticated risk management processes receiving wide application to food safety problems within New Zealand and around the world. One example is the now widely used Hazard Analysis and Critical Control Point (HACCP) approach to food safety; a preventive approach with a focus on designing and verifying safe and robust production procedures, rather than relying on finished product inspection (Ropkins & Beck, 2000).

The 1990's and 2000's saw major advances in the government's application of food safety policy, with increasing emphasis placed on the scientifically based processes of risk management (Flynn, 1999). Reliance on end product testing was being phased out and a new reliance on process control and management techniques such as HACCP was emerging. The responsibility for ensuring the production of safe food shifted from the government to farmers, processors and retailers. The responsibility of the government was still to assure quality, but without having to be deeply involved in the quality control and hands-on management of food businesses. By making food safety the primary responsibility of businesses the intention was to reduce direct government intervention without compromising food safety. This shift to science based quality management programs coupled with the implementation of 'user-pays' policies provided the foundations for the modern New Zealand food safety model. Today, food safety regulation is primarily coordinated by the Ministry for Primary Industries (MPI) (formerly the New Zealand Food Safety Authority (NZFSA)). MPI is responsible for the implementation and enforcement of food safety regulation in New

Zealand. MPI works closely with Food Standards Australia New Zealand (FSANZ), a joint Australia-New Zealand food safety agency, on issues such as the setting of Microbial Limits, environmental contaminant Maximum Limits (MLs) and labelling and composition standards. MPI also works closely with the Codex Alimentarius commission (CAC or Codex); the international body established by the Food and Agriculture Organization (FAO) and the World Health Organization (WHO) and tasked with coordinating and streamlining international food safety standards. In addition, MPI liaises with the food safety agencies of major trading partners, especially the USA's Food and Drug Administration (FDA) and the European Union's European Food Safety Authority (EFSA), on issues of mutual interest.

3.4 Modern Food Safety Concepts

3.4.1 The Farm-to-fork chain

The farm-to-fork chain, also known as the farm-to-table and paddock-to-plate chain, is an important concept in food production and safety. It refers to the different stages in the production of food, including growing, harvesting, processing, storage, sale and preparation (Figure 5). All food has been through some form of farm-to-fork chain, whether it be a simple chain where a farmer might grow and eat his own produce, or a more complex chain like the large scale production of processed meat products.

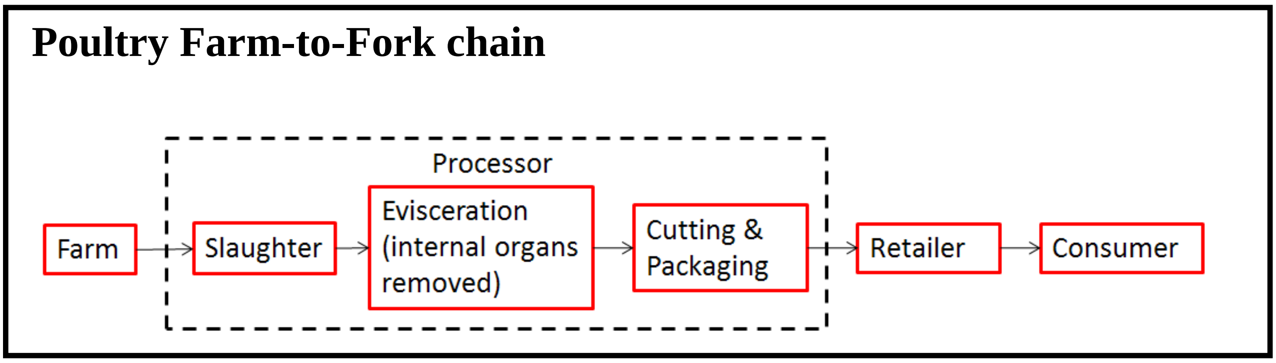


Figure 55: Simplified poultry farm-to-fork chain

Current best practice is to treat an entire farm-to-fork chain as a continuum rather than as individual production steps, with the continuum ending at the consumer's fork (CAC, 2007b, 2011). In the context of food safety risk management, farm-to-fork chains will likely be considerably more complex and detailed than the simple example given in Figure 5. Each stage of production will itself contain a number of smaller sub-stages or processes. In practice, any step or individual process where there is the potential to control a food safety hazard should be explicitly considered as part of the farm-to-fork chain (CAC, 2007b). This leads to the related concept of Hazard Analysis and Critical Control Point (HACCP).

3.4.2 Hazard Analysis and Critical Control Points (HACCP)

HACCP is a systematic and preventive approach to food safety. Seeking to identify physical, chemical or pathogenic hazards in the farm-to-fork chain which may result in an unsafe finished food product, HACCP aims to redesign or adjust the production process so that these food safety risks are consistently reduced to an acceptable level (see Section 3.4.4). HACCP focuses on the production process rather than the end product, and can be employed at any stage in the food production chain, including processing, packaging and preparation.

After its successful implementation by NASA to ensure the safety of astronaut food, HACCP was widely adopted internationally (Ropkins & Beck, 2000). Originally used only in the most high risk industries such as poultry and beef, HACCP is now the standard food safety methodology employed by food producers in the developed world.

HACCP General Principles

Before an HACCP plan is formulated and implemented a clear picture of a product's farm-to-fork chain must be constructed, including all steps where a food hazard might be controlled or minimised. From here HACCP systems follow seven general steps (CAC, 1997):

1. **Conduct a hazard analysis**: Determine the food safety hazards present in a system and identify the preventive measures that can be applied to the production process to control these hazards.

2. **Identify critical control points**: A critical control point (CCP) is a step in a food manufacturing process at which a food safety hazard can be prevented, eliminated, or reduced to an acceptable level.

3. **Establish critical limits for each critical control point:** A critical limit defines the minimum or maximum value to which a biological, chemical or physical hazard must be controlled in order to reduce risk to an acceptable level.

4. **Establish critical control point monitoring requirements:** Monitoring activities are required to ensure that the HACCP plan is working as intended at each of the critical control points.

5. **Establish corrective actions:** These are actions to be taken when monitoring detects a deviance from established critical limits.

6. **Establish procedures for ensuring the HACCP system is working**: Verification is a crucial step and ensures that the systems are successful in consistently achieving the production of safe food products. Verification procedures include activities such as a review of HACCP plans, CCP records, critical limits and microbial/chemical sampling and analysis.

7. **Establish record keeping procedures:** New Zealand regulations require that all businesses operating an HACCP plan maintain certain documents, including their written HACCP plan, monitoring records of critical control points, critical limits, verification activities, and the handling of deviations to critical limits during processing (Lee, 2012).

HACCP in New Zealand

HACCP plans are generally required for larger food businesses and those involved in the production and sale of higher risk products. The primary food safety legislation that is administered by the MPI is the:

- Food Act (1981)
- Animal Products Act (1999)
- Agricultural Compounds and Veterinary Medicines Act (1997)
- Wine Act (2003)

Across this legislation various core control tools are employed; Risk Management Programmes (RMPs) for producers of animal products, Product Safety Programmes (PSPs) for dairy operators and Food Safety Programmes (FSPs) and Food Control Programmes (FCPs) for most other food businesses. Not all food businesses are required to operate HACCP based programmes yet, but those that do must have them registered and approved by

MPI. Once an HACCP plan has been approved it is placed on an MPI register and will be subject to regular audits. Ongoing monitoring and record keeping are essential conditions of operating a HACCP plan in New Zealand, and non-compliance can result in the revocation of an operator's license to operate (NZFSA, 2010a).

3.4.3 Good Operating Practice (GOP)

GOP describes a range of actions that contribute to the production of safe food. It details the practices that a food business should follow in order to create a suitable and hygienic environment for the production of food. These procedures are generally facility wide initiatives covering aspects such as:

- Personal hygiene
- Staff training protocol
- Cleaning and sanitation
- Equipment maintenance
- Allergen management
- Facility layout and construction
- Pest control
- Labelling
- Waste management

In this context GOP is used as a blanket term also covering Good Hygienic Practice (GHP), Good Agricultural Practice (GAP) and Good Manufacturing Practice (GMP).

The difference between GOP and HACCP

GOP is a more general set of usually qualitative procedures to be implemented facility-wide, regardless of the food item being produced or hazards being managed. GOP tends to address food safety in a more indirect manner by creating an overall clean and safe environment for the production of food, as well as engraining the importance of hygienic practices into the culture of an organisation.

HACCP is a more precise and scientific methodology for addressing the remaining individual hazards associated with specific food products or processes once GOP has been implemented. HACCP protocols are usually quantitative in nature and involve the control of measurable parameters (e.g. temperature & pH) at the Critical Control Points in the production line, where food safety hazards can be best addressed.

HACCP requirements are usually far more detailed, and food or hazard specific. A large food production facility producing a number of different food items may operate under one set of GOP guidelines but may have a number of HACCP 'plans'; one for each production line. For higher risk businesses, including most producers of animal products, GOP must be implemented in conjunction with HACCP methodologies (MPI, 2014b).

3.4.4 Acceptable Level of Risk (ALR) and Appropriate Level of Protection (ALOP)

Appropriate Level of Protection (ALOP) or Acceptable Level of Risk (ALR) is defined by the World Trade Organization (WTO) as "*the level of protection deemed appropriate… to protect human, animal or plant life or health*" (WTO, 2013). It is the level of protection from food safety hazards that is deemed appropriate by a society (or its food safety authority).

Any hazard can be described in relation to the complementary concepts of safety and risk (Griffin, 2012). High levels of safety are associated with low levels of risk, and vice versa. Although the WTO considers ALR and ALOP to be equivalent concepts (WTO, 2014), Griffin argues that these two concepts address complementary sides of the same concept (Griffin, 2012). A regulatory body has the right (and duty) to establish an appropriate level of protection from a given hazard. In practice, this is achieved by determining a level of acceptable risk following a risk assessment, and evaluating the options available to reduce this risk to an acceptable level. Griffin argues that it is more practical to work with the concept of risk, as it is more measureable than the concept of safety, which is often more subjective and abstract (Griffin, 2012). For example, in the EU the ALOP is said to be achieved when risk is reduced to a *"negligible level"*. Australia defines its ALOP as *"reducing risk to a very low level, but not to zero"* and in the US there needs to be a *"reasonable certainty of no harm"* (DAFF, 2008). These statements are not clear regarding what the corresponding, and usually quantitatively expressed, Acceptable Level of Risk might be in each case.

During the interview process several food safety academics, regulators and industry representatives made a distinction between these two terms, referring to ALOP as a qualitative statement of the level of risk (or protection) to be accepted for a given hazard(s), with ALR being the explicitly stated numeric level of risk used to operationalise the ALOP for a given food/hazard combination (e.g. a 1 in a million chance of contracting a disease per portion of food consumed). The literature is somewhat inconsistent with the use of these terms. In addition to the debate regarding whether ALOP and ALR are interchangeable, ALOP has been implied to be a universal level of protection from all hazards for a given society (DAFF, 2008). In contrast, other literature implies that ALOP should be food/hazard specific (NZFSA, 2010b). Whether it is more appropriate to define one ALOP for each food/hazard pair, as is implied in the New Zealand Risk Management Framework (NZFSA, 2010b), or one universal ALOP across all food/hazard pairs for a given society, is subject to debate. However, the general consensus from the literature and stakeholder interviews is that ALR is most commonly a quantitative, unambiguous expression of risk associated with an end product, reflecting the sometimes qualitative and sometimes universal ALOP. ALR is an important concept as it is this figure that is often used to determine the chemical and microbial criteria used in HACCP systems and international standards that food businesses must adhere to.

3.4.4.1 Determining Acceptable Level of Risk

No food can be truly risk free and thus defining an acceptable level of risk is essential for setting guidelines and standards regarding the safety of food products (and many other activities). Setting acceptable risk requires placing a numerical value on risk(s); explicitly and, ideally transparently, defining a cut-off point above which a risk becomes unacceptable. Hunter and Fewtrell (2001) outline a number of commonly used methods for defining when a risk is acceptable (Hunter & Fewtrell, 2001). In summary, a risk might be considered acceptable if:

- It falls below some arbitrarily defined probability (e.g. 1/1,000,000)

- It falls below some level that is currently tolerated
- It falls below a certain attributable fraction of the total disease burden in a community
- The cost of reducing the level of risk would outweigh the costs saved
- Additional expenditure would be better allocated to other areas or on reducing other risks
- Public health (or other experts) say it is acceptable
- Society 'says' that a risk is acceptable (or do not say that it is unacceptable)
- Politicians decide that it is acceptable (or bargain with stakeholders)

These approaches can be broadly classified as either those that emphasise formal analysis and expert opinion, such as the economic or disease burden approaches, or those that emphasise some sort of political bargaining process. These different approaches may signal a conflict between objectively and subjectively approaching the quantification of risk, with the implication being that methods based on formal economic and epidemiological data and expert judgement will be more scientifically exact. However, there are still significant uncertainties around many of these processes that are yet to be resolved (e.g. the determination of utility weights in Cost-Utility Analysis or the value of a statistical life in Cost-Benefit Analysis). On the other hand, a weakness of the political bargaining approach is that the outcome is not always the most appropriate level of risk, but rather a level that is acceptable to most. This is known as satisficing, and can lead to inefficient outcomes, especially for vulnerable stakeholder groups (e.g. small businesses) whose interests may not be adequately represented during the decision-making process (Hunter & Fewtrell, 2001). MPI food safety express a preference for taking an analytical approach to setting acceptable risk in its Risk Management Framework (NZFSA, 2010b).

The decision regarding how to determine acceptable risk and what to set it at is that of the MPI, who often utilise considerable expertise from industry and other agencies such as the Ministry of Health and, in the case of internationally traded foods, organisations like Codex and the food safety authorities of New Zealand's trading partners.

In New Zealand acceptable risk and risk reductions, in the context of food safety and public health, may be expressed as a tolerable incidence of reported disease; the maximum acceptable number of cases of illness due to a particular hazard over a specific time (e.g. X cases per 100,000 people per year). In some situations the risk per edible portion of a food is used to communicate acceptable risk. Measurement of the societal impact of a foodborne disease using Health-Adjusted Life Years (E.g. DALYs) as a comparative unit is an increasingly popular method of comparing risks from disparate sources when deciding on a desired level of consumer protection. MPI often express their public health goals as percentage improvements over current rates of illness (NZFSA, 2010b). These goals are then converted into industry targets, in the form of chemical or microbial criteria, which specify the maximum tolerable concentration of a chemical or pathogen in the end product. This provides well understood and easily assessed standards for food businesses to meet if public health goals are to be achieved. Food businesses then have a responsibility to meet these standards, and may themselves set standards for their suppliers to help achieve this.

MPI emphasises the setting of food safety goals that reflect continuous improvements in consumer protection over time (NZFSA, 2010b). Systematic application of a risk management framework allows regulators to monitor progress (from surveillance data) and modify targets as needed. If continuous improvement in public health is not achieved, these targets can be adjusted, or the authorities can look to employ additional forms of intervention (NZFSA, 2010b).

3.5 International Food Safety Legislation and Organisations

3.5.1 Codex

The Codex Alimentarius Commission (CAC or Codex) is the international body tasked with coordinating and streamlining international food safety standards (Stewart & Johanson, 1998). Established in 1961-1962 by the FAO and WHO, it is the most influential international food safety organisation in the world, with over 185 member countries, including New Zealand. With input from member countries, Codex produces food safety guidelines regarding issues such as:

- Food labelling
- Food additives
- Maximum Residue Limits for chemical contaminants
- Microbial Risk Management (MRM)
- Allergens
- Genetically Modified foods
- HACCP principles
- Codes of hygienic practice for specific industries (e.g. poultry)

The work of Codex is carried out by subsidiary bodies, grouped along the lines of commodity committees (e.g. Meat Hygiene), general subject committees (e.g. Food Additives), ad hoc task forces and regional coordinating committees. There is an executive Codex committee who coordinate the strategic activities of Codex, and have final decision-making power within the commission (FAO, 2003).

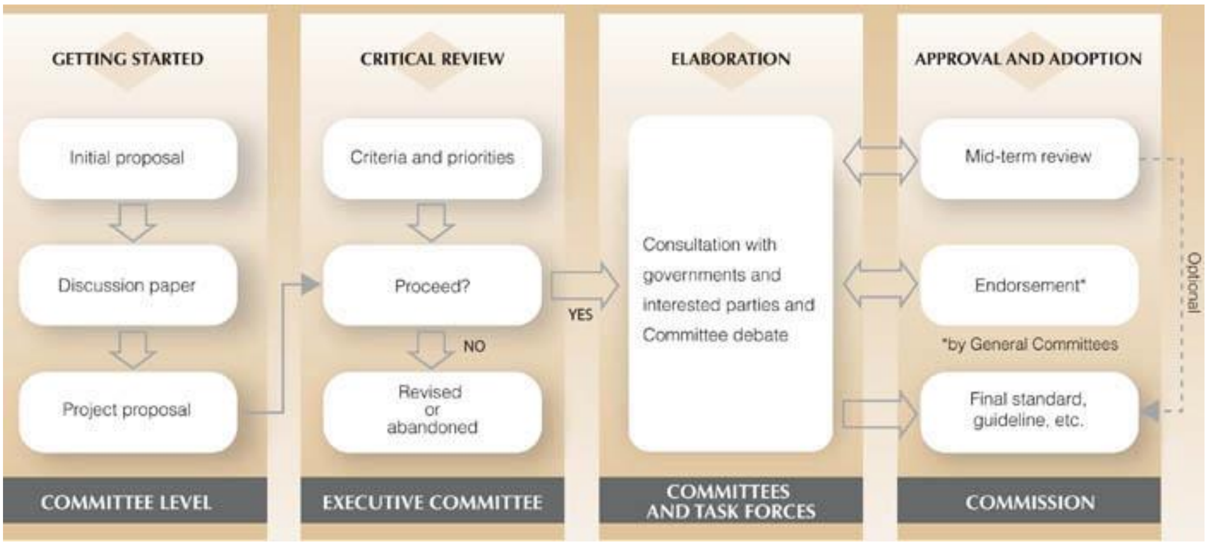


Figure 6: The Codex standards process. Diagram from (Codex, 2014)[5].

[5] http://www.fao.org/docrep/008/y7867e/y7867e05.htm (accessed January 2014)

While Codex guidelines are only recommendations for voluntary adoption by member countries, they often serve as a basis for national legislation (NZFSA, 2010b; Stewart & Johanson, 1998). The Codex Commission also plays an important role as a forum for member countries to discuss contemporary food safety issues and how they should be addressed.

Codex works to advance the objectives of the WTO Sanitary and Phytosanitary (SPS) agreement and the Technical Barriers to Trade (TBT) agreement. Codex and these two agreements are mutually reinforcing and are intended to provide a strong framework of food related standards and regulations at the national and international levels.

The Sanitary and Phytosanitary (SPS) Agreement

The main goal of this international agreement is to streamline global food safety regulations by encouraging members to base much of their food safety policies on international (i.e. Codex) standards and guidelines. Signatories of the SPS agreement are required to:

- Move towards operating a risk-based approach to food safety and basing food safety policy on the best available domestic and international science

- Ensure that food safety measures are no more restrictive than required to meet their appropriate level of protection (ALOP)

- Accept other countries' food safety measures as 'equivalent' if it can be demonstrated that an exporting country's food safety measures meet the importer's appropriate level of protection. This is known as the concept of 'equivalence'. New Zealand regulators and exporters interviewed in this research emphasised the importance of this concept in streamlining international trade.

The SPS agreement allows members to set their own standards so long as they abide by the above principles.

Technical Barriers to Trade (TBT) agreement

Many food standards relate to product description (e.g. labelling, packaging, consumer information). The TBT agreement aims to establish fair and practical international regulations that reduce the ability of countries to use these standards as trade barriers. Setting regulations arbitrarily can lead to their being used as an excuse for protectionism. The TBT agreement attempts to ensure that regulations are not creating unnecessary obstacles to trade while also providing members with the right to implement measures to achieve legitimate food safety objectives (i.e. the protection of human health).

New Zealand participation

New Zealand is a signatory to the SPS and TBT agreements and has been an active member of Codex since its inception in 1962 (NZFSA, 2009b). New Zealand's participation is coordinated by MPI, who play an active role in setting Codex standards by attending meetings of Codex committees, participating in a wide range of codex working groups (MPI, 2010/2011), leading international standards work in specific areas such as Meat and Milk Hygiene (NZFSA, 2009b) and collaborating with other Codex members to further areas of mutual interest. New Zealand places huge importance on the work of the Codex Commission and agrees especially with its position of facilitating increased international trade by streamlining global food safety standards and basing such standards on sound science (NZFSA, 2009b, 2010b).

One of MPI's strategic objectives is to actively contribute to the development of Codex standards. MPI implements this goal by coordinating with Codex and focusing its limited resources and expertise on research into the issues most relevant to New Zealand (e.g. Meat and Dairy issues) (NZFSA, 2009b).

3.5.2 Food Standards Australia New Zealand (FSANZ)

Established in 1995, FSANZ is a bi-national regulatory agency responsible for formulating food standards for Australia and New Zealand. FSANZ is responsible for setting Microbial Limits, environmental contaminant Maximum Limits (MLs), labelling and composition standards (NZFSA/FSANZ, 2008). However, New Zealand is able to abstain from adopting FSANZ standards and set its own standards if appropriate ("Food Standards Australia New Zealand Act," 1991; NZFSA/FSANZ, 2008).

Although both countries have mutual agreement on most food safety issues through FSANZ, there is no official requirement to consult with each other on Codex issues. In practice however, New Zealand and Australia co-operate on a wide range of Codex issues. MPI works with FSANZ, just as it works with Codex, to pool resources and expertise on a range of food safety issues that are important to both countries (NZFSA, 2010b).

FSANZ is not required to adopt Codex standards, but does try to harmonise its regulations with Codex standards wherever possible (FSANZ, 2013).

3.5.3 International Approaches to Food Safety Decision-Making

Historically, cost-effectiveness and cost-benefit analysis have been the most commonly employed methods for selecting risk-management options from a pool of several potential alternatives (Fazil et al., 2008). Most countries with modern food safety risk-management systems require some form of economic analysis of major food safety policies/interventions prior to their implementation (Hoffmann, 2010). These analyses usually focus on quantifying and comparing the financial costs of implementation with the likely health benefits of each intervention scenario. In OECD countries, the financial costs of regulation are generally calculated in accordance with the basic principles of applied welfare economics, and decomposed into direct compliance costs, indirect costs to consumers and producers, and the social costs of market readjustment (Hoffmann, 2010).

In the US, major food safety regulations require formal evaluation by way of cost-benefit analysis before becoming law (Executive Order 12866, 1993). In the EU, Health Adjusted Life Year (HALY)[6] metrics, such as the Disability-Adjusted Life-Year (DALY), are widely accepted and regularly used to support decision-making through techniques like Cost-Effectiveness or Cost-Utility Analysis[7] (Hoffmann, 2010). This approach is also in line with the World Health Organization’s Global Burden of Disease methodology (WHO, 2009) and WHO/FAO guidelines regarding the development and evaluation of food-safety standards (FAO/WHO, 2002). Although permitted for use in regulatory analysis in the US (U.S. OMB, 2003), HALYs are less commonly used in food-safety intervention analysis in the US and are not a recommended approach (IOM, 2006). One potential reason for this is due to the difficulty of monetising HALY’s for use in cost-benefit analysis (Hoffmann 2010).

---

[6] See section 6.3.2 for an explanation of HALY metrics.

[7] Cost-Effectiveness and Cost-Utility Analysis are economic evaluation techniques that do not require the monetisation of benefits like Cost-Benefit Analysis does. See Section 4.3 for a further discussion of Cost-Effectiveness and Cost-Utility Analysis.

Besides the explicit consideration of financial and public health criteria, many food-safety regulatory agencies will also evaluate the social impacts of a regulation or intervention. Such impacts include those with implications for animal welfare, sustainability, cultural minorities, social distribution of benefits and consumer perceptions (Fazil et al., 2008, Henson et al., 2008, König et al., 2010).

It is unclear exactly which criteria are utilised in different intervention evaluations conducted by different regulatory bodies. Furthermore, risk management frameworks do not always specify how performance on each criterion is to be quantified or measured.

One common issue with international risk management frameworks is their lack of detail on how information regarding the performance of different interventions on various criteria is to be synthesised into an overall decision. When utilising public health metrics, such as HALYs, and/or financial cost data, techniques such as CBA and CEA provide a measure of performance (in the form of some sort of costs to benefits ratio). However, when additional criteria are included in decision-making, as many frameworks propose, this synthesis of performance data can become subjective, with decisions potentially lacking transparency and consistency.

3.6 Groups Involved in Food Safety Regulation

From the perspective of the New Zealand government, there are three major groups directly involved in maintaining food safety, each with separate roles (EU, 2000):

- The Government
- Independent/third party auditors and verifiers
- Food businesses subject to regulation

3.6.1 The government – The Ministry for Primary Industries (MPI)

MPI is charged with regulatory control of food safety. It sets the legal parameters within which the food safety risk management system functions, as well as approving industry developed food safety systems and the third-party verifiers of these systems.

Under the current regulatory environment there are three tiers of government (Flynn, 1999);

- Ministerial/Parliamentary
- Policy development and advice
- Regulatory

These three levels generally perform their activities independently of the others, although there is significant communication between them. The Regulatory level is where the greatest amount of activity takes place, but it is at the Ministerial/Parliamentary level that a food safety issue has the greatest profile.

Ministers within the government help decide on strategic policy directions and provide recommendations for law to Parliament. This happens directly in the form of Acts[8] or indirectly in the form of government regulations[9].

The policy formulation branch of MPI food safety provides advice on the need for, and form of, government intervention to ensure food safety risks are managed. The senior leadership team at MPI food safety, led by the Director General, source advice from Principal Advisors, who are experts in their various fields of food safety (e.g. Toxicology) and act as the interface between the policy development and regulatory arms of MPI food safety.

[8] An Act is a law passed by Parliament and may designate an administrative agency (Ministry) of government to implement its provisions.

[9] A regulation is a specific description of what a Ministry requires as a means of implementing an Act or ensuring compliance. A regulation generally has the force of law unless it is found to extend beyond the jurisdiction of the Ministry that created it.

The regulators' tasks include approvals or accreditations, standard setting and compliance and enforcement. Their job is to operationalise the policies agreed on by the other two tiers or government. In the context of food safety, the main jobs of the regulators can be summarised as (NZFSA, 2008b):

- Approval of risk-based food safety programs (e.g. Food Safety Programmes or Risk Management Programmes) for food businesses
- Accreditation of independent verifiers and auditors
- Implementation of New Zealand (and sometimes joint New Zealand-Australia (FSANZ)) standards
- Negotiating with New Zealand's trading partners regarding the acceptance of food safety standards (e.g. MRLs, equivalence agreements, export certificates)
- Providing or facilitating technical and scientific input into the formulation of laws and regulations
- Coordinating New Zealand's activities with other food safety agencies, including FSANZ and Codex
- Being accountable for food safety assurances that domestic and internationally agreed food safety standards are being adhered to, and that food safety issues are being adequately managed
- Investigation into breaches of food safety agreements/standards, and determining the appropriate action in response to such breaches
- The enforcement of food safety regulations, including the prosecution of offending parties when deemed necessary
- Responding to food safety emergencies and recalls

3.6.2 Third party verifiers

Third party agencies are accredited by the regulatory tier of the government and are required to perform specific functions with regards to ensuring that food businesses comply with laws and regulations. They will have different roles depending on the food sector within which they operate, and the risk categories of the foods they are dealing with. A summary of verifiers main activities include (Flynn, 1999):

- The inspection of food businesses to assess whether they are meeting their food safety obligations
- Alerting the food safety authorities (MPI) when regulatory or legal requirements are not being met by food businesses
- Assessing the continued validity of specific risk management plans in a dynamic safety environment

Verifiers are more closely aligned with the government than the food industry. The government has not specified a maximum number of verifiers. In theory this allows industry operators to select verifiers from a pool of providers who must compete for contracts. However, the structure and small size of some New Zealand industries means that sometimes operators will have access to only a single verifier. In some cases verification activities can be undertaken by the regulators, such as the MPI food safety Verification Services (VS). VS operate as a business group within MPI food safety, with a significant degree of organisational separation between it and the other groups within MPI food safety.

Most industries in New Zealand operate under modern HACCP based legislation and their food safety programmes are required to be audited by third party verifiers. However, many smaller food businesses operate under the Food Hygiene Regulations 1974, which are not based on the principles of HACCP and are highly prescriptive by comparison. These regulations are also the responsibility of MPI.

3.6.3 Regulated food businesses

Food businesses have a requirement to:

- Develop and implement appropriate risk based systems for managing food safety risks under the relevant legislation, or comply with the Food Hygiene Regulations 1974 until these are phased out.

- Maintain compliance with their registered food safety systems

- Maintain adequate engagement with verifiers/auditors

- Ensure that all members in the business are aware of their specific roles in maintaining food safety and complying with relevant food safety agreements and laws, and that specific food safety related tasks become engrained in a business's process control program.

These three groups perform complementary roles which, when combined, are intended to operate as an efficient and robust system for protecting the health of domestic and foreign consumers, as well as enhancing New Zealand's position as a reputable and trusted exporter of safe and unadulterated food products (Flynn, 1999). Today, this modern food safety model is applied across all food sectors.

3.6.4 Perceived benefits of the current food safety model

The modern regulatory model has shifted primary responsibility for providing safe food to the food producers, allowing them more flexibility in managing risk but also making them more accountable for their actions. A view amongst relevant interviewees in this research was that this extra autonomy has resulted in many food businesses, especially large producers of higher-risk products, becoming highly involved and competent in many of the intricate and scientific aspects of risk-assessment and food safety management, with a greater understanding of the range of food safety hazards faced within their respective industries. Exporters indicated that the additional knowledge and expertise they have acquired through the continual improvement and refinement of their food safety processes has helped them produce food products that now consistently exceed minimum requirements imposed by foreign food safety authorities. In some cases this has allowed them to attract new customers with stringent food safety demands, including reputable international supermarket, hotel and fast-food chains. Interviewees agreed that New Zealand's adoption of scientifically based and internationally accepted food safety management methodologies (e.g. HACCP) and robust verification and assurance systems has, in addition to increasing the safety of food, facilitated access to new international markets and allowed New Zealand to reinforce its reputation as a trusted exporter of safe and suitable food. Exporters interviewed in this research expressed a desire to continually improve this global image through the continual improvement of food safety processes.

3.6.5 Levels of intervention

At this point in the research it is important to define what exactly is meant by the word 'intervention' in the context of food safety. In a broad sense, an intervention is any activity, procedure or technology which is intended to reduce the risk associated with a particular food safety hazard. In modern food safety systems, including those in New Zealand, there are generally three levels of intervention:

- Regulatory level intervention
- Industry level intervention
- Consumer level intervention

Regulatory level intervention

These interventions are mandated or implemented by food safety regulators. New Zealand regulators prefer a non-prescriptive approach to ensuring food safety and generally mandate outcomes rather than specific technologies or control measures, thereby providing food producers with a degree of flexibility regarding how such outcomes are operationalised. These outcomes generally come in the form of maximum limits and other product specifications that are associated with 'safe' food. Examples of this type of intervention include Maximum Residue Limits (MRLs) and microbial limits. Many of the key decisions made by regulators involve setting limits and specifications that reflect an acceptable level of risk (ALR – see Section 3.4.4). MPI also has the power to mandate specific technologies and processes for the control of certain hazards, but prefers to employ this power only as a last resort when food businesses have been unable to address a food safety issue through self-regulation or voluntary measures.

Industry level intervention

Industry level interventions are generally specific technologies (e.g. irradiation) or processes (e.g. biosecurity measures) implemented by food businesses for the control of food safety hazards. These interventions usually involve a physical process applied directly to the food at some point in the farm-to-fork chain by frontline workers or machinery. Businesses have a requirement to comply with the standards set by MPI, but are usually allowed some flexibility regarding how this is done. There are often a number of technologies available for the control of a particular hazard, and deciding which technologies to employ will be a key decision for food businesses looking to address a food safety risk. Food businesses interviewed indicated that technologies are typically evaluated against a number of important criteria.

Consumer level interventions

This type of intervention generally involves the promotion of consumer behaviours, usually the appropriate preparation and storage of food, which will minimise their risk of foodborne illness. Consumers are often the last line of defence against foodborne illness, and preparation and consumption can be efficient Critical Control Points for addressing a food safety risk (e.g. by thorough cooking of food). Stakeholders interviewed in this research broadly agreed that although consumers have significant potential to reduce the foodborne burden of disease, as a whole they could not be relied on to consistently and effectively implement many consumer level interventions.

3.7 The Risk Management Framework

The MPI Risk Management Framework (RMF) is designed to identify and manage the complex array of current and future food safety issues in New Zealand (NZFSA, 2010b). The RMF details an iterative process for managing risks, and consists of four interrelated components:

1. Preliminary risk management activities
2. Identification and selection of risk management options
3. Implementation of risk management decisions
4. Monitoring and review

Each of these steps is broken down into several individual tasks. The framework takes into account scientific developments, the practical experience of industry and input from other stakeholder groups.

The RMF is designed to cover all the different types of foodborne hazard that MPI is charged with managing, and allows for a reasonable degree of flexibility regarding how this can be done. The RMF emphasises the importance of quality science and a multidisciplinary approach to food safety. It is in line with international best practice and Codex principles (FAO, 2007b, FAO/WHO, 1997) and, in particular, emphasises HACCP principles and the need to consider the whole farm-to-fork continuum when managing risk.

Intended to provide a basis for dealing with situations in which the scientific basis for carrying out risk assessments varies widely, the RMF legitimises the use of taking a precautionary approach to food safety when necessary (described further in Section 3.8.2). It explains the importance of considering uncertainty in risk assessments and communicating this uncertainty to relevant stakeholders, including the public. It includes an obligation to engage with stakeholders during all phases of the risk management process as much as is practicable. The framework is used by MPI, who view the framework as a suitable guide for priority setting and for defining its public health and market access goals.

The RMF acknowledges the need to consider the work of international food safety bodies such as Codex during the standards setting process, and endeavours to align its methods and standards with those of Codex to the greatest extent possible. Step one and two in the RMF are generally the two steps where key decision-making processes take place, and involve answering one key question at each step:

Step 1: Which foodborne hazards/risks should MPI focus its resources on addressing?

Step 2: What are the most appropriate options for addressing/managing specific food safety issues?

3.7.1 RMF Step 1: Preliminary Risk Management Activities

Key Question: Which foodborne hazards/risks should MPI focus its resources on addressing?

The objective of Step 1 in the RMF is to generate a detailed understanding of the current food safety issue(s) faced by New Zealand consumers and businesses. The output of Step 1 is often a ranked list of issues to be given priority in further risk management activities. Food safety issues are often initially ranked according to either the amount of illness they cause or their potential to affect New Zealand's access to foreign markets. There are other criteria which may also be considered during risk-ranking. These criteria may consider aspects such as the number of deaths or the risk to vulnerable population cohorts arising from a particular hazard (NZFSA, 2010b).

3.7.1.1 Prioritisation according to disease burden

MPI continually emphasised in interviews (and in their documents) that their primary focus is the protection of consumer health. A significant amount of resources have been allocated to research prioritising such food safety issues, especially microbial threats, which are responsible for the majority of New Zealand's acute foodborne illness.

The MPI Science Group utilises the Disability-Adjusted Life Year (DALY) metric to rank microbial food safety hazards, allowing the combination of morbidity and mortality into a single unit representing the aggregated 'healthy life lost' due to a particular hazard or food (Lake, 2006). A key strength of this methodology is that it successfully integrates the incidence of a particular disease with its severity, as well as combining non-fatal and fatal outcomes, giving a more complete picture of the burden of disease than other popular metrics such as the incidence or number of years of life lost due to premature mortality (YLL).

This methodology has been in development for the last decade and is continually being refined. It was used as recently as 2011 to rank the burden of disease caused by six microbial hazards that are often foodborne (Cressey, 2012).

A future goal of MPI is to utilise this model to rank chemical hazards alongside microbial hazards. However, incorporating chemical threats into the methodology is likely to be challenging. Although it is possible to source incidence and severity data for the cancer and non-cancer health outcomes associated with chemical hazards, these outcomes usually occur after long term exposure, and determining the proportion of these outcomes attributable to chemical exposure, rather than other lifestyle or genetic factors, is difficult (Lake, 2006).

This process, although still in development, helps provide MPI with a clear idea of which hazards are making the most significant contribution to the foodborne burden of disease, thus helping regulators decide where to focus their efforts in order to reduce this burden.

Attribution Models

Attribution modelling refers to utilising mathematical techniques to expose and identify the relative importance of different transmission routes for microbial and chemical hazards (Lake et al., 2011). MPI and other agencies conduct a significant amount of attribution modelling, the results of which can give regulators an understanding of which exposure pathways are the major contributors to the overall disease burden associated with hazards such as *Campylobacter* (Müllner, 2009; Müllner et al., 2009). This type of modelling is useful for assisting regulators with deciding which food-hazard pairs to focus their efforts on. A more detailed description of this type of modelling can be found in Lake et al., (2011).

Risk Profiling

Risk profiles collate all relevant information regarding a food-hazard combination (e.g. *Campylobacter* in poultry) (MPI, 2014c). They give contextual background information regarding the possible risks associated with each pair, and will generally contain information regarding:

- How a food may become contaminated with a hazard (i.e. where in the farm-to-fork chain can/does contamination occur)
- What the critical control points in the farm-to-fork chain are, and where contamination is likely to be most effectively managed
- The number of consumers that are likely to be exposed to a hazard due to the various exposure pathways

- The potential health effects associated with exposure to a hazard

- Possible control measures and their availability. This may include information regarding specific environmental conditions under which a hazard is eliminated (e.g. pH, temperature, humidity, radiation) or specific risk management options that have been, or could be, implemented by stakeholders to reduce risk (e.g. adopt an *"all-in-all-out"* approach at the farm to control the between-flock spread of *Campylobacter* (Lake, Hudson, et al., 2007)).

- Any gaps in the research and suggestions of where further research would be beneficial in further understanding the risks associated with each food-hazard combination and how to manage them

MPI commissions risk profiles for some higher-risk food-hazard combinations. The majority of New Zealand risk profiles focus on bacterial pathogens coupled with either meat, seafood or dairy products, but other food-hazard pairs such as caffeine in energy drinks have been profiled (MPI, 2013g).

Quantitative Risk Assessment

Quantitative Risk Assessment (QRA) is the process of placing a numerical value on the risk of illness. For example, the chance of dying from campylobacteriosis in New Zealand has been reported as one in a million, or 0.0001% (Shaw, 2013). Recommended by the Food and Agriculture Organization, the World Health Organization and the Codex Commission (FAO/WHO, 1995b), QRA is a systematic and science based procedure for assessing the society-wide risks of illness (or worse) associated with a potentially contaminated food. MPI conducts comprehensive risk assessments for higher risk food-hazard pairs. For example, MPI's Science group have developed a farm-to-fork computer model of *Campylobacter* in the poultry food chain (Lake, Cressy, et al., 2007). The model describes the exposure of New Zealand consumers to *Campylobacter* from consuming contaminated poultry meat and estimates the number of new illnesses based on the daily number of people consuming fresh poultry. Such models have the ability to help identify Critical Control Points where contamination can be best managed, and can simulate potential risk management options (Lake et al., 2013). The model contributed to the comprehensive *Campylobacter* Risk Management strategy developed in 2006 to help select the interventions that would cut foodborne campylobacteriosis rates in half over the following five years (MPI, 2013b; NZFSA, 2010a).

These are some of the key tools available to MPI during Step 1 of the RMF, when there is a need to understand as much as possible about the major food safety issues faced by MPI. They are invaluable tools which allow MPI to help rank and prioritise issues according to their impact on consumer health, so that actions can be taken to address those risks most relevant to New Zealand consumers.

The cost-effectiveness and quality of the science associated with potential interventions are two important criteria that are considered during decision-making. MPI may prioritise issues for which there are efficient and scientifically backed interventions available for their management.

3.7.1.2 Prioritisation by Market Access

In addition to protecting health, MPI have a responsibility to improve opportunities for New Zealand exporting industries and food businesses. MPI runs a programme known as the Market Access Work Programme (NZFSA, 2010d).

MPI is seeing an ever increasing number of requests from food businesses for market access work, which generally involves MPI liaising with the food safety authorities of New Zealand's trading partners to discuss and formalise specific requirements that exporters must meet to access overseas markets. These requirements are known as Overseas Market Access Requirements, or OMARs (MPI, 2013f), and detail aspects of trade such as:

- Acceptable packaging, labelling and branding of exported food
- Transport arrangements, including the type of equipment to be used and methods of food storage. For example, OMARs might require the use of certified refrigerated containers and tamper-proof seals during transit.
- How verification activities are to be conducted

OMARs are not always necessary for New Zealand food businesses to export to a particular market, but the bargaining power of MPI can help to establish requirements, in the form of OMARs, that are more achievable by exporters and more in line with standard practice in

New Zealand. OMARs help to streamline the exporting process, especially for smaller exporters. MPI has implemented a new initiative known as the Market Access Prioritisation Process (NZFSA, 2010d) to "*objectively, transparently and consistently assess requests for significant market access activities*"; those activities that require substantial investments in time, money or expertise, often involving new or novel products or technologies. This work assists MPI with the allocation of the discretionary portion of its market access work programme resources.

Coordinated by directors from the Market Access, Assurances and Imports, and Export Strategy groups, the process involves New Zealand exporters submitting proposals detailing aspects such as:

- The market access problem that they are facing
- What they believe needs to be done to solve the market access problem
- The consequences for their business if their market access application is declined
- Any unique requirements imposed by the market they wish to export to
- Who would benefit from their proposal being accepted
- Suggestions on how these market access activities could be funded
- Any legal, political or trade issues that might prevent their market access work from proceeding
- The specific monetary costs and benefits to their business/industry if their proposal was successful
- Any risks and uncertainties with the information (especially the cost/benefit information) presented

Applications go through a three step decision-making process at MPI:

Stage 1: Assesses the acceptability of proposals, determining whether the proposal is within MPI's area of responsibility and strategic direction. In order for an application to proceed to stage two there must also be sufficient evidence provided of a market access issue or opportunity.

Stage 2: Applications are ranked in a spreadsheet model utilising a semi-quantitative Multi-Criteria Decision Analysis (MCDA) type methodology (Figure 7), where applications are assessed according to a range of weighted criteria, including Industry commitment to proposal, financial benefits, environmental, public health and animal welfare issues and the availability of resources to successfully implement a proposal.

Stage 3: An MPI panel will consider the outputs of the stage two ranking process, as well as additional aspects raised by panel members. This discussion focuses on assessing a proposal's ability to generate short and long term financial benefits and future negotiation leverage, as well as assessing the likelihood of the proposal being successfully implemented. Once the panel has made its decision regarding which proposals to pursue they will publish generic summaries of the applications as well at their rationale for their decision(s).

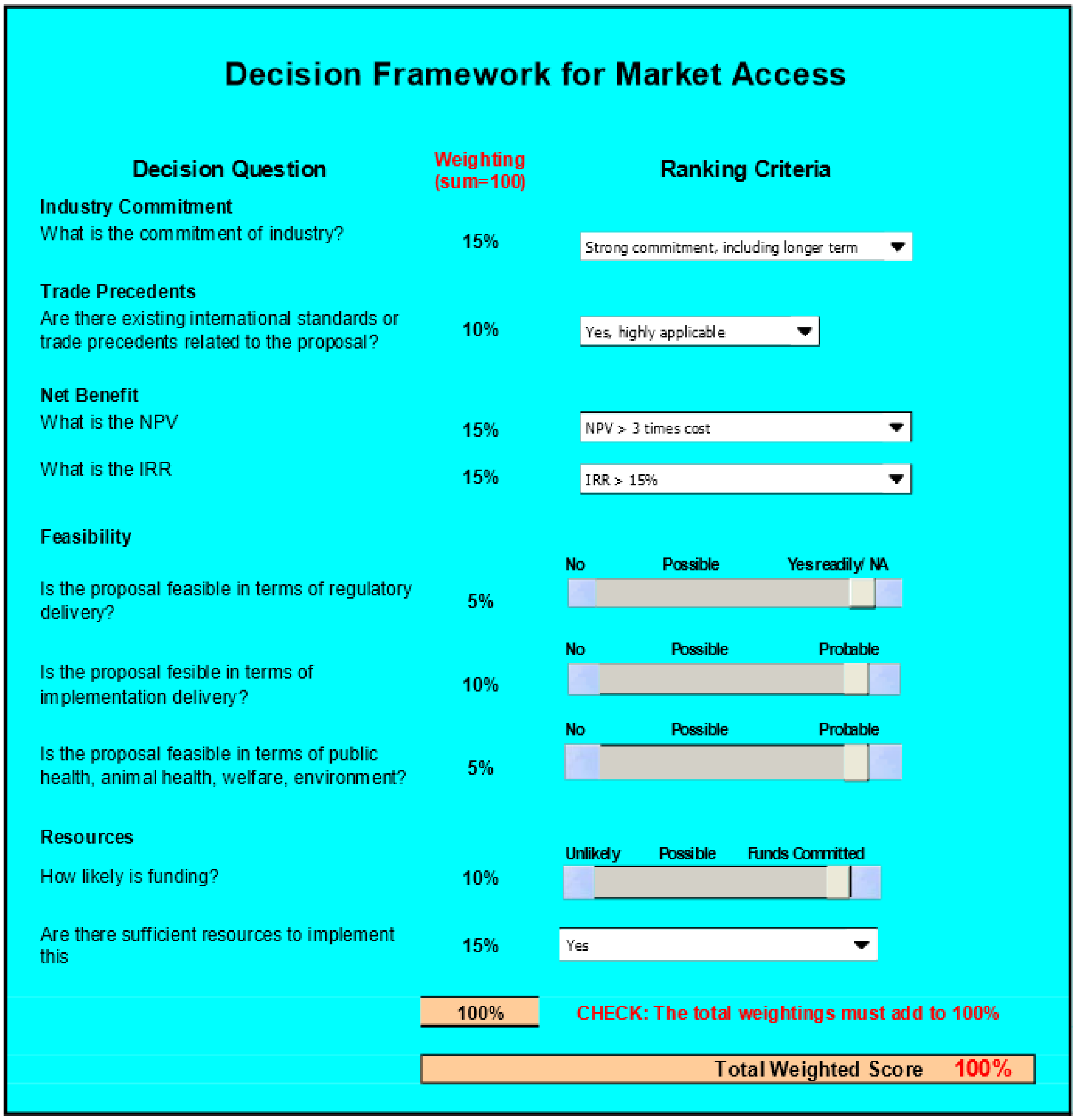


Figure 7: MCDA spreadsheet for market access prioritisation. From (NZFSA, 2010d).

3.7.2 RMF Step two: Identification and selection of risk management options

Key Question: What is the most appropriate option(s) for addressing/managing the food safety issues prioritised in step 1?

This is a second critical decision-making stage in the RMF, and is the focus of this thesis. The last 20 years have seen a shift in the way food safety risks are handled as the government has moved away from a hands-on prescriptive approach to managing food safety risks towards mandating outcomes (see Section 3.3). However, MPI still has the authority to prescribe specific processes or technologies if it deems necessary. MPI generally has three main categories of control measures available to it:

- Prescriptive control measures
- Chemical and microbial limits
- Consumer education campaigns

3.7.2.1 Prescriptive control measures used by MPI

These measures generally specify how certain steps in a farm-to-fork production chain must be conducted or how a particular food safety risk must be managed. These measures can be qualitative or quantitative in nature. They are nearly always designed in collaboration with industry and are generally based on scientific evidence and industry experience. Not all items within a COP are mandatory. The documents are structured to show mandatory requirements, strongly recommended procedures and other guidance material.

Codes of Practice (COPs) are the primary tool used by the MPI for administering prescriptive controls. COPs reflect acceptable industry agreed best practice and essential components of Good Operating Practice (GOP). They are designed to be used by food businesses as a template to develop their own GOP and HACCP based food safety programmes. The mandatory requirements reflect essential components of a food safety system that cannot be reliably achieved by any other means, while the remainder of the COP provides general guidelines and recommendations, allowing some flexibility in the way they are implemented. COPs are not produced for all industries, usually only the more complex or high-risk industries.

An example is the poultry industry Code of Practice for processors (NZFSA, 2007), developed in conjunction with the Poultry Industry Association of New Zealand (PIANZ)[10] in response to the escalating campylobacteriosis epidemic of the mid 2000's (NZFSA, 2008a, 2010a). This code provides general guidelines and recommendations, coupled with mandatory requirements necessary to ensure adequate food safety. The mandatory requirements in the COP detail:

- Procedures for the maintenance and cleaning of equipment
- How dropped or contaminated product is to be disposed of
- Procedures for dealing with birds that are dead-on-arrival and requirements for operators to set acceptable maximum limits for the number of dead-on-arrivals
- Animal welfare requirements during slaughter
- Conditions regarding how long birds are bled for
- Minimum limits for the amount of water to be used during scalding of birds and requirements for how often scald tanks are to be emptied and cleaned
- Biosecurity measures to be taken by workers
- Requirements for microbial testing of carcasses
- Requirements relating to the use of antimicrobial chemicals at different stages in primary processing

Utilising prescriptive controls to address food safety issues is not the preferred approach of MPI, but these measures are sometimes necessary, especially for higher risk industries like poultry, where there have been food safety issues that the industry was unable to completely address through self-regulation or voluntary measures.

---

[10] PIANZ is the organisation representing 99% of the poultry meat producers in New Zealand. PIANZ has the primary goal of promoting and protecting the interests of the New Zealand poultry industry and has the authority to interact with MPI on behalf of its members. The activities of PIANZ include working with the government to formulate food safety standards and Codes of Practice. Website: http://www.pianz.org.nz/about.

3.7.2.2 Chemical and Microbial Limits

MPI has moved towards an 'outcomes based approach' to food safety, where standards are set regarding the allowable levels of contaminants in consumed food, and these must be met by food businesses through their HACCP-based plans. The majority of food safety issues are of a microbial or chemical nature. MPI, in line with internationally agreed best practice, focuses on managing these threats with three main types of standard:

Microbial Limits

Microbial Limits are food-specific microbial levels that are unlawful to exceed. Maintained by MPI (MPI, 2014d), these limits are usually expressed in Colony-Forming Units (CFU)[11] per gram (or ml for liquids) of food.

New Zealand produced or consumed food will need to comply with all relevant microbial limits.

The process for setting limits is closely modelled on Codex guidelines and documents, drawing from the Codex 'Working Principles for Risk Analysis' (FAO, 2007a) and 'Guidelines for the Conduct of Microbial Risk Management' (CAC, 2007b) documents, as well as other relevant documents from Codex, Australia, Britain and the USA.

When setting these limits close attention is paid to core Codex principles, summarised below (CAC, 2007b):

- The protection of human health is the primary reason for setting limits
- Microbial limits must be scientifically justified using *"sound science"*
- Risk management options should be proportionate to the level of risk
- Limits should not be more restrictive of trade or technological innovation than required to achieve the Appropriate Level of Protection (ALOP)

---

[11] CFU is an estimate of viable bacterial or fungal numbers (i.e. cells or groups of cells that are not dead and are able to reproduce to form a new 'colony').

Codex does not produce Microbial Limits, but has begun to release recommended limits for some higher risk food-pathogen combinations such as *Listeria monocytogenes* in ready-to-eat foods (CAC, 2007a). Limits are likely to differ between countries due to differing preferences regarding the risk/benefit trade-offs associated with different limits. The setting of these limits at levels that reflect the ALOP is a key task for food safety authorities.

Maximum Residue Limits (MRLs)

MRLs specify a maximum allowable residue of a particular agrochemical (e.g. pesticide, fertiliser, veterinary medicine or hormone) in a crop or animal product at the point of sale. Usually measured in mg/kg, the milligrams of chemical per kilogram of food, MRLs do not have a toxicological basis, but are set by estimating the maximum possible concentration of an agrochemical that will occur in produce if the agrochemical is used as intended, in accordance with Good Agricultural Practice (GAP) (the farmers' equivalent of GOP) (Shaw, 2013). GAP recommends how and when an agrochemical should be used, and when crops and animals can be harvested in order to ensure that residues are kept to a minimum. GAP protocol usually differs between countries due to the different array of crops and animals farmed, as well as differing environmental conditions (e.g. soil chemistry, weather), resulting in different requirements regarding the use of agrochemicals. The process recommended by Codex for setting GAP is internationally recognised and takes into account differences between countries. Even though GAP may differ between countries or sectors, farmers that develop and follow GAP according to their country's guidelines can be reasonably confident that their produce will not contain residues that exceed international MRLs, provided their domestic food safety authorities are competent and up-to-date. Importing countries have the right to reject produce which exceed MRLs.

Produce imported into New Zealand must comply with one of three sets of MRL standards:

- New Zealand Food Standards (pursuant to the NZ Food Act 1981), administered by the MPI's Agricultural Compounds and Veterinary Medicines (ACVM) group
- Internationally agreed reference standards
- In the case of produce imported from Australia, the Australia New Zealand Food Standards Code (The Code)

New Zealand food exporters must comply with the MRLs set by the importing country. These MRLs may vary from country to country, but are becoming increasingly streamlined.

Maximum Limits (MLs)

Maximum Limits are very similar to MRLs except they refer to environmental and accidental contaminants such as heavy metals rather than agricultural chemicals that are intentionally used as part of the farming process. MLs are produced by Codex but the decision regarding whether to adopt them is conducted by FSANZ. FSANZ will generally adopt Codex MLs unless they feel that they are not appropriate for Australia and New Zealand. In this case FSANZ will set Australia and New Zealand specific MLs in a process very similar to the MRL setting process where, based on toxicological and country specific GAP and consumption data, limits are set so that the estimated daily intake must not exceed the acceptable daily intake.

3.7.2.3 Why are some limits different to Codex limits?

Codex produces international standards relating to the maximum acceptable levels of chemicals in foods. However, sometimes these limits will differ between countries, due to different economic and environmental conditions, farming/processing practices and consumption patterns.

For example, New Zealand's ACVM group sets MRLs for agricultural and veterinary chemicals according to Good Agricultural Practice. In New Zealand there is a different mix of pests and diseases to other countries, as well as different environmental conditions like weather and soil make-up. As a result, New Zealand's Good Agricultural Practice (GAP) will be different from GAP in some other countries. This could mean that some New Zealand farming procedures will vary considerably from those of its trading partners, including the types and doses of agricultural chemicals used. In this case there may be a need to develop New Zealand specific MRLs to reflect these differences.

Also, because the types of foods New Zealand consumers eat are different from those in other countries, maximum levels established by Codex will not always be relevant to New Zealanders. For example Codex has established a maximum level for cyanide in cassava flour, which is a staple food in some countries. New Zealanders (and Australians) by comparison consume very little cassava flour and so FSANZ have not set a ML for cyanide in cassava flour (FSANZ, 2013).

New Zealand is generally accepting of most chemical limits set by Codex and adopts them wherever possible (NZFSA, 2003). New Zealand has been more accepting of Codex limits than some trading partners like the US and EU (WHO/FAO, 2002).

3.7.2.4 Notionally Zero Risk for Chemicals

Microbial contamination of food can occur at almost any stage in the farm-to-fork chain. It is widely accepted that it is practically impossible to eliminate all potentially harmful micro-organisms from all foods, all the time. This is why the idea of acceptable risk is most commonly used when referring to microbial hazards. Accepting that producers will very rarely be able to guarantee that their food is completely free from harmful pathogens, standard practice is to reduce these microbial risks to an acceptable, non-zero level. For many pathogens, only a few cells, particles or spores are needed to cause illness, or to multiply to a level where illness is likely. Completely eliminating the risks from microbial pathogens would require guaranteeing that all food is 100% free of such contamination at the point of consumption, a task which is not practical at present.

In contrast, chemical contamination primarily occurs at the farming stages of production (provided GOP is being adhered to). Chemicals cannot grow or reproduce like microorganisms and so the chemical levels present at harvest are generally the maximum concentrations that will be found in the final food product. Also, unlike microbial hazards, the negative health effects associated with chemical contamination are directly related to the dose. Just because a chemical hazard is present does not mean there is a risk. It is possible to specify a daily level or dose for chemicals below which there is, in theory, a completely zero risk to consumers.

For these reasons most food safety authorities worldwide, including MPI, adopt a *"notionally zero"* risk policy regarding chemical contaminants in food (FAO/WHO, 1995a). Current international risk-management procedures are, in theory, able to practically and reliably eliminate the health risks associated with known chemical contaminants. This is not yet possible for microbial contamination.

3.7.2.5 Food safety interventions aimed at consumers

Consumers are usually the last line of defence when it comes to preventing foodborne illness. The actions taken by the consumer between purchasing and consuming food often have the potential to complement and enhance, or completely reverse risk-management actions taken by at earlier stages in the farm-to-fork chain. This is especially true with regards to microbial contamination of food, where the way food is stored and prepared by the consumer can have implications for the level of risk.

Consumer education campaigns have the potential to be a highly effective and efficient strategy for combating food safety risks by altering consumer behaviour. Consumer education campaigns are generally used for one of three reasons:

- To alter consumer behaviours with respect to the storage, preparation and consumption of higher-risk foods and to increase consumer awareness of real and significant food safety risks. This type of intervention is used where consumers have an opportunity to mitigate a food safety risk themselves but are failing to do so. This could be because consumers are uninformed or apathetic towards a particular risk. An example of such an intervention is the NZFSA and Ministry of Health (MoH) 'Foodsafe Partnership', which aimed to improve consumer hygiene practices regarding the storage and preparation of food by running an annual multi-media campaign during Food Safe week (Winnard, 2008).

- To allay consumer fears regarding perceived rather than real risks (i.e. to dispel food safety 'myths'). Interventions are often used during times when there have been high profile food safety incidents (e.g. food safety scares overseas), which have resulted in consumers overestimating the food safety risks associated with a food they are (or were) eating. An example of this is the MPI response to the 2013 horse meat in beef products scandal in Europe, where the MPI promptly released a fact-sheet on its website providing easy to understand evidence to allay consumer concerns. This type of campaign is sometimes funded by affected industries directly. Interviewees broadly agreed that the media and lobby groups often help to perpetuate these myths or embellish food safety issues because they have an agenda or are sensationalist. Such groups may also have a poor understanding of risk (Shaw, 2013).

Such campaigns may also be utilised to discourage consumers from altering their consumption habits away from healthy foods that are perceived as ‘risky’ (e.g. fish perceived to contain mercury). Often they are not so much intended to improve food safety, but could be considered ‘damage control’, designed to reduce consumer fear and distrust in certain food products or industries and the food safety authorities themselves. They attempt to reduce the economic losses to food businesses due to consumers altering their consumption patterns away from perceived ‘risky’ foods. In some cases these campaigns will be used by public health authorities to help discourage consumers from altering their consumption patterns away from healthy foods.

- In response to food safety ‘emergencies’. If a food safety issue is serious or has the potential to affect many, the authorities can use a variety of media avenues to alert customers of recalled products or other food safety emergencies.

MPI and MoH have educational material on their websites devoted to consumer food safety education. MoH focuses on educational information that is likely to address real food safety issues (e.g. *Campylobacter* in poultry, *Listeria* in ready-to-eat foods) rather than perceived issues (e.g. horse meat, aspartame). The MPI website contains information regarding both real and perceived food safety risks, with the goal of protecting both the health of New Zealanders and the economic interests of the New Zealand food sectors.

Which type of control is most appropriate?

MPI will generally opt to implement measures that result in the minimum amount of intervention necessary to achieve a particular goal. MPI places an emphasis on moving away from traditional prescriptive approaches to food safety towards an outcomes based approach such as the setting of chemical and microbial limits. Education campaigns are used by MPI (and MoH) when the food preparation and consumption stages in the farm-to-fork chain are seen as Critical Control Points (e.g. the Cooking stage in the preparation of poultry meat) which is often the case for microbial hazards. Consumer interventions are also used when consumer misperceptions regarding food safety issues have the potential to negatively affect consumption patterns and/or revenue generated by domestic industries.

3.8 The Role of Science in Decision-Making

The primary goal of MPI is to protect the health of New Zealanders. MPI interviewees repeatedly emphasised that all risk management activities, domestic and global, should be based on *"sound science"* to the greatest extent possible. The use of quality science is essential if regulators are to accurately estimate the risks associated with a certain food safety issue or the risk reductions achievable by risk management actions. MPI conducts a significant amount of science in house, but also draws on a number of other sources, both locally and internationally.

The task of ensuring that MPI decisions are informed by robust and up-to-date science lies with the MPI Food Safety Science Group. This group is able to provide quality, timely and practical scientific and technical advice to MPI decision-makers and the food industry. The group consists of a small number of highly qualified scientists from a range of disciplines. Their activities focus on scientific evaluations and risk assessments (Slorach, 2008).

Additionally, the Science Group is tasked with coordinating a relatively small amount of social science. Typically, such research has focused on discovering consumer understanding of food safety issues such as *Campylobacter* (ESR, 2008) and Folic Acid fortification of bread (Kalafatelia, 2011). MPI also conducts surveys to track the temporal change in consumer perceptions regarding a variety of more general food safety issues (MPI, 2010). MPI are uncertain about how consumers respond to advice and messages based on science, and are looking to increase the amount of consumer-focused social science they conduct (NZFSA, 2009c). They emphasise that future initiatives should still include eliciting consumer perceptions regarding food safety 'issues' such as irradiation, but should also move towards generating an understanding of the perceptions of farmers and processors, as well as an understanding of current food hygiene and handling practices by all players in the farm-to-fork chain (NZFSA, 2009c).

The small size of the Science Group means that MPI's demand for scientific expertise cannot be met by the MPI Science Group alone. MPI has opted to source a considerable amount of scientific support from external agencies such as the Institute of Environmental Science and Research (ESR) and Massey University, whose activities involve conducting detailed risk-assessments, attribution modelling and other specialised tasks. Other external science groups that MPI liaises with include AgResearch, NIWA, MoH, FSANZ and the Codex Commission.

3.8.1 Assessing the evidence

When MPI is assessing risk and evaluating risk management options there is often insufficient evidence to make an absolute judgement regarding the 'best' option. Sometimes MPI will be faced with an array of different and conflicting pieces of scientific information. Each study or piece of science is considered an individual piece of evidence, and MPI sees it as its job to amass, sort and assess the validity of each piece of evidence in order to synthesise all the information into a scientifically justifiable decision.

MPI has adopted an evidence-based approach for systematically reviewing and assessing the quality of existing research (NZFSA, 2011). This is similar to the Evidence Based Medicine approach utilised by the highly regarded Cochrane Collaboration (CCNET, 2012).

The MPI approach to reviewing and grading evidence is modelled closely on guidelines released by the American Agency for Healthcare Research and Quality (AHRQ) (NZFSA, 2011). This requires the systematic review of evidence in a reproducible manner, and ensures this by utilising two key systems:

1. For grading the quality of an individual piece of evidence

2. For rating the "*strength*" of an entire body of evidence

Grading the quality of individual pieces of evidence

Evidence comes in a variety of forms, some of higher quality than others. MPI has adopted the concept of a hierarchy of evidence proposed by Petticrew and Roberts (Petticrew & Roberts, 2003) and recommended by the American Agency for Healthcare Research and Quality (AHRQ, 2002). This hierarchy grades evidence from highest to lowest quality as follows:

1. Meta-analyses and systematic reviews

2. Randomised Control Trials

3. Non-randomised trials

4. Cohort or case-control studies
5. Cross-sectional studies
6. Case series
7. Individual case reports

In cases where there is an extreme shortage of evidence MPI may consider what it calls *"anecdotal evidence"* such as non-clinical case reports and even lower quality forms of evidence such as animal studies, opinions, editorials and In vitro research.

MPI sources evidence from published and peer reviewed literature, preferring to use evidence from leading global food safety and health agencies such as:

- The World Health Organization (WHO)
- The Food and Agriculture Organization (FAO)
- European Food Safety Authority (EFSA)
- United States Food and Drug Administration (FDA)
- United States Environmental Protection Agency (EPA)
- Food Standards Australia New Zealand (FSANZ)
- United Kingdom Food Standards Agency
- International Agency for Research on Cancer (IARC)

MPI generally presumes that any assessments conducted by these organisations constitute systematic reviews or studies of methodological quality equivalent to systematic reviews, the highest level of evidence available under the Hierarchy of Evidence (Harris, 2001).

Most systematic reviews and meta-analyses will allocate scores to these different types of evidence, or apply some form of 'quality filter', only accepting evidence above a certain level in the evidence hierarchy. The higher the piece of evidence sits in the hierarchy the greater the weight it is allocated in the final conclusion.

Grading the quality of an entire body of evidence

There are three main criteria against which MPI assesses the quality of a body of evidence:

1. The aggregated methodological quality of the individual studies. For example, ten randomised control trials represent a better quality body of evidence than ten case studies.

2. Quantity of the individual studies: This assesses the number of studies of each type, the sample sizes used and the effect sizes of each study.

3. The consistency between studies: A group of studies whose results are relatively consistent will have considerably greater 'strength' than a body of evidence with disparate results.

The specific mathematical procedures for implementing these principles may differ between reviews, but the general procedure of reaching conclusions by scrutinising the evidence on these criteria will be consistent across systematic reviews conducted by, and for, MPI food safety. The above process is very similar to the process of formal Meta-Analysis, which is one of MPI's preferred sources of evidence. Interviews with food safety academics and regulators indicated that MPI's preference is to systematically and independently consider all of the empirical studies surrounding a particular topic rather than the conclusions reached by individual members of the scientific community. MPI has its own experts who make judgements and recommendations to decision-makers once they have assessed the evidence themselves. These decision-makers have discretionary decision-making power. Several farming and industry representatives interviewed felt that MPI's decision-making processes could be more transparent, as they did not always understand MPI's rationale behind some of its decisions.

3.8.2 The Precautionary Principle

The precautionary principle, or precautionary approach, gives decision-makers the right to err on the side of caution in the absence of a scientific consensus regarding a food safety issue. The principle allows policy makers to make discretionary decisions in situations where there is the possibility of harm from taking a particular course of action or making a certain decision when conclusive science on the matter is lacking. It implies that there is a social responsibility to protect the public from exposure to harm when scientific investigation has found a possible but inconclusive link between a food and an adequately serious health condition. These protections can be relaxed if further scientific findings emerge that provide sound evidence that no harm will result (EU, 2000).

The precautionary principle is mentioned in the Risk Management Framework (NZFSA, 2010b) and has also been officially incorporated into specific legislation such as the Hazardous Substances and New Organisms Act 1996 (MfE, 1996). However, MPI acknowledges its obligation to exhaust all scientific avenues before employing the precautionary principle, usually as only a temporary solution while gaps in the scientific knowledge are filled (NZFSA, 2010b). MPI does require that food businesses follow a precautionary approach when dealing with potential food recalls and food safety incidents (MPI, 2013a), and especially when it comes to matters of biosecurity (e.g. preventing the spread of pests and diseases) (MPI, 2008). There are limited documented cases of MPI's large scale implementation of the precautionary principle to food safety issues. However, one such example given by MPI is its highly conservative stance regarding *E. coli* O157 standards for exported meat, stating that *"a single detection in exported product might trigger a worst-case reaction from trading partners"* (NZFSA, 2010b).

In a way the precautionary principle is at odds with MPI's core principle of science-based decision-making. Interviews with MPI employees (and other stakeholders) indicated that, although stakeholders understand the rationale for using the principle and concede that it serves as a sometimes necessary tool for managing food safety risks in an environment where there is often insufficient science, it can also be a considerable source of frustration for exporting nations like New Zealand. The use of the precautionary principle by the European Union, where the application of the precautionary principle is a statutory requirement (EU, 2000), was cited by meat exporters and regulators as a considerable frustration. A common example given was the EU's blanket ban on all hormone treated beef from all countries in the 1990's, which resulted in disruption to the New Zealand beef export industry.

MPI acknowledges the benefits of having the precautionary principle in its toolbox, but disagrees with its use as an unnecessary and unjustified international trade barrier (MPI, 2013e).

3.9 Consultation

MPI welcomes feedback from anyone regarding any food safety issue, but often only actively seeks consultation when the choice of risk management option to address a particular issue is not clear, or when a proposed action by the MPI may have significant or unintended consequences for a particular food sector, industry or stakeholder group.

MPI generally seeks consultation on specific issues by releasing consultation papers, available on the MPI food safety website (MPI, 2013c). These documents will usually detail the food safety issue requiring action as well as a proposed solution. They will outline the scientific basis for MPI's proposed action, with references.

Submissions can include any form of feedback regarding MPI's proposed actions, whether it is positive, negative or neutral. There is no standard time period given for submissions, but MPI generally allows stakeholders 1-6 weeks to make submissions. Submissions are then collated, classified and analysed by the appropriate working group(s) within MPI. MPI will then often release an *"analysis of submissions"* document on their website, outlining the different stakeholder submissions and providing responses and explanations or scientific justification for such responses.

Industry engagement in the consultation process is often, but not always, coordinated through the many New Zealand industry representative groups such as the Poultry Industry Association of New Zealand (PIANZ), the Seafood Industry Council (SeaFIC), the Meat Industry Association (MIA) and the Plant Market Access Council (PMAC). These groups represent the vast majority of food businesses in their sector and have been given authority by their members to engage with the government on their behalf. In addition to releasing consultation documents, MPI will also hold formal meetings with industry regarding issues that it deems important and industry specific. MPI sees these meetings as highly important, aiding MPI in maintaining robust and transparent processes for the sharing of information between regulators and industry and the development of practical risk management options. These meetings are held as often as necessary, with meetings occurring more frequently during times when there are significant, industry specific food safety issues that need to be addressed. For example, MPI met with the poultry industry twice in 2007, at the height of the campylobacteriosis epidemic, with the goal of developing new microbial limits for

Campylobacter in poultry, improved practices for the sharing of scientific information, and to develop and refine the wording of the updated poultry processing Code of Practice. An MPI interviewee mentioned that *"it is highly important that regulators and industry stay on the same page when it comes to understanding and managing food safety risks"*.

MPI conducts surveys regarding consumer attitudes and behaviours to gain a better understanding of consumers and plan more effective initiatives for protecting consumer health. These surveys include assessing consumer attitudes regarding a range of contemporary food safety topics at different points in time, as well as surveying consumer food safety practices such as hand-washing, food preparation and salt intake.

The MPI endeavours to utilise and incorporate stakeholder feedback into its decisions wherever possible, but has no obligation to do so. Little is published about how stakeholder feedback is incorporated into the decision-making process.

3.10 Conclusion

With a very strong focus on the scientifically based principles of risk analysis and management, MPI is considered a world leader in regards to food safety regulation. MPI generally sets performance standards and targets for food businesses, allowing a degree of flexibility in the way businesses manage food risks, rather than mandating prescriptive control measures. The MPI Risk Management Framework is a comprehensive yet flexible framework used for addressing a range of food safety issues. MPI's preference is to base its decisions on the highest quality and most up-to-date science, sourcing much of its science from reputable domestic and international food safety organisations. MPI also actively cooperates with international organisations and seeks to influence international standards by providing high-level science in areas of significance to New Zealand, such as meat and dairy research. MPI has limited social science capacity but has plans to increase the amount of social science they conduct. Little is known about consumers and how they respond to science based food safety messages, and this can sometimes be a barrier to the successful implementation of food safety interventions, especially those intended to alter consumer behaviour. Food safety decision-making at MPI is a flexible process, with ultimate decision-making power lying with the food safety minister. MPI considers information from a range of sources during decision-making, but remain flexible regarding how this information is utilised in decision-making. MPI’s Risk Management Framework is a comprehensive and robust framework for the evaluation of food safety interventions. MCDA is not new to MPI, but does not appear to be utilised for intervention evaluation. MCDA may be able to enhance the intervention evaluation process by following a structured approach for the systematic and transparent consideration of a wide range of criteria.

Chapter 4: Overview of MCDA and MCDA Models

4.1 Introduction

Multi-Criteria Decision Analysis is a methodology designed for evaluating a set of competing options on multiple, often conflicting, criteria. This chapter focuses on describing the general MCDA methodology before outlining some of the methods available for conducting the different steps of MCDA. Although the focus is on MCDA, this chapter begins with a description of Cost-Effectiveness Analysis (CEA), the main alternative to MCDA in the field of HTA[12] and food safety intervention evaluation. This familiarises the reader with the basic process of CEA and how it differs from MCDA, as both methodologies are discussed in detail in later chapters.

4.2 Methodology

This chapter seeks to address research questions 4 and 5:

Research Question 4: What is Multi-Criteria Decision Analysis and how can it improve decision-making?

Research Question 5: What are the different types of MCDA model and what are their key differences?

The internet, university research repositories and the databases Google scholar, Science Direct, Medline and PubMed were searched during the first half of 2013 for academic articles that focus on describing and explaining the MCDA process and the specific models that MCDA encompasses. This literature search also sought literature describing Cost-

[12] Health Technology Assessment is a multi-disciplinary field of policy analysis that studies the medical, social, ethical, and economic implications of development, diffusion, and use of health technology (INAHTA, 2014)

Effectiveness Analysis (CEA), especially in the context of food safety decision-making. An emphasis was placed on comprehensive research papers and guidance documents produced by government organisations and research institutions. The review of this literature and the process of writing this chapter guided me in selecting a small group of MCDA models that may be appropriate for food safety decision-making in New Zealand. It is these models which are constructed, tested and critiqued in Chapter 8. This chapter has a strong focus on MCDA but begins with an overview of CEA; a methodology widely used in health technology assessment and food-safety intervention evaluation, and often seen as an alternative to MCDA (Claxton, 2013).

4.3 Cost-Effectiveness Analysis

Cost-effectiveness analysis (CEA) is a type of economic analysis used to compare the relative efficiency of two or more courses of action. CEA is different from Cost-Benefit Analysis (CBA) because it allows the expression of effectiveness in natural units rather than a monetary value. CEA is often used in the field of health services, where it may be undesirable to monetize health effects. In CEA a ratio is calculated where the denominator is a measure of the gain in health (e.g. years of life, cases prevented, DALYs/QALYs) and the numerator is the cost associated with this health gain.

$$CER_i = \frac{Cost\ (\$)_i}{Effectiveness_i}$$

The Cost-Effectiveness Ratio (CER) for intervention i is calculated as the dollar cost associated with the implementation and maintenance of intervention i divided by its estimated effectiveness at reducing the burden of disease.

The CER represents the cost required to achieve one 'unit' of effectiveness, where, in the context of food safety intervention evaluation, a 'unit' could represent:

- 1 prevented DALY (see Section 6.3.2)
- 1 QALY gained (see Section 6.3.2)
- 1 death prevented

- 1 case of illness prevented
- A percentage reduction in bacterial numbers on food
- A percentage reduction in the prevalence of a disease in a human or animal population
- Many more metrics for assessing effectiveness

In the context of Health Technology Assessment, a commonly used outcome measure is Disability or Quality-Adjusted-Life-Years (DALYS or QALYs). Some refer to this type of analysis, where effectiveness is measured in Health-Adjusted-Life-Years, as Cost-Utility Analysis (Mangen, de Wit, et al., 2007; PHARMAC, 2012). Cost-effectiveness analyses are often visualized on a cost-effectiveness plane consisting of four-quadrants (Figure 8). Typically, outcomes plotted in Quadrant I are more effective and more expensive; those in Quadrant II are more effective and less expensive; those in Quadrant III are less effective and less expensive and outcomes in Quadrant IV are less effective and more expensive. Many organisations consider an intervention cost-effective if the cost-effectiveness ratio is less than a particular figure, often termed the cost-effectiveness threshold (dotted line in Figure 8) (Metcalfe et al., 2012).

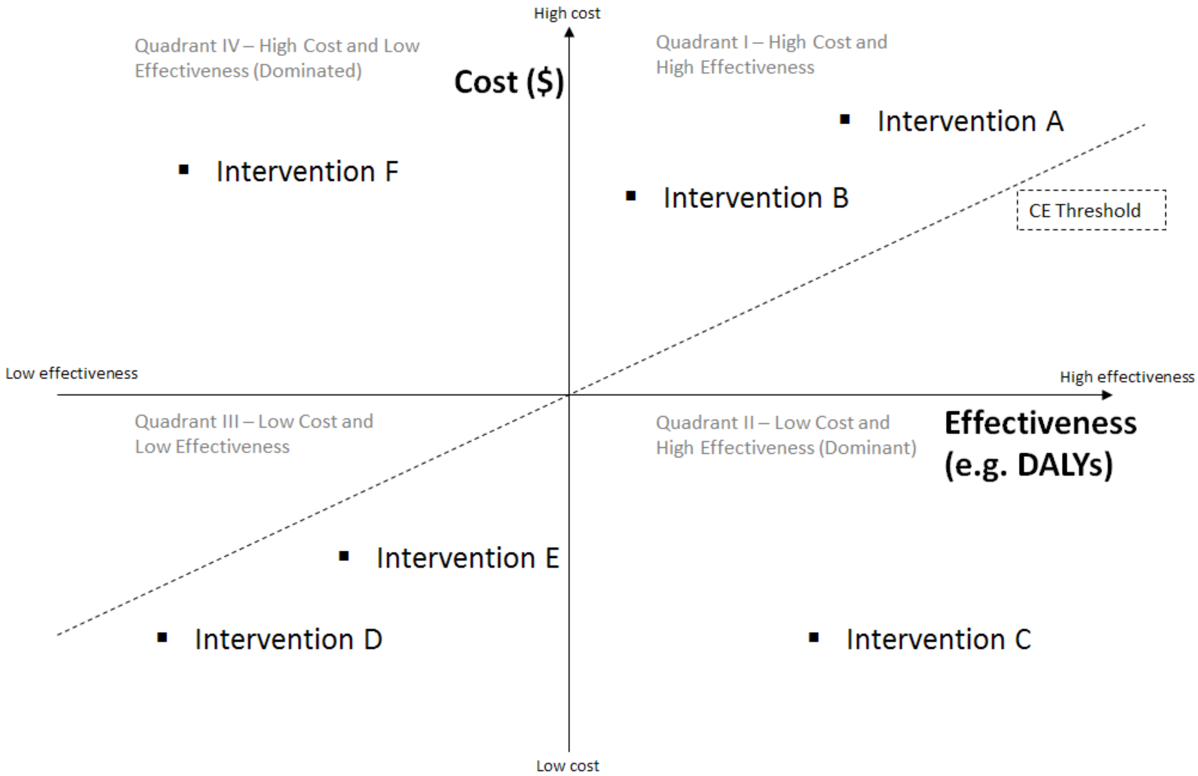


Figure 8: Hypothetical Cost-effectiveness plane. Interventions C, D and E all have CERs that are under the Cost-Effectiveness threshold (dotted line), Intervention C is the most efficient due to its high performance and low cost. Although interventions A and B are relatively effective, their high costs result in high CERs; cost-ineffective if applying the Cost-Effectiveness threshold. Intervention F has a high cost and low effectiveness, making it the least efficient, and dominated by all other interventions when assessed according to their CERs.

There are no official cost-effectiveness thresholds in New Zealand regarding the evaluation of health interventions. In Australia a cost-effectiveness threshold of approximately $50,000 per DALY is used as a rough guide for preventive health interventions (Lopert, 2009; Vos et al., 2010). In New Zealand, government health agencies such as PHARMAC and the Ministry of Health frequently use CEA to support decision-making (MOH, 2014; PHARMAC, 2012). PHARMAC sees CEA (CUA) as a useful decision-support tool, producing an official CUA guide for use in Pharmacoeconomic analysis (PHARMAC, 2012). However, PHARMAC emphasise that cost-effectiveness is only one of a number of important criteria that they base decisions on (PHARMAC, 2014).

4.4 Multi-Criteria-Decision-Analysis

MCDA lends itself well to situations where many criteria have a role in guiding decision-making and these criteria are in conflict with each other. Food safety decision-making is an example of such a situation where there is essentially a need to maximise the level of food safety for consumers while minimising financial costs and other negative impacts associated with interventions or regulations. The objective of MCDA is to improve decision-outcomes by allowing participants to explore the solution space of a problem in an interactive and systematic way.

The key features of MCDA are:

- The identification of options to be compared
- Determining a set of criteria to evaluate the options against. This includes deciding how performance on these criteria will be measured or assessed
- Assessing or measuring the performance of each option on each criterion. This is known as generating 'performance scores'
- Weighting the criteria according to their importance
- Aggregating weighted performance scores for each option and determining the 'best' option(s)

MCDA is theoretically founded on the principles of utility theory, where an option is preferred if its expected utility is greater than those of its competitors (Raiffa & Keeney, 1976). MCDA methodologies are able to systematically integrate heterogeneous information, measured on a range of technical, economic, social and political criteria. MCDA models are used in a myriad of decision contexts (Behzadian et al., 2010; Linkov et al., 2007; Vaidya & Kumar, 2006) and are intended to assist human decision-makers with systematically, consistently and transparently evaluating competing options when there is a large amount of complex information to consider.

All MCDA models require the explicit consideration of the various options and their contribution to a common goal(s), assessed on a clearly defined set of criteria. Most formal

MCDA techniques also utilise an explicit weighting system representing the relative decision-maker or stakeholder preferences for the different criteria.

There is a large and growing body of MCDA techniques, primarily due to the many different decision-making situations that have been identified as appropriate for MCDA, as well as differences in the time and data availability, analytical expertise and administrative requirements in various decision contexts (Guitouni & Martel, 1998). Differences between these methods are mostly due to variation in the way that data is aggregated to provide an overall indication of the performance of the different options.

A key feature of MCDA is its reliance on the judgements of decision-makers/stakeholders in establishing objectives and criteria, estimating the relative importance of the different criteria and, in some cases, estimating the performances of the options on the different evaluation criteria. Although MCDA applications often incorporate quantitative data, there is always a subjective element involved with the establishment of criterion weights, and even the selection of MCDA method to use.

4.4.1 Main Advantages identified in the literature

The most commonly stated advantage of MCDA is its ability to incorporate criteria which are unable to be included in techniques such as CBA due to difficulties with their quantification and/or monetisation. MCDA can potentially allow a more comprehensive analysis of a decision-making situation from a range of perspectives for this reason.

MCDA is designed to be an iterative and highly participative approach, with an emphasis on discussion and information sharing; although this may not always happen in practice. Promoting discussion and deep-thinking regarding many aspects of the analysis, through the MCDA process, can lead to a greater understanding of a problem, and thus improvements to the way it is formulated, modelled and solved.

The formal mathematical processes behind the modelling of a decision-making situation allow for the provision of an 'audit trail' (Bruggemann & Carlsen, 2012; Claxton, 2013; Fischer, 2009); a record of the decision-making process. This trail can be used as a tool for the explicit communication of a problem, its potential solutions, and, in particular, the rationale behind the decision-makers choice of solution. The complexity of this trail will depend on the specific MCDA model used.

Although MCDA is sometimes regarded as having a greater degree of subjectivity than techniques such as CBA and CEA, input data such as criterion scores and weights can be modified during sensitivity and variability analysis to help identify options which perform particularly well or poorly under a range of model parameters, thus providing decision-makers with greater confidence around the model outputs.

4.5 Steps in MCDA

MCDA can be described as an eight step process:

1. **Establish decision context:** This involves clearly defining the problem and the principal objectives of the analysis, as well as identifying the key stakeholders. As new stakeholders are identified they can also provide input. This step outlines the purpose of the analysis and the nature of the outputs. Is the primary objective of the analysis to explore the preferences of stakeholders and gather/share information? Or to produce a definitive 'best' option(s), shortlist or ranking?

2. **Identify the alternatives to be analysed:** The list of alternatives may be an output of considerable discussion and information sharing between stakeholders. Ideally, any stakeholder is able to suggest an alternative. This step may include an initial screening process, where options which are infeasible or unanimously unacceptable are culled. Ultimately the decision-maker(s) will have control over which alternatives are included in the formal analysis. In some cases those commissioning the analysis may already have determined the set of alternatives to be analysed.

3. **Identify the criteria against which the alternatives will be appraised:** Collectively these criteria will need to adequately assess the overall performance of each option. They should be able to capture all relevant positive and negative effects of each alternative.

4. **Assess the performance of the alternatives on each criterion:** Performance 'scores' can be qualitative 'scores' or quantitative measurements. Ideally, stakeholders will agree on the values of the performance scores, but this is not always possible, and there are procedures for coping with this situation (see Section 4.8).

5. **Allocate weights to each criterion, reflecting their relative importance in the final decision:** There are many ways this can be done (detailed in Section 4.7). Most MCDA models are not prescriptive regarding the elicitation of weights. This step in the analysis is often a source of disagreement between stakeholders (DCLG, 2009).

6. **Calculate overall weighted scores (global scores) for each option:** Often referred to as the Multi-Criteria Aggregation Procedure (MCAP), this is the step that separates most MCDA models, and there are a large number of such procedures. By this stage participants must have made a decision regarding the specific MCDA model they are going to use, which is likely to be a difficult decision without guidance from someone familiar with the different types of MCDA.

7. **Examine results:** This is a crucial step in any MCDA. Participants should never blindly accept the outputs of their model. Time should be taken to understand exactly why a model has produced the results that it has. This step should involve significant discussion regarding the interpretation of the results and whether they make sense to participants and decision-makers. The transparency offered by many MCDA models can assist this process, and the provision of an 'audit trail' can allow users to reverse engineer model outputs. Users may examine the underlying logic behind their model and this will assist either in its validation or in highlighting flaws in the model/model selection. At this point a decision might be made to redo the analysis with new options, criteria, scores or weights.

8. **Conduct sensitivity analysis on the results:** This involves altering model inputs such as criterion weights or performance scores to see if model outputs (i.e. rankings) are significantly affected. Stakeholders can discuss any uncertainties or voice any disagreements they have with current model input values, and these can be tested to determine whether such uncertainties or disagreements will significantly affect the outputs of the model. This stage can identify inputs to which the model is insensitive so that further analysis can focus on those inputs which do have potential to affect the model recommendations.

When identifying criteria (Step 3) it is preferable that decision-makers define criteria for which agreeable measurements can be determined or estimated, either qualitatively or quantitatively. These criteria should exhibit a number of properties (Bartolini & Viaggi, 2010; de Montis et al., 2005; Donoso, 2008; Raiffa & Keeney, 1976):

Completeness (or exhaustiveness): A set of criteria is complete if it is sufficient in indicating the degree to which the primary objectives, as defined by participants and/or decision-makers, are met. Incomplete criteria lists mean that the modelling process will be unable to consider potential impacts of the different alternatives that are relevant to decision-making.

Operational: Because MCDA is intended to assist decision-makers in choosing the best course of action, the choice of criteria must be useful for this purpose (Raiffa & Keeney, 1976). Each of the criteria, along with its proposed unit of 'measurement', should be relevant to decision-making and well understood by participants. The criteria included should be able to communicate the strengths and weaknesses of the alternatives to other stakeholders.

No double counting: Some refer to this property as *"non-redundancy"* (Raiffa & Keeney, 1976; Tony et al., 2010), but others have a different interpretation of non-redundancy (see below) (Donoso, 2008). There should not be any overlap between two different criteria. That is, two criteria should not be measuring the same thing in whole or in part.

Non-redundancy: If all options show the same level of performance on a particular criterion then that criterion will be redundant in the sense that it will not aid the decision-maker(s) in deciding between the different alternatives.

Mutual Independence of preferences: The performance of an alternative on one criterion should not be dependent on its performance on any other. Stakeholders should be able to allocate scores to an alternative on a particular criterion without requiring knowledge of its performance on any other criterion.

Minimal Size: It is desirable to keep the number of criteria as small as practically possible. This reduces data requirements and the 'elicitation burden' on the stakeholders/decision-makers (Karvetski et al., 2011; Montibeller et al., 2008).

The process of establishing objectives and criteria is a critical step in MCDA. It should be an inclusive and explicit process, with sufficient time and flexibility to allow for modifications if decision-makers feel necessary. Having identified the relevant criteria in a decision-making situation, measuring or estimating the performance of each alternative on each criterion is another crucial step. Sometimes there will be quantitative data or modelling techniques available for this task. An overview of potential metrics and models in the context of food safety intervention decision-making is covered in Chapter 6. When there is no accepted metric available for the quantification of performance on a criterion, or there are insufficient data to allow for reliable quantification, a value-judgement must be made. Many real-world applications of MCDA incorporate a combination of quantitative data and expert opinion. An advantage of MCDA over informal judgement is that criterion scores or measurements can be generated by specialists (e.g. lawyers, accountants and toxicologists) and not necessarily decision-makers themselves, who may not be experts in all aspects of a problem. This is useful when decision-makers have limited time, resources or technical expertise. This separation between decision-makers and those generating performance scores can help reduce bias and can provide decision-makers with more objective data to base their decisions on than if they had estimated performance scores themselves.

The final step before performance scores are aggregated is usually to assign weights to the various criteria reflecting their relative importance in the decision outcome. This weighting of criteria is perhaps the greatest source of uncertainty, primarily because of its subjectivity, as decision-makers assess the relative value of a criterion based on individual preferences and values (Donoso, 2008; Spackman et al., 2000). Section 4.7 contains a review of common methods for weighting criteria.

Once criteria scores and weights have been determined it is common to present this data in a performance matrix.

4.5.1 The Performance Matrix

A feature common to virtually all MCDA models is the performance matrix (Spackman et al., 2000). A performance matrix contains all information regarding the performance of each option on each criterion and, when relevant, the weights that have been allocated to the different criteria. Performance scores are often numerically expressed (see the example in Figure 10), but can also be expressed qualitatively as verbal ratings (e.g. low, medium, high), or as a binary variable (yes/no).

		Option A		Option B		Option C	
Criteria	Criteria Weights	Score	Weight x Score	Score	Weight x Score	Score	Weight x Score
Criterion 1	1	3	3	3	3	3	3
Criterion 2	2	2	4	1	2	2	4
Criterion 3	3	1	3	3	9	2	6
Total			**10**		**14**		**13**

Figure 9: A simple performance matrix

Figure 9 shows an example of a very basic performance matrix where three competing options (columns) are being assessed on three criteria (rows), with all performance scores and weights numerically expressed. An aggregated (global) performance score for each option is found by multiplying individual performance scores on each criterion by the corresponding criterion weight and summing these numbers. The 'best' options are identified as those with the highest global score. In the example above Option B achieves the highest global score and is the most preferred option according to the model. This example utilises one of the simplest types of MCDA model, the Linear Additive Model, where all performance scores are combined into a single number, the global performance score, which is used to rank or group options. Many MCDA models utilise more complex procedures to generate results. The main types of model are described in Section 4.6.

Some methods require that all performance scores be expressed on a consistent, homogeneous scale (e.g. 0 to 100). This is the case for Linear Additive Models, where the expected utilities (scores) are aggregated across all criteria into a global score known as a Single Synthesis Criterion (SSC) (Guitouni & Martel, 1998). For these methods all criteria scores, often measured using a diverse array of metrics, are transformed into a dimensionless number and mapped onto normalised homogeneous scales (see Weighted Sum Model examples in Sections 4.6.2 and 8.5.1). These homogeneous scales create a 'level playing field'; ensuring that all criteria are treated equally prior to criteria weighting. This remapping of scores or measurements onto homogeneous scales allows for the meaningful integration of formal measurements with qualitatively estimated criterion 'scores' (e.g. data from Likert scales can be combined with cost and effectiveness 'data'). Methods that do not produce a Single Synthesis Criterion do not require this remapping of criteria performances onto homogeneous scales prior to aggregation, and will allow for the aggregation of criteria performances measured on heterogeneous scales, including natural units. This is an advantage of these methods, although the mathematics behind their aggregation procedures is often more complex (see the PROMETHEE example in Section 8.5.3)

4.6 Common MCDA Models

4.6.1 Multi-attribute utility theory (MAUT)

Multi-Attribute Utility Theory (MAUT), based on the work of von Neumann and Morgenstern, Savage and Keeny and Raiffa (Von Neumann & Morgenstern, 1947, Savage, 1954, Raiffa & Keeney, 1976), seeks to generate a mathematical function that estimates the overall utility *(U)* of each option in a way that is consistent with the fundamental axioms of rational choice[13] (Tversky & Kahneman, 1986). MAUT assumes that "*there exists a real value function U defined on the set A of feasible actions, which the decision maker wishes, consciously or not, to examine*" (Munda, 1995). *U* essentially represents the overall utility of an option, calculated as the sum of the utilities across all criteria.

MAUT assumes that there is a range of competing options to choose from, that only one of them must be chosen now, and due to uncertainty about what the future holds, different options will have potentially different utilities to the decision-maker(s), dependent on the future state. For example, the utility associated with investing in a new irradiation plant to eliminate foodborne pathogens from meat and poultry products may depend on a number of factors, including the future consumer acceptance of such a technology and the future energy costs required to run such a plant.

The expected utility of any option is estimated by:

1. Identifying all possible future states (j) that could be seen as relevant in a particular decision context, and generating an estimate of the probability (pr) of each future state occurring

2. Estimating the utility (u_{ij}) of each combination of option i and future state j

---

[13] This type of decision research is referred to as Normative, where a focus is placed on how rational individuals *should* make decisions rather than how individuals *do* make decisions (descriptive research). Human beings do not always act in a rational manner.

3. Generating the global performance scores; the probability weighted average of all the outcome utilities for each option i:

$$U_i = \sum_{j=1}^{n} pr_j u_{ij}$$

Where:

U_i is the overall utility (global score) associated with option i;

u_{ij} is the estimated utility of option i if future state j occurs;

pr_j is the estimated probability of future state j occurring.

MAUT is classified as a type of MCDA method known as 'compensatory' because good performance on one criterion can fully compensate for poor performance on another (Linkov et al., 2006).

By formally incorporating uncertainty into model building from the outset by attempting to account for all possible future eventualities, MAUT can be a potentially demanding method to use. Users must assign probabilities and utilities to the range of outcomes that may occur following implementation of any of the options. Thus, there can be very large data requirements for the calculation of U in MAUT, which potentially makes it inappropriate for all but the largest of projects where there is significant uncertainty (as determined by decision-makers) and the consequences of a 'bad' decision especially severe. MAUT has been widely applied to decisions involving the construction or decommissioning of nuclear facilities, the reprocessing and storage of nuclear waste (Kim et al., 2007) and nuclear emergency management (Papamichail & French, 2013), although model recommendations have not always been accepted by those commissioning the analyses (DCLG, 2009).

MAUT assumes that individuals act in accordance with the axioms of rational human behaviour, which may not necessarily be true. Such assumptions may be valid when analysing problems with a minimal 'human element', (such as the handling of nuclear materials (Spackman et al., 2000)). However, from a food safety perspective, there is generally a large human element, most notably the consumer perception of risk-management options. There is a body of research detailing how consumers can behave in a seemingly irrational manner (Bruhn, 1998; Cox & Evans, 2008; Draper & Green, 2002; Dunphy & Herbig, 1995; Rimal et al., 2001; Siegrist, 2008; Verbeke et al., 2007; Yeung & Morris, 2001). MAUT may not be appropriate for analysing problems where key stakeholder groups do not conform to these axioms.

4.6.2 Weighted Sum Model (WSM)

The Weighted Sum Model is a straightforward MCDA model, and probably the most popular (Bartolini & Viaggi, 2010). Its transparency and intuitive appeal make it an attractive choice for users wishing to familiarise themselves with MCDA and its key procedures. A feature of this model is that all criterion scores for a particular option are combined into a single number, a global performance score (P), which is the sum of criterion performance scores (p) multiplied by their respective criterion weights (w):

$$P_i = \sum_{j=1}^{n} w_j p_{ij}$$

Where:

P_i is the global performance score associated with option i

p_{ij} is the score of option i on criterion j

w_j is the weight allocated to criterion j

This type of model is the most widely utilised and has a proven track record of generating reliable and effective recommendations under a range of decision-making contexts (Mabin, 2006).

There are two key inputs in the weighted sum model; criterion weights and performance scores. Performance scores can be quantitatively measured using existing metrics, or qualitatively assessed. Criteria weights can be generated in a number of ways (see Section 4.7).

The WSM requires the scoring of all options on homogeneous interval scales so that performances can be aggregated across criteria. One possibility is global scaling, which assigns a score of 0 to a hypothetical worst level of performance and 100 to a hypothetical best possible level of performance. Options are then allocated scores between 0 and 100 reflecting their performance relative to the two hypothetical end-points. Another option, local scaling, associates a score of 0 with the performance level of the worst performing option in the currently considered set of options, and 100 with that which performs best. Again, the remaining options are scored relative to these two extreme points. This process is completed for all criteria so that all performances are represented as numbers on homogeneous 0-100 interval scales.

4.6.3 Analytic Hierarchy Process (AHP)

A very popular method, the Analytic Hierarchy Process also operates as a linear additive model, where global performance scores for each option are calculated as the sum of the weighted criterion scores. Originally developed by Thomas Saaty (Saaty, 1980), AHP attempts to derive criterion scores and weights based on pairwise comparisons between criteria and the various options. Decision-makers are required to assess the importance of each criterion or option relative to every other.

The fundamental inputs into AHP consist of the participant's answers to a series of general questions taking the form of 'How much do you prefer Option A over Option B on criterion X' and 'How important is criterion X compared to criterion Y'. The decision-maker(s) must answer these questions for every possible pair of interventions on every criterion, and every possible pair of criteria, to generate performance scores and criterion weights. Participants are usually asked to express their preferences on a nine-point semantic scale (Saaty, 2001) (Table 2).

Rating	Description
1.00	Equally Preferred
2.00	
3.00	Moderately Preferred
4.00	
5.00	Strongly Preferred
6.00	
7.00	Very Strongly Preferred
8.00	
9.00	Extremely Preferred

Table 2: Analytic Hierarchy Process 9-point rating scale. Values 2, 4, 6 and 8 represent intermediate values between the five anchor points above.

If, for example, when asked the question 'How much do you prefer option X over option Y on criterion Z?' the decision-maker believed that option Y is, in fact, 'Strongly Preferred' over X, then the reciprocal of the index value is assigned; in this case, 1/5.

It is assumed that, for example, if the preference of A over B is 5 then the preference of B over A *must* be $1/5$. Because of this, for an AHP with n options it will be necessary to conduct $½n(n-1)$ pairwise comparisons for each criterion. For example, the pairwise comparison matrix for establishing the relative preferences between three options on a single criterion might look like that in Figure 10.

	Option 1	Option 2	Option 3
Option 1	1	2	5
Option 2	1/2	1	2
Option 3	1/5	1/2	1

Figure 10: Example of AHP pairwise comparison matrix for three options. The decision-maker would only need to specify the highlighted (grey) preferences.

After these preferences have been elicited and the corresponding matrix produced for each criterion, the criterion performance scores of the options can be synthesised. There are two main ways this can be done. Saaty proposes determining these preferences by calculating the elements in the eigenvector associated with the maximum eigenvalue of the matrix (Saaty, 2004). Alternatively, these preferences can be well approximated using the Geometric Mean approach (Kumar & Ganesh, 1996) by:

1. Summing the values in each column

2. Normalising each of the numbers in the matrix by dividing by its respective column total

3. Calculating the mean normalised score for each row

Figure 11 shows what this procedure could look like in Microsoft Excel:

	A	B	C	D	E	F	G
1		**Original Performance Matrix**					
2							
3			Option 1	Option 2	Option 3		
4		Option 1	1	2	5		
5		Option 2	1/2	1	2		
6		Option 3	1/5	1/2	1		
7		Sum	1.70	3.50	8.00		
8							
9		**Normalised Performance Matrix with Calculated Criterion Scores**					
10							
11			Option 1	Option 2	Option 3	Criterion Score	
12		Option 1	=C4/C$7	0.57	0.63	0.59	
13		Option 2	0.29	=D5/D$7	0.25	0.28	
14		Option 3	0.12	0.14	0.13	=AVERAGE(C14:E14)	
15		Sum	0.41	0.71	1.00	1.00	
16							

Figure 11: AHP calculation of criterion scores for three options and a single criterion

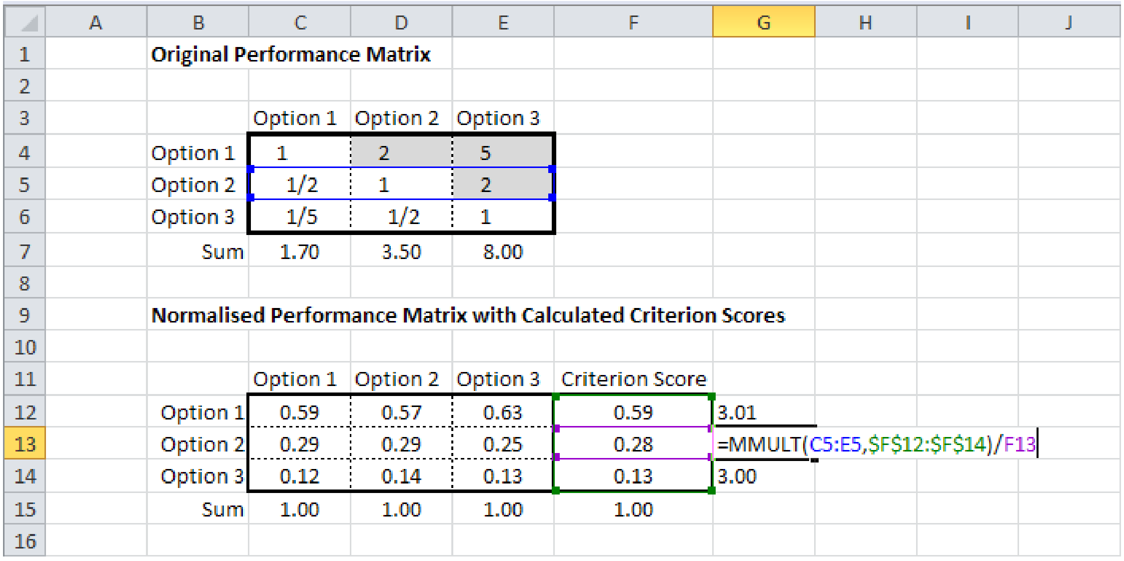
Criterion scores for each option on a single criterion are calculated as the average of the normalised row scores, which should sum to 1 (or 100%). In the above example Option 1 is most preferred on this criterion with the highest score of .59. In a full AHP analysis, this process would be completed for each criterion.

The procedure for calculating the criterion weights follows exactly the same process, where the decision-maker(s) must specify their preferences for each pair of criteria.

Once this is completed each option's criterion scores are multiplied by their respective criterion weight and summed to generate a global performance score for each option, which can be used to rank the options. This aggregation procedure is the same as that used by the WSM. The difference between the WSM and AHP is in the elicitation of criteria scores and weights.

Consistency of Preferences: In order for an AHP analysis to make sense the pairwise comparisons elicited from the participant(s) must be sufficiently consistent with one another. If Option A is twice as attractive as Option B and Option B is twice as attractive as Option C then option A should be four times as attractive as Option C. When creating AHP Saaty acknowledged that it would be near impossible for human decision makers to be absolutely consistent with their preferences, especially in more complex decision-making contexts. Saaty thus allowed a certain level of inconsistency during the elicitation of preferences, as measured by his consistency ratio (CR). A process for calculating the CR is as follows:

- **Step 1:** Multiply each pairwise comparison matrix by its respective criterion score vector and divide by associated criterion score for each option (cells G12-G14)



	A	B	C	D	E	F	G	H	I	J
1		**Original Performance Matrix**								
2										
3			Option 1	Option 2	Option 3					
4		Option 1	1	2	5					
5		Option 2	1/2	1	2					
6		Option 3	1/5	1/2	1					
7		Sum	1.70	3.50	8.00					
8										
9		**Normalised Performance Matrix with Calculated Criterion Scores**								
10										
11			Option 1	Option 2	Option 3	Criterion Score				
12		Option 1	0.59	0.57	0.63	0.59	3.01			
13		Option 2	0.29	0.29	0.25	0.28	`=MMULT(C5:E5,$F$12:$F$14)/F13`			
14		Option 3	0.12	0.14	0.13	0.13	3.00			
15		Sum	1.00	1.00	1.00	1.00				
16										

Figure 12: AHP assessing consistency Step 1

- **Step 2:** Compute average of the values calculated in Step 2 (denoted $\boldsymbol{\lambda max}$).

	A	B	C	D	E	F	G	H	I
1		**Original Performance Matrix**							
2									
3			Option 1	Option 2	Option 3				
4		Option 1	1	2	5				
5		Option 2	1/2	1	2				
6		Option 3	1/5	1/2	1				
7		Sum	1.70	3.50	8.00				
8									
9		**Normalised Performance Matrix with Calculated Criterion Scores**							
10									
11			Option 1	Option 2	Option 3	Criterion Score			
12		Option 1	0.59	0.57	0.63	0.59	3.01		
13		Option 2	0.29	0.29	0.25	0.28	3.00		
14		Option 3	0.12	0.14	0.13	0.13	3.00		
15		Sum	1.00	1.00	1.00	1.00	=AVERAGE(G12:G14)		
16									

Figure 13: AHP assessing consistency Step 2

λmax = 3.006

- **Step 3:** Compute consistency index (CI):

$$CI = \frac{(\lambda\text{max} - \text{n})}{(\text{n} - 1)} = \frac{(3.006 - 3)}{(3 - 1)} = 0.003$$

Where n is the size of the original comparison matrix

- **Step 4:** Compute consistency ratio (CR) by dividing the Consistency Index by an appropriate Random Index. The Random Index is the average Consistency Index for a very large number (e.g. 50,000) of randomly generated pairwise comparison matrices. Saaty proposes the following table of Random Indices for size n matrices (Saaty, 2004):

n	1	2	3	4	5	6	7	8	9	10
Random Index	0	0	.52	.89	1.11	1.25	1.35	1.40	1.45	1.49

Table 3: AHP random indices for use in assessing the consistency of pairwise comparison matrices

$$Consistency\ Ratio = \frac{Consistency\ index}{Random\ Index} = \frac{0.0028}{.52} = 0.0057$$

In this example the Consistency Ratio is the calculated Consistency Index divided by the Random Index for three options (.52). It is generally accepted that a Consistency Ratio equal to or below 0.1 indicates sufficient consistency (Finan & Hurley, 1997). If the CR is greater than this threshold then the pairwise comparisons within the matrix will need to be revised. One way to do this is to identify the most inconsistent judgment in the matrix and determine the range of values to which that judgment can be changed in order to reduce the inconsistency. Participants can then be asked if they are willing to revise this preference to a plausible value in that range. If unwilling, the analyst can ask the same question with the second most inconsistent judgment and so on. Identifying which preferences are most inconsistent can be a difficult task however (see discussions in Section 8.5.2.3).

4.6.4 Outranking methods

Comparing alternatives in any decision-making context can sometimes highlight a delicate issue regarding the strength or type of preference between two options. From a decision-making perspective there are a number of different relationships that can occur between two alternatives:

- **Strict preference:** an option is clearly preferred over another
- **Some Preference:** an option is preferred over another but the degree of preference is small or uncertain
- **Indifference (no preference):** two options have equal expected utility and are equally preferred

- **Incomparability:** competing options cannot be easily compared. One option may outperform another on some criteria while being outperformed on others (see Section 8.5.3.3 for examples and discussion of incomparabilities).

Although most MCDA models can easily incorporate the first three of these relationships in one way or another, outranking models are more explicit in defining the type of relationship between two options on different criteria. Outranking models can identify options that are said to be 'incomparable', or 'difficult to compare', whereas most other methods are incapable of identifying such relationships.

Developed by Bernard Roy and colleagues (Roy & Vanderpooten, 1996), outranking methods utilise the concept of dominance. An option is said to dominate another when it sufficiently outperforms it on at least one criterion (as defined by thresholds – see below) while not being sufficiently outperformed itself on all other criteria. All options are assessed in terms of the degree to which they outrank others and the degree to which they are outranked themselves. As mentioned above, a unique feature of this method is that is it often possible for two options to be classified as 'incomparable', meaning that the model will be unable to establish any clear outranking relationships between the two options. This may occur when two options have such different combinations of performances on the set of criteria that it becomes difficult to determine the better option. Having this possibility incorporated into the model allows the formal analysis of the options to continue without having to label the two options as 'indifferent', remove one option from the analysis completely, or 'force' a ranking of the options, which is what would occur under most other methods.

Another unique aspect of the outranking approach is its use of 'preference' and 'indifference thresholds', where decision-makers are required to specify exactly how much an option must outperform another on each criterion before it can be considered 'preferred' or 'strictly preferred', with any difference below these thresholds indicating indifference or weak preference between the two options on the respective criterion.

Indifference threshold: This figure indicates the minimum required difference between two options on a specific criterion in order for the better performing option to be regarded as preferred to some degree over the other. Any difference in performances below this threshold and the model will treat the two options as having equal performance on that particular criterion (i.e. an indifference relationship).

Preference threshold: Represents the required difference between two options in order for the better performing option to be regarded as 'strictly preferred' (the maximum degree of preference possible) on a criterion. Differences less than the preference threshold value but greater than the indifference threshold will still give a degree of preference (‘Some preference’ - see above).

Unlike most other MCDA models, outranking methods do not produce a ‘Single Synthesis Criterion’ and do not require the mapping of data inputs onto homogeneous rating scales prior to aggregation. Although there may be subtle differences between the different models of the outranking ‘family’ of MCDA models, the general approach is that pairwise comparisons are conducted (automatically and in the background – unlike AHP) between every pair of options on every criterion. The models consider the indifference and preference thresholds and assign one of the aforementioned preference relationships to every pairwise comparison. These preference relationships can be aggregated across all criteria for each option and the output of the models are usually expressed in terms of the positive and negative ‘flows’ for each alternative:

Positive flow: Calculated as the sum of all pairwise comparisons in favour of an option. Positive flow quantifies the degree to which an option is preferred over all other options (i.e. the degree to which it ‘outranks’ or ‘dominates’ all others)

Negative flow: Calculated as the sum of all pairwise comparisons where the option under consideration is the less preferred of the two. Negative flow quantifies the degree to which all other options are preferred over the alternative of interest.

Some outranking methods provide recommendations by analysing these flows individually. Such methods might consider two options, one with high positive and negative flows, and another with low positive and negative flows, as incomparable. Other methods will combine these two flow figures into a ‘net flow’, and use this number to generate (or ‘force’) a complete ranking of all options.

Section 8.5.3 details the specific principles and mathematics behind PROMETHEE, a popular MCDA model from the outranking family of models.

4.6.5 Fuzzy MCDA

It is not uncommon for decision-makers to face decisions where there is a lack of quantitative data for the generation of performance scores. In some cases decision-makers may only have qualitative judgements to base their decisions on. One method which attempts to cope with the imprecision of data inputs in decision-making is Fuzzy MCDA, which utilises 'fuzzy sets' in order to make recommendations (DCLG, 2009).

Fuzzy sets attempt to incorporate the concept that the language used by individuals when discussing problem situations is subjective and rarely precise. Fuzzy sets try to capture this imprecision through the idea of a 'membership function', where each option is assigned a numerical 'degree of membership' to a particular set of, for example, 'effective' options, usually between 0 and 1, with 1 denoting full membership to a particular 'set'. MCDA models of this type attempt to aggregate these fuzzy numbers into a somewhat definitive ranking or recommendation. Sometimes 'fuzzy weights' are also used to specify preferences for the different criteria.

Although Fuzzy MCDA models been discussed in the academic literature for over two decades (DCLG, 2009), they have received little application to real world decision-making and have been mostly confined to theoretical applications. Such methods appear more difficult to understand and implement than others, and lack the high degree of transparency that other methods can offer (DCLG, 2009). Although these methods have not received much practical application yet, they have recently been applied to food safety decision-making in Europe (Mazzocchi et al., 2013; Ragona, Mazzocchi, & Alldrick, 2012; Ragona et al., 2011). Fuzzy MCDA may be a useful method for situations where uncertainty is high and very little quantitative data is available, but is perhaps unsuitable for applications where quality numerical data is available for use as model inputs. At present the literature has not established that these methods offer a critical advantage over less complex methods.

4.7 Weighting of Criteria

Most MCDA models are not explicit regarding how criteria weights should be elicited from participants. With a notable exception being the Analytic Hierarchy Process, weighting of criteria is typically treated as a separate activity in MCDA models, with the focus of most models being the aggregation (rather than generation) of model inputs. There are several methods of generating weights and the most common are outlined below.

Direct Rating

Direct rating is one of the most straightforward methods for assigning weights (and criteria scores). Direct rating involves participants assigning a rating (score) to each criterion, reflecting its overall importance in decision-making. Typically participants will be asked to assign ratings to the least and most important criteria first, and then assign ratings to the remaining criteria relative to these two end points (Koen, 2008). Although ratings, in theory, can take any value, participants are often asked to assign ratings within a predetermined range, with the 0-100 range being a popular choice (Mabin, 2006). Unlike 'local' scales, which are used for assigning performance scores, and require that the worst and best performing options receive the minimum and maximum possible scores respectively, direct rating for criteria weights does not share this requirement. If this requirement was imposed on criteria weighting then the least important criterion would receive a weight of zero, essentially removing it from the analysis. Once all criteria have been assigned a rating it is usual for these ratings to be converted into proportions or percentages for subsequent use in aggregation. Tools such as the Visual Analogue Scale (VAS) can also be used to visually elicit direct ratings (Tong, 2014).

Pairwise Comparisons

Participants are asked to compare alternatives to one another using pairwise comparisons. They will be required to articulate, usually quantitatively, their relative preferences for each pairing. For example, participants may be asked a series of questions in the form of 'how much more important is criterion X over criterion Y'. If the answer to such a question is, for example, "5 times more important" then criterion X should be given a weighting five times that of criterion Y. This is the weighting procedure prescribed by the original AHP method (Saaty, 1980), and is rarely utilised in methods other than AHP and its variants.

SWING Weights

SWING weighting considers the ranges over which the criteria performance scores lie when estimating weights (Von Winterfeldt & Edwards, 1986) and, because of this, criteria performance scores must first be generated before SWING weighting can proceed.

Participants are asked to select the criterion where a move (swing) from the current worst to best performance score would have the most value. This criterion is assigned the highest score (usually 100). Participants then select the next criterion where a 'swing' from the current least to best score would bring the most value and this criterion is assigned a score lower than 100 indicating its importance relative to the previously considered criterion. This process continues until all criteria have been assigned scores. These scores are then normalised and represented as percentages or proportions (Sullivan, 2012). SWING weighting is a novel approach because it considers intervention performances during the weighting procedure. This process often results in higher (lower) weights for criteria where performance scores take a wide (narrow) range of values, when compared to the other weighting approaches (Spackman et al., 2000). For example, financial cost may be a very important criterion to decision-makers and, in other weighting techniques this would be reflected by a high criterion weight. However, if the costs of the set of alternatives all lie within a narrow range then SWING weighting may result in a low weight for 'cost' because the cost difference between the least and most expensive option might be viewed by decision-makers as negligible.

Simple Multi-Attribute Rating Technique (SMART)

With SMART, weights are elicited in a two stage process similar to the direct rating technique (Edwards, 1977; Edwards & Barron, 1994). Initially, participants rank criteria according to their importance in decision-making. Following this initial ranking participants assign a fixed number of points (usually 10) to the lowest ranking criterion. Then, more than 10 points are allocated to the penultimate criterion to reflect its importance relative to the lowest ranked. This process of assigning increasing amounts of points to the next highest ranked criteria continues until all criteria have been assigned a number of points. Criteria scores are then normalised to obtain the final weights in the forms of proportions or percentages.

Discrete Choice Experiments (DCE)

Discrete Choice Experiment methods present participants with a number of hypothetical situations where they must choose between two or more hypothetical 'options', each with different attributes (i.e. different combinations of criteria scores) (Ryan & Farrar, 2000). Such experiments are able to derive the preferences (criteria weights) of participants by analysing the decisions they make. A number of different statistical methods can be used to estimate the weights, as well as the willingness-to-pay (if cost is included as a criterion) and the willingness to trade one ‘attribute’ for another. Discrete Choice Experiments are resource intensive to conduct and require a sufficient sample size in order to produce results which might be generalisable (Rose, 2011).

4.8 Group Decision-Making

MCDA literature often refers to ‘the decision-maker’ as if MCDA models were intended to cater to a sole decision-maker with ultimate decision power. In practice such single ‘decision-makers’ are rare and it is far more common for a group of individuals to collectively assume responsibility for a decision instead (Belton & Pictet, 1997). Belton and Pictet identify three main options for conducting group MCDA:

Sharing

Sharing aims to obtain inputs by consensus; through discussion of the different stakeholder views and negotiation of an agreed set of inputs for subsequent use in a specific MCDA model. It addresses the disagreements between group members and attempts to reconcile these through explicit discussion of their cause.

Typically ‘sharing’ would involve a workshop or focus-group environment where members can openly present their perspectives of the problem and its solution. Geographical or time constraints often make such workshops impractical and, if this is the case, there is the potential to facilitate group decision-making through the use of teleconferencing or new online MCDA tools (Hansen & Ombler, 2014).

This method is seen as the most preferred way to conduct group MCDA as, although inputs are rarely deterministic, they can be treated as constants in the analysis if all members of the group are able to agree on their values.

Aggregation

Aggregation involves eliciting preferences (i.e. weights or scores) from each individual in the decision-making team in isolation and then aggregating (e.g. averaging) this data before continuing with the analysis. This approach is not ideal as it does not necessarily encourage discussion or information sharing. Although both ‘Sharing’ and ‘Aggregation’ acknowledge the disagreements between group members, a fundamental difference between the two is that 'Aggregation' forces the mathematical reconciliation of these differences without explicitly discussing their cause. This lack of discussion is a clear disadvantage of this method.

Comparing

It may be impossible to reach an absolute consensus when using the 'Sharing' approach above. The 'Comparing' approach requires that group members share as much information as possible to justify their individual preferences in an attempt to reach a consensus, but does not require that the group reach such a consensus. Multiple MCDA analyses can be conducted using a range of different input values and these results can facilitate further discussion and may highlight inputs that have little or no effect on the results and thus do not require further discussion. Comparing involves the significant discussion also present in the 'Sharing' method, but without requiring the reconciliation of all differences.

4.9 MCDA in the Food Safety Literature

MCDA is receiving increasing attention in the food-safety world. There is a relatively new and growing body of MCDA food-safety literature, and the methodology is being officially recognised and used by organisations like the FAO and WHO. MCDA has been more widely applied to problems relating to the prioritisation of hazards/risks rather than the evaluation of interventions.

Henson et al., have developed a *"multi-factorial risk prioritisation framework for foodborne pathogens"* utilising four key criteria:

1. **Health impact:** measured in DALYs and COI
2. **Market Impact:** in dollars: The financial losses arising from changes in supply or demand due to a pathogen. This criterion considers the current size and value of the domestic and export food markets associated with a particular pathogen and considers the potential for that pathogen to cause major disruption and financial losses to these markets.

3. **Consumer perceptions:** measured on Likert-type scales[14]
4. **Social sensitivity:** measured on Likert-type scales: This criterion considers any unique or disproportionate effects that a particular pathogen may have on, for example, elderly, cultural minorities, infants or part of an industry (e.g. small operators).

Caswell uses this same framework to build a case for expanding the conventional cost-benefit analysis to include the additional criteria mentioned above (Caswell, 2008).

Following on from this work Ruzante et al, further develop and apply this framework to the prioritisation of high profile food-pathogen combinations (Ruzante et al., 2010). This study utilised the PROMETHEE MCDA model, implemented using Decision Lab (Visual Decision Inc., Montreal, QC, Canada), a specialised MCDA software.

Donoso, for his Master's thesis (Guelph University – Canada), used specialised software to apply five MCDA methods to rank common food-pathogen pairs according to these same four criteria; public health, market level impact, consumer risk perception and acceptance and social sensitivity (Donoso, 2008). The methods studied were:

- Multi-Attribute Utility Theory
- Analytic Hierarchy Process
- ELECTRE III (outranking)
- PROMETHEE (outranking)
- NAIDE (Novel Approach to Imprecise Assessment and Decision Environments)

[14] Likert-type scale: An ordinal scale used in questionnaires to gauge a respondent's level of agreement with a particular statement. For example a five point scale might consist of five possible answers: strongly disagree; disagree; neither agree nor disagree; agree; and strongly agree.

Donoso suggests that the Outranking methods (ELECTRE III and PROMETHEE) are most appropriate for the prioritisation of food-pathogen hazards, citing them as providing *"a structured, expandable, credible and interactive framework for democratic decision-making that accommodates the values, the time availability, and the analytical skills of the multiple decision makers that may participate in the microbial hazard prioritization decision problem"*

Henson and Masakure (2011) propose a similar framework for the evaluation and selection of options for enhancing Sanitary and Phytosanitary (SPS) coverage. The framework assesses options on the following criteria, and aggregates them using PROMTHEE:

1. **Up-front investment:** Measured in dollars
2. **Maintenance costs:** Assessed on a semantic 1-3 rating scale
3. **The change in the value of exports:** Measured in dollars
4. **The change in agricultural productivity:** Assessed on a semantic 1-4 rating scale
5. **The change in public health:** Assessed on a semantic 1-4 rating scale
6. **The change in environmental protection:** Assessed on a semantic 1-4 rating scale
7. **The impact on poverty in the implementing country:** Assessed on a semantic 1-4 rating scale
8. **The impact on vulnerable groups:** Assessed using a binary/dummy variable (yes/no)

Henson and Masakure recommend the use of structured stakeholder focus-group-like workshops or the Delphi approach for eliciting scores. An example of Henson's recommended weight elicitation process involves stakeholders anonymously assigning percentages (summing to 100%) to each criterion, reflecting their overall importance in the analysis. The percentages are then averaged and presented to stakeholders who can debate their validity and make adjustments based on a group consensus before proceeding.

The Common Market for Eastern and Southern Africa organisation (COMESA) and the World Health Organization have implemented a modified version of this framework to the problem of strengthening Phytosanitary controls in the floriculture sector in Uganda

(Byanyima et al., 2013). They use specialised MCDA software to implement an unspecified outranking model (most likely PROMETHEE) in a real and comprehensive analysis. The output of their analysis is a short list of capacity building options to be given priority.

The World Health Organization have also used MCDA for the worldwide ranking of 24 foodborne parasites using the following qualitatively assessed criteria (WHO/FAO, 2014):

1. **Number of global illnesses**
2. **Global distribution**
3. **Acute morbidity**
4. **Chronic morbidity**
5. **Percentage resulting in chronic effects**
6. **Mortality**
7. **Increasing illness potential**
8. **Trade relevance**
9. **Socio-economic impact**

Expert judgement was used to generate numerical scores and weights for these criteria, and these inputs were aggregated using the Weighted Sum Model, resulting in a ranked list of parasites in order of *"importance"*.

Fazil et al, in a Canadian project, have published a hypothetical example of how PROMETHEE could be applied to the selection of food safety risk management options (Fazil et al., 2008). They propose four base criteria:

1. **Quantity*Quality of research regarding effectiveness:** The proportion of research articles supporting the effectiveness of a risk management option weighted by the quality of the research. For example, randomized control trials are given the greatest weighting while cohort studies are weighted much lower. Systematic reviews are proposed to gather this information from existing effectiveness studies.

2. **Effectiveness:** The percentage reduction in contamination, obtained from the literature and/or through quantitative modelling
3. **Cost:** in dollars
4. **Ease of implementation:** this data would likely require its own qualitative assessment involving consultation with experts

Ragona and Mazzocchi et al, have developed a MCDA spreadsheet model known as SCRYER[15], which allows the comparison of up to three interventions on a comprehensive set of criteria (Mazzocchi et al., 2013). Users qualitatively score each option on up to 60 criteria (organised into 14 groups) using Likert-type scales. Utilising the concept of positive and negative flows (see Section 4.6.4), the model uses a very complex fuzzy-outranking algorithm to compare each intervention in a pairwise fashion and arrive at a partial or complete ranking. The model has been tested on several case studies already, including interventions to address Mycotoxins in cereals (Ragona et al., 2011), dioxins in the EU (Ragona, Mazzocchi, & Rose, 2012), and two others in New Zealand which have not been published.

SCRYER is somewhat similar to the Australian developed Risk Ranger software for excel, which is used to help rank/prioritise potential food safety issues/interventions according only to their estimated potential to cause human illness (Ross & Sumner, 2002). Users answer 11 semi-quantitative questions regarding the severity, location and the risks associated with the different issues being analysed. Risk Ranger is recommended by the FAO/WHO as a useful decision support tool and is used by Australian and other international food safety authorities (Anders & Schmidt, 2011).

MPI currently use a simple MCDA spreadsheet tool to assist with their market access prioritisation work (NZFSA, 2010d). The spreadsheet considers financial criteria, feasibility, commitment from industry, trade precedents and the amount of resources already available to implement such work. This model is discussed in more detail in Section 3.7.1.2.

---

[15] SCRYER: Something or someone who can foretell the future using a crystal ball or other reflective object or surface. http://en.wikipedia.org/wiki/Scryer

4.10 Conclusion

MCDA is an iterative and inclusive process which can be a suitable decision-support methodology when there are multiple conflicting objectives and numerous stakeholders. MCDA is as much (if not more) about the process of arriving at an answer as it is about the answer itself, and should be a thought provoking experience from start to finish for all involved.

There are many MCDA models in existence today. The main types of MCDA model have been described in this chapter. The application of some MCDA models has been limited to experimental or academic purposes. However, some have received widespread application to real problems and have proven track-records in other areas such as logistics, manufacturing, engineering, politics and education. Some models have the potential for widespread application to food safety decision-making in New Zealand, especially the evaluation and selection of risk-management options. This chapter provides a basis for the selection, construction, testing and critique of a small group of models in Chapter 8.

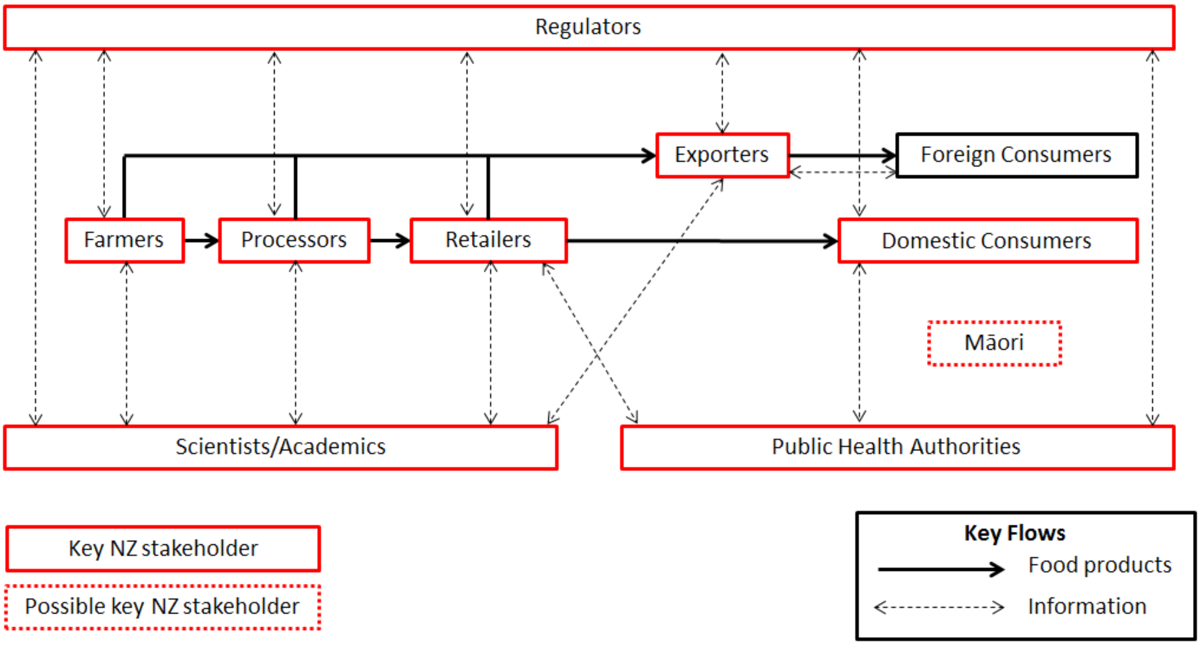
Chapter 5: Stakeholder Identification and Classification

5.1 Introduction and Summary

An essential first step in MCDA is the selection of key stakeholders to participate in the analysis (Spackman et al., 2000). The 'stakeholder' is an important concept to consider in large scale decision-making. However, the notion of the stakeholder is sometimes taken for granted in that many studies utilise the word 'stakeholder' without establishing a clear definition. Subtle differences in meaning can have a substantial influence on the selection of groups to involve in the formulation and solution of a problem. Establishing a clear stakeholder definition is an important initial step in complex decision-making involving multiple groups.

Food safety decision-making is conducted by different groups at different stages in the farm-to-fork chain. The outcomes of decisions will usually affect parties other than the decision-maker(s). For stakeholders considering the implementation of a food safety intervention it is important to understand how it may affect other parties such as consumers, and how these other parties may affect its implementation. Interventions which appear to be logical choices to some may be completely unacceptable to others, and this difference in sentiments may lead to a failed implementation in an environment where cooperation between different stakeholder groups is necessary. For large scale decision-making it is important that the groups that have the potential to contribute to the successful (or unsuccessful) implementation of an intervention are identified and their perspectives considered when evaluating risk management options.

This chapter examines the concept of the stakeholder in food safety decision-making, and reviews some of the ways that stakeholders have been classified in the broader stakeholder literature. Furthermore, based primarily on the results of key informant interviews, this chapter proposes a number of key stakeholder groups with regards to food safety decision-making. These groups and the key relationships between them are summarised in Figure 14.



Note: The nature of Māori communication with other stakeholders is unknown

Figure 14: Key stakeholder relationships in the farm-to-fork chain.

5.2 Methodology

This chapter seeks to answer research question 6:

Research Question 6: Who are the key stakeholders in food safety intervention decision-making and how can they be classified according to their role in food safety intervention decision-making?

Relevant food safety literature was reviewed, as well as broader stakeholder oriented literature such as the Stakeholder Analysis and Soft Systems Methods (SSM) literature. There are multiple definitions regarding what constitutes a stakeholder in the literature and these definitions were reviewed, and the most appropriate definition selected for use in this research. The following sections outline different stakeholder definitions from the literature before justifying the selection of the definition used for this thesis.

This thesis required a methodology for identifying and classifying the key New Zealand stakeholder groups in food safety decision-making so that key informants from them could be sampled. Google Scholar, Medline, ProQuest, PubMed and Science Direct were searched for relevant articles relating to food safety decision-making and approaches for the identification and classification of stakeholders (Appendix C contains keywords used). Additionally, a search of domestic (e.g. MPI) and international (e.g. Codex) food safety organisation websites was conducted. New Zealand university research repositories were also searched for relevant theses.

Reviewing food safety and broader stakeholder oriented literature identified how stakeholders might be identified and classified. However, some of the classifications used in the literature appeared too broad, while others seemed too detailed.

Interviews were subsequently conducted to identify which stakeholder groups from the literature should be considered 'key stakeholders' in the context of this research. Initial interviews were conducted primarily with industry representatives, regulators and food safety experts from large food businesses such as meat processors. Each informant was asked, amongst other questions (covered in later chapters – see Appendix A):

- **Whose responsibility is food safety in NZ?**
- **What is your role in the provision of safe food?**
- **Who are the key stakeholders when it comes to food safety and food safety decision-making in NZ?**
- **Is there anyone else within or outside of your organisation/industry who you think I should also talk to?**

Informants were asked to identify other key informants for interviewing, both within and outside of their industry and 'stakeholder group'. This sampling methodology could be described as a combination of purposive (Sekaran, 1992) and snowball sampling (Goodman, 1961), where initial participants were identified and selected based on their ability to inform the research questions and subsequently asked to identify other participants that they felt could provide additional insight. Stakeholder groups were ultimately considered as 'key' stakeholder groups if more than three participants mentioned them during the interview process. 55 interviews were conducted with key informants in total. The number of interviewees in each group is presented in table 4.

Stakeholder group	Interviewees
Regulators	7
Public Health Authorities	2
Scientists/Academics	5
Farmers	7
Processors	17
Retailers	13
Exporters	13
Maori	4

Note: All exporters also belonged to one of the other stakeholder groups

Table 4: Number of interviewees from each stakeholder group

Groups that were mentioned by less than four participants were excluded from the final list of key stakeholders (e.g. transportation companies and *"wharfies"*). It was decided to exclude these groups from the interview process because, although they have a role in ensuring safe food, interviewees suggested their role is generally very simple and relatively small (e.g. move food from A to B under agreed and easy to control conditions) when compared to the roles of the other stakeholder groups. Additionally, these groups also have very little, if any, involvement in decision-making, and their exact role in the farm-to-fork chain is generally dictated by one of the larger stakeholders in the chain (e.g. producers or customers).

These interviews, together with information from the literature, were helpful for clarifying the roles of the various stakeholder groups.

Consumers were a key stakeholder group identified consistently in the literature and by the vast majority of interviewees. Currently there are no organisations that are able to speak on behalf of consumers regarding food safety preferences/perceptions. Engagement with consumers in this research would have required a different research approach to that adopted for this thesis and this is the primary reason why a decision was made not to seek direct input from consumers in this research. The time and resources required to interview a representative sample of consumers, which would have most likely required a significant number of surveys, questionnaires and/or focus groups, was another reason why consumers were excluded from this component of the research. Furthermore, some food safety literature suggests that the criteria consumers use to evaluate risk and risk reduction strategies are context specific (Bruhn, 2007; Frewer, 2009; Houghton et al., 2006; Mørkbak et al., 2010;

Siegrist, 2008). That is, consumers may utilise different criteria to evaluate different food safety interventions depending on the nature of the intervention and the hazard being addressed. Attempting to establish a broad set of criteria for consumers may not be entirely meaningful when assessing individual technologies. Some research has suggested that consumer research regarding food processes or technologies be tailored to the specific technology and context in which it will be used (Frewer, 2009). Although consumers were not included in the qualitative research component of this thesis for the above reasons, this thesis acknowledges their high importance as stakeholders in food-safety decision-making. Food-safety intervention decision-making should include consumer input, gathered through an appropriate, thorough and systematic procedure.

5.3 Results

5.3.1 What is a stakeholder?

The stakeholder concept is not always explicitly considered in MCDA or food safety intervention evaluation studies. A number of authors have dealt with the type of MCDA problems where decisions are made by multiple stakeholders. Often these authors use alternative words such as, *"parties", "players", "groups", "actors",* or *"participants"* (Banville et al., 1998; Spackman et al., 2000). For example, Keeny attempts to satisfy the requirements of a Multi-Attribute Utility Theory (MAUT) within a decision-making context involving multiple stakeholders by utilising focus groups (Keeney, 1992). More recently, Bartolini and Viaggi (Bartolini & Viaggi, 2010) emphasise the difficulty surrounding the identification and inclusion of multiple stakeholders in food safety decision-making and suggest ways that group inputs can be gathered and formally incorporated into an MCDA (Bartolini & Viaggi, 2010). However, these studies generally do not elaborate on what exactly constitutes a stakeholder.

Many MCDA studies treat the concept of the stakeholder as transparent and with a single universal meaning. Although on the surface this seems to pose no problems, with the definition of a stakeholder being somewhat intuitive as anyone with a 'vested interest' in a decision problem, subtle differences in the way a stakeholder is defined can have consequences in such complex decision-making situations. For example, Savage et al, define stakeholders as "*individuals, groups, and other organisations who have an interest in the* *actions of an organisation and who have the ability* to *influence it"* (Savage et al., 1991). In the context of this thesis, such stakeholders could include the government or industry representative groups. Alternatively, a stakeholder can have a vested interest when the way a problem is formulated or solved has an impact on them; they are not directly involved in the decision-making process but stand to gain or lose in the process. Consumers and small food business owners might fit this definition. Finally, a stakeholder can also be defined as an individual or group who is able to simultaneously influence and be influenced by a decision-making process. For example, very large food businesses may fit this category. Thus, a stakeholder can be involved in a decision in one of three ways:

- By influencing the decision-making process
- By being affected by the formulation or decision
- By both influencing the decision-making process and being affected by it

Mason and Mitroff define a stakeholder as anyone who falls into one of these three categories (Mason & Mitroff, 1981). This is the most comprehensive and inclusive definition of a stakeholder and typically seems to be what is meant in the literature when the word 'stakeholder' is used without any further explanation. Food safety decision-making, especially at a regulatory level, will typically involve groups from all three of the above categories (e.g. regulators, consumers, and large food producers). Thus Mason and Mitroff's definition of a stakeholder is the most appropriate definition to use in this context, and is adopted for use in this thesis.

5.3.2 Classification of stakeholders

From one perspective stakeholders can be classified according to the level or participation they have in the problem formulation and outcome (Banville et al., 1998). Not all stakeholders in a problem situation are participants in the decision-making process. In many cases this would not be ideal or even possible. For example, it could be difficult to involve consumers in the scientific evaluation of technical data during the setting of Maximum Residue Limits and microbial performance targets. Vroom defines three 'types' of stakeholder involvement in decision-making (Vroom, 1974). Firstly he defines the autocratic process (*"type A"*), where stakeholders are 'observed' without any formal consultation or involvement in decision-making. Secondly, consultative (*"type C"*) processes involve decision-making where stakeholders are consulted on some level, but the final decision remains that of the individual or organisation responsible for addressing the problem. Finally, Vroom defines group processes (*"type G"*), where the decision is a collaborative effort, conducted usually by a group comprised of a subset of stakeholders assembled around the problem. In the context of food safety decision-making, the setting of general performance targets and levels by MPI, or the selection of specific food safety interventions designed to comply with these targets by food businesses, will be type A or C processes. Although consultation is heavily emphasised in MPI documents such as the Risk Management Framework (NZFSA, 2010b), there is no legal requirement for regulators or food businesses to systematically incorporate stakeholder feedback into their decisions.

Checkland's Soft Systems Methodology (SSM) provides another systematic approach for real problems involving social, cultural and psychological elements (Checkland, 2001), as is the case with many food safety decision-making problems. As part of SSM Checkland has defined the mnemonic CATWOE (Customer, Actors, Transformation, Worldview, Owner, and Environment) to describe a problem situation in relation to the different stakeholder perspectives taken. Within this framework Checkland defines three types of stakeholder:

> **Customers** – those who will benefit or suffer from the outcome of a decision. For this research the most obvious customers are consumers of food, whose health may be affected by any changes in the level of risk associated with food products. Food businesses are also customers as food safety decisions have the potential to affect their profitability.
>
> **Actors** – those who are responsible for implementing the decision and carrying out the activities that make a system work. There are two main actors in food safety decision-making in New Zealand, corresponding to the two main levels of intervention decision-making (regulatory and industry). In New Zealand regulators are responsible for setting broad targets and for creating the regulatory framework within which food businesses are given flexibility to implement the risk management options most appropriate for them.
>
> Public health authorities could also be classified as actors, as they help to deliver consumer focused food safety interventions such as education campaigns.

Owner(s) – those that have ultimate control over a system and who have the ability to change its measures of performance.

Regulators have ultimate control over food safety decision-making in New Zealand. MPI sets the legal framework under which all food businesses must operate. Additionally, MPI sets targets and maximum levels for hazards in food, reflecting an acceptable level of risk.

Clearly there are a number of ways that stakeholders could be classified in this research, and those described above are useful for identifying broad stakeholder groups. However, none of these classification systems are specific to food safety decision-making. In constructing a comprehensive picture of how food safety decision-making is conducted by, or how it affects, the various stakeholders, it would be over-simplistic to utilise any of the broad classification systems reviewed above. In particular, grouping all food businesses together into one stakeholder group could be misleading as there are different types of food business, most of which have very different roles in food production and sale, with potentially very different approaches and requirements when it comes to food safety decision-making. It is more appropriate to classify stakeholders according to the specific roles they assume in ensuring the safe production and sale of food.

Suggestions for food safety specific stakeholder classifications can be drawn from food safety oriented literature and official food safety documents. Although virtually all of this literature does not explicitly assign a meaning to the word ‘stakeholder’, it is apparent that use of the term in the majority of studies fits Mason and Mitroff’s definition, with ‘stakeholder’, including those who can affect the formulation or solution of a problem and/or are affected by the outcome of the decision (Mason & Mitroff, 1981). Particularly relevant pieces of literature include guidelines from the Codex Alimentarius Commission (CAC, 2005) and the Ministry for Primary Industries’ Risk Management Framework (NZFSA, 2010b). From an international standard setting perspective, Codex guidelines classify stakeholders as (CAC, 2005):

- Relevant government ministries (i.e. Regulators)
- Consumer groups
- Food Industries – further classified as:

    - Producers (i.e. farmers)
    - Processors
    - Importers
    - Exporters
- Academia – further classified as
    - Universities and professional bodies
    - Leading research institutions
    - Recognised experts
- Public health professionals

From a New Zealand regulatory perspective MPI provides a non-exhaustive list of stakeholders as (NZFSA, 2010b):

- *The minister for Food Safety*
- *Food Standards Australia New Zealand*
- *Food industry associations and forums*
- *Food Businesses*
- *Representative community groups and individual consumers*
- *The scientific and academic community*
- *Trading partners*
- *International food safety agencies*
- *Third party agencies who undertake evaluation, verification and enforcement roles such as territorial authorities and public health units*
- *Other government agencies*

Both of these lists are extensive. In contrast to the problem from the general stakeholder literature of having classifications that are too broad, this food safety specific literature has perhaps suggested too many categories. The goal of this research is not to generate an exhaustive list of stakeholders, but rather to identify a list of 'key' stakeholders. Interviews with key informants were utilised to assess which of the groups mentioned during this review should be considered 'key stakeholders'.

5.3.3 Results of Interviews

From key informant interviews nine key stakeholder groups were identified as having significant involvement in providing safe food in New Zealand. There was a general consensus amongst interviewees regarding the appropriateness of these groups, which are described below.

Food Businesses

Most interviewees agreed that the term 'food business' was too broad to be appropriate during the detailed evaluation of specific risk-management options. 'Food business' is a broad term encompassing all businesses directly involved in the production or sale of food. There are a large number of ways that 'food businesses' could be sub-grouped. Interviewees generally classified these food businesses according to their main role in the farm-to-fork chain. These groupings provide a more specific and useful classification, yet are broad enough to apply to most food production chains.

Farmers

Farmers grow produce which will then enter processing or sometimes be sold directly to food retailers or the public. Farm produce could be plant or animal. In the context of food safety the primary responsibility of farmers is to adhere to Good Agricultural Practice (GAP) and maintain appropriate biosecurity standards. GAP provides guidelines regarding the use of agricultural compounds and veterinary medicines so that chemical residues in the foods they are growing will be kept to acceptably low levels. For example, farmers of plant produce have an obligation to use only approved agricultural chemicals such as fertilisers, and only at levels which are in accordance with GAP, so as not to produce residue levels which violate Maximum Residue Limits (MRLs). Farmers of animal produce also have an obligation to keep residues such as growth promotants and antibiotics to acceptable levels. Adhering to biosecurity processes helps to ensure that any microbial contamination of food products is kept to a minimum. Farmers and farming representatives interviewed indicated that adhering to GAP was relatively easy compared to the task of preventing the introduction or spread of harmful pathogens (e.g. *Campylobacter* or *E.coli*) in farmed food products.

Processors

Processors process or manufacture food products which are then sold to retailers or the public. Processors often place strict food safety requirements on their suppliers to ensure the raw materials they receive are of the highest quality, allowing them to meet their own food safety obligations to their customers. Processors have a number of important responsibilities in maintaining food safety standards. One of the most basic yet important responsibilities of processors is their adherence to Good Operating Practice (GOP –Section 3.4.3 contains a more detailed description of GOP), which is a general set of guidelines and procedures for maintaining a clean and sanitary workplace environment. Additionally, processors must ensure that any chemicals, processing aids or food additives are used responsibly and in accordance with regulations. This will generally involve ensuring that adequate training has been provided to workers regarding the use of different types of chemicals, as well as the appropriate calibration and monitoring of machinery and the production process as a whole. For processors of all products, especially animal products such as raw meat and poultry, there is a requirement to conduct processing in a way that minimises any microbial contamination and cross-contamination of the final product. This minimisation of microbial contamination is often the main food safety focus for processors of higher-risk animal products. Processors also have a duty to prevent the contamination of their foods with foreign objects such as loose bolts and screws from machinery. For processors, all food safety hazards should be effectively managed through their mandatory HACCP plans (Chapter 3.4.2 explains HACCP in detail).

Food Retailers

In the context of this research retailers include all businesses whose primary activity is the sale of food. Food retailers include a range of different businesses, from corner dairies to large supermarkets and wholesalers; small restaurants to fast food chains. Some interviewees made a distinction between retailers and restaurants, with restaurants selling food as already prepared and ready for consumption, usually on premise, and retailers vending food that generally requires further preparation by the consumer before consumption. While most interviewees recognised that there was a clear difference between these two types of business, they also recognised that in practice many straddle both. For example, supermarkets operate primarily as 'retailers' in the conventional sense of the word, yet often supply a variety of ready-to-eat meals. Similarly, many restaurants will also sell food products requiring further preparation by the consumer at a later time. Many interviewees felt that although restaurants and retailers were clearly two different types of food business, their roles in providing safe food to the end consumer were very similar and thus they could be grouped together for the purpose of this research. From a food safety perspective, the primary responsibility of this group is to store, handle and prepare food in a safe manner. In particular, it is crucial that these businesses safely manage cross contamination of high-risk foods such as raw poultry, as well as ensuring that, where relevant, such foods are sufficiently cooked to eliminate any pathogens that may be present. Additionally, worker hygiene is crucial for this group, especially for workers handling foods which will not be subsequently cooked. In many instances retailers are the last line of defence against foodborne illness. Some foodborne illness outbreaks stem from this group, often due to poor worker hygiene (Greig et al., 2007). Many larger retailers such as supermarket and fast food chains are required to formulate their own HACCP plans. Smaller businesses such as cafes and dairies will often opt to follow a more prescriptive set of guidelines, the Food Hygiene Regulations 1974, produced by the Ministry of Health (NZFSA, 2009a).

Exporters

Exporters include any food businesses who sell their products overseas. Members in this group will also belong to one of the above categories. Interviews identified a need to allocate exporters their own classification, as often the roles, requirements and priorities of exporters are very different to those of similar businesses operating on the domestic market. Aside from meeting their obligations under the other relevant categories, exporters have an obligation to consistently meet the minimum food safety standards of the country, region, or customer they wish to export to. Interviews indicated that many New Zealand exporters market their products as premium quality and exceed the minimum requirements of New Zealand's trading partners. Exporters also have an obligation to uphold New Zealand's international reputation for providing high-quality, safe and unadulterated food products. International food safety incidents associated with New Zealand food are rare but have the potential to do widespread damage to this reputation. The 2013 Fonterra botulism scare is a good example of such an incident (Adams, 2013).

Other Stakeholders

Other stakeholder groups mentioned during the interview process were very similar to key groups mentioned in the literature. These groups have been classified in this research as:

Regulators

The Ministry for Primary Industries sets the legal parameters under which most New Zealand food businesses must operate. MPI has ultimate control over the implementation of food safety interventions in New Zealand. MPI's primary obligation is to the consumers of New Zealand food, with the protection of human health being its number one priority. This emphasis on protecting health was continually emphasised in interviews, and is echoed in key MPI documents such as their Risk Management Framework (NZFSA, 2010b). Secondary to the protection of human health, MPI has a goal to help facilitate overseas market access for New Zealand food exporters (NZFSA, 2010b). Local councils also play a role in food safety regulation by licensing, registering and monitoring new and existing food businesses. Chapter 3 provides a far more detailed overview of how food safety regulation is conducted in New Zealand.

Public Health Authorities

In New Zealand the Ministry of Health (MoH) employs regional Public Health Units (PHUs), whose role is, among other things, the investigation of food safety outbreaks and incidents, as well as reporting of infectious disease surveillance data. The EpiSurv database, administered by ESR on behalf of MoH, is a key tool in monitoring diseases relevant to food safety.

These groups also help to formulate and deliver consumer focused food safety interventions such as education campaigns and, based on frontline experience, are able to offer practical advice regarding the sources of, and solutions to, certain food safety issues. This practical experience complements the work of scientists and regulators in the formulation of risk-management options.

Scientists and Food Safety Academics

Scientists and food safety academics provide science input into high-level decision-making. Clearly, this is a broad group, as there are many aspects of food safety in which there are experts. This group was mentioned by many interviewees as making a significant contribution towards managing foodborne illness. Scientists and food safety academics/experts assist with the development and validation of new control measures and, through their research activities and publications, also help to facilitate open discussion and debate regarding a variety of issues which are relevant to food safety decision-making.

Consumers

Consumers are perhaps the most obvious stakeholders in food safety decision-making. Consumer health may be affected by any changes in the level of risk associated with food products as a result of decisions. Additionally, consumers will be affected by any variations in product choice, quality or price associated with food safety decisions.

Although consumers generally do not play a formal role in food safety decision-making, they are able to participate en masse through the purchasing choices they make. If an intervention has the potential to cause a change in consumer purchasing behaviour then this should be considered during decision-making (Section 6.3.4 further details how consumer perceptions are relevant to the other stakeholders mentioned in this chapter).

Consumers can play a positive role in ensuring food safety by purchasing, storing and preparing food in a safe manner; although many interviewees felt that consumers as a group cannot and should not be relied on to do this.

Māori

Although only 2 interviewees explicitly named Māori as key stakeholders, many Māori have unique cultural beliefs and preferences regarding the way food is produced and prepared. Māori academics whom I consulted regarding this aspect of my research indicated that eliciting Māori viewpoints towards food safety intervention evaluation and selection would involve a substantial amount of additional time, resources and expertise in order to be considered genuine and meaningful consultation. I was unable to take on such additional research. I did speak to a few Māori academics and researchers. These interviewees indicated that Māori may wish to incorporate traditional knowledge and practices into contemporary life, and to consider traditional knowledge systems (e.g. Matauranga) alongside western science during formal decision-making.

5.4 Conclusion

This thesis focuses on methods available for the multi-criteria evaluation of food safety interventions from multiple stakeholder perspectives. However, before any data is collected or any formal modelling conducted it is essential that relevant stakeholders are clearly identified so that subsequent data collection represents the full spectrum of stakeholder views. Nine potentially key stakeholder groups were identified in this portion of the research:

- Regulators
- Public health authorities
- Food safety scientists/academics
- Consumers
- Māori
- Food Businesses (further classified as):
    - Farmers
    - Processors
    - Food retailers
    - Exporters

The categorisation of stakeholders in this context is a complex task, and the above categories are not mutually exclusive. There are some stakeholders who will fall under two or more of these categories and will thus assume the roles and responsibilities of multiple groups. Taking an inclusive approach to defining stakeholders allows for the identification of a wide range of stakeholders so that subsequent data collection and analysis can generate a comprehensive list of criteria and thus capture the full range of potential implications associated with competing risk-management options. An inclusive and systematic approach to data collection in such complex decision-making environments can help facilitate greater buy-in by all stakeholders and a more complete understanding of the different perspectives of the different groups. The research methodology adopted for this thesis is not appropriate for eliciting consumer preferences and this is the primary reason why consumers are not formally consulted in this research. However, this thesis does not dispute the 'key stakeholder' status of consumers, nor does it deny the importance of considering the consumer voice during intervention decision-making. Such stakeholder groups are not always included in the decision-making process. However, it is still important to sufficiently understand and consider their views, as they have the power to affect the successful implementation of different intervention schemes. A genuine attempt should be made to understand and incorporate the views of all key stakeholder groups during food safety decision-making, whether or not these stakeholders are formally included.

Chapter 6: Evaluation Criteria and Measurement – Literature Review

6.1 Introduction

This chapter provides an overview of the criteria utilised in food safety intervention evaluation studies, as well as the most widely used methods for their measurement or quantification. This chapter is intended to help address Research Question 7:

Research Question 7: What criteria are relevant to key New Zealand stakeholders during the evaluation of food safety interventions, and how can performance on these criteria be measured or quantified?

The results of this research will help guide the selection of criteria for use in Chapter 8's case study.

6.2 Methodology

A literature search of the databases Google Scholar, ScienceDirect, Medline and PubMed was conducted over the six months of January to June 2013 (see Appendix C for key search terms). Official documentation from large food safety organisations and regulators was also sought, as well as relevant theses and research papers produced by universities and other research institutions. The search focused on identifying and selecting literature regarding the ex-ante evaluation of food safety interventions, especially research that utilised systematic procedures for the quantification of performance on an explicitly defined set of criteria. References from relevant papers were also examined to identify additional papers that met these selection criteria.

Financial costs/benefits and health benefits (e.g. DALYs) have been identified in this research as important criteria. The most popular methods of quantifying these criteria have been outlined by reviewing a number of key food safety Cost-Benefit and Cost-Effectiveness studies. Relevant government documents such as New Zealand treasury CBA guidelines, MPI's Risk Management Framework and PHARMACs Cost-Utility Analysis guidelines were also reviewed to identify relevant criteria and methods for their measurement.

The following sections detail the criteria identified during this literature review process, and the methods used for their quantification.

6.3 Results

6.3.1 Financial Compliance Costs

These are the monetary costs required for the implementation and maintenance of an intervention, including the costs for new equipment, extra labour, additional training, process verification and other administrative costs. These costs primarily affect food businesses but may also affect regulatory and public health organisations. Antle outlines three methods for the systematic estimation of these costs, all with potential advantages and disadvantages (Antle, 1999):

Accounting approach: A number of firms are surveyed to generate estimates of the extra costs required to achieve compliance. This approach is straightforward and can accommodate a high level of detail, but collecting data from only a few firms means that the entire industry can be misrepresented. This approach can also fail to capture the cost differences between firms of different sizes, operating at different efficiencies, if a range of firms are not surveyed. The accounting approach provides no insight into the underlying cost function for the firm(s) and thus provides no information regarding the potential for firms to adjust to the regulation (E.g. product innovation, technological advancement) (Ragona & Mazzocchi, 2008b). This approach is utilized by Ollinger et al., to estimate the costs of HACCP regulation to meat and poultry processing plants (Ollinger et al., 2004).

Economic-Engineering approach: Detailed engineering data is combined with input cost data to generate a model of the production process, which can then be used to estimate a parametric cost-function for a firm. Changes in costs can then be modelled by changing the input parameters. This approach is time consuming and costly, and to capture the heterogeneity of different plants a model must be constructed for each business (or type of business) (Ragona & Mazzocchi, 2008a). This approach may be easier to employ in New Zealand relative to other large regions such as the USA or EU, due to its small size and the comparatively much smaller number of plants that would require modelling.

Econometric approach: A parametric cost-function is modelled using aggregated industry data. Econometric models can use very large data sets to model an entire industry. These models are based on actual observed behaviour of firms and thus reflect actual choices of managers in the past. It can use real information on how firms responded to previous regulation, revealing how their costs of inputs and prices of outputs have changed as a result, in order to make inferences regarding their future behaviour. This approach lacks the plant level detail available in the two methods above, but could well represent an industry as whole (Lawson et al., 2007). This procedure has been utilised in New Zealand by Cao et al, to estimate the costs to seafood producers of the mandatory HACCP implementation required under the Animal Products Act 1999 (Cao et al., 2003; NZFSA, 2009a)).

6.3.2 Effectiveness

Effectiveness refers to an intervention's ability to reduce the burden of foodborne illness and/or the contamination of food products[16]. The literature identifies a number of ways for measuring effectiveness. The measure of effectiveness utilised in the literature depends on the type of hazard being addressed and the stakeholder perspective taken. Different evaluation techniques correspond to different measures of effectiveness.

Effectiveness measured as a reduction in contamination

Many cost-effectiveness studies take an industry (usually processor) perspective and present results as reductions in contaminant or prevalence levels.

[16] Reducing the contamination of food does not necessarily guarantee a reduction in the burden of foodborne illness.

Reduction in (livestock) prevalence

In the context of food safety, prevalence refers to the proportion of an animal group found to be harbouring a particular microbe or disease. Estimating prevalence reductions is a popular approach used in studies evaluating interventions at the farm or processor level. For example, Baptista et al., and Lawson et al., have modelled the reductions in carcass *Salmonella* prevalence associated with various plant level control strategies in Danish pork abattoirs (Baptista et al., 2006; Lawson et al., 2009).

Reduction in contaminant numbers

Some studies, usually taking an industry perspective, use the achievable reductions in pathogen numbers during production as a measure of effectiveness. For example, Jensen et al estimate the logarithmic reductions in *Salmonella* numbers resulting from a variety of interventions in meat processing plants (Jensen et al., 1998). Measuring reductions on a logarithmic scale, rather than as percentage reductions, is generally seen as preferable when pathogen reductions can span many orders of magnitude. Reductions can be assessed between any two stages in the farm-to-fork chain, but most studies evaluate the reductions that can be achieved during the primary processing stage or during certain sub-stages of primary processing. From a public health perspective it is preferable to estimate the potential reductions that could be realised at the point of consumption.

Effectiveness as a reduction in disease incidence

The effectiveness of an intervention can be reported as an expected reduction in disease incidence (Sears, 2011; Tappero et al., 1995). That is, a reduction in the expected cases of a disease in a given time period (i.e. a reduction in the rate of new cases), generally expressed as the number of cases per 100,000 population. Comparing baseline incidence estimates with the reduced incidence expected under a new intervention is one way to express the effectiveness of an intervention. For the ex-ante evaluation of interventions, estimating effectiveness in this way requires a dose-response model to estimate how potential pathogen reductions translate into a reduced risk of illness, and thus a reduction in disease incidence. This method expresses incidence in a way that is relevant from a public health perspective and gives an idea of how an intervention might contribute to a reduction in the foodborne burden of disease (Havelaar et al., 2007).

Reduction in outbreaks

Havelaar and Bräunig et al, suggest that the effectiveness of interventions targeting microbial diseases such as *Campylobacter* could be better 'measured' by the expected reduction in the number of outbreaks, instead of the overall number of cases (Havelaar et al., 2007).

Effectiveness as a Health-Adjusted Life Year

Health-Adjusted Life Year (HALY) metrics such as the Quality-Adjusted Life Year (QALY) and Disability-Adjusted Life Year (DALY) have become increasingly popular in health technology assessment (HTA). HALYs combine the morbidity and mortality associated with a health condition into a composite metric. Provided data is available regarding the duration and severity of a health outcome, HALY metrics can give a good indication of the effectiveness of an intervention from a public health perspective. The DALY and QALY are the two HALY metrics most commonly used.

The Disability-Adjusted Life Year (DALY) measures the number of years of healthy life lost due to a health condition. It aggregates the Years of Life Lost (YLL) due to premature mortality with the Years Lost due to Disability (YLD), in order to estimate the burden of disease associated with a disease, disability or other health problem:

$$DALY\ Burden = YLL + YLD$$

Total YLL for a health condition is calculated as the sum of the expected residual life-expectancies of each deceased individual; the number of additional years they would have been expected to live had they not been affected by the health condition under study:

$$YLL = \sum_{i=1}^{n} LE_i$$

Where:

n = number of deceased individuals

LE_i = expected residual life expectancy of individual i at time of death

Residual life expectancy information is usually available from period life tables, produced through the analysis of historical data regarding the age of death for a range of different cohorts (StatsNZ, 2014).

YLD is calculated as the product of the number of incident cases of a disease (I), a disability weight reflecting the severity of the disease (DW), and the average duration of the disease (D).

$$YLD = I \cdot DW \cdot D$$

Disability weights are an essential factor for establishing DALY estimates for different disease states. This weight is a value that estimates the (reduction in) quality of life associated with particular health state (e.g. having a broken bone, a mental health condition or being ill with ‘food poisoning’). Disability weights are presented as a number between 0 and 1. The larger this weight, the lower the quality of life. A disability weight of 0.4, for example, would indicate a 40% reduction in the quality of life relative to being in full health. If an individual lived with this health state for one year they would be classified as having lost 0.4 DALYs, or 0.4 years of ‘healthy life’.

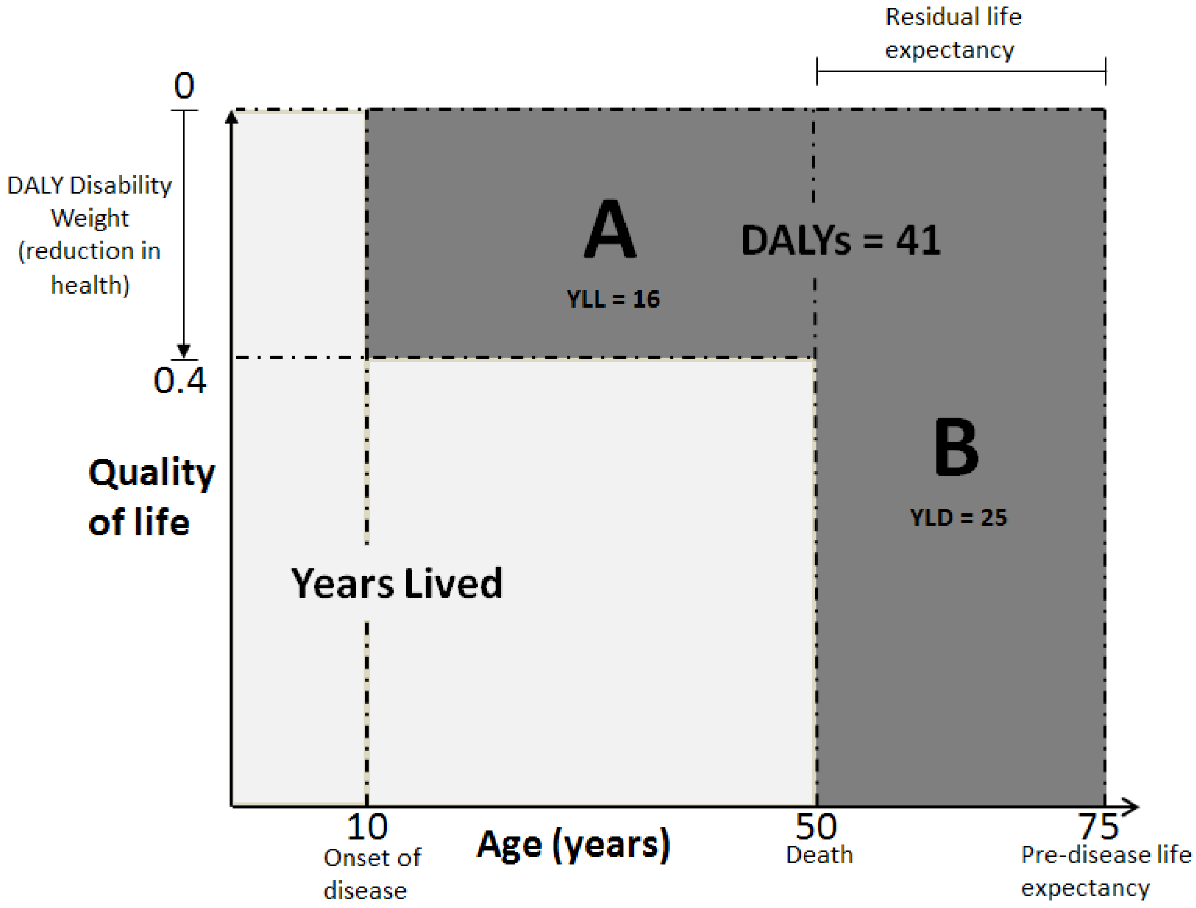


Figure 15: DALY Example: After living until age 10 at full health an individual is diagnosed with a disease. The individual lives for another 40 years, although at a reduced quality of life (40% reduction from full health). The individual dies at age 50, 25 years earlier than would have been expected had they not been diagnosed with the disease (i.e. 25 years residual life expectancy). This represents 25 years of healthy life lost through death (YLL) and 16 (40x0.4) years of healthy life lost through disease (YLD). The total DALY burden for the individual is:

$$DALY\ Burden = YLL + YLD$$

$$DALY\ Burden\ =\ 25\ +\ 16\ =\ 41\ DALYs$$

The aggregate burden of a disease for a population is the total number of DALYs attributable to the disease across the population of interest.

There are different methodologies used to generate disability weights. Generally weights are elicited from a panel of experts (e.g. medical professionals) using the Visual Analogue Scale (VAS), Time Trade Off (TTO) or 'Equivalent numbers' approach (Haagsma, 2010; Stouthard et al., 1997). The DALY approach can estimate reductions in the burden of disease by comparing the estimated burden of disease under each intervention scenario with the status quo (Mangen, Havelaar, et al., 2007).

In contrast to DALYS, which measure the amount of life *lost*, QALYs measure the *gain* in health associated with an intervention. QALY values are a combined estimate of the added life expectancy and the improved quality of life associated with an intervention. The process for calculating QALYs is similar to that of DALYs. 'QALY weights' are the QALY equivalent of the DALY's disability weight, but generally signify an improvement (rather than a reduction) in quality of life.

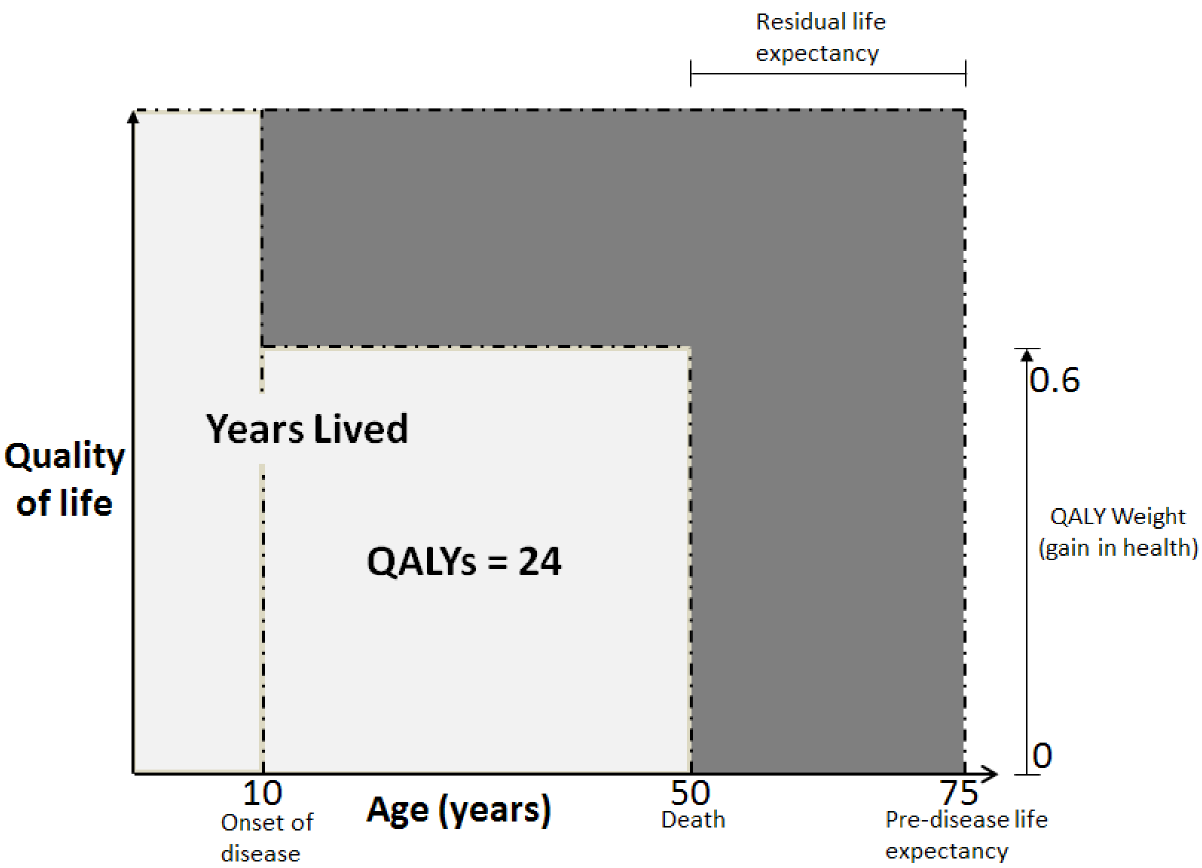


Figure 16: QALY Example: After living until age 10 at full health, an individual is diagnosed with a life-threatening disease. Without life-saving treatment (which the patient receives), the patient would die at age 10. The treatment extends the patient's life by 40 years, although at a reduced quality of life (40% reduction from full health). The patient dies at age 50, 25 years earlier than would have been expected had they not been diagnosed with the disease (i.e. 25 years residual life expectancy). The treatment is able to *add* 24 (40x0.6) QALYs to their life.

Disability and QALY weights are generally elicited from a panel of judges. Haagsma identifies three main considerations regarding the elicitation of such weights (Haagsma, 2010):

1. Who should make up the panel of judges (e.g. patients, medical professionals, members of the public)?

2. Which valuation method should be used to elicit the weights?

3. Should the panel of judges be given generic or disease-specific information regarding the health states being assessed?

Aside from the DALY and QALY being conceptually different (i.e. health loss vs gain), there are usually methodological differences that set them apart also. These differences often arise from different approaches used to address the three concerns above (see Haagsma (2010) for a more thorough description of the methodological issues associated with developing QALY and DALY weights).

The 1996 Global Burden of Disease (GBD) study, conducted by the WHO, utilised the DALY metric, primarily because conceptually it is the more logical choice for assessing disease burden (QALYs measure health gained rather than disease burden) (Murray et al., 1996). This sparked further development of the DALY methodology in the area of foodborne illness; in particular, the development of disability weights for a number of foodborne illness outcomes (Bonsel et al., 2003; Haagsma et al., 2008; Havelaar & Melse, 2003; Melse et al., 2000). Drawing on this international research, ESR adopted the DALY metric for use in its 2002 risk ranking project to rank selected microbial hazards according to their burden of disease (Cressey, 2004).

The DALY has since been used to estimate the effectiveness of microbial interventions in New Zealand (Lake et al., 2013) and overseas (Mangen, de Wit, et al., 2007), and has also been used in MCDA frameworks for prioritising food safety hazards overseas (Caswell, 2008). The DALY is the more popular of the two HALY metrics for both food safety burden of illness and intervention evaluation studies in New Zealand and worldwide. The adoption of the DALY by the WHO, followed by the significant amount of work developing the DALY methodology for use in food safety research, are key reasons for the DALY's popularity and continued use in food safety research.

Although HALY's hold some key advantages over other metrics used for the quantification of (or reduction in) burden of disease, they are not without their critics. Several potential issues accompany the use of HALY metrics for the allocation of health resources. Many of these issues are of an ethical nature.

HALY metrics support a utilitarian approach to resource allocation. Although this approach may maximise the overall health for a population, the distributional implications of this efficiency-focused approach may be problematic for certain subgroups, who may not be viewed as the best investments in the HALY calculus (Gold et al., 2002).

Interventions that extend the life of disabled or sick people avert fewer DALYs (add less QALYs) than interventions extending the life of those who are not (Anand & Hanson, 1997). HALYs discriminate against individuals with pre-existing disabilities or limited treatment potential. In a similar way, prioritisation by HALYs assigns less value to interventions that save the lives of older people, as they have lower residual life expectancies (Robberstad, 2005). Whether this is a valid assumption is debated. Further to this inherent age weighting, some analyses assign higher weights to years lived at a young age, or during productive years (Barendregt, 1996). Anand and Hanson disagree with this approach and argue for a standard intrinsic valuation of human life, regardless of age (Anand & Hanson, 1997).

Nord et al., argue that HALYs do not reflect society's general preference of giving priority to those worst off (e.g. on the basis of poor health, financial status, or other socioeconomic factors). Similarly, prioritisation by HALYs does not distinguish between life extending and life improving interventions, and does not explicitly reflect societal preferences for saving lives over improving quality of life (Nord, et al., 1999).

Although HALY metrics can be a powerful tool, it may be problematic to prioritise health interventions according to their HALY benefit alone.

Effectiveness as a financial benefit

Cost of illness (COI)

The COI approach estimates and aggregates the (reduction in) medical costs and forgone income due to lost productive work time associated with an illness. In general, the COI procedure requires estimates of disease incidences under different scenarios (i.e. baseline vs intervention). The annual number of cases is divided into severity groups according to the amount of medical care they require, the duration of illness and even the number expected to result in death. The medical and lost productivity costs are then accounted for each group and aggregated to give a total COI figure. Lake, et al have utilised this approach as part of their project to estimate the burden of foodborne disease in New Zealand (Lake, Cressey, et al., 2010). The COI approach is intuitive and easy to understand.

Disadvantages include the difficulty and often high uncertainties surrounding the classification and coding of disease outcomes and the associated costs, as well as its exclusion of less easily measured disease outcomes, such as pain and suffering. Like the HALY approach, there can be disagreement regarding the values of parameters used in the methodology, such as the values of treatment costs and lost wages. Using the COI approach alone may lead to underestimating the benefits of food safety interventions, as it does not account for the (reduction in) suffering and loss of life associated with different health conditions which, although difficult to monetise, is of significant value to individuals (Buzby & Roberts, 2009). COI is utilised in Caswell's MCDA framework for prioritising foodborne hazards (Caswell, 2008).

The COI methodology can be used to estimate the reduction in costs associated with different intervention scenarios. By estimating the financial costs associated with the (reduced) burden of disease under different intervention scenarios, and comparing these costs to the existing financial burden of disease, an estimate of the financial benefits associated with each intervention can be generated. Mangen et al., have used this approach to estimate the monetary benefits associated with different interventions to control *Campylobacter* in the Dutch meat chain (Mangen, de Wit, et al., 2007).

Willingness to Pay (WTP) and Willingness to Accept (WTA)

In the economics literature two similar principles are often employed to generate monetary values of 'non-market resources' such as health and food safety. These principles are Willingness to Accept (WTA) and Willingness to Pay (WTP). WTA is the minimum monetary amount an individual would be willing to accept to give up a resource, while WTP is the maximum they would pay to keep that resource. WTA and WTP figures are usually generated through choice modelling techniques where individuals are surveyed and offered a choice between two or more hypothetical options involving monetary trade-offs for non-market resources (Abelson, 2003). This type of study has been widely used in America to assess the potential financial benefits of various risk reduction strategies (Buzby & Roberts, 2009; Huang et al., 2006). WTP and WTA methods are also recommended in New Zealand CBA guidelines as appropriate methods for valuing non-market resources for subsequent input into CBA; although there are no specific guidelines regarding how these figures should be generated (Treasury, 2005). A main advantage of these methods is that they place dollar figures on impacts and resources, which is appealing to many in government and business, who are able to compare disparate policy options (e.g. health and transport policies). Another advantage is that they attempt to reflect actual preferences for risk reduction and include a valuation of pain, suffering and other intangible costs. A disadvantage is that the validity and consistency of WTP and WTA valuations can be highly uncertain and are sensitive to different populations and risk types (Buzby & Roberts, 2009) as well as question format (Ragona & Mazzocchi, 2008b).

6.3.3 Market Impacts

Domestic Market Effects

Food safety regulations have the potential to cause a variety of unintended economic effects. These effects include shifts in the supply and demand for food products. Partial Equilibrium Models can be used to estimate these effects for single markets. Unnevehr et al., expand this idea to produce a Multi-Market Equilibrium Model to quantify the effects of HACCP regulation in the US meat and poultry industries (Unnevehr et al., 1998). Utilising own and cross-price elasticity data for different meat products, the model predicts the new market equilibriums post HACCP, resulting in an estimate of the financial costs to firms. General Equilibrium Models expand on this idea even further to consider financial interactions between all sectors of an economy. This type of model can estimate the distributional impacts of regulations on various stakeholders in the farm-to-fork chain, such as food businesses and consumers (Ragona & Mazzocchi, 2008a). Such models can be complex and have significant data requirements.

Equity of costs/benefits

Regulations can result in disproportionate costs for the different businesses required to implement them. Small businesses will generally have less ability to implement new interventions that require substantial capital investment, and they may not achieve the economies of scale that large businesses enjoy. Sometimes there are substantial variations in the way competing businesses conduct their operations, and taking a one-size-fits-all approach to food safety regulation may have disproportionately unfair implications for some stakeholders. Antle uses an econometric model and plant level census data to estimate the impact of mandatory HACCP regulation on the variable costs for different sized processing plants in the US (Antle, 2000). Ragona & Mazzocchi also suggest the use of Partial and General Equilibrium Models for estimating such effects (Ragona & Mazzocchi, 2008b). The direct accounting method (see Section 6.3.1) has been used by Crutchfield et al., and Nganje et al., to estimate the costs of HACCP regulation for different sized meat and poultry processing plants in the US (Crutchfield et al., 1997). Onal et al., takes a Linear Programming (LP) approach to illustrate the need for American pork processing plants to consolidate their production into *"larger, more economically efficient units"* in order to stay competitive under (then) new HACCP regulations (Önal et al., 2000).

International Market Impacts

Interventions targeting exported foods can have an effect on bilateral trade. It is possible that trade could be impaired because an importing nation has a problem with an exporter's method of intervention (Buzby, 2003). For example, an importer may have restrictions on products which have been exposed to certain chemicals, drugs or hormones (Johnson, 2010). Similar to the way they are used to estimate domestic market effects, equilibrium models can be used to simulate the global market effects of different protection measures. Orden et at., have utilised such a model to simulate the effects of the 2002 Russian ban on US poultry exports (Orden et al., 2002).

Today, a main reason for trade impacts is due to price changes. In such a competitive global economy, relatively minor changes in export prices can have large effects on the quantities demanded. A common way of modelling bilateral trade is by using gravity models (Ragona & Mazzocchi, 2008a). A gravity model uses econometric data from two trading countries, regarding variables such as price indices, exchange rates, trade data, GDP, geographical location, colonial ties etc., and 'fits' a parametric trade equation to this data, which can then be used to estimate the effects of varying certain parameters, such as price or food safety policy. Gravity models have been especially utilised to estimate the effects of modern food safety regulations (e.g. HACCP) imposed by developed countries on developing nations (Anders & Caswell, 2009; Chen et al., 2008; Otsuki et al., 2001). Li, et al., have conducted an extensive meta-analysis of studies utilising the gravity model approach for assessing the implications of international food safety policies on bilateral trade (Li & Beghin, 2012). Disadvantages of gravity models include their data requirements and, being macroeconomic models, they can only provide aggregate results, which may not necessarily be applicable to individual firms (Ragona & Mazzocchi, 2008b).

6.3.4 Consumer perception and choice

Consumers base their purchasing decisions on a number of product attributes. Consumption changes in response to price changes are relatively easy to model using, for example, equilibrium models (see Section 6.3.3). In contrast, estimating demand changes resulting from variations in other product attributes such as safety, quality, taste, convenience and nutrition are much harder.

The most common way to quantify these changes is using stated preference techniques and attempting to generate an aggregate consumer preference structure. This generally involves surveying a group of consumers and asking them questions relating to how important they regard non-price attributes relative to price. Some of these methods place a dollar value on these other attributes (i.e. Willingness-to-Pay studies). For example, some New Zealand consumers are willing to pay more for fresh chicken than frozen chicken despite fresh chicken containing (on average) far greater levels of bacterial contamination (ESR, 2008).

Contingent Valuation (CV) is a popular WTP method for valuing non-market product attributes such as safety (or perceived safety). Generally, consumers are surveyed regarding the amount they would be willing to pay to obtain a specified increase in a desirable product attribute (e.g. safety, taste, convenience) or reduction in a negative attribute (e.g. fat or sugar content).

Conjoint Analysis (CA) is a similar method in which respondents are asked to rate or select their preferred product from a hypothetical range of products with different and clearly defined combinations of product attributes, in order to make inferences regarding the monetary values and relative importance of these different attributes. This method has been used by Halbrendt et al., to construct a preference model for pork based on three attributes (fat reduction, price and production technology) (Halbrendt et al., 1994).

Experimental auctions are another popular method for generating WTP figures associated with product attributes. In CV and CA, participants understand that they are in a hypothetical situation, which can result in biased results. Experimental auctions attempt to overcome this issue by utilising real products and real money.

Hedonic Pricing is another method, different from those mentioned above. Hedonic Pricing relies on actual historical price and consumption data to estimate a *"hedonic function"* (Ragona & Mazzocchi, 2008b), relating the overall price of food products to their specific product attributes. The modelling procedure can be relatively simple and can estimate the marginal contribution of each product attribute to the overall price. Hedonic pricing is mentioned in New Zealand treasury CBA guidelines as a suitable revealed preference method for valuing non-market resources (Treasury, 2005).

6.3.5 Quality of Science

The Codex Commission emphasise the importance of basing decisions on high quality evidence and utilising "*all newly generated data in the evaluation and review of risk management decisions*" (FAO, 2007a). The New Zealand Ministry for Primary Industries has also identified a strong desire for basing food safety decision-making on "*the best available science*" (NZFSA, 2010b). MPI does, where appropriate, follow systematic processes for assessing bodies of evidence, with a particular preference for basing decisions on studies of high methodological quality such as systematic reviews and meta-analysis (NZFSA, 2011). However, assessing the quality or sufficiency of the scientific evidence in favour of a particular risk management option is a flexible process and decisions regarding the nature and quantity of evidence required are generally dependent on the specific hazards and risks involved (NZFSA, 2010b).

Fazil et al., explicitly consider the "*weight of evidence*" in their food safety MCDA framework (Fazil et al., 2008). They do this by defining a strength-of-evidence (SoE) index, taking into account the number of studies addressing the effectiveness of an intervention, their methodological quality and the ratio of favourable to non-favourable studies. Information on these three criteria is synthesised into a single number representing the overall weight of evidence associated with each intervention. This process could perhaps be thought of as a type of simple meta-analysis. This study is unique as it is the only food safety MCDA framework identified in this review which quantitatively assesses weight-of-evidence in its own standalone criterion.

6.3.6 Environmental impacts

Environmental impacts of food safety interventions and regulations are considered important by most regulatory bodies, including those in New Zealand (Treasury, 2013). The effect of policies such as allowing or restricting certain agricultural chemicals and the setting of Maximum Residue Limits can bring real environmental effects through alterations to farming or production methods. Such effects can be very hard to monetise, but methods are available for their non-monetary quantification. Ragona and Mazzocchi suggest Life-Cycle Analysis (LCA) as a particularly relevant method for eliciting the effects on the environment of food safety interventions (Ragona & Mazzocchi, 2008b). Life-Cycle Analysis involves constructing an inventory flow model showing all of the relevant inputs and outputs that might have an effect on the environment at each stage of production. For example, water, energy and raw material inputs, as well as releases into the environment (e.g. pollution) should all be included in the model. An assessment of the effects of these flows on the environment is then conducted and any implications for the environment are presented as results. This type of analysis could be used to rank intervention options in terms of their effect on the environment. Criticisms of this approach include its inability to monetise environmental effects, which often leads to their exclusion from CBA. Additionally, LCA has substantial data requirements and the costs of conducting such an analysis may not be seen as justified, especially if 'environmental impact' is not a criterion given priority in the intervention evaluation and selection process.

6.3.7 Ease of implementation

Ease of implementation, also referred to as 'practicality' or 'practicability', generally considers the unique characteristics of an intervention that can allow (or prevent) its smooth and timely implementation. Both MPI and Codex acknowledge that interventions need to be practical to be successfully implemented (FAO, 2007a; NZFSA, 2010b). Ease of implementation will likely be examined on a case-by-case basis (Fazil et al., 2008).

Fazil et al., consider ease of implementation *("Practicality")* as an important standalone criterion in their food safety MCDA framework (Fazil et al., 2008), but acknowledge that it is a *"relatively subjective measure…that clearly needs expert and stakeholder input, particularly from those that will be implementing the intervention"*. Nevertheless, they consider it important to incorporate these subjective measures into decision-making in a systematic and transparent process. To them, a practical intervention will involve minimal infrastructure changes, minimal alterations to current industry practices and minimal changes to *"labour intensity"*. They propose assessing practicality on a 5-point Likert-type scale (Low … Very high).

6.3.8 Other criteria

Technology and innovation

Henson and Caswell and Ragona and Mazzocchi suggest that process standards (e.g. specific mandated interventions) can impede research and innovation in the associated food sector (Henson & Caswell, 1999; Ragona & Mazzocchi, 2008b). In contrast, product standards (e.g. maximum limits) can give firms greater freedom regarding the technologies they use to meet them, which could lead to a greater level of research and development into promising technologies or interventions. Ragona et al., include this as a criterion in their SCRYER[17] MCDA model by qualitatively evaluating the severity, scale, likelihood and uncertainty of the intervention's impact on research and innovation (Ragona et al., 2011). This criterion is also mentioned in the MPI RMF as one which is considered during the *"identification and selection of risk management options"* process (NZFSA, 2010b).

Animal welfare

Animal welfare is a qualitatively assessed criterion in the SCRYER model of Ragona et al., (Ragona et al., 2011). It is also mentioned by the FAO as a possible criterion to consider during the design of food quality systems (Kreiger, 2007). This review located no studies which had systematically incorporated this criterion into the intervention evaluation or decision-making process.

---

[17] Definition: SCRYER- To see or predict the future by means of a crystal ball.
http://en.wikipedia.org/wiki/Scrying

6.4 Measurement Issues

In practice there will usually be criteria for which there is no natural way of quantitatively measuring. In this case there will be a need to qualitatively assess the performance of interventions on these criteria if they are to be included in decision-making. Most MCDA models utilise some sort of rating scale to subjectively assess these criteria. The two most common are discussed below.

Likert type-scales

Likert-type scales are a very popular psychometric scale used in questionnaires and other situations where participant preferences are to be estimated. It is the most widely utilised approach for scaling responses in survey research. Likert-type scales involve participants assigning a level of preference for an action or a level of agreement with a statement on a symmetrical agree-disagree scale. For example, a typical 5-point Likert-type scale may look something like:

1. Strongly agree
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Strongly Agree

Whether items on such scales should be considered as ordinal or interval data is subject to some debate in the literature (Brown, 2011). Likert-type scales are traditionally treated as interval scales, with the difference between each successive category assumed to be equidistant. Well-designed scales should present a symmetry of (ideally seven) clear and well-defined linguistic response categories about a central, usually neutral, midpoint, with the choice of response categories inferring equidistance (Allen & Seaman, 2007). When this is the case the scale is more likely to behave similar to a quasi-normal, interval-level measurement scale, even though ordinal categories are being used. Likert-type scales are most commonly treated as interval scales and are analysed with descriptive statistics such as the mean and standard deviation, as well as inferential statistics like correlations and factor analysis. There are several pieces of literature supporting this approach for well-designed scales (Baggaley & Hull, 1983; Maurer & Pierce, 1998; Vickers, 1999). Such scales have proved very useful in situations where variables are unable to be quantitatively measured. Such rating systems are commonly used in MCDA analyses (Fazil et al., 2008; Ragona et al., 2011; Sauian, 2010), and have been explicitly incorporated into the Analytical Hierarchy Process (AHP) family of MCDA models (Saaty, 1980).

Direct rating scales

Direct rating scales are another main type of scale utilised to assess criteria for which there is no natural unit of measurement available. The concept of direct rating is simple; decision-makers or experts rate a given option on a numerical interval scale, usually 0-10 or 0-100, with a higher number generally indicating better performance on a particular criterion.

A local scale is one on which the best performing option currently included in the analysis is assigned the top possible value (e.g. 100 on a 0-100 scale) with the worst performing option assigned the value of zero. All other options are scaled somewhere between these two numbers based on their relative performance compared to these end points. There is always at least one option that achieves the maximum possible score and at least one achieving the minimum possible score. Local scales are seen as easier to work with when dealing with qualitative assessments, as the user is not required to consider hypothetical end-points for the scale, as is required with global scales (Mabin, 2006).

A global scale is one on which hypothetical best and worst values from the decision-makers own judgements, experience or aspirations are assigned the maximum and minimum possible values. From here, all options under evaluation are rated on the scale relative to these hypothetical extremes. Usually none of the currently available options will receive the minimum or maximum possible scores. Global scales are more useful for analyses where new options may be introduced into the analysis at a subsequent stage, as scores will not need rescaling if the new option(s) scores higher (lower) than the current best (worst) performing option.

6.5 Conclusion

This section has provided an overview of the main techniques utilised in food safety intervention evaluation studies for quantifying a range of potential criteria. Most techniques have unique advantages and disadvantages. For many stakeholders there is a real desire to employ techniques which can attach a dollar figure to certain impacts. However, many of the potential impacts associated with food safety interventions are difficult to reliably measure, and especially place a dollar value on, even when utilising some of the advanced techniques mentioned in this chapter. Even for well-designed and widely utilised techniques there is often a lack of sufficient data to allow meaningful and reliable quantitative estimation. For regulators, macro level models will likely be highly useful to help generate an estimate of the aggregate effects of policy or intervention option. However, regulators may also wish to model the distributional effects of such options, which requires the use of micro level models. For individual firms or sectors, macro level models may not be overly useful as it is the effects at the individual firm level which are of most concern, and quantitative estimates of these effects can only be achieved through micro level models. Sometimes the only practical option may be to qualitatively measure some criteria on, for example, Likert-type scales, or using direct rating methods. Such scales and rating methods have been widely utilised in a range of MCDA applications. Decisions regarding how to measure certain criteria will likely depend on the stakeholder perspective, data availability and the required level of detail and precision.

Chapter 7: Criteria Highlighted by Stakeholders

7.1 Methodology

This chapter details the criteria that were identified as relevant to food safety intervention decision-making by interviewees. Interviewees were identified and classified according to the methodology described in Section 5.2. They included members from government, food businesses, academia, public health and Māori. The goal of this chapter is to help address research question 7:

Research Question 7: What criteria are relevant to key New Zealand stakeholders during the evaluation of food safety interventions, and how can performance on these criteria be measured or quantified?

Semi-structured interviews were used to address this question. Interviewees were recruited using a combination of purposive and snowball sampling (detailed in Section 5.2), where participants were identified based on their ability to provide insight into this research, and subsequently asked to identify other individuals or organisations that could potentially contribute to this research. Each interviewee was asked the same nine questions outlined in the participant information sheet (Appendix B). The questions most relevant to this Chapter are questions 1, 2, 3 and 4 from the information sheet:

1. *What criteria would you consider relevant when evaluating potential food safety interventions?*
2. *How would you go about 'measuring' the performance of an intervention on each criterion? (I.e. what key data would you require for each criterion?)*
3. *Are there any additional criteria which should be included in these evaluations but are often excluded?*
    a. *Why are they excluded?*

b. *How could/should these extra criteria be measured or estimated?*

4. *Which criteria (if any) are the most important in your opinion? (I.e. which should be given the most weight in the final decision?)*

Interviewees were asked additional follow up questions to help clarify their answers and to provide examples where possible. In order to provide interviewees with an idea of the type of information I was seeking, examples and possible interpretations of two criteria that might be relevant to their decision-making processes were provided; financial cost and intervention effectiveness. Participants were asked if these criteria were, in fact, relevant to their decision-making processes, and were subsequently questioned about additional criteria.

Participants were not prompted for answers (e.g. additional suggestions of criteria or measurement) unless there was some confusion surrounding the question being asked. Interviews ranged in length from around 15 minutes to over three hours, with the average interview lasting 30-45 minutes. Interviews were conducted until what is described as the point of saturation (Glaser & Strauss, 1967), where the information generated from interviews, especially the list of criteria, became repetitive, with further interviews revealing little additional information. The number of interviews required to reach this point of saturation was 55. Most interviews were conducted over the telephone (n=44), with the remainder conducted as face-to-face interviews (n=8) or as email exchanges (n=3).

The following sections have been organised by stakeholder group (identified in Chapter 5), with the criteria relevant to each group identified and detailed. In answering interview question 4:

"*Which criteria (if any) are the most important in your opinion? (I.e. which should be given the most weight in the final decision?)"*

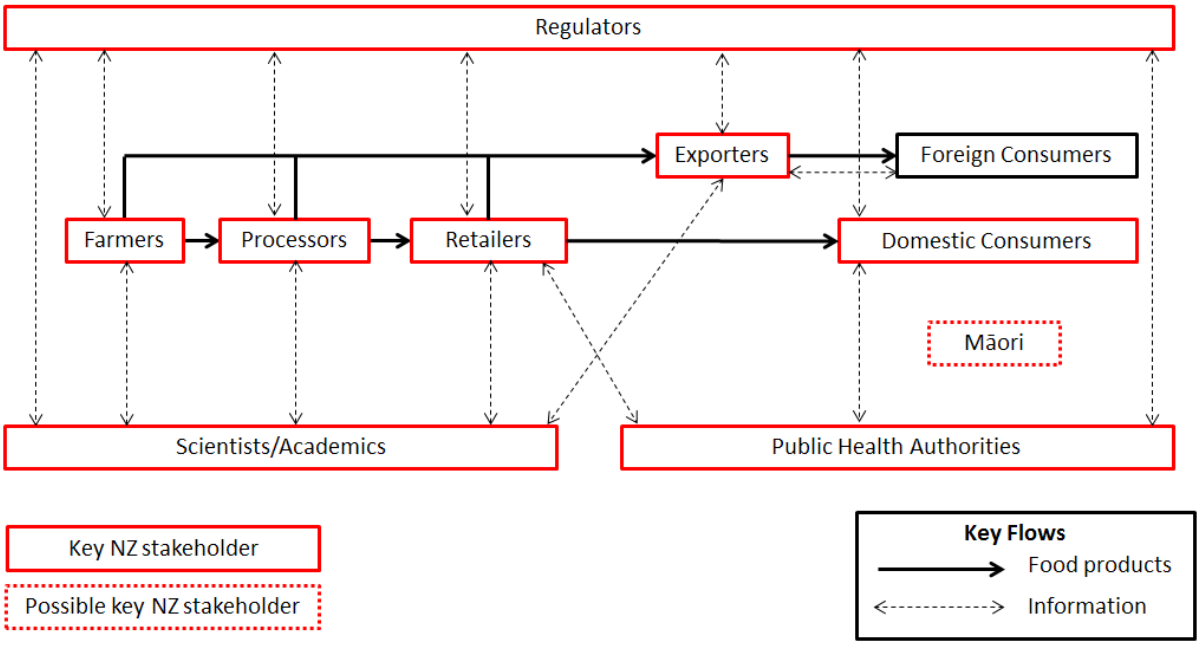
Most interviewees were able to identify the two or three most important criteria from their perspective, but were usually unable to provide a definitive ranking of all criteria. Participants within the same stakeholder group sometimes gave different rankings for criteria, omitted criteria altogether, or found it very difficult to rank criteria that they felt were all crucial to decision-making. I have ordered the criteria in each section according to the responses to question 4, as well as my own perceived level of emphasis placed on the criteria by interviewees. These orderings should be seen as general indications of the amount of emphasis placed on criteria by interviewees, rather than definitive rankings. The general agreement amongst the different groups was that although some criteria may be weighted more heavily than others, all have an important role to play in decision-making when relevant.

Each of the following sections begins with a brief summary of the criteria relevant to interviewees within each stakeholder group, which is then followed by a more detailed explanation of why each criterion is relevant to that group.

7.2 Results

Interviews were used to identify the criteria relevant to different stakeholder groups during the evaluation of food safety interventions. These stakeholders included:

- Regulators
- Food Businesses (further classified as):
    - Famers
    - Processors
    - Food retailers
    - Exporters
- Public health authorities
- Food safety scientists/academics
- Māori



Note: The nature of Māori communication with other stakeholders is unknown

Figure 17: Key stakeholder relationships in the farm-to-fork chain *(Figure is a repeat of Fig 14 for convenience)*

The criteria identified by these stakeholders are summarised as:

Effectiveness: Refers to how effective an intervention is at reducing contamination or eliminating the risk associated with a food safety hazard. There are a number of different ways to measure effectiveness, depending on the stakeholder perspective taken and the type of hazard being addressed.

Financial Cost: The direct costs of implementing and maintaining an intervention in NZ$.

Market Access: Some interventions may have implications for exporters wishing to meet the standards of foreign regulators or customers, and thus increase (or maintain) export volumes or value. An intervention could facilitate additional market access through the meeting of international standards that were previously unmet. Alternatively, an intervention could restrict access to some degree. There are generally two types of market access requirements most relevant to exporters; those imposed by the official regulatory body of an overseas trading partner and those imposed by an international customer such as a large fast food, hotel or supermarket chain. The latter requirements are usually stricter than those of the regulatory bodies of the countries within which they operate. Both types of requirements usually come in the form of product requirements (e.g. MRLs, Microbial limits, temperature limits etc.), but sometimes also include process requirements (i.e. detailing exactly how the product specifications are to be met).

The ‘Market Access’ criterion is distinguished from ‘Consumer Perceptions’ (in the international context) in that it covers the minimum requirements that an exporter *must* meet in order for its products to be accepted into a foreign market, as opposed to the preferences of overseas consumers that an exporter *chooses* to meet (or not to meet). Both criteria can have an impact on the level of foreign consumption of New Zealand food, but ‘Market Access’ refers only to the mandatory requirements imposed by foreign food safety regulators or customers. Foreign markets and customers impose concrete standards on the food products coming from New Zealand and these are market access requirements. There is however, some degree of flexibility given to exporters regarding how these standards are met, and it is these flexible aspects of production which have the potential to affect overseas consumer perceptions, and are captured by the ‘Consumer Perceptions’ criterion.

Consumer perceptions: The potential for an intervention to be perceived in a positive or negative manner to the extent that it may alter consumer purchasing behaviour. Additional to regulatory bodies assessing the safety of New Zealand food, consumers may further scrutinize New Zealand’s products and production processes on a number of criteria. Like domestic consumers, foreign consumers may make purchasing choices based on how safe they perceive products to be, and these choices, unlike the choice of regulators to allow products into the country, may not be based on any scientific determination of risk. That is, despite a regulatory body determining that a New Zealand product is safe, consumers may still consider it ‘risky’ for one reason or another, and may avoid it. Furthermore, consumers may also make purchasing choices based on non-safety attributes such as perceived quality, organic status and environmental impact (e.g. carbon footprint). In short, regulatory bodies may not always care so much how New Zealand food is made safe, so long as it meets the product standards they impose. However, consumers may have certain negative or positive perceptions of the processes used to make food safe and may alter their purchasing behaviour accordingly.

Ease of Implementation: ‘Ease of Implementation’ has a various meanings depending on the stakeholder perspective taken. Some have described ease of implementation as *“practicality”* or *“practicability”*. To most interviewees an easy to implement intervention is auditable, easily integrated with current (e.g. factory) infrastructure, requires minimal training or new expertise and will not slow production or delivery times. ‘Ease of implementation’ is probably the most subjective criterion of those described in this chapter. Although some interviewees included aspects of other criteria when describing this criterion (e.g. safety), in the context of this research I am defining ‘Ease of Implementation’ as the unique characteristics of an intervention which can affect how easily it can be implemented, but are not covered by other criteria.

Quality or Suitability: Suitability refers to the non-safety characteristics of a food that make it suitable for its intended purpose. There is the potential for an intervention to cause physical changes to the food it is applied to, including alterations to organoleptic properties such as taste, texture, smell and colour. Such changes may be viewed by consumers or customers as a reduction in quality, and they may reject it (or demand lower prices). Some interventions may also improve the quality and/or suitability of a product. This criterion is differentiated from ‘Consumer Perceptions’ in that it involves real, physical changes to food as a result of an intervention, rather than perceived changes.

Quality of Science: ‘Quality of Science’ refers to the quantity and quality of the evidence in support of the effectiveness of an intervention in a real world setting.

Equity of Costs: The distribution of costs across the various stakeholders when implementing an intervention. Interventions are sometimes implemented by one stakeholder group yet bring benefits for stakeholders further down the farm-to-fork chain. Some interviewees indicated that the costs of interventions should be distributed across the stakeholders who stand to benefit from them. Most agreed that the equitable distribution of costs could be achieved if stakeholders implementing the intervention(s) could recoup their costs through higher prices.

Equity of Benefits: Refers to the distribution of health benefits when implementing an intervention and whether an intervention carries unique health benefits for certain groups. Decision-makers may wish to prioritise interventions that address health inequalities or reduce risks for vulnerable populations such as pregnant females and the elderly.

Safety: The potential for an intervention to carry additional risks of illness or injury to those implementing and maintaining it.

Animal welfare: The potential for an intervention to cause changes to the standard of animal welfare during the farming stages of food production.

Cultural impact: The ability of an intervention to have unique implications for specific cultural groups (e.g. Māori).

The following table provides a quick reference guide to the criteria identified by the stakeholders interviewed as part of this research. Although there may be other criteria relevant to these stakeholders, the table below contains only the criteria which were explicitly mentioned by interviewees.

	Cost	Effectiveness	Market Access	Consumer Perceptions	Quality of Suitability	Ease of Implementation	Quality of Science	Equity of Costs	Equity of Benefits	Safety	Animal Welfare	Cultural Impact
Regulators	✓	✓	✓	✓			✓	✓				
Farmers	✓	✓		✓		✓		✓		✓	✓	
Processors	✓	✓		✓	✓	✓	✓			✓		
Retailers	✓	✓		✓	✓	✓						
Exporters	✓	✓	✓	✓	✓		✓					
Public health authorities	✓	✓		✓		✓	✓		✓			
Scientists/Academics	✓	✓		✓			✓					
Māori												✓

Table 5: Summary of the criteria explicitly identified as relevant by interviewees within the different stakeholder groups.

The following sections detail the relevance of these criteria from the perspective of these different stakeholder groups.

7.2.1 Regulators

Summary

Regulators whom I interviewed emphasised that their primary goal is the protection of human health. From a public health perspective, quantifying effectiveness in financial terms or using HALY metrics was viewed as appropriate. Interventions aimed at chemical hazards were said to be effective if they were able to consistently meet relevant MRLs, which should reflect a notionally zero lifetime risk to consumers.

Interviewees from MPI all agreed that market access is very important to all exporters and that interventions should enhance, and not restrict, market access.

Quality science establishing intervention effectiveness was identified by MPI interviewees as one of the most important criteria to regulators.

Financial cost was also acknowledged by these interviewees as being relevant to, but not the focus of, nearly all intervention decision-making.

Consumer perceptions were mentioned as important for MPI and other organisations to consider when contemplating different interventions, as they have the potential to affect the successful implementation on an intervention. These interviewees acknowledged that predicting consumer responses to interventions was difficult.

Equity of costs was also mentioned by a couple of MPI interviewees. They felt that there should not be any disproportionate costs to any stakeholders when implementing new interventions.

Effectiveness

Interviewees from MPI identified effectiveness as a crucial attribute of any food safety intervention. Protecting the health of consumers is MPI's number one focus and from a public health perspective there is a preference to assess effectiveness using metrics that give an indication of the actual health benefits realised by consumers. For microbial hazards 'Effectiveness' can be measured as reductions in the incidence of certain foodborne diseases. Going one step further, measurement of the reduced burden of disease could be measured in Health-Adjusted Life Years (HALYs) such as DALYs or QALYs, which give meaningful estimates of disease burden and allow comparison of risks from disparate sources (see Section 6.3.2 for further discussion of HALY metrics). DALYs were preferred over QALYs by this group, due to the DALY methodology being more developed and widely used in food safety research, both domestically and internationally.

Rather than expressing effectiveness using a public health metric, the government may wish to express effectiveness as a dollar amount associated with a reduction in medical expenses and lost productivity due to foodborne illness. Additionally, pain, suffering and lost (or gained) life may also be monetised. In the past the government has assigned a value to the burden of foodborne disease by monetising DALY estimates and using Value of a Statistical Life-year (VSL) estimates (Gadiel & Abelson, 2010). The benefit of monetising health gains from interventions is that health policies can be compared to policy options in other sectors (e.g. transport). It is also possible to utilise both HALY and financial Cost of Illness estimates when communicating disease burden (Lake, Cressey, et al., 2010).

Interventions aimed at chemical hazards generally involve regulators setting maximum limits on chemicals in food in accordance with international limit setting practices; with producers modifying their farming and processing practices accordingly to meet them. From the perspective of regulators, and in the context of public health, an effective MRL is one that should, in theory, result in a *"notionally zero"* lifetime risk to consumers (see Section 3.7.2.2). A chemical intervention is seen as effective by regulators if it can consistently achieve chemical residue levels below relevant MRLs.

Market Access

The MPI has a strong interest in facilitating and enhancing overseas market access for New Zealand food exporters as part of the wider government plan to grow the New Zealand economy. It is primarily the responsibility of exporters to make sure that their products comply with international standards, but MPI will support exporters in streamlining the export process where possible.

MPI has suggested that changes to market access could be quantified as a dollar figure indicating the possible financial benefits (losses) of increasing (reducing) market access. Semi-quantitative or qualitative assessments of these financial gains or losses were suggested as a possible option when information is lacking.

Quality of science

MPI places huge emphasis on the need for high-quality evidence before taking action against a particular food safety issue. Although MPI has no current definition or criterion regarding what it calls *"sound science"*, it does endeavour to draw upon the best available evidence when making any food safety decision (see Section 3.8).

For many high profile food safety issues, the government has required systematic reviews, either conducted abroad or within New Zealand, by trusted organisations, before taking action. Interviewees indicated that MPI's desire for *"unequivocal"* evidence has resulted in MPI opting not to follow major trading partners in addressing some potential food safety issues. For example, despite the US and EU ban on baby bottles containing Bisphenol-A (BPA), New Zealand has opted not to ban these due to the evidence against BPA being inconclusive (NZFSA, 2010c). Interviewees from MPI stated that MPI's preference is to systematically and independently consider all of the empirical studies surrounding a particular topic, rather than the conclusions reached by individual members of the scientific community. MPI has its own experts who will make judgements and recommendations to decision-makers once the evidence has been assessed.

Interviewees indicated that assessing the sufficiency of evidence in support of a certain action is conducted on a case-by-case basis and depends on a number of factors such as the potential consequences of making an incorrect decision. In the words of one interviewee, the level of evidence required depends on *"how high the stakes are"*.

Financial Cost

The financial cost of implementing an intervention was identified as relevant by MPI interviewees, who believe that an intervention should not overly restrict the ability of food businesses to make a fair profit. Although there are minimal costs to MPI associated with setting new performance targets or limits for industry, interviewees revealed an interest in maintaining the competitiveness of New Zealand food industries and the health of the economy in general; recognising that over-regulation could stifle growth. One interviewee suggested that MPI might prefer to see as many criteria as possible monetised and aggregated into a cost-benefit ratio. Cost-Benefit Analysis is widely used and understood by government decision-makers, and interventions that require large capital expenditures will be given greater consideration if they have favourable cost-benefit ratios.

Consumer perceptions

Interviewees from MPI indicated that MPI is concerned not only with protecting the New Zealand *"clean and green"* brand overseas, but its own MPI brand too. MPI strives to be seen as one of the most advanced, efficient and trusted food safety agencies in the world. One interviewee from MPI stated, in relation to the MPI brand, that *"we have a lot going for us, we have low corruption, we're honest, and we tailor interventions to our specific risks… we need to protect this image"*. New Zealand places very high standards on its foods, and when a food business fails to meet these standards MPI has a duty to intervene, not only to protect the health of consumers and the reputation of responsible food producers, but also to strengthen MPI's reputation for taking quick and effective action when required. This reputation is important for maintaining the trust of overseas trading partners, the New Zealand public and other stakeholders.

MPI interviewees indicated that they are sometimes forced to divert resources away from significant issues towards debunking food safety *"myths"*, or allaying consumer fears regarding what they see as perceived, or relatively minor issues.

Interviewees from MPI expressed a preference for interventions that would not be perceived negatively by consumers, especially in overseas markets. There was agreement that increasing the public's understanding of new interventions is important to help consolidate expert and public assessments of risk and reduce the chances of unintended consumer reactions. However, low consumer awareness and a general lack of interest and engagement from the public regarding scientific issues was seen as an obstacle. Increasing MPI's social science capacity was seen by interviewees as an important avenue for understanding more about the consumer.

Equity of costs

Interviewees expressed a need to allow flexibility where possible regarding the way different types and sizes of food businesses are allowed to implement regulations. These interviewees acknowledged that different business will not have equal capacity to implement changes in the same way or at the same rate. Interviewees generally agreed that large businesses often have a greater ability to incorporate a range of new intervention measures, whereas smaller businesses may not have the capacity to implement the same exact interventions as larger businesses, and may not have the resources available to alter their production processes to meet new standards as rapidly as larger businesses. Interviewees indicated that this issue requires a balancing act from MPI, who want all operators to conform to new standards as rapidly as possible, but without placing potentially unfair and unachievable demands on the less able.

7.2.2 Farmers

Summary

Effectiveness was important to the farmers and farming representatives interviewed. Measuring reductions in on-farm prevalence of diseases was agreed as being appropriate for measuring the success of many interventions aimed at microbial and parasitic hazards in livestock. If interventions resulted in food that consistently met MPI's standards (e.g. MRLs) then they were generally seen as effective.

Financial cost was important to all interviewees in this group.

The fair distribution of intervention costs were mentioned as important by three farmers, who felt that farmers sometimes bore a disproportionate share of intervention costs.

Most farmers agreed that consumer perceptions were important to consider in intervention evaluation. A couple of farmers felt that consumer perceptions were not always based on quality evidence, and this was a concern for them.

Ease of Implementation was important to all farmers interviewed, who are generally under significant time pressures and expressed a desire for interventions that placed minimal additional burdens on them and their staff, and that were easily integrated with existing farm infrastructure.

Two farmers mentioned Safety as an important criterion. Interventions that carried additional risks that could not be managed were unlikely to be implemented by these farmers.

Two farmers of livestock expressed a desire for interventions that improve, and certainly not compromise, the welfare of their animals.

Effectiveness

Farmers and farming representatives agreed that the effectiveness of an intervention was the most important criterion. Assessing the effectiveness of on-farm interventions in livestock was suggested as being most appropriately measured by the ability to prevent the spread of a microbe or parasite throughout a group of animals, thus reducing the number of carrying animals, referred to as *"on-farm prevalence"*, usually expressed as a percentage.

The effectiveness of interventions aimed at reducing microbial contamination of fruits and vegetables could be assessed by their ability to reduce the public health burden in consumers, or the number of detections or rejects associated with contaminated harvests.

Farmers of all types also agreed that, for chemical residues, any intervention which consistently met relevant MRLs would be considered effective.

Financial cost

Farmers and farming representatives indicated that cost would be a substantial factor when deciding whether to implement any new on-farm intervention.

Equitable distribution of costs

The fair distribution of costs was seen by most farmers as an important criterion. These farmers felt that on-farm interventions could be implemented and financed solely by farmers and yet brought benefits for stakeholders further down the farm-to-fork chain. This could put potentially unfair additional financial pressures and workloads on farmers. One farming representative from a meat industry emphasised that systematic costings of on-farm interventions are seldom done. One farming representative mentioned that large customers had the potential to use their market power to unfairly *"squeeze"* farmers into lowering costs or meeting other requirements such as food safety standards. Farmers also felt that there should be scope to customise interventions to the different types of farm. Interviewees agreed that interventions where costs could be recouped through higher selling prices at the farm gate would be fair.

Consumer perceptions

As consumers become more interested and aware of how their food is produced they have an increasing say regarding which farming practices are acceptable. Consumer attitudes towards issues such as animal welfare and chemical residues seem to be putting pressure on farmers to adopt more 'acceptable' farming practices. Three farmers expressed frustration that consumer segments held strong views regarding some farming practices, yet had a limited understanding of how they actually worked and why they were being used. Some farmers felt that the media and lobby groups were misinforming the public and perpetuating negative and *"one sided"* views of farming practices. Overall, farmers agreed that it was becoming increasingly important to consider how the public may react to any on-farm food safety intervention during the intervention evaluation phase. One farming representative suggested that qualitative research (e.g. *"focus groups"*) could provide an indication of how consumers may react to a new intervention.

Ease of implementation

Farmers felt that interventions should be relatively easy to implement and practical for their needs. The most common examples given by farmers regarding what constitutes an easy to implement intervention included:

- Interventions should not put significant additional labour/time pressures on farmers
- Farmers should have the expertise and knowledge to successfully implement an intervention, or should be able to with minimal training
- Interventions should easily integrate with existing farm infrastructure

Some representatives of vertically integrated industries felt that the farm often may not be the best place to implement food safety interventions. There were a number of reasons given for this:

- Farms are often spread out and do not have experts on hand to *"measure everything and check pathogen levels"*, whereas processing plants and factories can achieve economies of scale with their food safety monitoring activities.
- For more complex interventions farmers *"are not adequately trained to implement (them)…they probably don't have the skill set required"*.
- Farms do not have the same level of control over production processes as factories.
- Farmers have a range of duties and tasks to tend to on the farm and are not available to constantly monitor food safety threats.

The assessment of whether an intervention is practicable from the perspective of farmers would likely need to be handled on a case by case basis, as different farmers have different requirements and ideas regarding this criterion.

Safety

The safety of interventions was mentioned as important by two farmers. Some interventions are potentially hazardous (e.g. require the use of dangerous chemicals or carry a risk of physical injury from livestock). Hazardous interventions should be avoided where possible. Additional health and safety risks that could not be minimised to an acceptable level would likely exclude an intervention from selection.

Animal welfare

Farmers of livestock expressed a preference for interventions that could improve, and certainly not compromise, animal welfare. Interviewees that mentioned this criterion generally felt that MPI had a genuine interest in maintaining high animal welfare standards for New Zealand livestock. Interventions that met the MPIs animal welfare standards would likely be viewed by most farmers as acceptable. A smaller organic animal farmer felt that MPI’s animal welfare standards were inadequate, and this farmer would not be implementing interventions that reduced their own animal welfare standards if at all possible, even if they complied with MPI standards.

7.2.3 Processors

Summary

Effectiveness is perhaps the most important criterion to processors. For interventions aimed at addressing microbial hazards, the achievable reduction in pathogen numbers was viewed as the most appropriate way to gauge effectiveness. If an intervention could consistently produce food that met MRLs and other standards then it would be considered effective; although many processors expressed a desire to exceed these standards.

Cost was considered by processors during intervention evaluations, but was not necessarily a focus of their evaluations.

Quality science was mentioned by almost all processors interviewed. Many processors, especially larger ones, go to great lengths to establish and monitor the effectiveness of their food safety systems.

Consumer perceptions were mentioned by processors as having an influence over the interventions they implement. Considering the consumer voice during decision-making was seen as important. However, consumer preferences were mentioned as being very difficult to accurately elicit or estimate.

The quality and suitability of their final products was explicitly mentioned as important by over half of the processors interviewed. Many of these processors market their products as premium products and indicated that it would be highly unlikely that they would implement interventions which compromised the quality or suitability of their foods.

Ease of implementation was also mentioned by over half of the interviewees in this group. First and foremost, interventions should not slow production down by, for example, creating bottlenecks or requiring a significant amount of additional maintenance, calibration or monitoring. Interventions should be easily integrated with existing infrastructure and should require minimal additional training and expertise.

Safety was mentioned by processors as an important criterion to consider. Additional health and safety risks to employees would likely be unacceptable.

Effectiveness

Effectiveness was a highly relevant criterion to all interviewees in this group. Some made a distinction between effectiveness and efficacy. In the medical literature effectiveness refers to how well an intervention works in practice while efficacy refers to how well an intervention works under *"ideal circumstances"* (Gartlehner et al., 2006). Some processors referred to effectiveness as the ability of an intervention to control a food safety risk in a *"commercial setting"*, and efficacy as its ability to control a food safety risk under controlled conditions, often in a lab or controlled trial. Several representatives from large organisations were reluctant to accept that interventions for which there was evidence of effectiveness overseas would also be effective in the New Zealand commercial environment. Differences in New Zealand environmental conditions, genetic variations in our plants, animals and pathogens, soil makeup, animal feed types, farming methods and processing plant layouts were reasons given why these interventions might not be as relevant to New Zealand processors. A representative from the meat industry talked about interventions being very effective overseas but potentially less so for the New Zealand industry. *"Our meat is a lot cleaner* (than some overseas meat) *to begin with"* he said, stating that there is less potential for New Zealand *"gains"* with some of these interventions: *"If you apply an intervention to a dirty product then it will seem more effective than if you applied that same intervention to a much cleaner product"*. In contrast, a small number of processors felt that the process for implementing interventions which had been proven successful overseas should be more straightforward and should not always require domestic trials. One stated that *"we must accept that validated operating parameters overseas are valid over here"*. The need for domestic trials of interventions appears to be context specific. In some cases processors seem happy to implement overseas interventions without trials, but most processors interviewed indicated that trials would be necessary before full scale implementation.

Most processors agreed that the most appropriate way to measure the effectiveness of an intervention aimed at controlling microbial contamination was by its ability to reduce the number of pathogenic organisms on (or in) the final product. For example, in the poultry industry, the effectiveness of interventions aimed at controlling *Campylobacter* is measured as the logarithmic reduction in bacterial numbers they can achieve[18].

Although chemical hazards generally have a less obvious effect on human health than microbial hazards, processors and manufacturers still have an obligation to keep chemical residues in food within agreed limits. Gauging the effectiveness of interventions to control chemical hazards is relatively easy, as if a product exceeds specified residue limits it is considered unacceptable and usually unfit for purpose. An intervention aimed at controlling a chemical hazard either consistently achieves residue levels below specified limits and is thus considered effective, or does not, and is considered ineffective.

Processors of a variety of products were, in addition to being concerned about microbial and chemical hazards, concerned about physical hazards appearing in the final product. For example, bolts from machinery, pieces of packaging and insects were examples of physical hazards which were sometimes discovered in their final products. The type and frequency of

---

[18] Reductions are measured on a logarithmic scale because, when working with large reductions in pathogen numbers, it is often easier to refer to the orders of magnitude of a reduction rather than, for example, the percentage reduction. Often reductions of many orders of magnitude are required in order to make an appreciable impact on public health and it is thought that, for example, a 5-log reduction is more easily comprehended than a 99.999% reduction in pathogen numbers.

these hazards will vary across the different types of processor. Some suggested estimating the success of interventions aimed at reducing the presence of these physical hazards by their ability to reduce the number of complaints associated with specific physical hazards. For example, a small goods processor suggested estimating the effectiveness of installing an X-ray machine at the end of a production line by the potential reduction in the number of complaints due to metal fragments in their products.

Financial Cost

The cost of an intervention was highlighted as an important criterion for all stakeholders in this category. Although not the most important criterion, especially for processors who were also exporting their products, cost would always be considered when assessing potential interventions.

Quality of Science

High quality evidence was extremely important to almost all processors interviewed, as any intervention that does not consistently achieve agreed standards will likely result in rejected product. The industry process for collecting sufficient evidence of effectiveness for new food safety technologies generally begins with a review of reputable, internationally published research, or through information sharing with international industry members. If research indicates that a new intervention could work in the New Zealand commercial environment then a small scale industry trial may be conducted, sometimes with help from the government. Often the government requires industry trials before a new intervention can be approved for commercial use. In big industries such as meat and dairy, larger trials will usually follow a successful pilot trial, with the intervention adjusted and refined along the way. If an intervention is effective in large trials then it will usually be considered appropriate for full scale implementation. This is a flexible process and will not always be followed; it will be dependent on the intervention being studied and the quality of existing research. Some interventions may not require industry trials to establish effectiveness and existing evidence might be sufficient, with any variations in New Zealand commercial or environmental conditions being inconsequential. For example, there is overwhelming international evidence to demonstrate the effectiveness of irradiation at controlling microbial hazards in food, and there are no differences in New Zealand production processes that could mitigate this effect. Industry would likely not require trials to further establish the effectiveness of irradiation (but may conduct trials for other reasons). In contrast, an intervention such as adding a bacteriophage to animal feed would likely require extensive trials to demonstrate its effectiveness in New Zealand. The effectiveness of such an intervention is likely to be sensitive to subtle differences in New Zealand farming conditions.

Consumer perceptions

There was a consensus amongst interviewees that consumers will be adverse to some new technologies or processing methods that they do not understand, or feel conflict with their beliefs. Irradiation was a common example of a food technology that, despite being highly effective, would be unacceptable to many consumers and would therefore not be implemented unless this perception could be altered.

Many stakeholders in this category expressed frustration at the public's perceptions of risk often bearing little relationship to any scientific determination of risk. These stakeholders place a very heavy focus on science when it comes to food safety decision-making, but consumer risk-perceptions can undermine efforts to base decisions on scientific data alone. There is a desire to employ technologies such as irradiation, but this is unlikely to occur in the near future because of consumer resistance.

Some processors stated that they (along with the MPI) were often *"forced"* to divert resources away from significant issues and towards debunking food safety *"myths"*, or allaying consumer fears. Examples given of food safety 'myths' included the widespread use of steroids in animal farming, poultry consumption causing antibiotic resistance in humans, irradiation making food radioactive and the massive overstating of the dangers associated with artificial sweeteners such as aspartame.

Brand image seems extremely important to processors of food products, who indicated that they would be unlikely to adopt any intervention that would damage their image in any way. Furthermore, most have a desire to collectively protect the image of their respective industries. This is especially true for higher risk industries (e.g. animal products). Industry representative groups such as the Poultry Industry Association of New Zealand (PIANZ), Meat Industry Association (MIA) and the Seafood Industry Council (SeaFIC) have the job of representing the interests of their members. This includes helping to protect the reputation of their industries as a whole by penalising members who operate in a manner which could damage the industry image.

Most industry representatives expressed frustration at the businesses that are not meeting expectations and causing potential damage to their industry image. Examples included homekill operators selling their (potentially unsafe) products on the black market and poultry and egg producers who consistently fail to meet animal welfare and hygiene standards. Industry representatives revealed that considerable information sharing regarding food safety issues and strategies often occurs between competitors in order to protect the overall industry image. In the words of one representative, *"If a bunch of people get sick from eating poultry, it doesn't matter who produced it, it will affect the whole industry"*. The poultry industry in particular has emphasised that there is one hundred percent transparency within industry when it comes to the sharing of food safety information. This information sharing and pooling of resources has been effective at helping to refine interventions to reduce the incidence of campylobacteriosis in New Zealand (Sears, 2011).

Although representatives of different processing industries indicated a desire for interventions which were not *"controversial"*, and would not result in damage to their industry's image, a couple noted that a small amount of 'damage' would quite possibly only be short term because *"consumers are forgetful"* and consumer attitudes towards products are not always reflected in their purchasing and consumption habits.

Processors generally did not have suggestions regarding how to measure alterations to brand image, and many felt this was one of the greatest sources of uncertainty when trying to assess the implications of food safety issues, emergencies and alterations to production and marketing processes. Several expressed a desire for a method of reliably quantifying alterations to brand image as a dollar figure, but conceded that this is difficult and subjective. A few suggested using expert judgement to estimate alterations on a qualitative (e.g. Likert) scale. Some also specifically indicated that they use focus groups and other qualitative research to estimate how consumers may react to a particular product or process change, but also highlighted that this method was not always reliable or representative of the general public.

Quality and Suitability

Often a processor's brand image relies on consistently delivering products that have a particular taste, along with other organoleptic properties. Manufacturers of branded goods such as chocolate, potato chips and beverages would be especially unlikely to accept interventions that might noticeably alter the physical characteristics of their products in any way.

Ease of implementation

Ease of implementation was identified by most processors as an essential component of any intervention. Although different processors had different ideas regarding what constituted an easy to implement intervention, in general, they were suggested as those that:

- Are easily incorporated into existing (e.g. factory) infrastructure. A practical intervention should be able to fit inside existing plants. The poultry industry, for example, cited the mandatory freezing of all fresh poultry as an effective, yet highly impractical intervention from their perspective, due to its requirement of massive freezing units; with current processing plants having insufficient land space.

- Should not slow production in any significant way by, for example, requiring more frequent maintenance or inspection of certain machinery.

- Have the expertise available to successfully maintain an intervention. Processors emphasised that the day-to-day operation and maintenance of a new intervention should be able to be conducted by current employees, with a minimal amount of additional expertise required to keep the intervention performing optimally. Where there is a requirement for additional expertise, that expertise should be readily available in the different geographic locations where an intervention is being implemented.

Processors indicated that the evaluation of 'Ease of Implementation' would be done on a case by case basis and that interventions could be ruled out as impractical for a large number of reasons, only some of which are described above.

Safety

Safety was mentioned by most processors. Some included it as part of 'Ease of implementation' but many considered 'Safety' as a stand-alone criterion. Interventions implemented at processing plants can often involve hazards such as potentially dangerous machinery, harsh chemicals, vapours and falling objects. There was a strong desire to implement interventions that did not expose workers to additional workplace hazards. Some additional risk seems acceptable, provided it could be managed through adequate training and health and safety protocols. Some processors were extremely proud of their health and safety records and indicated that any interventions which resulted in a greater number of workplace accidents would be unacceptable.

7.2.4 Food Retailers

Summary

Effectiveness was important to all interviewees in this group. Most felt that interventions would be considered effective if they consistently complied with regulations. Many in this group outsourced the formulation of their food safety plans to third parties, who provide advice regarding which interventions are most appropriate according to a range of criteria.

Maintaining the quality and suitability of their food was extremely important to interviewees in this category. This was especially true for fast food outlets, supermarkets and businesses selling what they referred to as *"premium"* products. Most indicated that an intervention should improve, and not compromise, quality and suitability.

Similarly, consumer perceptions were also extremely important to most of these interviewees. This was especially true for fast food and retail chains, which have strong brand images to maintain. Almost all interviewees in this group indicated that they would not implement interventions that had the potential to cause negative consumer perceptions.

Cost was important to all interviewees in this category, although less so to large businesses.

Ease of implementation was mentioned as important by a few in this category, who expressed a desire for interventions that did not slow production, and that required a minimal amount of additional training, space and new equipment.

Effectiveness

Food retailers and restaurant chains considered effectiveness as one of the most important criteria. Most participants in this category believed they already had good processes in place for ensuring food safety. Large retailers generally sought external expertise and advice regarding intervention selection and the formulation of food safety (e.g. HACCP) plans. Those that utilised these third party providers agreed that their advice was of a high quality and helped to ensure that the food they sold complied with regulations, and should therefore be considered safe. Many admitted that they did not have detailed knowledge regarding the specifics of interventions before implementation, and generally put their trust in third party food safety agencies to provide systems which would be effective, as well as acceptable regarding a number of other criteria.

Quality or Suitability of product

Some in this group emphasised that their business relied on delivering products with certain characteristics (e.g. flavour and appearance), and interventions which altered these characteristics would be unacceptable. Fast food chains emphasised the need to maintain the organoleptic properties of the foods they produce, as these are the qualities that loyal customers demand. The brand image of many of these businesses is intimately tied to the unique products they produce, and even subtle changes to these products could be enough to impact their brand. There are a number of high profile examples of large brands causing widespread consumer backlash after implementing new processes which resulted in changes to the organoleptic properties of their unique products.

Consumer perceptions

This was identified as an extremely important criterion to almost all retailers, who are highly susceptible to consumer backlash should a food safety incident occur. There are numerous examples of food safety issues damaging the brand of a restaurant or retail chain, and in some cases, food safety incidents can completely ruin an operator by destroying consumer confidence in their food.

Many felt there was limited scope for implementing a wide range of food safety interventions at the retail level when compared with, for example, the processing stage of production. This group generally felt that their suppliers (e.g. processors and farmers) had an obligation to provide them with food that should already be relatively safe, and that their job was to maintain this high level of safety, mostly through correct storage and preparation. Most could not think of any interventions commonly used by retailers that might elicit negative consumer perceptions, but agreed that they would be reluctant to implement such interventions if they were available.

Interviewees in this group agreed that consumer perceptions were very hard to measure or estimate and this was often one of the greatest sources of uncertainty associated with any changes a business might make to their products or processes. Some respondents identified a desire to estimate the monetary effects of any changes in brand image or consumer perceptions. Others indicated that a qualitative judgement was the only way to estimate the implications of consumer perception changes.

Financial Cost

The financial cost of an intervention was identified as an important criterion to all interviewees in this group. Some retail and restaurant chains are very large and have the ability to implement more costly interventions in the name of ensuring food safety (and protecting their brand); while for smaller operators financial cost may play a much larger role in decision-making.

Ease of implementation

There were several suggestions regarding what constituted easy to implement interventions from the perspectives of the different interviewees in this group. These interventions were generally viewed as those that do not require much additional space, large or complex equipment, labour or training, and do not slow production or compromise the convenience of the final food product.

7.2.5 Exporters

Summary

Most exporters were also processors and so the criteria mentioned in Section 7.2.3 will also be relevant to most of these businesses. Additionally, 'Market Access' was mentioned by these interviewees as being by far the most important criterion to their businesses. Many exporters felt it was not always sufficient to meet domestic standards in order for an intervention to be considered effective, as some foreign customers place additional demands on their products. Most exporters wished to meet the standards of as many trading partners as possible in order to remain flexible regarding who they export to. Interviewees in this group desire interventions that enhance market access.

Market Access

'Market Access' is the most important criterion to the exporters I interviewed. Many exporters felt that their products were very safe already and that interventions were sometimes intended not so much to provide any real health benefit, but to meet the often strict standards of overseas markets.

Foreign trading partners have their own standards when it comes to the importation of New Zealand foods. These standards include maximum limits on microbes and chemicals and, in some cases, process standards relating to the way food is to be transported, labelled and made safe (e.g. some markets refuse chemically treated products). Some New Zealand exporters feel that these requirements can be based on the unjustified use of the precautionary principle (See chapter 3.8.2) overseas; where food safety standards may be used as protectionist measures.

Many high volume export industries such as the beef and lamb industries wish to remain flexible regarding which markets they can export to, because global demand for New Zealand exports can fluctuate. For these exporters the type of intervention they implement is often *"dictated"* by the trading partner with the strictest import requirements. For example, hormonal growth promotants for use in beef are legal in New Zealand and are extensively used in Australia and the US, two core trading partners of New Zealand. However, the EU, another major consumer of New Zealand beef, has a ban on hormonal growth promotants in beef. As a result, virtually all (~99%) New Zealand beef exporters opt not to use hormones in order to ensure continued access to European (and other) markets. This gives New Zealand exporters the flexibility to adjust to global market shifts at short notice, selling their products where they will generate the greatest return. As well as a ban on hormones in beef, the EU generally has stricter import requirements regarding microbes and chemicals on a range of foods (e.g. the use of some chemical sprays would affect trade with Europe). Many exporters choose to meet these requirements, and thus the requirements of more lenient trading partners by default. A similar situation exists in New Zealand with halal meat, and almost all meat produced is halal in order to maintain flexible access to domestic and international halal markets.

The alternative to this approach is to tailor individual interventions to specific markets and, for example, use chemical sprays for US destined beef and not for EU beef. This is not ideal as it restricts the ability of exporters to respond to increased EU demand for beef. Also, global consumption preferences mean that different cuts of meat from the same animal will often be exported to different countries. This means that tailoring interventions to different trading partners may require application at the end of processing, after a carcass has been separated into individual cuts; potentially limiting opportunities to efficiently address food safety issues at an earlier point in the farm-to-fork chain.

This situation of allowing one market to 'set' food safety standards regarding the production of food for all is common in large export industries, which would need to operate parallel production or processing lines in order to cater to individual markets, leading to decreased flexibility and potentially higher production costs. The general consensus amongst high volume exporters is to implement one set of production processes that are acceptable to all major trading partners, allowing flexibility in meeting world demand.

For lower volume or very high-value exporters (e.g. some shellfish), it seems more viable to tailor production processes to meet the needs of emerging or niche overseas markets. For example, consumers in certain markets in Asia and the Middle East are willing to pay a premium for the absolute highest quality food products like seafood and poultry. The high value of such food products justifies the setup of flexible or multiple production processes that can be tailored to meet the exact needs of these overseas markets. One shellfish industry representative revealed that *"certain shellfish (growing) areas and processors are listed for US but not EU destined shellfish… we* (New Zealand shellfish growers) *cater for certain markets with shellfish because it's more of a niche product"*.

Sometimes costly food safety interventions are implemented in order to add value to a product and appeal to overseas niche markets. These are interventions that would not be feasible for application on the domestic market, or for the majority of overseas markets, because the demand and willingness to pay for such products is not present. An example of this can be seen in the New Zealand poultry industry, which has typically operated as a closed domestic market but has, in recent years, focused effort towards growing its exports and appealing to niche overseas markets. The poultry industry has been particularly successful at accessing high-value markets in the Middle East by applying an intervention known as High Pressure Processing (HPP), which sterilises food at high pressure, extending its shelf-life and allowing it to be marketed as chemical and additive free. Certain poultry processors run two production lines; one high volume production line catering mostly to the domestic market and relying primarily on chemical washes to reduce microbial contamination; and another lower volume production line catering to the demands of high value overseas markets by utilising the HPP intervention and completely sterilising this exported poultry without the use of chemicals. This poultry processor is able to justify the expense of setting up a parallel production line due to the product's high value and predictably expanding demand within these overseas niche markets.

Not only can certain countries demand food produced in a specific way, but so too can large domestic or international businesses such as fast food or high-end supermarket or hotel chains. Some New Zealand exporters cater to these large customers, and interviewees emphasised that having New Zealand exports meet the standards imposed by the regulatory bodies of trading partners does not necessarily ensure they will meet the standards of large customers, who often have their own stringent criteria. MPI generally sets its standards so that exports will meet the requirements of the countries New Zealand trades with, but sometimes exporters will need to exceed these standards to meet the needs of big customers. Examples of such international customers include McDonalds, Burger King, and Tesco's supermarkets (UK).

Exporters suggested that market access implications associated with a new intervention could be estimated by the dollar amount that an exporter stands to gain or lose through its implementation. However, sometimes there will be insufficient data to accurately estimate such financial implications of an intervention. Other suggestions include rating the intervention on a qualitative or semi-quantitative scale indicating the general magnitude of these financial gains or losses.

The overwhelming consensus among exporters is that restrictions on market access would be unacceptable. Any restrictions would need to be minor and offset by reduced operating costs or increased access to new or existing markets.

Effectiveness

Effectiveness was a key criterion highlighted by exporters of New Zealand food. Generally, exporters felt that so long as an intervention allowed their products to consistently meet the food safety specifications of their trading partners and customers, it would be considered effective. Reductions in pathogen numbers and the control of chemical hazards could be measured in much the same way as is done by processors (See section 7.2.3). An intervention will either consistently meet required specifications or will not, and this will determine whether the associated food product can be exported.

Quality of Science

Exporters emphasised the need to back interventions with the best available evidence. Interventions need to consistently meet the requirements imposed by overseas trading partners and customers. Failing to meet these requirements, even once, can have major consequences for an exporter and the wider export industry. This group identified a preference for trialling interventions which had been highly successful overseas and/or had a substantial amount of research of high methodological quality establishing evidence of effectiveness. Exporters views on this criterion were very much in line with those of processors (see Section 7.2.3).

Suitability/Quality

Exporters generally market New Zealand products as premium quality with an emphasis on freshness. Most exporters mentioned that they would be unwilling to adopt interventions which could have any negative effect on the quality or suitability of a food product.

Consumer perceptions

Much like other food businesses, exporters are concerned with the perceptions that consumers have of their products and processes. Most exporters interviewed were aware that the production processes they utilise to make their food safe may have negative implications if overseas consumers take issue with them. In contrast, there is also an opportunity for exporters to appeal to international consumers by adopting interventions which might be received well overseas (e.g. environmentally friendly interventions). Although there may not always be tangible evidence of any difference in product safety, foreign consumers were reported as perceiving some New Zealand foods as safer than foods from other countries. This is an example of consumer perceptions working in favour of New Zealand exporters. Most interviewees attributed this to New Zealand's "*Clean and Green*" and "*100% Pure"* image overseas. Exporters indicated that collectively protecting this image was, in addition to protecting their own brand and industry image, crucial.

Financial Cost

Exporters agreed that the cost of an intervention was always going to be considered in the decision-making process. However, most did not feel that cost was one of the most important criteria, as the financial benefits of increasing (or maintaining) the export quantities of New Zealand products or accessing new markets are likely to outweigh the costs of implementing most interventions.

7.2.6 Public health authorities

Summary

Interviewees in public health were primarily concerned with the degree to which an intervention can improve the health of consumers. Suggestions for the measurement of effectiveness included DALYs and reductions in disease incidence.

Cost was mentioned by interviewees in this group as important, as there are limited financial resources available for implementing food safety initiatives by food safety authorities.

High quality science was viewed as very important. Systematic reviews and meta-analyses were mentioned as ways of assessing the evidence surrounding an intervention, and this could be followed up with pilot trials if necessary.

Interviewees all agreed that consumer perceptions played a major role in the successful implementation of food-safety interventions. From the perspective of public health experts there is a need to find a balance between providing consumers with food safety information and minimising consumer alarm, especially regarding healthy foods.

Ease of implementation was mentioned by all interviewees as important. From the perspective of these interviewees, such interventions involve effective communication, timely implementation and are able to be widely understood by consumers.

Equity of benefits was mentioned by all interviewees. These individuals felt that it may be appropriate to prioritise interventions that addressed health inequalities, or reduced risks in higher-risk populations such as the young, elderly, pregnant or immunocompromised.

Effectiveness

Like regulators, those in public health are primarily concerned with the degree to which an intervention can actually reduce the burden of disease. An interviewee in public health suggested that the use of the DALY methodology could allow consistent and meaningful rankings of risk reduction methods. Other suggestions for measuring effectiveness included estimating the change in disease incidence. One interviewee in this category felt that, from a public health perspective, it was inappropriate to monetise health and aggregate this information into a cost-benefit figure. The monetising approach raised moral and methodological questions for this interviewee.

Financial Cost

Interviewees agreed that interventions implemented by public health authorities should be cost-effective, as they have limited funds to allocate to food safety initiatives.

Quality of Science

From a public health perspective, interventions should be based on the best available evidence. In line with the approaches used by MPI and industry (see Section 3.8.1), systematic reviews and meta-analyses, coupled with pilot trials, were suggested as preferred approaches for establishing the effectiveness of an intervention. These interviewees seemed more accepting of the use of the precautionary principle than other stakeholders.

Consumer Perceptions

Public health experts agreed that certain interventions have the potential to cause alarm amongst the public and this should be avoided if possible. There is an *"art"* to delivering public health messages to the public and this can be a tricky task. Some foods carry real health benefits but also come with risks, and educating the public about these risks without causing some to switch away from a food completely can be a balancing act. Fish was an example given of a food that carries benefits but also a certain level of risk (e.g. mercury toxicity), and communicating the risks (and risk reduction strategies) associated with fish consumption without causing consumers to alter their consumption habits (potentially towards more unhealthy alternatives) was difficult.

Ease of implementation

Those in public health suggested that an easy to implement intervention will involve good communication between all relevant stakeholders and will be widely understood. Potentially controversial interventions could include early communication with the consumer or target populations to increase buy-in and consumer understanding. In the case of food safety emergencies, communication and action (e.g. regarding food recalls) needs to be rapid and well organised. Quantifying performance on this criterion would need to qualitatively assessed on a case by case basis.

Equity of benefits

Interviewees suggested that interventions that reduced risks for at-risk sub-populations could be given priority. At-risk groups mentioned were the very young, elderly, immunocompromised, pregnant females and those of lower socioeconomic status. The example of listeriosis in pregnant females was given as an example of a preventable illness that can cause a significant amount of lost life (due to perinatal deaths) in a high risk group. Assessing performance on this criterion would likely involve a qualitative judgement from relevant public health experts.

7.2.7 Food Safety Scientists/Academics

Summary

Effectiveness was the number one criterion to all interviewees in this group. They broadly agreed that it was appropriate for different stakeholders to measure effectiveness using the metric(s) most relevant to them.

Quality science was extremely important to all interviewees in this group. Most interviewees agreed that MPI and many large food businesses consider the most up-to-date and high quality evidence when evaluating interventions. These individuals broadly agreed with the use of the precautionary principle when there was a lack of evidence.

These individuals acknowledged that cost would be a factor in virtually all food-safety intervention decision-making.

Even though consumers may not always base their preferences on scientific evidence, food safety academics/experts agreed that consumer preferences should be considered during decision-making.

Effectiveness

Effectiveness was a key criterion amongst food safety academics and experts. This group acknowledged that different measures of effectiveness would be relevant to different stakeholders. For example, for processors of animal products, the reduction in pathogen numbers on final product was seen as being the most relevant measure of effectiveness for microbial interventions; whereas for those interested in how these reductions would translate into a benefit for public health, incidence reductions or HALY estimates were suggested as being a more relevant measure of effectiveness.

Experts also agreed with processors that a distinction could be made between efficacious and effective interventions, and that industry pilot trials were often necessary to establish sufficient evidence of effectiveness in the New Zealand commercial setting.

Quality of Science

Interviewees agreed with other stakeholders that MPI generally considers the best available evidence when considering when and how to intervene when a food safety issue arises. There was less agreement regarding the conclusions that MPI draws from this evidence. There are several high profile examples of New Zealand food safety academics interpreting pieces of evidence differently to MPI and strongly disagreeing with MPI decisions. Examples of issues that have divided scientists include the safety of aspartame, A1/A2 milk and GM foods. While some experts on both sides of these debates feel the science is conclusive enough to make a judgement, others feel that the science is not developed enough to make definitive judgements and that more research is required to justify either position.

All participants in this category accepted that the precautionary principle had its place in protecting the health of consumers, although there was some disagreement regarding when the principle should be utilised. Some participants indicated a desire to employ the precautionary principle only as a last resort, while others felt that the principle has a greater role to play in protecting public health, especially in the long term. One expert in particular felt that the precautionary principle should be more heavily employed to protect consumers from new and emerging chemical hazards such as endocrine disrupting chemicals (EDCs), as very little is known about the health implications of long term exposure to these chemicals, especially in regards to the *"cocktail effect"* of a number of different chemicals being consumed over time.

Financial Cost

Scientists and food safety experts agreed that cost would be considered as part of any decision-making process.

Consumer perceptions

Scientists and academics agreed that consumer perceptions of risk often differed from those of food safety experts and risk assessors. There was a consensus that this could undermine efforts by regulators and food businesses to address a risk, and could force these groups to address perceived rather than real risks.

Some interviewees believed that trying to alter consumer perceptions and behaviour was too difficult, while others felt that it was possible with a concerted effort from all stakeholders. Some interviewees felt that, in the short term, negative consumer perceptions of some food safety issues and practices (e.g. irradiation) is something to be accepted and worked-around, rather than changed.

7.2.8 Māori

Summary

Māori academics whom I consulted in this research expressed a desire for genuine consultation with Māori during intervention decision-making. The consideration of cultural implications of interventions was important to them, as was the formal consideration and acknowledgement of traditional indigenous knowledge and beliefs.

Cultural sensitivity

Traditional Māori beliefs emphasise the interconnectedness of all things in the universe. Māori, being the Tangata Whenua (people of the land), see the 'health'[19] of their traditional lands as being vital for the production of healthy food, the maintenance of traditional family structures and the health of the individual. The concepts of sustainability, spiritual wellbeing, identity, participation, interconnectedness and equity are central to traditional Māori life. Traditional Māori could be described as environmentalist and share many of the same concerns as western environmentalist groups, including as a strong opposition to genetically modified foods (Frewer, 2009). There seems to be a strong preference for sustainable, natural and unadulterated food, produced with a minimum of processing aids.

There is a desire for traditional Māori knowledge, known as Matauranga[20], to be used alongside western science in many New Zealand decision-making contexts, including food safety decision-making. The use of Matauranga in decision-making is perhaps more important than western science to some Māori.

While many Māori in New Zealand may not give much attention to traditional Māori concepts and will be comfortable with the 'western' methods of scientific inquiry and decision-making, there is a general consensus that the Māori world is experiencing a resurging desire to incorporate traditional knowledge and practices into contemporary life, and for these two knowledge systems to be acknowledged as equally valid, and officially incorporated into New Zealand policy decision-making.

---

[19] Health, from a traditional Māori perspective, takes a more holistic view than the traditional western view of health, which focuses primarily on physical health. The traditional Māori view of health includes not only physical and mental health, but also the spiritual health of the individual, family (whanau), and environment (rohe). Within this traditional Māori paradigm, an individual cannot possibly attain optimal health unless the land and family structures are healthy too. Preserving the Mauri (~spiritual life-force) of all living and non-living things is a central concept in attaining full health.

[20] Matauranga involves the knowledge, comprehension and understanding of all things in our world. Often used synonymously with wisdom, Matauranga includes present day knowledge as well as traditional knowledge passed down from generation to generation by Māori elders.

The cultural impacts associated with a potential new food safety intervention or technology could be assessed through genuine consultation with Māori.

One Māori academic suggested the possible use of a tool known as the Mauri Model (Morgan, 2006), a MCDA Model, designed by Māori engineer and scholar Te Kipa Morgan, which could quantitatively rank different policy options, interventions or technologies according to traditional Māori criteria. The Mauri Model has received application to a number of real problems (Awatere et al., 2008; Hikuroa et al., 2010; Kepa Brian Morgan et al., 2012; Kraus, 2013; Morgan, 2008, 2011; Peacock, 2011; Peacock et al., 2012; Platia, 2012). My limited research regarding Māori viewpoints towards food safety interventions suggests that Māori may wish to evaluate different options using their own methodologies and criteria.

7.3 Quantification

This chapter has identified a list of key criteria which are relevant to the key stakeholder groups involved in the evaluation, selection and implementation of new food safety interventions. A number of criteria already have well established methodologies available for their quantification. For example, Effectiveness, Cost and Quality of Science are all systematically quantified by at least some stakeholders. These criteria are more likely to be systematically assessed and explicitly considered during decision-making. Some of these criteria have only one logical metric for their measurement, while others have a number of alternative approaches; often depending on the specific stakeholder perspective and/or hazard being managed. An example of such a criterion is 'Effectiveness'. The effectiveness of microbial interventions could be measured in pathogen, incidence or DALY reductions, or as a financial benefit. Although reliant on a number of assumptions, there are methodologies available for converting between these four units of measurement, provided data is available. Decision-makers are potentially able to display the same information in four different ways, with stakeholders utilising the unit of measurement most relevant to them. There is no single measure of effectiveness that is relevant to all stakeholders, and presenting information using multiple but analogous measures, as described above, would likely increase the relevance and stakeholder buy-in associated with any decision-making framework or model.

This chapter has also identified a number of other criteria which, although important, do not seem to have any well-established methods for their quantification. These are the criteria which can often be overlooked during decision-making, or handled in an informal or inconsistent manner. Because there is often a lack of data regarding these criteria, and no obvious unit of measurement available for their quantification, any MCDA model utilising these criteria would likely require the use of qualitative rating systems. An MCDA model utilising the criteria highlighted in this chapter would likely require a mixture of quantitative, semi-quantitative and qualitatively expressed criteria.

7.4 Conclusion

This chapter has identified a core list of criteria that are relevant to key stakeholders involved with food safety decision-making and intervention implementation.

Effectiveness was a key criterion identified by all groups interviewed. Any intervention needs to be consistently successful at reducing the risk associated with a particular hazard to an acceptable level. In the case of chemical hazards, the MRL approach is seen as acceptable to all relevant stakeholders, and any intervention which consistently meets MRLs will generally be considered effective. Methods for assessing the effectiveness of microbial interventions will differ according to stakeholder perspective. Regulators often set microbial targets for foods and, in the short term, processors and exporters can assess the effectiveness of microbial directed interventions by their ability to meet these targets. However, MPI and many other regulators place a focus on continuous improvements in public health over time, which means that an intervention that meets the targets of today may not necessarily meet the targets of tomorrow. The most appropriate method for assessing the effectiveness of microbial interventions from an industry perspective is to estimate the reduction in the number of pathogenic organisms they can achieve. This can allow for a definitive ranking of competing intervention options and can allow operators to identify the intervention(s) that will remain effective if microbial criteria are tightened. Regulators interviewed agreed that this method is relevant and easy to monitor for processors and exporters. However, regulators often base these microbial criteria on their ability to improve the societal burden of foodborne disease. To do this, possible pathogen reductions are converted into estimates of a reduced disease incidence. Going one step further, these incidence reductions can be translated into Health-Adjusted Life Year (DALY or QALY) estimates. The DALY approach is preferred over the QALY approach because the DALY methodology is already well developed in the field of food safety decision-making, both domestically and internationally. Although converting health benefits into a dollar figure was suggested, it was not a popular option amongst most interviewees.

There was broad agreement that high quality science and strong evidence of effectiveness is important before implementing any new intervention. All stakeholders agreed with MPI's approach of basing decision-making on studies of high methodological quality such as systematic reviews and meta-analyses. Food businesses felt that, additional to high quality studies, commercial trials were generally necessary before approving an intervention for full scale implementation. MPI and food businesses, especially exporters, expressed a preference for establishing *"unequivocal evidence"* where possible, rather than employing the precautionary principle. Some academics and public health experts disagree with this position and believe that the precautionary principle could be more heavily employed to protect public health in the absence of conclusive evidence. However, all groups appreciated that MPI has a tough job balancing the protection of health and the facilitation of market access in such a food export driven economy. All stakeholders agreed that MPI generally does a very good job at protecting consumer health. There was agreement that MPI considers the highest quality evidence during decision-making, although some MPI decisions have divided scientists. There was no clear consensus regarding what constitutes sufficient evidence to implement an intervention. This decision is likely to be context specific and handled on a case-by-case basis for most stakeholders.

All interviewees agreed that interventions could be perceived by both domestic and/or international consumers in a positive or negative light, and this had the potential to alter consumption habits. All relevant stakeholders agreed that consumer perceptions should be considered during decision-making. Brand image is particularly important to food businesses, but also food safety organisations such as MPI. Most stakeholders agreed that consumer perceptions had a large role to play in determining the success of a new food safety technology, but also agreed that very little was known about the way consumers assess risk, or how they respond to food safety information.

Market access was extremely important to regulators and exporters. Estimated changes to market access could be expressed as a dollar amount that exporters and/or other stakeholder groups stand to gain or lose from the implementation of a new intervention. In the absence of reliable information, these changes could be estimated on a semi-quantitative or qualitative scale.

Ease of implementation was important to all stakeholders involved in the direct implementation of interventions, but had different meanings to different interviewees, even within the same stakeholder group. However, relevant stakeholders generally agreed that easy to implement interventions required minimal additional physical space, training or expertise, and would not slow production. This criterion was probably the most troublesome to define and describe. There was no clear consensus regarding how to assess the ease of implementation of an intervention; this would likely require a case-by-case assessment.

Safety was a criterion important to stakeholders where potentially dangerous interventions could be implemented. Any increase in workplace accidents associated with a new intervention would be unacceptable.

Financial cost was important to all stakeholders. All food businesses acknowledged their role in continually improving the safety of the food they provide, and most were happy to accept that there were financial costs associated with this.

This chapter has reinforced the multi-criteria, multi-stakeholder nature of food safety decision-making. There are a number of criteria to consider other than cost and effectiveness during intervention evaluation. Different stakeholders consider different criteria and have different preferences regarding the way performance on these criteria is measured or quantified. The criteria identified and described in this chapter help guide the case study performed in Chapter 8.

Chapter 8: Model Construction, Testing and Critique

8.1 Introduction

This chapter investigates the suitability of three common MCDA models with regards to the problem of selecting food safety interventions from a pool of possible options. These models are the Weighted Sum Model (WSM), Analytic Hierarchy Process (AHP) and PROMETHEE. The construction and testing of these models on the historical New Zealand problem of addressing the high rates of campylobacteriosis associated with fresh poultry consumption in the mid-2000s is detailed. This chapter seeks to identify the main strengths of these models as well as any practical issues with their real-world application to food safety decision-making.

8.2 Methodology

After evaluating the provisional set of twelve criteria (identified in Chapters 6 and 7) for appropriate use in a test case study, by assessing them against the set of suitability criteria outlined in Section 4.4, versions of the Weighted Sum Model (WSM), Analytic Hierarchy Process (AHP) and PROMETHE models were implemented in Microsoft Excel 2010. These models were chosen because of their different characteristics, each belonging to a different 'family' of MCDA method, as well as their popularity in the literature, with each method having received application to a range of real problems. The complexity of their algorithms ranges from simple to more sophisticated and, in a similar way, the demands they place on those utilising them ranges from low to high.

The Weighted Sum Model was chosen because it is perhaps the most popular and rudimentary of MCDA models, having received wide application to a range of problems (Spackman et al., 2000). It has a simple algorithm and appears to place relatively low elicitation demands on the user.

AHP is a unique and very popular method, having also received wide application to an array of real problems. AHP utilises a unique scoring approach, which is seen as simple and intuitive by many (Spackman et al., 2000).

The PROMETHEE method is another widely used method and is a product of the unique 'outranking' school of thought. The underlying logic of PROMETHEE is quite different from the other two methods, and this is reflected in the mathematics behind the model. It has received application to a number of food safety studies, including the evaluation of risk-management options, and is perhaps the most widely applied MCDA method in food safety studies (Fazil et al., 2008; Henson et al., 2007; Henson & Masakure, 2011; Ruzante et al., 2010).

MAUT is a sophisticated and heavily studied method, but was excluded from this research because of its substantial data requirements. MAUT is generally applied to large problems, and requires estimates of the value and probability of all future states associated with each option. There are a large number of other MCDA models that could also have been implemented in this research. I decided to include the three methods above because the literature suggested that they showed promise in the context of food safety research. This research only allowed time for the implementation of a small selection of MCDA models and I have settled on the above three methods because I can better justify their inclusion, rather than justify the exclusion of other methods.

Models were tested on the problem of selecting the 'best' intervention(s) for addressing the New Zealand *Campylobacter* in poultry problem of the mid 2000s (referred to as the '*Campylobacter* case-study' throughout this chapter), the details of which are provided in 8.3. Test data consisted of quantitative data from a cost-effectiveness analysis conducted during this thesis (Lake, 2013) (included in Appendix D) and qualitative criteria scores and weightings developed after consultation with a food safety modelling expert familiar with the *Campylobacter* problem. Inputs for the WSM were generated in collaboration with this expert and this consultation helped guide my own estimates of model inputs for the other models. Stakeholders were not directly involved in the generation of the qualitative model inputs. This case study is intended as an illustrative example, with the results not necessarily reflecting stakeholder preferences. Sensitivity analysis was conducted on many of the inputs for the Weighted Sum Model, and Monte-Carlo simulation was conducted to explore the effects of multiple simultaneous changes to model parameters. Additional modifications were made to the original models to explore the effects of different model configurations on model outputs. These changes are detailed in the later sections of this chapter.

Decision-makers vs Participants

Throughout this chapter reference is made to the 'participants' of an analysis. Participants include all individuals who formally participate in the MCDA. Formally, participants are involved, to some degree, in all eight steps of a MCDA (see Section 4.4). I have decided not to use the word 'decision-maker(s)' when describing these individuals as not everyone who participates in an analysis will have decision-making power and, in practice, many MCDA analyses will include individuals who, other than providing their preferences/inputs, may have no final decision-making power whatsoever. Conversely, those with final decision-making power may not participate in an analysis at all. The two terms are not mutually exclusive however. Participants will always be involved in all steps of an analysis, especially establishing the decision context, the weighting of criteria and the interpretation of the results. Decision-makers are those who make the final decision, regardless of the level of participation they have had in the analysis. Ideally, all decision-makers would be participants, but this may not always occur.

8.3 Case Study – Campylobacter in Poultry

Campylobacter in New Zealand poultry has been studied in depth by a range of scientists, regulators and academics (Baker, 2005; Baker et al., 2007; Baker et al., 2006; ESR, 2007, 2008; Hudson et al., 2008; Lake, Cressy, et al., 2007; Lake et al., 2011; Lake, 2013; Lake, Hudson, et al., 2007; Lake, van der Logt, et al., 2007; Lake, Cressey, et al., 2010; MPI, 2013b; NZFSA, 2010a; Sears, 2011), although no studies utilising formal MCDA methodologies were located in this research. The models constructed in this chapter were tested on the historic problem of selecting the best intervention(s) to address the high New Zealand rates of campylobacteriosis in the mid-2000s, thought to be largely attributable to poultry meat (Sears, 2011). The author and supervisors of this thesis have previously contributed to a cost-effectiveness paper examining this problem (Lake, 2013), included in full in Appendix D. The models in this chapter offer an alternative approach to this cost-effectiveness framework by incorporating additional criteria in the evaluation of the same pool of interventions. Other than cost and effectiveness figures, which have been sourced from the aforementioned CEA paper, all other data utilised in the testing of models has been estimated after interviews with stakeholders and consultation with a food safety expert familiar with this particular problem. Rather than identifying a definitive solution to the *Campylobacter* problem, the intention of this analysis is to use test data to explore the potential for widespread application of a selection of MCDA models to New Zealand food safety decision-making.

8.3.1 The Problem

In the mid 2000s campylobacteriosis rates in New Zealand were high by developed country standards and were thought to be largely attributable to the consumption of poultry meat (Sears, 2011). There were a number of potential interventions studied in the 2013 CEA paper to reduce the burden of campylobacteriosis due to poultry. These interventions include:

- The elimination of thinning by introducing an *"All in, All out"* regime, where all birds are removed from a shed and sent to processing together
- Increased biosecurity on broiler farms
- Single use of catching crate modules
- Freezing of the entire poultry supply
- Continuous chemical treatment of broiler meat during primary processing
- Multiple improvements during primary processing. This was an attempt to assess the changes made by the New Zealand poultry industry in 2005 – 2007.
- Irradiation of poultry meat
- Consumer hygiene education
- Bacteriophage treatment of broilers

A more comprehensive description of these interventions can be found in the original paper, included in Appendix D. This paper produced the following table containing cost and effectiveness (in DALYs) data:

Intervention	**Reduction in Poultry Associated Disease Burden per Year** **(DALY) (% of poultry associated burden estimated for 2005)**	**Annual equivalent cost ($m)**	**Cost per DALY ($)**
On farm			
Broiler farms all in all out	331 (33)	8.9	27,000
Increased biosecurity on farm	165 (16)	4.5	27,000
Single use of catching crates/modules for thinning	82 (8)	0.75	9,100
Primary processing			
Freezing the proportion of poultry supply currently being sold as fresh product (75%)	947 (94)	8.4	8,900
Continuous chemical controls (chlorine) during primary processing	854 (84)	1.5	1,700
Multiple interventions during primary processing	854 (84)	1.0	1,200
'Irradiation'	1010 (100)	43.9	43,400
Consumer			
8) Hygiene education	8 (0.8)	0.16	17,900

In development			
9) Control of numbers of *Campylobacter* in faecal material by administration of bacteriophage prior to slaughter	993 (98)	2.97	3000

Table 6: Cost-effectiveness of interventions to control Campylobacter in poultry. From (Lake, 2013).

The situation of needing to integrate quantitative data with qualitative estimations of performance is typical of many food safety decision-making situations. There appears to be a need for a MCDA type methodology that can simultaneously accept both quantitative and qualitative inputs and arrive at defensible recommendations regarding which course of action is most appropriate to address a food safety issue. This chapter constructs a selection of popular and potentially suitable MCDA models and tests them on this *Campylobacter* case study.

8.4 Assessment of Provisional set of Criteria

In the previous chapters the problem of selecting a food safety intervention from a pool of possible options was defined and structured. Key stakeholders were identified and interviews with these groups highlighted a core list of criteria that could potentially be used to evaluate new interventions. These interviews also guided the choice of how to measure performance on each criterion. This list of criteria appears below. Comprehensive descriptions of these criteria can be found in Chapter 7.

Effectiveness: Measured in DALYs

Financial Cost: In $NZ

Market Access: Measured qualitatively, or in $NZ if information is available

Consumer Perceptions: Qualitatively assessed

Ease of Implementation: Qualitatively assessed

Quality or Suitability: Qualitatively assessed

Quality of Science: Qualitatively assessed

Equity of Costs: Qualitatively assessed

Equity of Benefits: Qualitatively assessed

Workplace Safety: Qualitatively assessed

Cultural Impact: Qualitatively assessed

Animal welfare: Qualitatively assessed

Before finalising the choice of criteria, they themselves should be evaluated against a range of attributes to establish their suitability for subsequent use in formal MCDA:

Completeness The final set of criteria should be able to adequately assess the overall performance of a new food safety intervention in terms of its ability to meet set objectives. At this stage it is important to think about whether any major indicators or categories of performance have been overlooked. The criteria used in this case study are those identified by various stakeholders during a comprehensive interview process (detailed in Chapter 7) and appear to capture the key positive and negative aspects of new food safety interventions.

Operational

The meaning of each criterion should be easily understood by stakeholders, with the effects of an intervention on each criterion easily understood also. Chapter 7 explains the meanings of the different criteria and, although some of the differences between criteria appear subtle, these differences are clearly defined. Ideally, criteria will be able to be measured using existing metrics, but this is not always the case and most criteria identified in this research have no natural unit of measurement; these criteria will be qualitatively assessed on some form of numerical or semantic scale. Criteria such as 'Cost' and 'Effectiveness' can be measured using existing metrics such as dollars and DALYs.

No double counting

Criteria should not overlap. That is, two or more criteria should not measure the same thing, either wholly or in part. Although there are criteria which may appear to overlap, there are distinct differences in their meanings that justify their inclusion. For example, the ‘Market Access’ criterion includes an intervention’s ability to affect access to a market through meeting (or not meeting) the specific food safety requirements (i.e. legislation) of food safety authorities or customers (not consumers), whereas ‘Consumer Perceptions’ captures an intervention’s ability to offend (or appeal to) consumers, regardless of whether or not it meets market access requirements. An intervention could be ‘legal’ in a foreign market yet completely offensive to consumers, or vice versa. Imported tomatoes that have been irradiated are completely legal for sale in New Zealand (they meet market access requirements), but this does not necessarily mean that consumers will purchase them, as there is often a negative stigma attached to irradiated products that makes consumers less likely to purchase them (ESR, 2008; Frenzen et al., 2000; Mehta, 2002).

In a similar way, the ‘Quality or Suitability’ criterion captures an intervention’s ability to cause real changes to the organoleptic (physical) properties of a food product, while ‘Consumer Perceptions’ includes only perceived rather than real changes to a food-product. Although these criteria do seem to cover similar impacts of food safety interventions there are important distinctions that separate them. A consumer education campaign may be able to address negative consumer perceptions of a food product but will be unable to reverse any physical changes to a food, and thus it is important to distinguish between real and perceived changes to a product through the use of both criteria.

Non-redundancy

If all interventions perform identically on a criterion then that criterion will be redundant in the sense that it will not aid the decision-maker in making a choice. Assessing the redundancy of criteria will need to be done on a case-by-case basis. For the *Campylobacter* in poultry example it is probable that the ‘Animal Welfare’, ‘Cultural Impact’ and ‘Equity of Benefits’ criteria are redundant because all options may score equally and these criteria will therefore in no way contribute to a decision one way or another. If this was indeed the case then these criteria could be removed from the analysis. This can be an important step in simplifying the analysis, especially with models that place a significant ‘elicitation burden’ on the participants.

Mutual independence of preferences

Most MCDA applications require that preferences associated with different alternatives are independent from one criterion to the next. If performance scores can be assigned for one criterion without knowing the scores for any other then criteria are said to be preferentially independent. The criteria outlined in this research appear to exhibit preferential independence. Assuming one has a clear understanding of the different criteria, there is no reason why performance scores for each criterion could not be assigned independently.

Minimal size

Keeping the number of criteria to a minimum while still capturing the key effects of an intervention offers the advantage of simplifying the analysis for all those involved. The assessment of whether all criteria identified in this research will be necessary for decision-making would need to be conducted on a case-by-case basis. It is likely that in most food safety decision-making situations it will be possible to eliminate at least one or two of these criteria. For example, food safety decision-making for vegan food products would likely render the 'animal welfare' criterion unnecessary.

8.5 Model Construction and Results

The following models were constructed in Microsoft Excel 2010. The analyses in this chapter allocated equal performance scores for all interventions on the 'Equity of Benefits', 'Safety', 'Animal Welfare' and 'Cultural impact' criteria. As such, these criteria violate the condition of non-redundancy (i.e. they are redundant) and were excluded from subsequent analysis. This section outlines the construction of each model, followed directly by the results and critique of each model's application to the *Campylobacter* case study.

8.5.1 Weighted Sum Model (WSM)

8.5.1.1 Construction

This implementation of the WSM requires the assessment of each option on a 0-100 local[21] scale for all criteria, with a higher score indicating better performance on the corresponding criterion. The model is accepting of quantitative data like cost and effectiveness figures, which are automatically mapped onto the 0-100 local scale. When there are no quantitative data available, decision-makers are required to directly rate the performances of options on a numerical scale. These direct ratings usually take a value between 0 (worst possible option) and 100 (best possible option), however this is not a requirement as the model automatically remaps all ratings onto a 0-100 local scale in the same way it does for quantitative inputs. In the *Campylobacter* case study there are quantitative data available regarding effectiveness (measured in DALYs) and financial cost (measured in annual cost in $NZm). Cost is a criterion to be minimised (i.e. lower costs indicate better performance) and because of this, once the costs are mapped onto the local scale their values are reversed so that the less costly options receive the highest criterion scores. This is consistent with the practice of selecting the option with the highest global performance scores (i.e. the sum of weighted criterion scores) (Mabin, 2006).

The model will accept weights generated using any of the weighting procedures outlined in Section 4.7. This case study utilised the 'direct rating' method described in Section 4.7, with the user assigning each criterion a score between 0 and 100, reflecting its relative importance in the decision-outcome. These scores are then normalised and a percentage weight is assigned to each criterion.

$$w_j = \frac{pw_j}{\sum_{j=1}^{n} pw_j},$$

[21] A local (rather than global) scale assigns the best performing option (of those being studied) on each criterion the maximum possible score and the worst performing option the lowest possible score, with all other options allocated scores relative to these two end points. Section 6.4 further explains these two types of scale.

Where:

w_j = normalised weight allocated to criterion j

pw_j= raw/unstandardised weight (out of 100)

It is the raw/unstandardised criteria weights *relative* to the others that determine the final normalised criteria weights (percentages). Allocating all criteria the same raw score, irrespective of what score that may be, will result in all criteria receiving equal weights.

Table 7 shows the raw data inputs into the WSM model while Table 8 shows the remapped criteria scores and normalised criteria weights.

Key
Quantitative data
Qualitative scores/weights

	Cost (NZ$m)	Effectiveness	Market Access	Consumer Perceptions	Quality or Suitability	Ease of implementation	Quality of Science	Equity of costs
Multiple Interventions	1.0	854	70	35	95	90	100	50
Chemical Treatment	1.5	854	70	35	95	90	100	50
Bacteriophage	3.0	993	70	35	50	80	50	50
Freezing	8.4	947	50	50	20	20	100	10
Single Use Crates	0.8	82	50	80	50	80	40	50
Hygiene Education	0.2	8	50	100	50	100	80	50
All-in-all-out	8.9	331	50	80	50	80	40	50
Increase Biosecurity	4.5	165	60	80	50	90	70	50
Irradiation	43.9	1010	0	10	95	0	100	50
Criteria Weights	100	90	30	60	60	70	90	25

Table 7: Raw data inputs for Weighted Sum Model

Key: Quantitative data (shaded) / Qualitative scores/weights (unshaded)

	Cost	Effectiveness	Market Access	Consumer Perceptions	Quality or Suitability	Ease of implementation	Quality of Science	Equity of costs
Multiple Interventions	98.1	84.4	100.0	27.8	100.0	90.0	100.0	100.0
Chemical Treatment	96.9	84.4	100.0	27.8	100.0	90.0	100.0	100.0
Bacteriophage	93.6	98.3	100.0	27.8	40.0	80.0	16.7	100.0
Freezing	81.2	93.7	71.4	44.4	0.0	20.0	100.0	0.0
Single Use Crates	98.7	7.4	71.4	77.8	40.0	80.0	0.0	100.0
Hygiene Education	100.0	0.0	71.4	100.0	40.0	100.0	66.7	100.0
All-in-all-out	80.0	32.2	71.4	77.8	40.0	80.0	0.0	100.0
Increase Biosecurity	90.1	15.7	85.7	77.8	40.0	90.0	50.0	100.0
Irradiation	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0
Criteria Weights	19%	17%	6%	11%	11%	13%	17%	5%

Table 8: Remapped data inputs for Weighted Sum Model (performance scores remapped onto local scales and normalised criteria weights)

8.5.1.2 Results

The WSM multiplies each performance score by its corresponding criterion weight and sums these weighted scores across all criteria to generate a 'global score' for each option. These scores are shown in Table 9 and Figure 18.

WSM Performance matrix for Campylobacter problem

	Cost	Effectiveness	Market Access	Consumer Perceptions	Quality or Suitability	Ease of implementation	Quality of Science	Equity of costs	Global Scores	Rank
Multiple Interventions	18.7	14.5	5.7	3.2	11.4	12.0	17.1	4.8	87.4	1
Chemical Treatment	18.5	14.5	5.7	3.2	11.4	12.0	17.1	4.8	87.2	2
Hygiene Education	19.0	0.0	4.1	11.4	4.6	13.3	11.4	4.8	68.7	3
Bacteriophage	17.8	16.9	5.7	3.2	4.6	10.7	2.9	4.8	66.4	4
Increase Biosecurity	17.2	2.7	4.9	8.9	4.6	12.0	8.6	4.8	63.5	5
Freezing	15.5	16.1	4.1	5.1	0.0	2.7	17.1	0.0	60.5	6
All-in-all-out	15.2	5.5	4.1	8.9	4.6	10.7	0.0	4.8	53.7	7
Single Use Crates	18.8	1.3	4.1	8.9	4.6	10.7	0.0	4.8	53.0	8
Irradiation	0.0	17.1	0.0	0.0	11.4	0.0	17.1	4.8	50.5	9
Criteria Weights	19%	17%	6%	11%	11%	13%	17%	5%		

Table 9: Weighted Sum Model performance matrix for Campylobacter case study

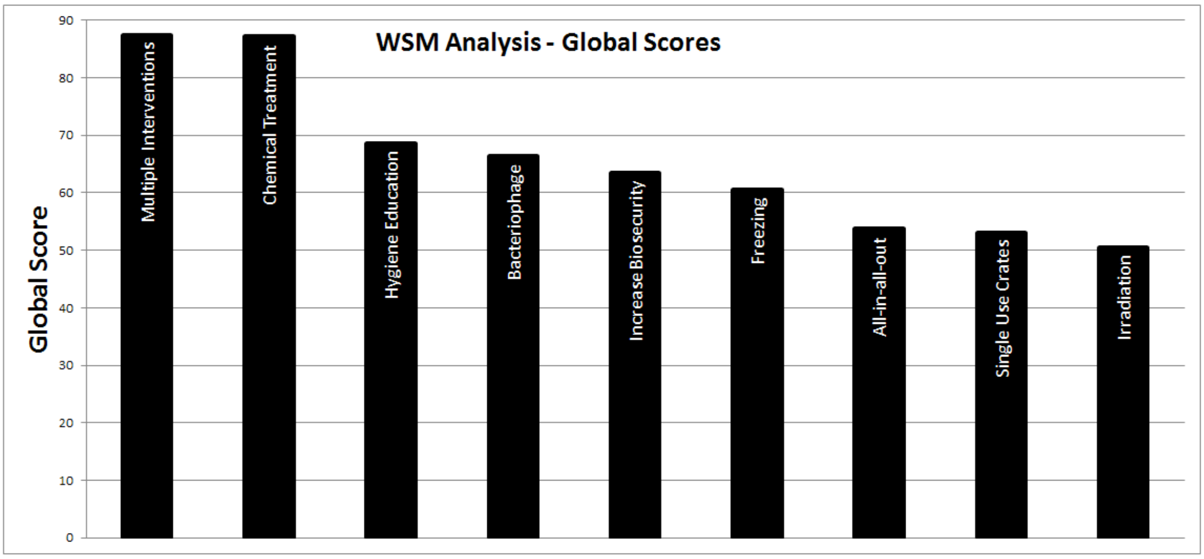


Figure 18: Global performance scores for Weighted Sum Model

Treating all scores and weights as static values identifies three potential clusters of interventions, with similar scores within each cluster. ‘Chemical Treatment’ (score 87.4) and ‘Multiple Interventions’ (score 87.2) stand out as top performers, while ‘Hygiene Education’, ‘Bacteriophage’, ‘Increased Biosecurity’ and ‘Freezing’ make up the middle cluster, with scores in the 60-70 range. Worst performers are ‘All-in-all-out’, ‘Single use Crates’ and

’Irradiation’, with scores in the low 50s. ‘Multiple interventions’ and ‘Chemical treatment’ both score very well due to their strong performance on all criteria except ‘Consumer Perceptions’; receiving a maximum score on the ‘Market Access, ‘Quality or Suitability’, ‘Quality of Science’ and ‘Equity of Costs’ criteria. Despite receiving a top score on the ‘Effectiveness’, ‘Quality or Suitability’, ‘Quality of Science’ and ‘Equity of Costs’ criteria, ‘Irradiation’ ranks last due to its score of zero on all remaining criteria.

Sensitivity Analysis

As cost is the only criterion separating the top two interventions (Table 9) it is clear that the annual cost of ‘Multiple Interventions’ would need to increase, or the annual cost of ‘Chemical Treatment’ decrease, by more than half a million dollars before ‘Chemical Treatment’ became the more preferred intervention. Table 10 shows the results of further sensitivity analysis; identifying the ranges over which criterion scores and weights can vary in isolation whilst preserving the intervention rankings. Many of the scores are able to individually vary across wide ranges without altering the current rankings. However, the value of this sensitivity analysis is limited as it assesses each change in isolation; the ranges in Table 10 apply only if other inputs remain static. Once two or more of these inputs change simultaneously these numbers no longer apply. Furthermore, these ranges refer to the preservation of the rankings of all nine interventions, including the lower ranking interventions, which are of less interest to us at this point.

Ranges over which intervention rank is preserved

	Cost (NZ$m)	Effectiveness	Market Access	Consumer Perceptions	Quality of Suitability	Ease of implementation	Quality of Science	Equity of costs
Multiple Interventions	0-**1**-1.5	842-**854**-inf	69-**70**-100	34-**35**-100	94-**95**-100	89-**90**-100	**100-100**	49-**50**-100
Chemical Treatment	1-**1.5**-inf	0-**854**-866	0-**70**-71	0-**35**-36	0-**95**-96	0-**90**-91	20-**100**	0-**50**-51
Bacteriophage	2.85-**2.97**-12	761-**993**-996	50-**70**	0-**35**	24-**50**	51-**80**	31-**50**	17-**50**
Freezing	0-**8.4**-23.9	553-**947**-1010	15-**50**-72	0-**50**-82	0-**20**-48	0-**20**-50	77-**100**	0-**10**-50
Single Use Crates	0-**.75**-19.1	0-**82**-123	9-**50**-53	18-**80**-85	0-**50**-54	21-**80**-85	0-**40**-42	0-**50**-57
Hygiene Education	0-**.16**-.29	0-**8**-1010	50-**50**-100	99-**100**	**50**-100	**100-100**	**80**-100	**50**-100
All-in-all-out	0-**8.9**-10.5	290-**331**-725	47-**50**-100	75-**80**-100	46-**50**-94	75-**80**-100	38-**40**-63	45-**50**-100
Increase Biosecurity	0-**4.5**-13.9	0-**165**-397	39-**60**-82	48-**80**-100	23-**50**-76	60-**90**-100	56-**70**-83	16-**50**-100
Irradaition	13.5-**43.9**-inf	0-**1010**	**0**-47	0-**10**-74	0-**95**-100	**0**-63	0-**100**	0-**50**-100
Criteria Weights	96-**100**	75-**90**	7-**30**-70	44-**60**-90	21-**60**-82	49-**70**-100	67-**90**-100	11-**25**-60

Table 10: Sensitivity Analysis of Weighted Sum Model (static values in bold)

8.5.1.2 Monte-Carlo simulation of multiple parameter changes

The primary goal of this analysis is to assess whether certain combinations of parameter changes could alter the rankings of the most preferred options.

Criteria Weights

In MCDA it is often the criteria weights that carry the most uncertainty or disagreement (DCLG, 2009). For this analysis an initial variability analysis was performed on criteria weights. Most weights were allowed to vary by 15 'units' either side of their static values (up to a maximum value of 100). These were modelled by uniform distributions in @RISK. Certain criteria were allowed to vary over a much wider range of values due to the range of stakeholder responses regarding their importance. These criteria were:

- **Market Access:** Currently only a small amount of poultry meat is exported, but there is a desire from some to expand on this (see Section 7.2). To these stakeholders 'Market Access' could receive a very high weighting, but to others unconcerned with expanding overseas, 'Market Access' may not be considered at all. Criteria weights for 'Market Access' were modelled as a *U(0,100)* distribution.

- **Consumer Perceptions:** Some stakeholders felt that 'Consumer Perceptions' was one of the most important criteria. Others felt that, although consumers perceptions were important, negative perceptions could be addressed and may only be an issue in the short term. Consumers may also base the majority of their purchasing decisions on other factors such as price, convenience and taste. Criterion weights for 'Consumer Perceptions' were modelled as a *U(50,90)* distribution.

- **Quality or Suitability:** Some stakeholders were accepting of a small reduction in the quality or suitability of their products while others emphasised that any reductions would be unacceptable. This criterion weight was modelled as a *U(60,100)* distribution.

Simulations used Latin hypercube sampling and were run for 100,000 repetitions.

Figure 19 shows that rankings are generally insensitive to even the large fluctuations in criteria weights outlined above. While the rankings of the seven lowest ranked interventions might be subject to some reordering as a result of these variations, the top two interventions still significantly outperform the other seven, even at their lowest values. The tornado plot in Figure 20 shows that the top intervention's ('Multiple Interventions') performance score is likewise insensitive to all modelled variations in criteria weights. Figure 18 also shows that it is possible that some of these modelled parameter changes could result in 'Bacteriophage' displacing 'Hygiene Education' as the third ranked intervention. Inspection of the performance matrix in Table 9 reveals that this is only likely to occur if the weight of 'Effectiveness' is substantially increased, as this is the only criterion where 'Bacteriophage' sufficiently outperforms 'Hygiene Education'.

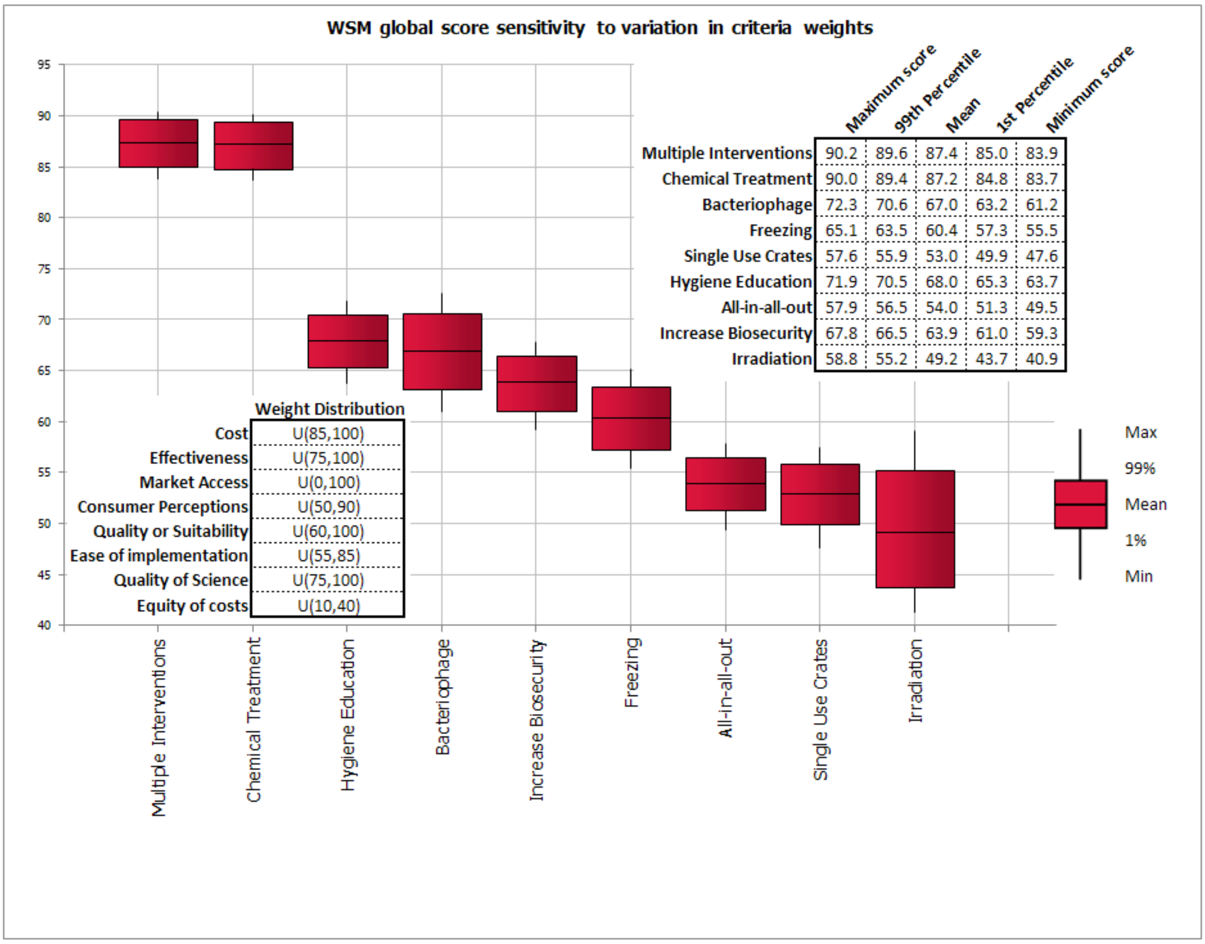


Figure 19: Variability analysis of Weighted Sum Model criteria weights

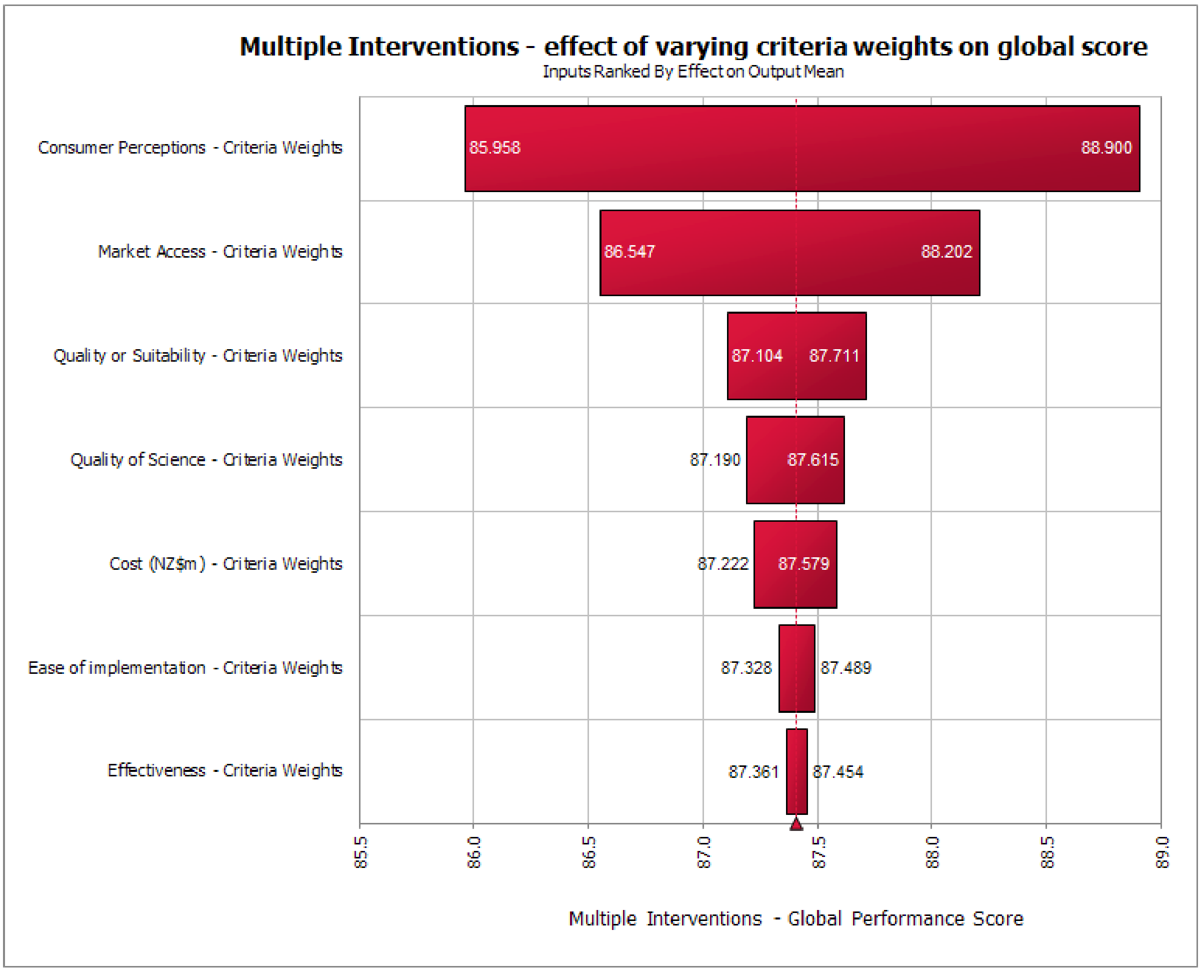


Figure 20: Tornado plot for 'Multiple Interventions'

The 'tornado plot' in Figure 20 shows the effects of varying the different criteria weights across their aforementioned ranges on the global score of 'Multiple Interventions'. 'Multiple interventions' is most sensitive to 'Consumer Perceptions' (although still reasonably insensitive, as it only varies by ~2% in either direction) and 'Market Access' criteria weights, and is highly insensitive to variations in the remaining criteria weights.

Sensitivity to criteria performance scores

In addition to the variations in criteria weights outlined above, Monte-Carlo simulation was also conducted for selected performance scores. The choice of which parameters to vary included those where there, based on consultation with stakeholders, appeared to be significant uncertainty surrounding their values. Additionally, performance scores of high ranking interventions were varied on the criteria with the highest weightings and on the criteria where there were a large range of values separating the top 5 ranked interventions, as changes in these scores have the greatest potential to alter the top rankings.

The uncertainties associated with these scores were modelled as PERT[22] distributions (Vose, 2007), with the 'most likely' values assuming their previous static values.

Cost: The costs of the top six interventions were allowed to vary from their static values by 50% in either direction.

Effectiveness: The effectiveness estimates of the top four interventions were allowed to decrease by 20% and increase by 10% (to a maximum of 1010 DALYs) from their static values. Additionally, the effectiveness of 'Hygiene Education' was assigned a maximum possible value of 80 DALYs; ten times that of its static value. This was to simulate the effects of a much greater uptake of the intervention by consumers.

Market Access: Performance scores for 'Multiple interventions', 'Chemical Treatment' and 'Bacteriophage' were allowed to deviate 20 points in either direction from their static values. This was to reflect uncertainty regarding the current and future acceptability of these interventions to overseas trading partners such as the EU. Performance scores for 'Freezing', 'Single Use Crates' 'Hygiene Education', 'All-In-All-Out' and 'Increase Biosecurity' were all allowed to increase up to the maximum possible score of 100. There is high uncertainty regarding the performance of these five interventions on the 'Market Access' criterion. Some of New Zealand's foreign trading partners may prefer more reliable, relevant (e.g. other than 'Hygiene Education') or less extreme interventions (e.g. other than 'Freezing'), although they are unlikely to have major objections to their use either.

Consumer Perceptions: Performance scores for 'Single Use Crates', 'All-In-All-Out' and 'Increase Biosecurity' were all allowed to increase up to the maximum possible score of 100.

---

[22] **Program Evaluation and Review Technique (PERT):** The PERT distribution is used exclusively to model 'expert' estimates by allowing them to specify a maximum, minimum and most likely value of a parameter. PERT distributions allow for the elicitation of more accurate expert estimates than uniform distributions, without placing unreasonable data requirements on participating experts, and are arguably a better alternative to the similarly used triangle distribution.

This reflects a potential consumer preference for these preventive and non-chemical interventions over potentially more controversial or less understood interventions like ‘Chemical Treatment’ and ‘Irradiation’.

Quality or Suitability: Performance scores for ‘Multiple Interventions’ and ‘Chemical Treatment’ were allowed to decrease from their static values to a minimum value of 50. These interventions can increase the shelf life of poultry. However, they require the use of chemicals that may affect the organoleptic properties of the meat, which could be viewed as a reduction in quality. A wide range of values captures this uncertainty. ‘Single Use Crates’, ‘Hygiene Education’, ‘All-In-All-Out’ and ‘Increase Biosecurity’ were allowed to increase to the maximum value of 100. This reflects that, although these interventions may not increase the shelf life as much as other interventions, they will arguably allow the meat to better maintain its natural organoleptic properties over other interventions.

Ease of Implementation: Criterion scores were not varied. This was to reflect a higher degree of confidence surrounding these scores. Based on interviews and consultation there was a strong consensus that ‘Freezing’ and ‘Irradiation’ would both be considerably more difficult to implement than the other interventions and this was reflected in the initial criteria scores.

Quality of Science: Criterion scores were not varied; this is to reflect a higher degree of confidence surrounding these scores, based on interviews and consultation.

Equity of Costs: Of those included, this criterion is perhaps the least relevant to the *Campylobacter* case study. All but ‘Freezing’ received equal scores in the initial analysis. During the interview process there was little concern from most interviewees regarding this criterion. A higher score for ‘Freezing’ in the original analysis may have warranted further analysis of this criterion. However, even if ‘Freezing’ received a top score on this criterion its ranking would increase by only one position.

The distributions described above are summarised in Table 11.

	Cost (NZ$m)	Effectiveness	Market Access	Consumer Perceptions	Quality or Suitability	Ease of implementation	Quality of Science	Equity of costs
Multiple Interventions	PERT(.5,1,1.5)	PERT(683,854,939)	PERT(50,70,90)	35	PERT(50,95,95)	90	100	50
Chemical Treatment	PERT(.75,1.5,2.25)	PERT(683,854,939)	PERT(50,70,90)	35	PERT(50,95,95)	90	100	50
Bacteriophage	PERT(1.49,2.97,4.45)	PERT(794,993,1010)	PERT(50,70,90)	35	PERT(50,50,100)	80	50	50
Freezing	PERT(4.2,8.4,12.6)	947	PERT(50,50,100)	50	20	20	100	10
Single Use Crates	0.75	82	PERT(50,50,100)	PERT(80,80,100)	50	80	40	50
Hygiene Education	PERT(.14,.16,.18)	PERT(6.4,8,80)	PERT(50,50,100)	100	50	100	80	50
All-in-all-out	8.9	331	PERT(50,50,100)	PERT(80,80,100)	50	80	40	50
Increase Biosecurity	PERT(2.25,4.5,6.75)	165	PERT(50,50,100)	PERT(80,80,100)	50	90	70	50
Irradiation	43.9	1010	0	10	95	0	100	50

Note: Numbers in paranthesis represent (from left to right) the minimum, most likely and maximum estimated values.

Table 11: Distributions used in monte-carlo analysis for simulating the effect of varying multiple performance scores

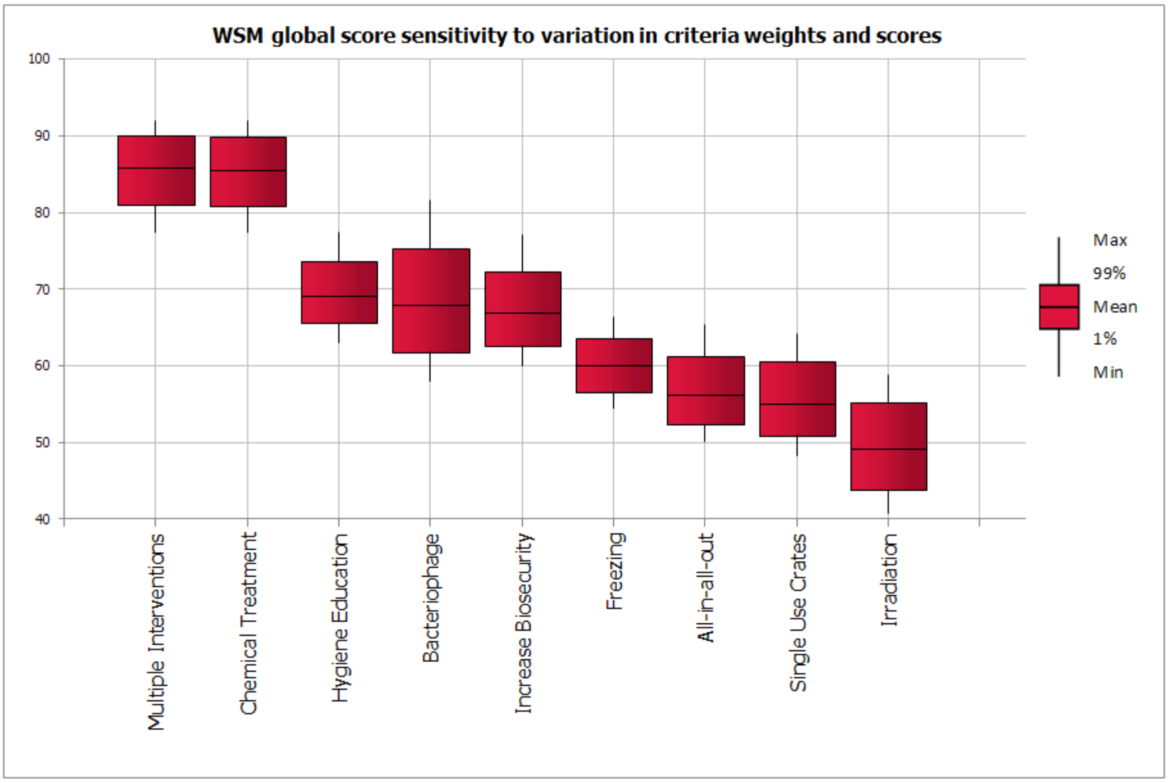


Figure 21: Variability analysis of Weighted Sum Model criteria weights and scores

Figure 21 shows that under the modelled combinations of criterion scores and weights 'Multiple Interventions' and 'Chemical Treatment' still clearly outperform the others. The original intervention rankings appear to be very robust and insensitive to combinations of changes to input parameters. This insensitivity to many parameter changes occurs in many MCDA analyses (DCLG, 2009).

8.5.1.3 Critique

Simplicity

Of the models studied in this research the WSM is the easiest to implement, use and understand. The mathematical operations required in the WSM make intuitive sense to most users with a basic understanding of mathematics and utility theory. In a decision-making context such as food safety decision-making, where there are a number of stakeholders with different levels of analytical ability, having these tools as simple and user-friendly as possible is a major advantage. One of the most commonly stated advantages of MCDA is its inherent transparency and ability to provide an 'audit trail' (Barcus & Montibeller, 2008; Belton & Stewart, 2001; Bertsch et al., 2006; Geldermann et al., 2009; Mayon-White & Wisudha, 2010; Phillips, 2008; Phillips & Stock, 2003; Redpath et al., 2004; Salo & Hämäläinen, 2010; Spackman et al., 2000; Young et al., 2001). This is especially true in contexts where there is a desire to involve laypeople or to communicate the underlying model logic and decision outcome to non-technical personnel. Not all MCDA models are transparent, with some models functioning as 'black boxes' (Bruggemann & Carlsen, 2012; Fischer, 2009), being understood only by the most competent analysts. The WSM is not one of these models.

Quantitative Inputs

A further advantage of the WSM is its ability to utilise quantitative data alongside expert or stakeholder input, qualitatively measured on user defined numerical or semantic rating scales. In food safety decision-making there are often data gaps for a number of important criteria which can only be filled by utilising expert judgements. However, there may also be quantitative data available on a number of other criteria, obtained through modelling or measurement techniques. The WSM can utilise both types of data, usually with little mathematical manipulation. Some other methods require all performance scores to be entered in a qualitative fashion. Mapping precise quantitative data, often obtained through costly and time consuming modelling activities, onto rating scales (e.g. a 9-point scale) results in a loss of precision. The WSM's acceptance of different data types, coupled with its straightforward and easily understood MCDA algorithm, could offer significant credibility in food safety decision-making contexts.

Remapping of inputs onto homogeneous scales

It is conventional with models of this type to allot a value score to options on an interval scale, usually 0-100 (Mabin, 2006; Spackman et al., 2000). This is to ensure that differences in scores have consistency between criteria prior to aggregation. There are generally two choices for this type of scale; local and global scales. Typically, both types of scale will generate the same rankings (Mabin, 2006). However, global scales are a more appropriate choice if one wishes to add new options to the analysis at a later stage, especially if these options are likely to have extreme (good or bad) performances on at least one criterion. However, utilising global scales requires additional subjective judgements regarding the hypothetical extremes for each criterion. Local scales are generally easier to work with as they require less guesswork, but if new options are introduced to an analysis with extreme values, the scores of the existing options may need re-evaluation relative to the new extreme(s) (Mabin, 2006).

Comparison with Cost-Effectiveness Analysis

Cost-Effectiveness Analysis (CEA) is a decision-support methodology used to evaluate health interventions, including food safety interventions (Mangen, Havelaar, et al., 2007). CEA ranks interventions according to their Cost-Effectiveness Ratio (CER), which can be thought of as the cost of achieving one 'unit' of effectiveness (E.g. one Disability-Adjusted Life Year) under the intervention scenario being studied. CEA and MCDA are often treated as alternative techniques (Baltussen & Niessen, 2006; Claxton, 2013; Tony et al., 2011). Therefore an analysis was undertaken using a WSM model with only two equally weighted criteria ('Cost', and 'Effectiveness'), and the resultant rankings compared with those from CEA. The results of this analysis are presented in Figure 22 and Table 12.

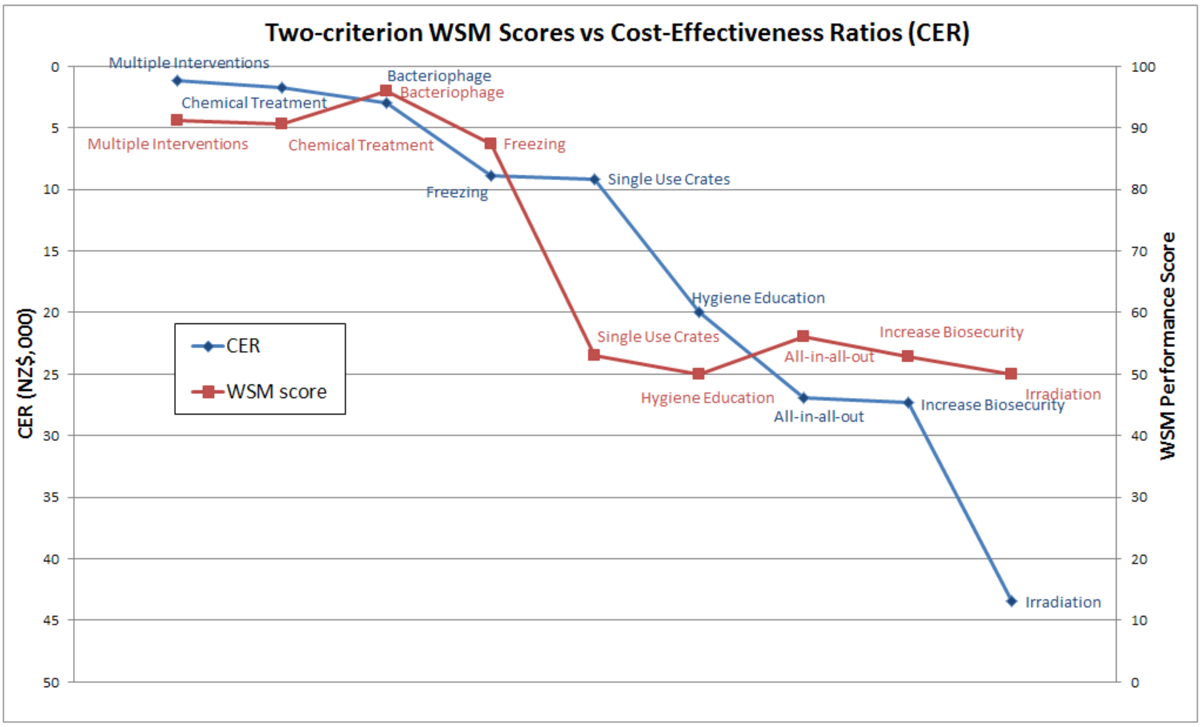


Figure 22: Cost-Effectiveness Ratios and 2-criterion WSM global scores

	CER	Rank
Multiple Interventions	$1,171	1
Chemical Treatment	$1,756	2
Bacteriophage	$2,991	3
Freezing	$8,870	4
Single Use Crates	$9,146	5
Hygiene Education	$20,000	6
All-in-all-out	$26,888	7
Increase Biosecurity	$27,273	8
Irradiation	$43,465	9

	WSM score	Rank
Multiple Interventions	91.3	2
Chemical Treatment	90.7	3
Bacteriophage	95.9	1
Freezing	87.4	4
Single Use Crates	53.0	6
Hygiene Education	50.0	8
All-in-all-out	56.1	5
Increase Biosecurity	52.9	7
Irradiation	50.0	8

Table 12: Cost-Effectiveness Ratios and 2-criterion WSM global scores

Although the rankings are by no means identical, both techniques place ‘Multiple-Interventions’, ‘Chemical Treatment’ and ‘Bacteriophage’ in the top group, with global scores markedly different from those of the remaining interventions. Both also rank ‘Freezing’ at number 4 and place ‘Increased Biosecurity’ and ‘Irradiation’ as the worst performing interventions. The major differences in these rankings occur for the 3 remaining interventions. ‘Single use crates’ moves from 5 to 6, ‘Hygiene Education’ from 6 to 8 and ‘All-in-all-out’ from 7 to 5 when ranked by their WSM scores rather than their CERs.

A major reason for the differences between the two analyses lies in the way the two methods aggregate their results. CEA produces a *ratio* of the cost to effectiveness. This ratio is completely independent of the performance of any other intervention. Conversely, the WSM produces a *weighted sum* of the criteria performance scores:

$$WSM\ Global\ Score = Weight_{Cost} \cdot Score_{\ Cost} + Weight_{Effectiveness} \cdot Score_{Effectiveness}$$

$$Cost\ Effectiveness\ Ratio = \frac{Cost}{Effectiveness}$$

Due to the differences in the way these two methods calculate overall performance, users cannot expect them to produce similar results, even if only 'Cost' and 'Effectiveness' criteria are included in both analyses. It may seem logical to assume that these two competing techniques should give similar results when identical input data is used. However, this is not a valid assumption.

The 'Bunching' Problem

One major difference between the CEA and 2-crtieron WSM scores from the analysis above lies in the performance of 'Irradiation', which for CEA is by far the worst (with a CER approaching twice that of the next intervention). This is not the case for 'Irradiation's' WSM score, which is equal to that of 'Hygiene Education' (even though 'Hygiene Education' is more than twice as cost-effective), and very close to the scores of some of the other interventions. Furthermore, the rankings of the top three interventions are different between the two analyses, despite the substantial differences in their cost-effectiveness. In the *Campylobacter* CEA study by Lake et al., the three most cost-effective interventions are (CER in parenthesis):

1. *Multiple interventions during primary processing ($1,200)*
2. *Continuous chemical controls (chlorine) during primary processing ($1,700)*
3. *Control of numbers of Campylobacter in faecal material by administration of bacteriophage prior to slaughter ($3,000)*

Note that the first ranked intervention is two and a half times more cost-effective than the third ranked intervention.

A reason for the difference between the two analyses is because the WSM scales the performance scores on each criterion relative to the maximum and minimum raw scores; a necessary procedure in order to represent criteria scores on homogeneous scales. Very large or small raw data values (e.g. 'Irradiation's' high cost) for one intervention can result in the remaining interventions receiving very similar criterion scores (i.e. 'bunching up' or clustering) when they are remapped onto a homogeneous scale, even if there are significant differences between these interventions when examined in a pairwise fashion.

In the WSM utilising a standard 0-100 local scale (global scales would give similar results) the worst performing option on each criterion is allocated a score of zero, with the best performer scoring 100, and all other options allocated values relative to these extremes. This process generated the performance scores in Table 13.

	Cost	Effectiveness
Multiple Interventions	98.1	84.4
Chemical Treatment	96.9	84.4
Bacteriophage	93.6	98.3
Freezing	81.2	93.7
Single Use Crates	98.7	7.4
Hygiene Education	100.0	0.0
All-in-all-out	80.0	32.2
Increase Biosecurity	90.1	15.7
Irradiation	0.0	100.0

Table 13: Remapped cost and effectiveness data on homogeneous scales (0-100 local scales)

When data is remapped onto the homogeneous scale, 'Hygiene Education' is allocated the highest possible score on the 'Cost' criterion as it is the intervention with the lowest cost. Conversely, 'Irradiation' is assigned a score of zero indicating that it is the most expensive intervention. All other costs are scaled relative to these two 'extremes'. Similarly, on the 'Effectiveness' criterion 'Irradiation' performs the best and 'Hygiene Education' the worst, with scores of 100 and 0 respectively.

On the 'Cost' criterion all interventions other than the two extremes have high scores in the 80's or 90's. This reflects the fact that their costs lie much closer to those of the best performing option ('Hygiene Education') than the worst performing option ('Irradiation').

‘Irradiation’ has a very high cost, many times that of the next most expensive intervention (‘Freezing’), and it is this cost that is causing the remaining performance scores to ‘bunch’ around the higher end of the scale. Looking at the performance scores on the ‘Cost’ criterion for ‘Multiple Interventions’, ‘Chemical Treatment’ and ‘Bacteriophage’ (the three most cost-effective interventions), all three have very similar (and very high!) performance scores of 98.1, 96.9, and 93.6 respectively, despite ‘Chemical Treatment’ having a cost one-and-a-half times that of ‘Multiple Interventions’ and ‘Bacteriophage’ having a cost three times that of ‘Multiple Interventions’.

If the two criteria are given equal weightings in the WSM, the global scores will be calculated as the average of the two performance scores. This results in ‘Bacteriophage’ ranking number one in the two-criterion WSM, despite its far less favourable cost-effectiveness ratio when compared with ‘Multiple Interventions’ (Table 14).

The cost of ‘Bacteriophage’ is three times that of ‘Multiple Interventions’, but because of the distortive effect of ‘Irradiation’ this is reflected by a score differential of only 4.5 (93.6 vs 98.1) on the ‘Cost’ scale. In contrast, ‘Bacteriophage’ outperforms ‘Multiple Interventions’ by only 16% in terms of its DALY benefit (993 DALYs vs 854 DALYs), but this is reflected by a score differential of 14 on the ‘Effectiveness’ criterion (98 vs 84). This is because there is comparatively much less distortion occurring on the ‘Effectiveness’ criterion. When the global scores of the two options are then calculated as:

$$Global\ Score = Weight_{Cost} \cdot Score_{\ Cost} + Weight_{Effectiveness} \cdot Score_{\ Effectiveness}$$

$$Global\ Score_{Bacteriophage} = 50\% \cdot 93.6 + 50\% \cdot 98.3 = \mathbf{95.9}$$

$$Global\ Score_{Multiple\ Interventions} = 50\% \cdot 98.1 + 50\% \cdot 84 = \mathbf{91.3}$$

‘Bacteriophage’ outranks ‘Multiple Interventions’, despite its much worse CER.

When ‘Irradiation’ is removed from the two-criterion WSM the scores on the homogeneous scales adjust to better reflect the true cost differences between the remaining interventions (i.e. they ‘unbunch’) and the results end up closely matching the results of CEA (Table 14 and 15).

Remapped Scores (Original)

	Cost	Effectiveness	Global Score	Rank
Multiple Interventions	98	84	91.3	2
Chemical Treatment	97	84	90.7	3
Bacteriophage	94	98	95.9	**1**
Freezing	81	94	87.4	4
Single Use Crates	99	7	53.0	6
Hygiene Education	100	0	50.0	8
All-in-all-out	80	32	56.1	5
Increase Biosecurity	90	16	52.9	7
Irradiation	0	100	50.0	8

Remapped Scores (Irradiation Removed)

	Cost	Effectiveness	Global Score	Rank
Multiple Interventions	90	86	88.1	**1**
Chemical Treatment	85	86	85.3	2
Bacteriophage	68	100	83.9	3
Freezing	6	95	50.5	4
Single Use Crates	93	8	50.4	5
Hygiene Education	100	0	50.0	6
All-in-all-out	0	33	16.4	8
Increase Biosecurity	50	16	33.1	7

Table 14: 2-criterion WSM scores with and without Irradiation

Distortion can occur when quantitative data is remapped onto homogeneous scales. Irradiation's very high cost inflates the other 'Cost' scores when remapped onto a homogenous scale (top table). When 'Irradiation' is included in the two-criterion WSM, the rankings are very different to the Cost-Effectiveness rankings. However, when 'Irradiation' is removed the distortion is heavily reduced and the global scores better reflect the differences in performance, with the rankings closely resembling the Cost-Effectiveness rankings (bottom table).

	CER	Rank
Multiple Interventions	$1,171	1
Chemical Treatment	$1,756	2
Bacteriophage	$2,991	3
Freezing	$8,870	4
Single Use Crates	$9,146	5
Hygiene Education	$20,000	6
All-in-all-out	$26,888	7
Increase Biosecurity	$27,273	8
Irradiation	$43,465	9

	WSM score	Rank
Multiple Interventions	88.1	1
Chemical Treatment	85.3	2
Bacteriophage	83.9	3
Freezing	50.5	4
Single Use Crates	50.4	5
Hygiene Education	50.0	6
All-in-all-out	16.4	8
Increase Biosecurity	33.1	7

Table 15: Results of the CEA and two-criteria WSM ('Irradiation' removed) utilising identical cost and effectiveness data

Distortion can be an issue for any MCDA methods utilising homogeneous scales. Because of the scaling of scores required to remap raw data onto these scales, with all interventions being scored relative to the minimum and maximum values, there will always be some degree of distortion. However, this distortion can be minimised by removing options with 'extreme' scores (i.e. 'Irradiation's' cost) on any one criterion. This will allow the criteria scores of the remaining interventions to better reflect the differences in their performance, and will allow for a more reliable and robust analysis.

Including Cost-Effectiveness in MCDA

If users wish for their model to explicitly consider cost-effectiveness then the cost-effectiveness ratios will need inclusion as a standalone criterion in the model. It may be appropriate to replace the 'Cost' and/or 'Effectiveness' criterion in the model with 'Cost-Effectiveness'. Table 16 and Figure 23 show the effect of utilising different combinations of 'Cost' (C), 'Effectiveness' (E) and 'Cost-Effectiveness' (CE) data in the original WSM analysis. The last columns show the cost-effectiveness ratios. In the WSM the 'Cost', 'Effectiveness' and 'Cost-Effectiveness' criteria were given equal weightings of 100, with all other criteria scores and weights remaining as they were in the original analysis.

	WSM Models						CEA	
	C + E	Rank	C + CE	Rank	E + CE	Rank	CER	Rank
Multiple Interventions	87.3	1	87.1	1	87.7	1	87.7	1
Chemical Treatment	87.1	2	86.6	2	87.4	2	87.4	2
Hygiene Education	67.4	3	77.7	3	59.0	5	59.0	5
Bacteriophage	67.0	4	63.4	6	67.4	3	67.4	3
Increase Biosecurity	62.6	5	65.9	4	53.0	6	53.0	6
Freezing	61.1	6	61.2	7	61.2	4	61.2	4
All-in-all-out	53.3	7	53.7	8	45.7	9	45.7	9
Single Use Crates	52.2	8	65.0	5	48.9	8	48.9	8
Irradiation	51.4	9	32.7	9	51.4	7	51.4	7

Table 16: Results of CEA and original WSM with different combinations of Cost, Effectiveness and Cost-Effectiveness input data

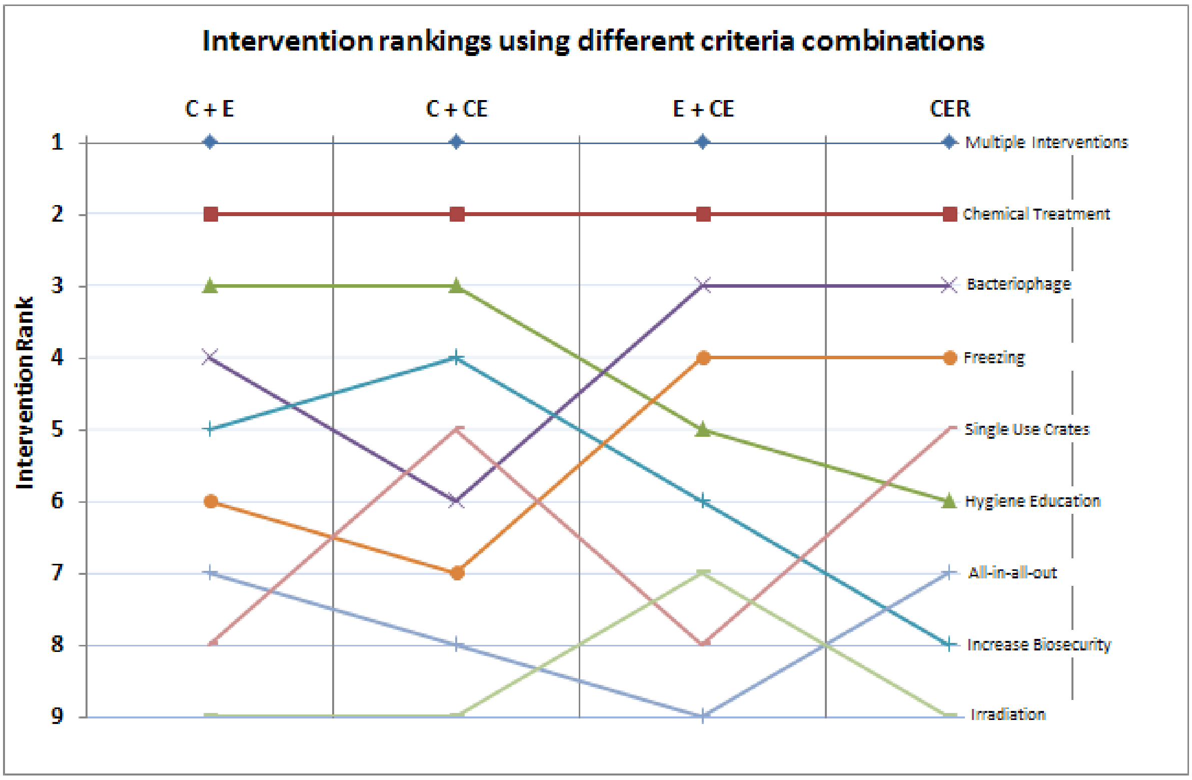


Figure 23: Results of CEA and original WSM with different combinations of Cost, Effectiveness and Cost-Effectiveness input data

For the *Campylobacter* case study the different model setups do not affect the rankings of the top two interventions but do influence the rankings of all other interventions. Replacing 'Effectiveness' with 'Cost-Effectiveness' has the effect of more heavily favouring the cost-effective interventions and less favouring interventions with high effectiveness. ‘Bacteriophage’, ‘Freezing’ and ‘All-in-all-out’ move down the rankings by one spot due to their poor performance on the ‘Cost’ criterion and because their high effectiveness is no longer explicitly considered in the analysis. In contrast, ‘Increase Biosecurity’ and ‘Single use crates’ move up by 1 and 3 places respectively due to their low cost and their very poor performance on the ‘Effectiveness’ criterion being no longer explicitly considered in the analysis. This situation of including ‘Cost’ and ‘Cost-Effectiveness’ while excluding ‘Effectiveness’ from the analysis seems unlikely given that the primary goal of most food-safety interventions is to reduce the burden of foodborne disease. However, this situation may occur if, for example, decision-makers are faced with a finite budget for addressing a food-safety issue.

Including ‘Effectiveness’ and ‘Cost-Effectiveness’ but excluding ‘Cost’ from the analysis has the effect of favouring the more effective and cost-effective interventions over those with a low cost. ‘Bacteriophage’, ‘Freezing’ and ‘Irradiation’ rank higher than in the previous two scenarios, due to their very high effectiveness being explicitly considered in the model while their high cost is not. Under this same scenario ‘Hygiene Education’, ‘Increase Biosecurity’, ‘Single Use Crates’ and ‘All-in-all-out’ all drop in the rankings. These are options with lower costs which are no longer being explicitly considered in the model. Furthermore, apart from ‘Irradiation’, these are the four interventions with the highest (worst) cost-effectiveness ratios.

‘Hygiene Education’ and ‘Increase Biosecurity’ are high performers on the ‘Cost’ criterion, and steadily slip to sixth and eighth place respectively as ‘Cost’ is given progressively less influence over the results in the four analyses. Both interventions perform worse when assessed according to their CERs than in the WSM analyses, due to their strong performances on ‘Consumer Perceptions’ and ‘Ease of Implementation’ being no longer considered in CEA.

The changes to the rankings of most interventions when moving from MCDA to CEA is a good example of how CEA fails to capture the big picture, and can result in rankings that do not reflect the complete preferences of participants.

8.5.2 Analytic Hierarchy Process (AHP)

8.5.2.1 Construction

A classic AHP model, proposed by Saaty (Saaty, 1980), was implemented in Excel 2010. This model requires the comparison and rating of every possible pair of interventions on every criterion. A 9-point rating scale is used for all pairwise comparisons (Table 17); including the generation of criterion weights (Section 4.6.3 has more detail on the generation and aggregation of inputs using AHP).

Rating	Description
1.00	Equally Preferred
2.00	
3.00	Moderately Preferred
4.00	
5.00	Strongly Preferred
6.00	
7.00	Very Strongly Preferred
8.00	
9.00	Extremely Preferred

Table 17: The standard AHP 9-point rating scale

Pairwise Comparison Matrix for Effectiveness

	Multiple Interventions	Chemical Treatment	Bacteriophage	Freezing	Single Use Crates	Hygiene Education	All-in-all-out	Biosecurity	Irradiation
Multiple Interventions	1	1	1/2	1/2	9	9	7	9	1/2
Chemical Treatment	1	1	1/2	1/2	9	9	7	9	1/2
Bacteriophage	2	2	1	2	9	9	8	9	1/2
Freezing	2	2	1/2	1	9	9	8	9	1/2
Single Use Crates	1/9	1/9	1/9	1/9	1	4	1/5	1/3	1/9
Hygiene Education	1/9	1/9	1/9	1/9	1/4	1	1/7	1/6	1/9
All-in-all-out	1/7	1/7	1/8	1/8	5	7	1	5	1/7
Biosecurity	1/9	1/9	1/9	1/9	3	6	1/5	1	1/8
Irradiation	2	2	2	2	9	9	7	8	1

Table 18: AHP pairwise comparison (input) matrix

In Table 18 the grey cells in the upper triangle of the matrix (36 comparisons) need completing for *each* of the eight criteria in the *Campylobacter* case study.

Normalised Comparison Matrix for Effectiveness

	Multiple Interventions	Chemical Treatment	Bacteriophage	Freezing	Single Use Crates	Hygiene Education	All-in-all-out	Biosecurity	Irradiation	Criterion Scores
Multiple Interventions	0.12	0.12	0.10	0.08	0.17	0.14	0.18	0.18	0.14	**0.14**
Chemical Treatment	0.12	0.12	0.10	0.08	0.17	0.14	0.18	0.18	0.14	**0.14**
Bacteriophage	0.24	0.24	0.20	0.31	0.17	0.14	0.21	0.18	0.14	**0.20**
Freezing	0.24	0.24	0.10	0.15	0.17	0.14	0.21	0.18	0.14	**0.17**
Single Use Crates	0.01	0.01	0.02	0.02	0.02	0.06	0.01	0.01	0.03	**0.02**
Hygiene Education	0.01	0.01	0.02	0.02	0.00	0.02	0.00	0.00	0.03	**0.01**
All-in-all-out	0.02	0.02	0.03	0.02	0.09	0.11	0.03	0.10	0.04	**0.05**
Biosecurity	0.01	0.01	0.02	0.02	0.06	0.10	0.01	0.02	0.04	**0.03**
Irradiation	0.24	0.24	0.40	0.31	0.17	0.14	0.18	0.16	0.29	**0.24**

Table 19: Normalised AHP matrix for 'Effectiveness'

Criterion weights are generated in the same way as performance scores, by completing a pairwise comparison matrix comparing every pair of criteria. The *Campylobacter* case study requires 36 comparisons for each of the eight criteria and a further 28 comparisons to produce criterion weights; a total of 316 comparisons.

Performance scores and weights are calculated according to the mathematical procedure described in Section 4.6.3. The primary output of this model is a single global performance score for each option, which can be used to rank interventions.

8.5.2.2 Results

The global score for each intervention is calculated as the sum of the criteria scores multiplied by their respective weights. Inputs and outputs for the AHP analysis are shown in Table 20 and the global scores are graphed in Figure 24.

AHP overall performance scores

	Cost	Effectiveness	Market Access	Consumer Perceptions	Quality of Suitability	Ease of Implementation	Quality of Science	Equity of costs	Global Score	Rank
Hygiene Education	0.37	0.01	0.19	0.35	0.05	0.27	0.09	0.12	0.19	**1**
Multiple Interventions	0.14	0.14	0.04	0.04	0.19	0.15	0.19	0.12	0.14	**2**
Chemical Treatment	0.09	0.14	0.04	0.04	0.19	0.15	0.19	0.12	0.13	**3**
Irradiation	0.01	0.24	0.01	0.02	0.34	0.02	0.21	0.12	0.13	**4**
Freezing	0.03	0.17	0.10	0.07	0.02	0.02	0.19	0.02	0.10	**5**
Single Use Crates	0.20	0.02	0.19	0.15	0.05	0.07	0.03	0.12	0.10	**6**
Bacteriophage	0.07	0.20	0.04	0.04	0.05	0.07	0.02	0.12	0.09	**7**
Biosecurity	0.05	0.03	0.19	0.15	0.05	0.16	0.05	0.12	0.07	**8**
All-in-all-out	0.03	0.05	0.19	0.15	0.05	0.07	0.03	0.12	0.06	**9**
Criteria Weights	27.5%	23.0%	2.9%	6.9%	6.9%	10.5%	20.5%	1.9%		

Table 20: AHP final performance matrix showing criteria scores, global performance scores and final intervention rankings

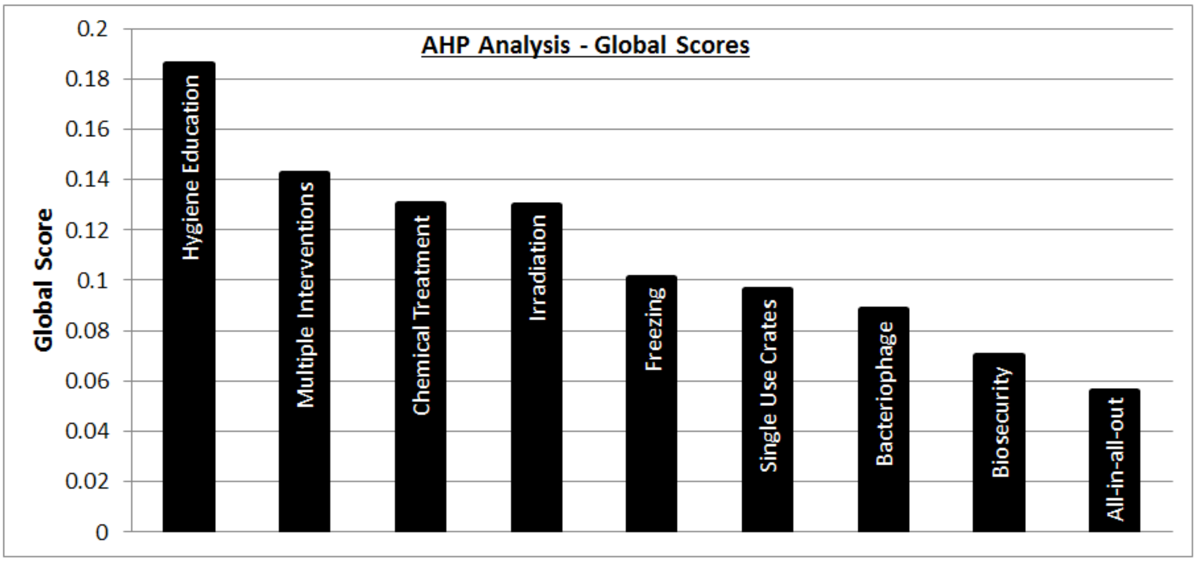


Figure 24: AHP global performance scores

In this AHP analysis 'Hygiene Education' ranks number one, followed by 'Multiple Interventions' and 'Chemical Treatment'. 'Irradiation' also scores highly in this analysis relative to the WSM. 'Hygiene Education' achieves a higher rank than in the WSM analysis because of its comparatively better performance score on the 'Cost' criterion. 'Irradiation' scores comparatively well due to its top performance on the 'Effectiveness', 'Quality or Suitability' and 'Quality of science' criteria.

8.5.2.3 Critique

Pairwise comparisons

One often stated strength of AHP is its ability to break a problem down into a number of easy to manage pairwise comparisons. Rather than assessing the performance of an option relative to all others at the same time, AHP compares each option to each alternative, one at a time, under the rationale that comparing only two options at any one time is a much easier mental task for decision-makers. There are a number of studies emphasising that this process of comparing only two options at a time is easy for, and attractive to, non-specialists (DCLG, 2009; de Montis et al., 2005; Spackman et al., 2000). Although this premise may be true, and would be useful for MCDAs involving few options and criteria, in problems with even a moderate number of criteria and options the number of required pairwise comparisons becomes large. In the *Campylobacter* case study there are nine options and eight criteria. This means that for each criterion the decision-maker is required to make 36 pairwise comparisons. The *Campylobacter* case study required a total of 288 pairwise comparisons to generate criteria performance scores, and an additional 28 to generate criteria weights. This could be a demanding task for participants, requiring a large investment of time and concentration in order to conduct a meaningful analysis. Such models are often referred to as placing a high *"elicitation burden"* on participants (Karvetski et al., 2011; Montibeller et al., 2008).

Consistency of preferences

Preference consistency requires that, for example, if Option 1 is twice as attractive as Option 2 and Option 2 is twice as attractive as Option 3 then Option 1 must be four times as attractive as Option 3. In order to satisfy this requirement decision-makers must ensure that the pairwise comparisons they make are sufficiently consistent, as assessed by the consistency indices (see Section 4.6.3). Although a decision-maker need only consider two options at a time when making pairwise comparisons, ensuring consistency requires that they remain cognisant of previous comparisons they have made, and possibly even the future comparisons they will make, within the comparison matrix they are completing. Ensuring sufficient consistency when completing pairwise comparison matrices can be a difficult task and consistency may not be achieved on the first attempt. If a comparison matrix is deemed to be inconsistent (as indicated by an unacceptable consistency index) without any additional information regarding where the inconsistencies lie, revising these preferences can be a difficult task. There is the question of which preferences are inconsistent and which preferences should be altered to achieve consistency? Without the use of specialised software to guide the user the task of revising pairwise preferences can be frustrating. Often revising some preferences to achieve consistency can lead to inconsistency in other preferences. This requirement of sufficient preference consistency before proceeding could lead to the participant making rushed or even arbitrary changes to their preferences in order to achieve consistency. Completing over 300 pairwise comparisons and achieving sufficient consistency within each of nine performance matrices was a challenging and time consuming task in this research.

This requirement also makes it more difficult to conduct sensitivity analysis. If conducting sensitivity analysis on the raw inputs (I.e. the pairwise comparisons), there is the issue of which preferences to alter and how to alter these in a way that maintains consistency. There are so many pairwise comparisons in this example that it is very difficult to identify the comparisons that are having the most influence over the results, and thus conduct sensitivity analysis on these inputs. Furthermore, if participants wish to vary their pairwise comparison values there is the problem of maintaining consistency while this is done. Altering one preference as part of sensitivity analysis may lead to inconsistency with other preferences. This problem is not insurmountable, but makes sensitivity analysis more difficult than with other models. An alternative approach is to vary the criteria scores and weights themselves (i.e. the outputs, rather than inputs, of the individual pairwise comparison matrices). This can be done in the same way as with the WSM, by allowing weights to vary over a range, with the remaining weights readjusting to ensure that they sum to 1 (Chang et al., 2007). However, this process could be viewed as causing a disconnection between the scores/weights and the pairwise comparisons used to generate them; deviating away from the pairwise comparison approach that defines AHP.

Specialised AHP software could be very helpful, perhaps necessary, to guide new users through such comprehensive AHP analyses. EXPERT CHOICE and HIPRE 3+ are examples of software packages that can help guide the analyses (Spackman et al., 2000). Also, eliminating potentially unnecessary criteria or, where there are quantitative data inputs, utilising a modified AHP incorporating normalised quantitative data can reduce the elicitation burden on participants (see Section 8.5.2.3 – 'Quantitative Inputs').

Another potential issue with AHP arises through the use of a finite rating scale (i.e. the 9-point scale) when analysing data with criteria scores that vary over wide ranges. For example, if Option A scores 6 times more highly than does Option B, which in turn scores 6 times more highly than does Option C, there is a limitation in AHPs ability to capture the fact that Option A should score 36 times more highly than does Option C if using a 9-point scale. The homogeneity axiom of the AHP suggests that all options being analysed should have criteria performances that are similar enough to be captured by the 9-point scale (Saaty, 2004). That is, no option should score more than nine times more highly than another on any criterion. The standard 9-point scale is unable to cope with options with attributes that vary over a very wide range of values as is the case with the 'Cost' criterion in the *Campylobacter* case study (Figure 25). One way of overcoming this problem involves grouping the options into *"homogeneous sets"*, where the best performing option in each set is no more than 9 times 'better' than the worst (Saaty, 2004).

Semantic rating scale

Another concern with AHP is that the link between the numerical values of the 1-9 rating scale and the corresponding verbal descriptions lacks a theoretical foundation. The verbal statements utilised in this scale are subjective and the interpretation of them is likely to vary between decision-makers. What does 'moderately preferred' mean? What is the difference between 'very strong preference' and 'extreme preference'? How does a decision-maker decide which verbal rating to use in their pairwise comparisons? In a context such as high-level food safety decision-making, where there is a need to involve a wide array of stakeholders, differences in the way participants complete the performance matrices may arise due to different interpretations of the verbal statements, rather than actual differences in participant preferences. There is the potential for different stakeholders who share identical preferences to complete their pairwise comparison matrices differently because of differences in their interpretations of words like *"moderate", "strong", "very strong"* and *"extreme"*. Conversely, stakeholders with different preferences could produce similar matrices for the same reason.

If using a bounded (e.g. 9-point) scale an individual's interpretations of different levels of that scale may seem dependent on the interventions being studied. If participants wish to introduce a new intervention into an analysis they may need to reconsider their pairwise comparisons, and thus alter their interpretations of the verbal statements from the scale. For example, consider the *Campylobacter* in poultry example. If participants were to conduct pairwise comparisons using a 9-point scale on the 'Cost' criterion for all interventions except 'Irradiation', they may determine that 'Chemical Treatment' (cost = $1.5m) is 'extremely preferred' to 'Freezing' (cost = $8.4m). However, if 'Irradiation' (cost = $43.9m) is subsequently introduced into the analysis the user would likely need to reconsider their interpretation of the word 'extreme', and would perhaps re-evaluate their initial pairwise comparison between 'Chemical Treatment' and 'Freezing', and assign it a lower level on the scale (Figure 25).

Matrix A

	Chemical treatment	Freezing
Chemical treatment	1	9
Freezing	1/9	1

Matrix B

	Chemical treatment	Freezing	Irradiation
Chemical treatment	1	9	A
Freezing	1/9	1	9
Irradiation		1/9	1

Figure 25: Homogeneity and a limitation of the 9-point scale

In Figure 25, if ‘Chemical Treatment’ is ‘extremely’ preferred over ‘Freezing’ (matrix A) and ‘Irradiation’ is subsequently added to the analysis and there is, in turn, an ‘extreme preference’ for ‘Freezing’ over ‘Irradiation’ (Matrix B), a problem arises when allocating a 9-point scale value between ‘Chemical Treatment’ and ‘Irradiation’ (Cell *A* in Matrix B). If consistency is to be achieved then Cell *A* requires a value of 81 (9x9). To achieve consistency, preferences could be revised or the scale extended; neither option is ideal. Such data violates the AHP axiom of homogeneity and is potentially unsuitable for use in a simple AHP analysis.

The standard 9-point scale also fails to accommodate situations where ratios between the performances might be less than two. That is, the scale cannot readily cope with situations where an option might be considered less than twice as ‘good’ on a criterion. For the *Campylobacter* case study there is no standard rating that can reflect the fact that ‘Chemical Treatment’ costs only 50% more than ‘Multiple Interventions’. To overcome situations where the ratio of preference is less than 2 it is possible to utilise scales between 1 and 2, using decimals (Saaty, 2004).

The problem of introducing new interventions with ‘extreme’ values and these interventions significantly influencing the performance scores of the remaining interventions is similar to the ‘bunching’ problem occurring with the use of 0-100 rating scales in the Weighted Sum Model (see Section 8.5.1.3). However, adjusting the analysis to cope with new interventions is a much more demanding task with AHP. AHP could cope with such situations without the need to revise preferences, but this may require the extension of the 9-point scale, and this would cause a disconnection between the numerical and verbal ratings on the scale. If a 9 on the scale corresponds to ‘extreme preference’ and the scale is subsequently extended or revised then under the new scale it is unlikely that a 9 will continue to correspond with a verbal rating of ‘extreme’. There are modifications to the AHP that will allow for the comparison of non-homogeneous options without the need to revise preferences or scales through; for example, organising options into hierarchies where, within each hierarchy, options can be evaluated using the 9-point scale (Macharis et al., 2004; Saaty, 2001; Saaty & Vargas, 2012). However, as the levels of the hierarchy increase, so does the difficulty and time required to complete judgements and synthesise weights (Alexander, 2012).

Rating	Description
1.00	Equally Preferred
2.00	
3.00	Moderately Preferred
4.00	
5.00	Strongly Preferred
6.00	
7.00	Very Strongly Preferred
8.00	
9.00	Extremely Preferred

Table 21: Standard AHP 9-point rating scale with corresponding verbal statements

There is also a potential issue with the validity of assuming that the ordinal verbal statements of the 9-point scale can be translated into numerical values on a 1-9 ratio scale, as is the case with AHP. Verbal judgements such as saying something is ‘moderately preferred’ are not usually thought of as ratio comparisons. However, this is exactly how the standard AHP treats such verbal ratings. When a decision-maker assigns a preference of, for example, ‘moderately preferred’ between two options on a criterion, this verbal statement corresponds to a value of 3 on the 9-point ratio scale, implying that the more preferred option performs three times as well as the less preferred option. Similarly, the scale implies that having an ‘extreme preference’ is akin to an option performing nine times better than another. Even the lowest possible degree of preference, other than being indifferent, implies that the more preferred option is ‘twice as good’ as its comparator option. To many, an option performing twice as well as another may not correspond to a verbal rating somewhere between ‘no preference’ and ‘moderate preference’. To some participants the verbal statements may not correspond well with the numerical ratios on the 9-point scale. For example, in a food safety decision-making context a decision-maker may be comparing two options in terms of their cost-effectiveness; it may be reasonable for a decision-maker to exhibit an ‘extreme’ preference for an option which has a cost-effectiveness ratio half that (i.e. twice as good) of a comparator, and yet this verbal rating would correspond to a score of 9 on the 9-point scale. This would result in the preferred option being treated as being nine times more cost-effective in the analysis.

Without adequate training of participants regarding how verbal statements are translated into numerical values and utilised in AHP models, the use of such scales could generate model inputs (and thus outputs) that do not correctly reflect the preferences of participants. The possible issues with the 9-point scale described in this section are not insurmountable. However, overcoming these issues can require significant modifications to the model/analysis, which adds a layer of complexity.

Quantitative inputs

AHP generates performance scores through the completion of pairwise comparison matrices. Each pairwise comparison must be allocated a value on the 9-point scale. For criteria for which there is quantitative data available this would require its conversion into a series of pairwise comparison values from the 9-point scale. This act of converting quantitative data into subjective performance scores represents a loss of information and an additional layer of uncertainty. In food safety decision-making, where there is often some quantitative data available for use in decision-making, there is a desire to keep analyses as precise as possible. This can be achieved through the use of MCDA models which can accept quantitative inputs with a minimal amount of mathematical transformation. Fortunately, the AHP is able to be modified so that it can accept quantitative inputs without the need to conduct subjective pairwise comparisons (de Montis et al., 2005). The output of the classical pairwise comparison approach is a set of normalised criterion performance scores summing to 1. If a decision-maker is able to provide the AHP analysis with normalised quantitative data in this same format then the aggregation procedure is able to utilise this information alongside the qualitatively assessed criterion scores for other criteria. Inputting quantitative data into AHP in this way can aid the analysis in a number of ways:

- By retaining a degree of precision in the quantitative data
- By minimising the elicitation burden on the user by reducing the number of pairwise comparisons

- By ensuring the criterion performance scores are completely consistent

Modified AHP overall performance scores

	Cost	Effectiveness	Market Access	Consumer Perceptions	Quality of Suitability	Ease of Implementation	Quality of Science	Equity of Costs	Global Scores	Rank
Multiple Interventions	0.13	0.16	0.04	0.04	0.19	0.15	0.19	0.12	0.15	1
Chemical Treatment	0.13	0.16	0.04	0.04	0.19	0.15	0.19	0.12	0.15	2
Freezing	0.11	0.18	0.10	0.07	0.02	0.02	0.19	0.02	0.12	3
Hygiene Education	0.14	0.00	0.19	0.35	0.05	0.27	0.09	0.12	0.12	4
Irradiation	0.00	0.19	0.01	0.02	0.34	0.02	0.21	0.12	0.12	5
Bacteriophage	0.13	0.19	0.04	0.04	0.05	0.07	0.02	0.12	0.10	6
Biosecurity	0.12	0.03	0.19	0.15	0.05	0.16	0.05	0.12	0.09	7
All-in-all-out	0.11	0.06	0.19	0.15	0.05	0.07	0.03	0.12	0.08	8
Single Use Crates	0.13	0.01	0.19	0.15	0.05	0.07	0.03	0.12	0.08	9

Table 22: Modified AHP performance matrix with global performance scores

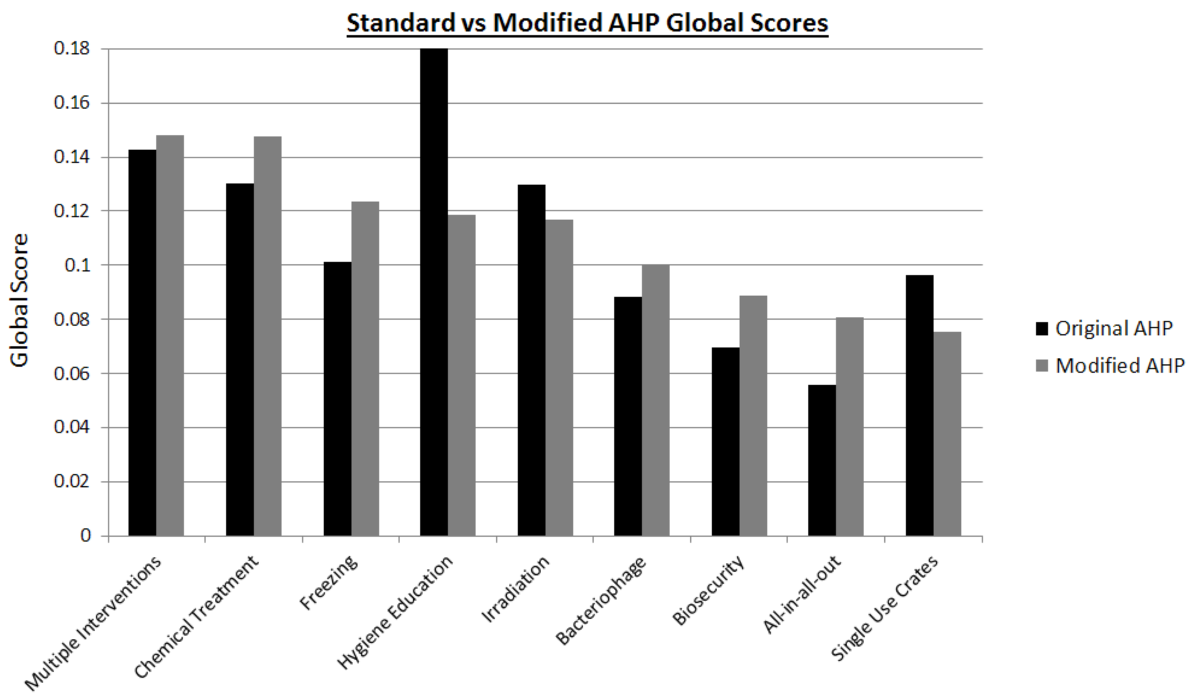


Figure 26: Comparison of standard and modified AHP performance scores

Table 22 shows the results of modifying the original AHP to utilise normalised cost and effectiveness data rather than requiring pairwise comparisons for these criteria. Figure 26 shows the difference between the original AHP and modified AHP global scores. This illustrates the potential difficulty with accurately representing quantitative data using the pairwise comparisons approach. In the original analysis a large amount of effort went into attempting to select pairwise comparisons which reflected the actual differences between the quantitative data for both the 'Cost' and 'Effectiveness' criteria. Despite this, there was still a significant difference between the results of the two analyses, resulting in the selection of 'Hygiene Education' (by a significant margin) as the highest scoring intervention in the original analysis. The results of the modified AHP are broadly in line with the results of the other models utilised in this chapter. However, similar to the WSM, incorporating quantitative data in a modified AHP analysis in this way does not avoid the 'bunching' problem (see Section 8.5.1.3). It may be wise to also rerun a modified AHP analysis after the removal of options with very high or very low values on at least one criterion, so as to minimise their distortive effects.

What differentiates the AHP from the standard Weighted Sum Model is its pairwise comparison approach to the generation of performance scores. Incorporating quantitative data in the above manner is, in a way, a hybrid of the WSM and AHP methodologies. If a model contained only quantitative data, incorporated in the above manner, the analysis would simply reduce to a Weighted Sum Model. AHP can be thought of as a Weighted Sum Model that utilises a unique procedure for the elicitation of performance scores and criterion weights. In a situation where there are a number of criteria that can be assessed using available quantitative data it would be wise for the decision-maker to question whether the AHP is the most appropriate method, and ask themselves whether a different technique (such as the Weighted Sum Model) might be more appropriate. AHP seems more appropriate for situations where the majority, if not all, data is of a qualitative nature.

Rank Reversal

A problem that can sometimes occur in an AHP is 'rank reversal', where the introduction of new options or the removal of existing options can change the relative rank of the original options (Alexander, 2012; Barzilai & Golany, 1994). An example can be seen in the *Campylobacter* case study, where the simple act of completely removing 'Irradiation' from the original analysis has the effect of reversing the relative rank of 'Single Use Crates' and 'Bacteriophage' (Table 23).

	Original Analysis		Irradiation Removed	
	Score	Rank	Score	Rank
Hygiene Education	0.19	1	0.21	1
Multiple Interventions	0.14	2	0.17	2
Chemical Treatment	0.13	3	0.16	3
Irradiation	0.13	4	-	-
Freezing	0.10	5	0.12	4
Single Use Crates	0.10	6	0.10	6
Bacteriophage	0.09	7	0.11	5
Biosecurity	0.07	8	0.08	7
All-in-all-out	0.06	9	0.06	8

Table 23: Illustration of rank reversal in the AHP case study

This problem was first reported by Belton and Gear and stems from AHP's failure to consistently relate scales of performance (measurement) to their associated weights (Belton & Gear, 1983).

The AHP is a very popular MCDA method which has been thoroughly studied and has received wide application in a large array of contexts other than food safety, especially in engineering, manufacturing, education, logistics and politics (Ho, 2008; Vaidya & Kumar, 2006). However, there are some largely unresolved problems that accompany the use of the AHP, generally arising from the demands it places on the decision-maker(s) coupled with the inherent subjectivity of basing comparisons on verbal ratings. Additionally, the possibility of rank-reversal reduces AHP's credibility and its potential for consistent and widespread use as a high-level food safety decision-support tool. However, the straightforward (albeit more time consuming) pairwise comparison approach is intuitive and easy to understand and use for problems where the number of criteria and options is small, and the differences in performance easily captured on the standard 9-point scale. The use of semantic rating scales, rather than quantitative scales, may actually be more appropriate, and the preferred approach, for problems where performance scores cannot be quantitatively assessed. This is often the case in food safety decision-making contexts, where there can be a lack of quantitative data, or no logical way to quantify some criteria. For decision-making situations such as the *Campylobacter* case study, where there is some quality quantitative data available but also a number of criteria that must be qualitatively assessed, the modified AHP, utilising normalised quantitative data in conjunction with the standard AHP pairwise comparison approach, could be an appropriate alternative to either the Weighted Sum Model or conventional AHP. This approach is indeed a hybrid of the two models and avoids the requirement of subjectively transforming quantitative data into semantic ratings, while also reducing the elicitation burden on participants. Additionally, not all criteria mentioned at the beginning of this chapter will be relevant to every food safety decision-making situation, and the AHP may seem far more appealing in situations where there are a small number of criteria and/or options to consider.

8.5.3 PROMETHEE

8.5.3.1 Construction

The PROMETHEE family of algorithms compares every pair of interventions on every criterion in order to establish the degree to which an intervention outperforms all others (known as its positive flow) and the degree to which an intervention is outperformed by all others (negative flow). An intervention scores well if it has a high positive flow and a low negative flow. The PROMETHEE I model generates a partial ranking of options based on these two flow figures. If an option has a higher positive flow and a lower negative flow than another it is said to outrank that option. However, if an option has a higher (lower) positive flow but also a higher (lower) negative flow then, under the PROMETHEE I methodology, the two options cannot be ranked and are said to be incomparable. If a PROMETHEE I analysis contains options which it defines as incomparable it is unable to clearly rank all options in the analysis, and this is why PROMETHEE I guarantees only a partial ranking (Brans & Vincke, 1985).

Different from PROMETHEE I, the PROMETHEE II algorithm combines the two flows into a 'Net flow' figure, and it is this number which is used to generate a complete (rather than partial) ranking of interventions (Brans & Vincke, 1985).

PROMETHEE converts quantitative data or user defined criterion scores into degrees of preference between every pair of interventions on every criterion. This is somewhat similar to AHP except PROMETHEE completes all pairwise comparisons automatically (and in the background). Participants must still assign a performance score to every intervention on every criterion, but are not required to explicitly consider any pairwise comparisons. Mathematically, pairwise preference relationships (different from performance scores) are expressed in the model in one of three categories:

0 = 'indifference': both options are equally preferred on a criterion

0 < 1 = 'some preference': The better performing option is preferred but not 'strictly preferred' (see below) over another on a criterion. This is represented in the model as a number between 0 and 1. The closer this number is to 1 the stronger the degree of preference.

1 = 'strict preference': The better performing option is considered 'strictly' or 'absolutely' preferred over another on a particular criterion, and is allocated the maximum possible value, represented by a '1', in the background calculations of the model.

Note: These 'preference relationship' values are intermediate outputs in the PROMETHEE models. Usually they will not be displayed and need not be viewed nor understood in order to participate in a PROMETHEE analysis.

Which of these categories a pairwise comparison will fall under is dictated by the indifference and preference thresholds; additional parameters to be elicited from users. These thresholds are a unique feature of outranking models.

Indifference threshold: This figure indicates the minimum difference required between two options on a criterion in order for the better performing option to be regarded as preferred to some degree over the other. Any difference in performances below this threshold and the model will treat the two options as having equal performance on that particular criterion (i.e. an indifference relationship). For example, if the difference in cost between two options is $100,000 and the indifference threshold is $200,000, the model will treat both options as having equal performance on the 'cost' criterion, despite this (estimated) cost differential.

Preference threshold: Represents the required difference between two interventions in order for the better performing intervention to be regarded as 'strictly preferred', and thus receive the maximum degree of preference possible on a particular criterion. For example, on the 'Cost' criterion a preference threshold value of $100,000 indicates that an intervention must cost at least $100,000 less than another for it to be strictly preferred. Differences smaller than the preference threshold value but greater than the indifference threshold will still give a degree of preference ('some preference' - see above).

Mathematically, the difference between two options a and b on criterion k can be represented as $d_k(a,b) = x_{ab,k}$. The degree of preference of option a over b on criterion k is calculated according to the following function:

$$Pr_k(a,b) = \begin{cases} 0, & if \;\; x_{ab,k} \leq q_k \\ \dfrac{x_{ab,k} - q_k}{p_k - q_k}, & if \;\; q_k < x_{ab,k} \leq p_k \\ 1, & if \;\; x_{ab,k} > p_k \end{cases}$$

Where: q_k and p_k are the indifference and preference thresholds respectively.

The interpretation of this function is as follows: when the difference between two interventions is less than the indifference threshold it is considered as negligible and the corresponding preference degree is set to zero. If the difference exceeds the preference threshold it is considered large and the preference degree is set to one (‘strictly preferred’ - the maximum value). When the difference is between the two thresholds an intermediate value is computed for the preference degree using a linear interpolation.

The positive flow, $\phi^{+}(a)$, and negative flow, $\phi^{-}(a)$, of option a is calculated as:

$$\phi^{+}(a) = \sum_{b \in A} \sum_{k=1}^{c} Pr_k(a,b) \cdot w_k$$

$$\phi^{-}(a) = \sum_{b \in A} \sum_{k=1}^{c} Pr_k(b,a) \cdot w_k$$

Where:

$A = \{a_1 \dots a_n\}$, the set of n alternatives to be considered

$K = \{k_1 \dots k_c\}$, the set of c criteria used to evaluate the options

$Pr_k(a,b)$ is the degree of preference of alternative a over b on criterion k

w_k is the weight associated with criterion k

Positive flow is calculated as the sum of all preference degrees in favour of option a, while negative flow is the sum of all preference degrees against. Positive flow quantifies the degree to which action a is globally preferred over all other options, and negative flow quantifies the degree to which all other options are preferred over a.

Using PROMETHEE I an option is said to outrank another if its positive flow is higher and its negative flow lower than another. Mathematically, a outranks b when:

$$\phi^{+}(a) > \phi^{+}(b)$$

and

$$\phi^{-}(a) < \phi^{-}(b)$$

PROMETHEE II combines the two flow numbers into a net flow:

$$\phi(a) = \phi^{+}(a) + \phi^{-}(a)$$

This allows a 'complete' ranking of options without the possibility of two options being classified as incomparable. Under PROMETHEE II option a outranks b when $\phi(a) > \phi(b)$. The PROMETHEE I and II models constructed in this research are able to accept quantitative or qualitative data inputs for any criterion. Qualitative data can be entered as a numerical score in a similar way to the Weighted Sum Model (see Section 8.5.1.1). Raw quantitative data (e.g. cost and effectiveness figures) can be entered into the model with zero transformation required (i.e. no requirement to map data onto homogeneous scales).

This PROMETHEE analysis used criteria scores and weights identical to those used in the WSM analysis, generated using 'direct rating' (see Section 4.6.4). Initial indifference thresholds were set at $1m per year for cost, 100 DALYs per year for effectiveness and 15 'points' for the remaining criteria. Preference thresholds were set at $2m per year for cost, 200 DALYs per year for effectiveness and 35 points for the remaining criteria. The input data for the PROMETHEE model are shown in Table 24.

Key
Quantitative data
Qualitative scores/weights

	Cost (NZ$m)	Effectiveness	Market Access	Consumer Perceptions	Quality or Suitability	Ease of Implementation	Quality of Science	Equity of costs
Multiple Interventions	1	854	70	35	95	90	100	50
Chemical Treatment	1.5	854	70	35	95	90	100	50
Bacteriophage	2.97	993	70	35	50	80	50	50
Freezing	8.4	947	50	50	20	20	100	10
Single Use Crates	0.75	82	50	80	50	80	40	50
Hygiene Education	0.16	8	50	100	50	100	80	50
All-in-all-out	8.9	331	50	80	50	80	40	50
Increase Biosecurity	4.5	165	60	80	50	90	70	50
Irradiation	43.9	1010	0	10	95	0	100	50
Indifference threshold	1	100	15	15	15	15	15	15
Preference threshold	2	200	35	35	35	35	35	35
Criterion weights	19.0%	17.1%	5.7%	11.4%	11.4%	13.3%	17.1%	4.8%

Table 24: PROMETHEE MCDA inputs for the Campylobacter case study

8.5.3.2 Results

Positive flow is calculated as the sum of all pairwise comparisons in favour of an intervention while negative flow is the sum of all pairwise comparisons against. Net flow is the sum of these two flow numbers. The three flow figures for each intervention in the PROMETHEE I and II analysis are presented in Table 25 and Figure 27.

	Positive Flow	Negative Flow	Net Flow	PROMETHEE II Rank
Multiple Interventions	19.5	3.92	**15.58**	1
Chemical Treatment	19.84	8.47	**11.37**	2
Irradiation	23.12	21.25	**1.87**	3
All-in-all-out	17.16	15.75	**1.41**	4
Increase Biosecurity	15.78	14.91	**0.87**	5
Bacteriophage	11.97	14.83	**-2.86**	6
Hygiene Education	6.6	14.09	**-7.49**	7
Freezing	16.25	23.9	**-7.65**	8
Single Use Crates	9.5	20.05	**-10.6**	9

Table 25: Calculated flows for PROMETHEE analysis of Campylobacter case study

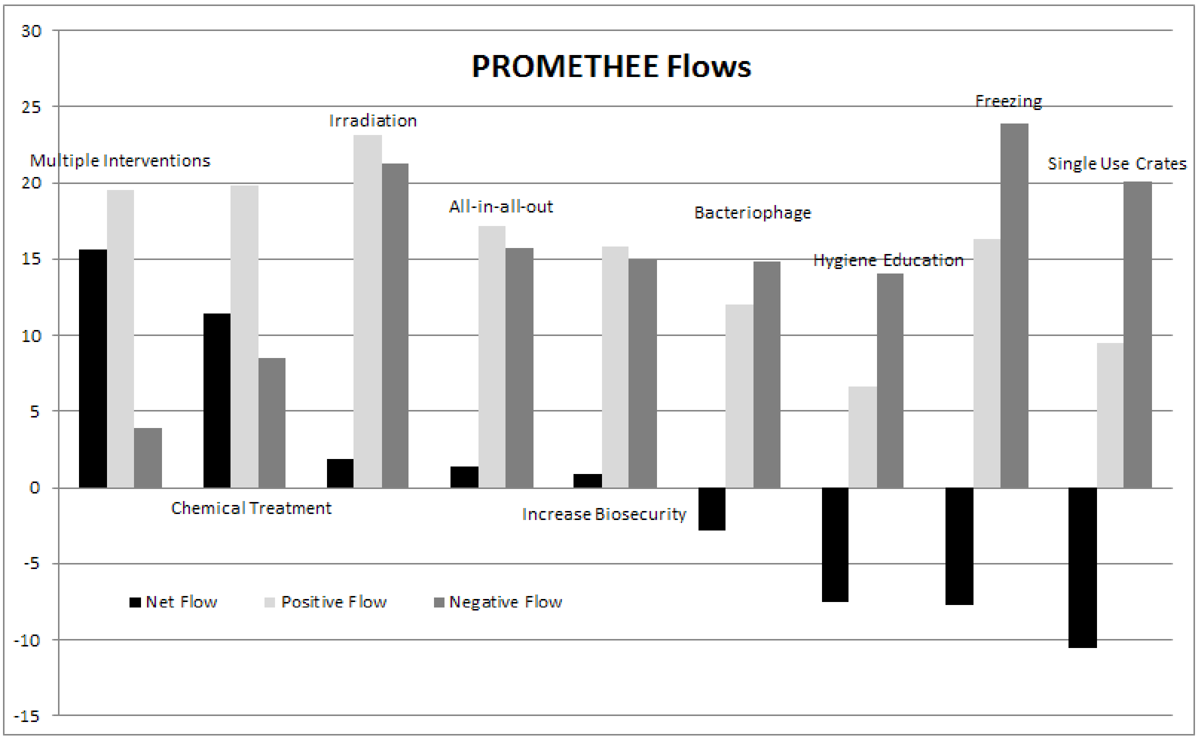


Figure 27: Intervention flows for PROMETHEE analysis of Campylobacter problem

Under the PROMETHEE I methodology an intervention can only be said to outrank another if it has both a higher positive flow and lower negative flow. If this type of outranking relationship cannot be established between two options then they are said to be incomparable. In this analysis 'Irradiation' outranks 'Freezing' but is incomparable when compared to any other intervention. This is the result of 'Irradiation' performing very well on some criteria such as 'Effectiveness', 'Quality of science' and 'Quality or Suitability', but very poorly on other criteria such as 'Cost', 'Market Access', 'Consumer Perceptions' and 'Ease of Implementation'. There are also several other incomparable intervention pairings in these results.

PROMETHEE II removes the possibility of incomparabilities by aggregating both positive and negative flows into a 'net flow'. Utilising a PROMETHEE II methodology, 'Multiple Interventions' and 'Chemical 'Treatment' outrank all other interventions. An interesting result is that 'Irradiation' is third ranked in this analysis. This is primarilly because it is less heavilly 'penalised' for its extremely poor performance on the 'Cost' criterion under PROMETHEE II when compared with the WSM. In this analysis all pairwise comparisons where the performance differential is greater than the preference threshold ($2m for 'Cost') are treated equally on the 'Cost' criterion. Whether the difference is equal to the preference threshold or many orders of magnitude greater is irrelevant, as all pariwse comparisons with a cost differential equal to or larger than the threshold will result in the maximum degree of preference for the better performing option. The cost of 'Irradiation' could decrease by over $30m or increase infinitely and, provided the preference threshold remained static, the model scores would remain static too. This could be viewed as throwing information away, but may not matter if participants uninimously support the preference threshold.

8.5.3.3 Critique

The PROMETHEE MCDA approach is a comprehensive and promising method which has been previously applied to food safety problems (Fazil et al., 2008; Henson et al., 2007; Ruzante et al., 2010).

Partial or Complete ranking

In the PROMETHEE I analysis 'Irradiation' is deemed incomparable with all other interventions except 'Freezing' (Figure 27). Irradiation's polarised performances on most criteria cause its high positive and negative flow values, rendering it incomparable with many interventions with more centralised performances (and thus lower positive and negative flow values). PROMETHEE I has difficulty ranking such options based on these flows and thus classifies them as incomparable. This does not necessarily mean that the two options cannot be compared, but rather the algorithm is suggesting that the two options might be so different, in terms of their performances on the different criteria, that it is difficult to select a clear 'better' option based on the information available. An advantage of this partial ranking approach is that it avoids the potentially premature selection of the 'best' option(s). Incomparabilities prompt the participants to further scrutinise the two options in question and investigate why they are being labelled as incomparable. This can facilitate a greater understanding of the characteristics of the different options and the preferences of participants. As a result the participants might, for example, reach the conclusion that the two incomparable options would make excellent complements, and this process could lead to the (re)formulation of new (existing) options that retain their strengths, but with fewer weaknesses.

In contrast, the complete ranking approach of PROMETHEE II does not identify incomparabilities and may lead to a less comprehensive understanding of the options under study, and perhaps a missed opportunity to combine options or design new ones. The concept of incomparability resulting from ambiguous positive and negative flow figures may be unappealing for some decision-makers who may prefer a simple and complete ranking of options. PROMETHEE II forces a complete ranking even when there are options which could be seen as incomparable. This could be viewed as an advantage to those who wish for simple and unambiguous model outputs, or a disadvantage to those who feel that 'forcing' a ranking is a missed opportunity to understand and discuss incomparabilities and potentially improve decision-making through doing so.

Natural Scales

Outranking algorithms like PROMETHEE have the advantage of capturing a more natural structure of a problem by not requiring the remapping of quantitative data onto homogeneous scales. Any data can be expressed numerically or verbally on its most natural scale (although the model will need to assign numerical values to verbal statements). Because of this, outranking approaches seem particularly suitable for problems for which there is an abundance of quantitative data available. In the *Campylobacter* case study there are 'Cost' and 'Effectiveness' data available for each intervention. PROMETHEE can accept this data without requiring any transformations. Expressing data in its 'natural' units can help participants with conceptualising the problem and the trade-offs they are willing to make. This also avoids the 'bunching' problem that can occur with homogeneous scales (see Section 8.5.1.3).

Adding new options

Because outranking methods readily accept data expressed in natural units, without requiring remapping onto homogeneous scales, it is easy to add new options to an analysis, even if these options assume extreme values on one or more criteria. Unlike the other models studied in this chapter, PROMETHEE criteria scores are completely independent of one another and adding or removing options has zero effect on other criteria scores.

Rank Reversal

Like the AHP, PROMETHEE and other similar outranking MCDA models (such as ELECTRE) have the potential to exhibit the rank reversal phenomenon (Macharis et al., 2004; Mareschal et al., 2008) (see Section 8.5.2.3 – rank reversal).

Thresholds

The idea of defining indifference and preference thresholds is unique to outranking methods. In contrast to the fully-compensatory Weighted Sum Model and AHP, outranking models like PROMETHEE are only partially-compensatory, meaning that very poor performance on one criterion cannot always fully compensate for excellent performance on another. It could be argued that this property adds a more realistic dimension to an analysis by capturing some of the political realities of food safety decision-making. This does indeed seem like a realistic feature of MCDA, as it is unlikely that even the most outstanding performance on any criterion could completely compensate for an intervention causing, for example, a sufficiently negative consumer backlash.

When considering outranking models there is the issue of assessing whether the use of thresholds is appropriate for food safety decision-making. Are there any criteria for which decision-makers will be completely indifferent between (small) differences in performances? What is the difference between 'some preference' and 'strict preference'? How do these preferences affect the model outcomes? Training participants in PROMETHEE methodologies and eliciting answers to such questions could be more difficult than for other MCDA models. The concept of thresholds and the mathematics in the model are more complex than other methods, and this may lead to a loss of transparency and understanding of the model. Additionally, participants also must decide whether to utilise one or both types of threshold, and must decide what values to set each at. Again, participants may have no clear idea of what to set these values at, or there may be disagreement regarding their values; although highlighting such disagreements can help with understanding the problem situation; a key intention of MCDA methodologies.

Indifference thresholds may also be useful for considering options with criteria performances that lie within another's range of uncertainty. For example, in the *Campylobacter* case study, if the cost estimates for the various interventions had a confidence interval of \$0.5m in either direction then an indifference threshold could be set at \$0.5m and this would result in any interventions with cost estimates within \$0.5m of one another being treated as equal on the 'Cost' criterion. This use of thresholds could be seen as a type of (in)sensitivity analysis, as variations to criterion scores may make no difference to the rankings if, for example, differences between criterion scores remain lower than the indifference threshold or greater than the preference threshold.

Although the use of thresholds could help add an extra dimension of realism to the modelling process, it may also add an extra dimension of uncertainty or disagreement.

8.6 Conclusion

This chapter has tested and critiqued some of the most widely applied MCDA models. The primary goal of this chapter was to conduct a preliminary assessment of each method for application to food safety intervention decision-making. All methods have their own advantages and disadvantages. The Weighted Sum Model is the simplest model to implement and use, holding an advantage over the other models in terms of its transparency. AHP constructs a linear additive model where recommendations are based on the sum of weighted criterion scores. However, the unique AHP approach to generating weights and scores is controversial. On the one hand the hierarchical conceptualisation of decision-making situations, coupled with its pairwise comparison approach, offer a systematic and structured methodology to aid decision-making. On the other hand there are issues with the subjectivity and validity of the scales used for pairwise comparisons and the requirement of 'homogeneous' data may restrict AHP's application in the food safety arena. AHP seems especially appealing for situations with limited options and criteria, and where there may be a lack of available quantitative data. Outranking algorithms such as PROMETHEE can help preserve a more natural problem structure by allowing the expression of criteria performance in natural units, and their partially-compensatory nature may be able to more realistically approximate human decision-making, especially in political contexts like high-level food safety decision-making (de Montis et al., 2005). However, outranking algorithms are also mathematically more complex than others and this reduces their transparency. Although they can add an extra dimension of realism to the analysis, the question of whether the different thresholds are relevant in different decision-making contexts, and what to set these thresholds at, places additional demands on the participants and adds additional layers of uncertainty to the analysis. The choice of model to utilise in decision-making is likely to depend on the objectives and preferences of the decision-maker(s), as well as the individual characteristics of the decision-making situation itself.

Chapter 9: Summary, Discussion and Conclusions

A range of chemical, microbial and physical hazards may occur in New Zealand food. Microbial hazards alone are estimated to cause over 2000 years of lost healthy life (Cressey, 2012) and $62m in medical costs and lost productivity annually (Gadiel & Abelson, 2010). Global food safety authorities are seeking improvements in the level of consumer protection. Such improvements will rely on technological innovation in the food production and service sectors coupled with a policy of evidence-based decision-making. Determining whether a specific technology is appropriate for a particular context will also need to follow a systematic and structured approach. This thesis proposes that MCDA is such an approach.

This thesis examined the nature of the foodborne hazards present in New Zealand. Chemical hazards are ubiquitous but are well managed through controls such as Maximum Residue Limits (MRLs) and Good Agricultural Practice (GAP). Microbial hazards are also ubiquitous but are considered harder to control than chemical hazards, primarily due to their ability to reproduce along the farm-to-fork chain. Addressing microbial hazards probably offers the greatest opportunity to reduce the current burden of foodborne disease.

A detailed review of the existing New Zealand food-safety regulatory environment was conducted as part of this research. This involved reviewing a wide range of documents, produced mostly by MPI and other regulatory organisations such as Codex, and supplemented by interviews with key New Zealand stakeholders, including MPI experts. MPI has two primary goals: firstly, to reduce the disease burden associated with foodborne hazards, and secondly, to facilitate improved market access for food businesses. MPI's different objectives mean that the evaluation criteria and decision-making process itself can vary with the type of decision being made. Interviewees generally agreed that New Zealand is at the forefront of food-safety decision-making, with a robust regulatory framework in place for managing foodborne hazards. MPI consider a number of criteria when evaluating food safety risk-management options, although how performance on these criteria is assessed, and how exactly information is utilised in decision-making, is sometimes uncertain.

This research has explored the New Zealand food safety intervention decision-making environment in detail, and investigated the potential for MCDA to contribute to current decision-making practices. In doing so, this thesis has made some key contributions to this field:

- By developing a system for the identification and classification of stakeholders into key stakeholder groups.
- By identifying a comprehensive list of criteria which are considered relevant to food-safety decision-making by key stakeholders, and investigating and identifying methods for their quantification. This was achieved through interviews with stakeholders, which provided direct information from key players involved in food-safety decision-making.
- By detailing the construction and testing of three very popular, although different, MCDA models, and identifying key strengths and weaknesses of each when applied to a realistic food-safety problem.

9.1 Stakeholder Classification

This thesis used qualitative research methods to explore what it meant to be a stakeholder with respect to food-safety decision-making. There are a number of stakeholders involved in food-safety decision-making, either through being in a position where they are able to directly or indirectly influence decision-making (e.g. MPI), or being in a position where they are affected by decision-outcomes (e.g. consumers and food businesses). A list of stakeholder 'types', classified according to their role in food-production and safety, was identified through interviews with New Zealand stakeholders. These stakeholder groups were identified as:

- Regulators
- Public health authorities
- Food safety scientists/academics
- Consumers
- Māori

- Food Businesses (further classified as):
    - Farmers
    - Processors
    - Food retailers
    - Exporters

9.2 Decision-Making Criteria

This research involved interviews with a wide range of stakeholders regarding the criteria relevant to food-safety intervention decision-making. This process generated a list of twelve criteria:

- Cost
- Effectiveness
- Market Access
- Consumer Perceptions
- Ease of Implementation
- Quality or Suitability
- Quality of Science
- Equity of Costs
- Equity of Benefits
- Safety
- Animal Welfare
- Cultural Impacts

9.3 MCDA

MCDA is receiving increasing attention as a suitable methodology for systematically evaluating competing options where there are multiple conflicting criteria and a number of diverse stakeholders. All MCDA should follow a similar process (outlined in Section 4.5). However, there are different ways of conducting some MCDA steps. In particular, there are a large number of models available for the aggregation phase; the process of converting model inputs, in the form of criteria scores and weights, into model recommendations. This thesis has reviewed and outlined the main classes of model, as well as the literature describing applications of MCDA to food-safety decision-making. This helped guide the construction and testing of specific models in this thesis.

This research constructed and tested three types of MCDA model: the Weighted Sum Model (WSM), Analytic Hierarchy Process (AHP) and PROMETHEE. These models were selected due to their having different characteristics and degrees of complexity, as well as being popular methods in the food-safety and HTA literature. The results of this testing are a key output of this thesis and can provide insight into how such models might be constructed and exploited in food-safety decision-making. The conclusions reached throughout this research phase can be classified into one of two broad groups:

- Those relating to MCDA as an entire process/methodology for decision-making
- Those relating to the specific models and mathematical procedures for generating numerical inputs and outputs

9.4 Lessons Learnt

This thesis has examined the potential for the use of MCDA in food-safety intervention decision-making. The key lessons learned throughout this research are summarised in the following sections. These lessons could serve as practical advice or recommendations for food-safety regulatory agencies, or any other groups or individuals interested in applying MCDA to food safety intervention decision-making problems.

9.4.1 MCDA is well suited to problems where there are conflicts between objectives and/or stakeholders

The *Campylobacter* case study involved the evaluation of technologies available for addressing the *Campylobacter* in poultry problem of the mid 2000s. This type of problem is faced by many food industries/businesses, who must evaluate and select interventions within MPI's regulatory framework.

This thesis identified a range of different stakeholders relevant to food-safety decision-making. Most decision-making will involve multiple stakeholder groups, opening up the potential for disagreement regarding aspects of decision-making, including the selection of evaluation criteria and their relative importance. Interviews with New Zealand stakeholders confirmed that such disagreements do occur. In particular, the primary objectives of regulators and food-businesses may conflict, with the primary focus of regulators being the health of consumers, and the primary goal of food-businesses being the generation of profit.

Conflicts during decision-making may also occur between MPI's own objectives. Primary objectives of MPI are (NZFSA, 2010b):

- *"The improved safety and suitability of food"*
- *"Effective government role in facilitating commerce and market access"*

These two objectives could be seen as being in conflict with one another. MPI must balance making food safer and making it easier for food-businesses to produce, export and sell their products.

MCDA appears to be especially suited to food-safety intervention decision-making because of its ability to cope with multiple stakeholders and conflicting objectives in an explicit and structured manner. This structure forces users to be clear about the trade-offs they are willing to accept, and provides a useful tool for understanding the preferences of others. MCDA aims to facilitate agreement between stakeholders through a deliberative process. Ideally, stakeholders agree on model inputs, including criteria scores and weights, before results are generated. This is not always possible however, and there may be differences which cannot be reconciled during an analysis. This thesis has reviewed a number of procedures available for coping with such situations in MCDA (see Section 4.8). Furthermore, sensitivity analysis can help identify situations where such differences will have no effect on the model results (e.g. rankings), and this can help direct further work towards reconciling those differences that do have the potential to affect the results. The Monte-Carlo simulation of model inputs, conducted as part of the *Campylobacter* case study, illustrated that MCDA results can be insensitive to large variations in model inputs (see Chapter 8.5.1.2).

9.4.2 Food safety decisions are true Multi-Criteria problems, and there are multiple ways to assess performance on each criterion

This thesis has demonstrated that food-safety intervention decision-making is a true multi-criteria problem. Twelve criteria were identified as being potentially relevant to stakeholders during this type of decision-making. Not only has this thesis showed that there are a range of criteria to consider during decision-making, it has also demonstrated that there are numerous ways to quantitatively or qualitatively assess the performance of interventions on different criteria. Decision-makers embarking on MCDA need not only to consider the range of criteria to include in their analysis, but also the most appropriate metrics for assessing performance on these criteria. The appropriate choice of measurement will depend on who is conducting the analysis as well as other constraints such as time, money and data availability. From a processor perspective, the most appropriate way to measure the performance of an intervention designed to reduce bacterial contamination of food might be the percentage reduction in bacterial numbers; but from a public health perspective the most appropriate measure might be the reduction in incidence or DALYs. Farmers may measure the performance of similar interventions in terms of their ability to reduce in the in-flock prevalence of the same bacteria. Stakeholders identified a desire to employ the metric most relevant to their own performance targets and goals (e.g. bacterial thresholds).

Where there is the potential to convert between metrics it may be possible to present such performances using multiple metrics so that each stakeholder can view the metrics most relevant to them. The *Campylobacter* cost-effectiveness paper completed during this thesis converted bacterial and prevalence reductions into DALY reductions because we were most interested in the degree to which an intervention could reduce the burden of foodborne illness. In theory, these ‘equivalent’ measurements (DALY and bacterial reductions) could be incorporated and presented within the same analysis to increase the relevance to different stakeholders.

Decision-makers need also to consider constraints on an analysis when deciding how to appropriately ‘measure’ performance. Many criteria require assessment using value judgements. In a real MCDA these assessments would most likely take the form of expert judgements on, for example, a Likert-type or 0-100 scale. For many criteria there may be significant time and financial costs associated with generating quantitative performance data, and for some interventions it may be extremely difficult to generate any meaningful and reliable quantitative data on these criteria. These are the criteria which can often be excluded from other methodologies because they cannot be easily or reliably quantified or monetised. Using ‘expert’ judgements is an alternative that stakeholders may wish to consider when quantitative data is unavailable or prohibitively resource intensive to generate. Although value judgements may not be the best method for generating criteria scores, incorporating such criteria in an analysis forces participants to explicitly consider the performances of interventions on these criteria, as well as the relative importance of these criteria. Excluding these criteria altogether eliminates the possibility of discussion and information sharing regarding these criteria.

9.4.3 The MCDA model selection phase should consider the goals and constraints of an analysis

The question of which MCDA model to apply and how comprehensive an analysis needs to be depends on the goals and constraints of the users/decision-makers. If the goal of an analysis is to select the ‘best’ alternative to address a food safety issue then a full and comprehensive MCDA analysis will ideally be conducted, including analysing and selecting a suitable MCDA model. Generally, such an analysis should be tailored to the requirements of decision-makers and the specific problem being addressed, with careful attention paid to each of the eight steps described in Chapter 4.5. Interviews with New Zealand stakeholders found that, for many stakeholders, especially large food businesses and regulators, there is a need for a systematic and comprehensive framework for evaluating alternative intervention schemes. However, some stakeholders also expressed a desire for a system for collating and sharing information regarding alternatives; for learning about stakeholders and how they structure their preferences in different decision-contexts. Such a system may not necessarily be used exclusively as a basis for decision-making, but as a process for collating intervention characteristics and eliciting and storing stakeholder preferences in an explicit way; allowing decision-makers to examine this data over time and better understand the other stakeholders involved with decision-making.

Different MCDA models may have implications regarding the transparency of an analysis and the ability to communicate ideas and results to stakeholders. Some MCDA models operate as ‘black box’ models, where sophisticated behind-the-scenes mathematics generates outputs in a way which could be difficult for non-specialists to understand. In contrast, there are also relatively straightforward models which use much simpler algorithms to produce outputs. Decision-makers can weigh-up the possible additional benefits of more complex models against the greater transparency and usability offered by simpler models. If the primary goal of an analysis is to identify the ‘best’ option then there is merit in investigating whether more sophisticated models and elicitation techniques may produce the most reliable outputs, especially in contexts where the consequences of a poor decision are significant. However, if the primary goal of an analysis is to explore, understand and communicate stakeholder views then it would be worth considering a model with a simple and easy to understand algorithm and elicitation process. MCDA is not only a useful tool for decision-making, but also for the communication of ideas.

Utilising complex MCDA models and elicitation protocols does not necessarily mean that widespread distrust and low stakeholder buy-in will ensue. Although such models can indeed reduce the transparency of results, it may be possible to compensate for this through adequate training of participants. Furthermore, it may not be necessary for all participants to understand the specific mathematics of the MCDA algorithms and elicitation procedures. It may be sufficient to provide participants with a laypersons overview of the underlying logic of these procedures.

This research found that stakeholders in food-safety decision-making often face significant time constraints when it comes to decision-making. Efficient use of time was mentioned as important by many interviewees in this research. Different MCDA models have different requirements regarding the data inputs they can accept, and thus the time required to complete an analysis. Linear Additive Models like the Weighted Sum Model require inputs expressed on homogeneous scales; AHP requires the pairwise comparison of all alternatives on every criterion; PROMETHEE methods often utilise indifference and preference thresholds. There are two main considerations when assessing the time required to implement and use these models. Firstly, there is the time required to train individuals in their use. Different models are likely to require different degrees of training. Furthermore, some stakeholders such as MPI experts or members of the decision-making teams of large food businesses may more easily grasp the concepts behind some MCDA models than other stakeholders. Secondly, there is the time required to generate or elicit data inputs for a model. Some models such as the Weighted Sum Model, and especially PROMETHEE, are naturally able to accommodate quantitative data, and there is, of course, a time cost associated with generating such data. In terms of eliciting value judgements from experts or other stakeholders, AHP is the most demanding of the models studied in this thesis. The requirement of $½ \cdot n_{alternatives} \cdot (n_{alternatives} - 1)$ reasonably consistent[23] pairwise comparisons for each criterion means that, for analyses with many criteria, AHP is likely to take much more time to elicit preferences than the other two methods. The AHP *Campylobacter* case study with 9 alternatives and 8 criteria required the elicitation of over 300 pairwise comparisons. However, some decision-makers may find the process of only considering two alternatives at a time appealing, and may see the extra time as a cost worth paying. Furthermore, specialised AHP software seems to be of particular benefit for guiding this process, especially for identifying and resolving inconsistencies. This would be an area worth investigating for those considering the AHP model.

In order to ensure that the requirements of an analysis are met, it is also important that the model selection phase considers model characteristics such as the types of input data they can accept, the outputs they produce and their underlying logic, Analysts need to consider the type of data they are working with. Whether they are working with quantitative, qualitative or mixed data inputs should affect which model they select. It may be unwise to select, for example, the AHP model if working with pure quantitative data inputs. Similarly, analysts may prefer their model to produce a complete ranking of interventions rather than a partial ranking (i.e. with ‘incomparabilities’), shortlist, or single ‘best’ option. Another characteristic that analysts may wish to consider is the degree of compensation between criteria. Some models, such as the WSM, will allow for full compensation between criteria. That is,

[23] As judged by the Consistency Ratio (see Chapter 4.6.3)

extremely good performance on one criterion can compensate for extremely poor performance on another. Other models will allow only partial compensation between criteria, or even no compensation at all. Section 4.6 and Chapter 8 detail many of the characteristics of the different models studied in this thesis. For those looking to utilise MCDA it would be prudent to consider whether the properties and underlying logic of the models they are considering are compatible with the requirements of their analysis and the preferences of decision-makers.

9.4.4 Understanding how a model has produced its results is important

An important part of the MCDA process is understanding why and how a model has generated its outputs. This is especially true for results that would appear to conflict with initial predictions. An example of such a result may be seen in the *Campylobacter* case study, where 'Hygiene education' scores reasonably well in the Weighted Sum Model (WSM) despite its extremely poor performance on the 'Effectiveness' criterion. Inspection of the WSM performance matrix (Figure 28) reveals that 'Hygiene Education' also scores relatively well on 'Quality of Science'. That is, there is strong evidence that 'Hygiene Education' has only a minimal effect on the *Campylobacter* disease burden. Poor performance on one criterion is 'compensated' by a high score on another, resulting in a higher ranking than might have been expected. This is the fully compensatory nature of the WSM methods[24].

[24] AHP is also fully compensatory

WSM Performance matrix for Campylobacter problem

	Cost	Effectiveness	Market Access	Consumer Perceptions	Quality or Suitability	Ease of implementation	Quality of Science	Equity of costs	Global Scores	Rank
Multiple Interventions	18.7	14.5	5.7	3.2	11.4	12.0	17.1	4.8	**87.4**	1
Chemical Treatment	18.5	14.5	5.7	3.2	11.4	12.0	17.1	4.8	**87.2**	2
Hygiene Education	19.0	0.0	4.1	11.4	4.6	13.3	11.4	4.8	**68.7**	3
Bacteriophage	17.8	16.9	5.7	3.2	4.6	10.7	2.9	4.8	**66.4**	4
Increase Biosecurity	17.2	2.7	4.9	8.9	4.6	12.0	8.6	4.8	**63.5**	5
Freezing	15.5	16.1	4.1	5.1	0.0	2.7	17.1	0.0	**60.5**	6
All-in-all-out	15.2	5.5	4.1	8.9	4.6	10.7	0.0	4.8	**53.7**	7
Single Use Crates	18.8	1.3	4.1	8.9	4.6	10.7	0.0	4.8	**53.0**	8
Irradiation	0.0	17.1	0.0	0.0	11.4	0.0	17.1	4.8	**50.5**	9
Criteria Weights	19%	17%	6%	11%	11%	13%	17%	5%		

Figure 28: WSM performance matrix for the Campylobacter case study

Sometimes it is difficult to predict results such as those described above, especially during the first iteration of an analysis. It is important to thoroughly inspect and critique the results of MCDA in order to understand how they have been produced. In the example above, some decision-makers may find the global score of 'Hygiene education' perfectly acceptable, while others may refuse to accept such a result due to 'Hygiene Education's' minimal effectiveness. Such a result should spark debate amongst participants regarding the contribution 'Effectiveness' and 'Quality of Science' should make to decision-making; essentially forcing participants to focus on the relative importance of all criteria. As a result, participants may wish to re-evaluate the weights allocated to some criteria. Provided it is conducted in an honest manner, the adjustment of model inputs is in line with the intended iterative nature of MCDA. This process has the potential to be abused however, and care should be taken to ensure that numbers are not being manipulated to produce a predetermined outcome.

9.4.5 Model configuration and elicitation protocols can significantly affect results

Decision-makers should be aware that, just as results are sensitive to the type of MCDA model used, results can also be highly sensitive to how these models are configured, and how inputs are gathered. In particular, this includes the type of scales used to score alternatives and the procedures for generating criteria weights. For example, those utilising homogeneous scales (e.g. global or local scales - see Section 6.4), where alternatives are scored relative to each other, should remain cognizant that alternatives with ‘extreme’ scores (i.e. much larger or smaller than the others) on one or more criteria may be unacceptably distorting the scores of the remaining alternatives (i.e. ‘bunching’ - see 8.5.1.3). In the Weighted Sum Model case study there is a significant distortive effect as ‘Irradiation’s’ very high cost inflates the ‘Cost’ scores of the remaining options, causing them to bunch at the top of the scale because they are all being scored relative to ‘Irradiation’ (see Figure 29). Distortion will be most prominent where there are ‘extreme’ scores on criteria with high weightings.

Remapped cost data onto homogeneous (local) scales

With Irradiation

	Cost Score
Hygiene Education	100.0
Single Use Crates	98.7
Multiple Interventions	98.1
Chemical Treatment	96.9
Bacteriophage	93.6
Increase Biosecurity	90.1
Freezing	81.2
All-in-all-out	80.0
Irradiation	0.0

Irradiation Removed

	Cost Score
Hygiene Education	100.0
Single Use Crates	93.2
Multiple Interventions	90.4
Chemical Treatment	84.7
Bacteriophage	67.8
Increase Biosecurity	50.3
Freezing	5.7
All-in-all-out	0.0

Figure 29: Example of distortion that can occur with homogeneous scales

Distortion can occur when quantitative data is remapped onto homogeneous scales. Irradiation’s very high cost inflates the other scores when remapped onto a ‘local scale’ (See Section 8.5.1.3). When Irradiation is removed from the analysis the distortion is reduced and the scores better reflect the true differences in performance between the remaining interventions.

If decision-makers decide, after the first iteration of an analysis, that alternatives with extreme scores on one or more criteria are not acceptable solutions to an issue, then it may be appropriate to rerun the analysis with these options excluded. As demonstrated in Section 8.5.1.3, this will minimise distortion and will allow the scales to better reflect the differences in performance of the remaining options. MCDA models which do not require the scoring of performances on homogeneous scales avoid this problem, although such models generally utilise more sophisticated aggregation algorithms to cope with the heterogeneous nature of criteria scales. Prior to selecting an MCDA model users should think about the degree of distortion they are likely to face through the use of homogenous scales, and whether this distortion is a price worth paying for the simplicity and usability associated with these scales and the MCDA models that rely on them.

9.4.6 Understanding the weighting procedure(s) is important

The type of method used for generating criteria weights must be well understood by decision-makers before proceeding with their use. Some of these methods, such as SWING weighting, rely on decision-makers knowing the performance scores of all interventions on all criteria before weighting can begin. SWING weights reflect the value of moving from the existing worst to best score on a criterion, and will generally allocate a higher weight to criteria where there is a large spread of values when compared with conventional direct rating methods. Because SWING weights are specific to a particular analysis they cannot be used in subsequent analyses without repeating the elicitation process. SWING weighting, although novel and arguably a superior weighting procedure for individual analyses, would be an inappropriate process for generating a 'standard' set of criteria weights for use in multiple MCDA studies (e.g. if regulators or large businesses wished to develop a standard set of criteria weights for use in future MCDA). In contrast, conventional weighting techniques such as the direct rating method can be conducted without requiring any knowledge of criteria performance scores. The criteria weights generated through direct rating (and similar techniques) remain static even if performance scores change, making them potentially usable in future MCDA.

9.4.7 Partial MCDA can still offer significant benefits

Although MCDA can provide structure to decision-making, there is also much flexibility in the way it is implemented. For some stakeholders wishing to take advantage of the MCDA methodology, but lack the time, resources or necessity to conduct a full MCDA, partial MCDA can still yield significant benefits. For example, many of the benefits of MCDA can be realised without the explicit aggregation of criteria scores and weights (e.g. through direct inspection of the performance matrix). There are many benefits of completing the initial phases of MCDA; conceptualising the problem; defining the stakeholders and criteria and generating performance scores. These steps involve a significant amount of discussion and information sharing between stakeholders, and are stages of the analysis where such information sharing can yield substantial benefits. Summarising and presenting all available information in the form of a performance matrix may be sufficient for stakeholders to reach a successful decision through a deliberative process. This research has shown that the performance matrix is able to summarise the data relevant to food-safety decision-making, including the outputs of CEA, CBA or other quantitative methodologies, as well as qualitative data. Such a process may be appropriate for decision-making contexts where the complete aggregation of performance scores, through the use of an aggregation procedure, is deemed unnecessary or inappropriate. Decision-makers may not be comfortable with using one of the many aggregation procedures, or may wish to retain a level of discretion when it comes to final decision-making.

Alternatively, stakeholders may not wish (or need) to involve other 'stakeholders' in decision-making. They may themselves have conceptualised the problem, developed a set of criteria and outlined a number of potential options for consideration. Situations where decision-making is completely internal can still benefit from MCDA procedures such as the explicit quantification of criterion scores, weighting of criteria, and the aggregation and analysis of global performance scores. An example of such a situation may occur, for example, when large processors are faced with a choice of several interventions to address a particular food safety issue. They may not have the need, time or resources to engage with all stakeholders, but may wish for a more comprehensive evaluation methodology than is offered by other methodologies such as CEA or CBA. Through explicitly considering and recording the compromises that decision-makers are willing to make, this process may prove superior to less structured decision-making processes that are reported to take place at some large food businesses. By providing an audit trail of historical decisions, such a process can allow organisations to analyse their own decision-making over time, allowing for the identification of strengths and weaknesses, and ideally facilitating improved future decision-making.

9.4.8 Consistency with decision-making is likely to require a consistent MCDA approach (e.g. model + configuration)

Stakeholders consulted in this research expressed a desire for a multi-criteria decision framework which can facilitate consistent and traceable decision-making. As demonstrated in this thesis, the choice of MCDA model and the way these models are configured can significantly affect the results of an analysis. If regulators or other large organisations intend utilising a methodology such as MCDA for decision-making, it may be beneficial to develop a standard MCDA protocol, detailing the process and setting some methodological guidelines/requirements. This could be similar to the New Zealand Treasury's Cost-Benefit Analysis Primer (Treasury, 2005). This could also benefit other stakeholders looking to implement MCDA methodologies themselves, who will be faced with a potentially overwhelming number of methodological choices when investigating and utilising MCDA.

Developing a standard MCDA protocol for food-safety decision-making could provide many benefits. Most importantly, it could allow decision-making to be conducted in a consistent manner, with each piece of evidence explicitly quantified and considered according to a standard protocol, leaving behind an audit trail which can assist in communicating the rationale behind decisions to stakeholders who may or may not have been formally included in decision-making. Having such a protocol in place, could assist decision-making by streamlining or eliminating tasks such as:

- The identification and inclusion of stakeholders
- The choice of criteria to include and how these should be measured
- The generation of criterion weights
- The selection of MCDA model
- Procedures for the analysis of results

Although tasks such as the determination of the alternatives, generation of performance scores and analysis of results will need to be conducted during every analysis, having a MCDA protocol in place could guide these steps and potentially eliminate much of the work required for the remaining steps, resulting in more consistent and transparent decision-making and a reduction in the cost of, and time required for, an analysis.

9.4.9 MCDA may be a more appropriate methodology than conventional methodologies for food-safety intervention decision-making

By identifying a comprehensive list of twelve criteria relevant to decision-making, this thesis has demonstrated the multi-criteria nature of food-safety intervention decision-making. In many cases it is difficult to quantify performance on these criteria, making it difficult or inappropriate to incorporate such criteria in more conventional methodologies like Cost-Benefit and Cost-Effectiveness Analysis. Although cost and effectiveness are considered in most decision contexts, they are certainly not the only criteria to consider. Using the results of CBA or CEA (e.g. Cost per DALY) as a sole basis for decision-making can lead to sub-optimal outcomes and potential ethical issues (some of which are discussed in Section 6.3.2). As demonstrated in this thesis, stakeholders involved in food-safety decision-making consider a range of other criteria important.

Decision-makers may deviate from the results of CEA because of other important criteria. This consideration of other criteria is often an "*afterthought*" of the CEA process (Baltussen, 2013), with the CEA component being the focus of an analysis. Treating 'other' criteria in this way makes it difficult to piece together the overall performance of an option. The advantage of MCDA is that all criteria are dealt with in a consistent and systematic manner through their explicit quantification and weighting within a single framework. This systematic consideration of all relevant criteria, with explicit mathematical procedures for the quantification of criteria scores and their subsequent synthesis into an overall measure of performance offers a potential advantage over existing domestic and international frameworks (described in Sections 3.5.3 and 3.7). This process can provide a degree of structure around the more subjective elements of decision-making, allowing for more consistent and justifiable decisions.

CEA is used frequently in the field of Health Technology Assessment (HTA). Many decision-makers already consider cost-effectiveness data alongside other criteria such as equity and fairness. PHARMAC, New Zealand's Pharmaceutical Management Agency, consider nine criteria when evaluating different pharmaceutical funding schemes (PHARMAC, 2012). MPI, in their Risk Management Framework, have outlined at least eleven criteria that can be considered in food-safety risk-management decision-making (NZFSA, 2010b). The National Health Committee[25] consider up to eleven criteria in the assessment of health technologies (NHC, 2013). Internationally, many regulatory agencies also emphasise the importance of considering multiple criteria (see Section 3.5.3). In a way, these organisations are already conducting a type of unstructured MCDA, only without the explicit quantification and synthesis of criteria performances that would occur in formal MCDA. This thesis makes a contribution to existing domestic and international decision-making frameworks by suggesting ways to structure aspects of intervention analysis, such as the generation and synthesis of performance scores, so that decision-making may be conducted in a more consistent and transparent manner.

MCDA as a methodology is gaining traction as a suitable alternative to CEA and CBA. Many researchers are now promoting MCDA as a better approach to resource allocation in healthcare. Similarly, MCDA is being applied to the evaluation of regulations and specific risk management options. Key international food safety organisations such as the Codex Commission, European Food Safety Authority and Food and Agriculture Organization are also recognising the value of MCDA approaches and are developing, using and recommending MCDA methodologies for prioritisation and risk management activities. MPI is also a user of MCDA for assisting with the prioritisation of their market access work. There is possible scope for expanding MPI's use of MCDA to other decision-making contexts, such as risk prioritisation and intervention evaluation.

CEA and MCDA are often considered as competing evaluation techniques (Claxton, 2013). This need not be the case, as both techniques can play a complementary role in decision-

[25] The National Health Committee is an Independent branch of the Ministry of Health. The NHC's work includes *"assessing new and existing (non-pharmaceutical) health and disability technologies, services, models of care and programmes" and "providing advice to the Minister of Health on their value for money and prioritisation"* (NHC, 2014)

making. MCDA is mentioned in the New Zealand Treasury Cost-Benefit Analysis Primer as being a useful alternative or complement to cost-benefit Analysis (Treasury, 2005). In a similar way MCDA could also be a useful alternative or complement to CEA, for use in Health Technology Assessment (HTA) or food-safety decision-making. This thesis has demonstrated how the results of CEA can be successfully incorporated within a wider MCDA framework, alongside other quantitatively and qualitatively expressed criteria.

9.4.10 Consumers are becoming increasingly important to include in decision-making

The natural sciences have a leading role to play in the development of food-safety initiatives and in the generation of evidence supporting their effectiveness. This sentiment was echoed by stakeholders during the interview process, especially regulators, large food businesses and academics. Most modern food-safety organisations and large food businesses have systems in place for assessing interventions against scientific criteria. However, it is insufficient to base some decisions on scientific criteria alone. Social science initiatives are potentially underdeveloped and/or underutilised by many stakeholders. Consumers play a pivotal role in the success of an intervention, but incorporating the consumer voice into decision-making presents challenges for decision-makers. Questions arise regarding whom to ask, what to ask, and how to ask, in order to gather the most reliable consumer preference data. There have been some quality New Zealand social science initiatives aimed at eliciting consumer sentiments towards specific food technologies (ESR, 2008; Frewer, 2009), but interviewees agreed that consumer behaviour generally remains difficult to predict. Such research is likely to be increasingly important as new and unfamiliar food technologies emerge. Stakeholders also agreed that maintaining the trust of consumers will be important. This is especially true for regulators, who need to be seen to be putting the health of consumers at the forefront of all decision-making. Open and ongoing dialogue with consumers seems crucial to building and maintaining this trust (Baker, 2009). Many stakeholders interviewed in this research see merit in utilising a consistent, transparent and inclusive approach to decision-making. MCDA could be a suitable tool for achieving this by involving stakeholders in the decision-making process and helping to communicate the rationale behind the decisions that are being made.

Consumers base their purchasing decisions on a range of real or perceived food characteristics, including taste, quality, nutrition, price, safety and the process/technology used to make food safe. Consumers generally have a large amount of information available regarding the characteristics of the foods they eat. Stakeholders agreed that consumer perceptions are important to consider when selecting interventions, as consumers can very easily substitute their food products for a competitor's, or another food group entirely, if they are dissatisfied in any way. Most stakeholders acknowledged that little is known about how consumers respond to new technologies, and predicting consumer responses to food-safety interventions is difficult, largely because consumers do not always behave in a consistent or science-driven manner. Interviewees, especially regulators and food businesses, expressed frustration at this misalignment between the behaviour of consumers and the views of scientists, which can make intervention selection a *"messy"* problem. Nevertheless, consumer perceptions should be considered during decision-making in order to understand how consumers may respond to different technologies.

Criteria such as 'Consumer Perceptions' can be resource intensive and time consuming to elicit and quantify, with the results of studies not always considered reliable by stakeholders. This research confirmed that this is one reason why consumers are not always formally consulted during decision-making. MCDA does not require that specific or sophisticated consumer preference elicitation techniques be used to incorporate a consumer voice in decision-making. MCDA as a methodology allows the expression of criteria performance (e.g. 'Consumer perception') on a wide range of scales and metrics. In the *Campylobacter* case study conducted in Chapter 8, 'Consumer Perceptions' were scored on a 0-100 scale. Such an approach is perhaps not the best way to 'collect' consumer data for use in decision-making. However, regardless of whether consumers are included as formal participants in MCDA, the simple act of including 'Consumer Perceptions' as a criterion encourages discussion of the criterion and forces participants to explicitly consider each intervention's performance with regards to consumer perceptions.

9.5 Recommendations for Food Safety Authorities

MCDA may be an attractive methodology for the evaluation of food safety policies and interventions by food safety authorities. The major advantage MCDA offers over many conventional decision-making approaches is that it allows for the structured and transparent integration of disparate information into overall measures of performance, and can assist with identifying the most preferred option.

Based on the research conducted as part of this thesis, I would make the following recommendations to food safety authorities looking to explore the use of MCDA as a decision-support tool:

- The identification and inclusion of key stakeholders in the MCDA process is essential in order to increase stakeholder buy-in and ensure that decision-outcomes are appropriate for a range of stakeholders
- It is important to include stakeholders in the generation of evaluation criteria, so that these criteria may capture the full range of potential positive and negative effects of the various interventions.
- Ensure that your MCDA model is compatible with your requirements, assumptions and data
- Do not make an MCDA analysis more sophisticated than necessary, or use a sophisticated MCDA model without justifying its selection over simpler models
- Try a few different MCDA models out before moving towards a consistent and structured MCDA methodology.
- Treat MCDA as an entire process, with an emphasis on stakeholder participation and information sharing, rather than just a mathematical aggregation procedure
- Investigate whether specialised MCDA software may be a worthwhile investment
- MCDA should be an iterative rather than linear process, with earlier stages of an analysis being revisited as necessary
- The results of MCDA should be scrutinised and questioned thoroughly before a decision is made

9.6 Limitations and Further Research

A wide range of stakeholders from the different stakeholder groups (identified in Chapter 5) have been interviewed in this research. Some of these groups, such as food producers, retailers and exporters have been represented well in this research, especially large processors and exporters of animal products. In contrast, I would have preferred to have interviewed more individuals in the ‘Regulators’ and ‘Public Health Authority’ categories. The lack of interviewees in these categories was due to a combination of low response rates to invitations to interview, and a low number of potentially suitable interview candidates (I.e. candidates with specific food-safety knowledge). There are few public health experts with the type of knowledge I was seeking and as a result this sample was low in comparison to other groups.

This thesis initially intended to canvas Māori preferences in the same way it has for other stakeholders. However, it became clear that using the same research protocol for Māori may not be appropriate. Consultation with Māori advisors suggested that doing a sincere and thorough job of this would require a different approach, and more time and resource than was available to me. Interviews identified disagreement between interviewees regarding whether Māori should be considered as a separate stakeholder group in food-safety decision-making, or whether they are adequately represented within other stakeholder groups. The small number of Māori academics that I consulted did feel that Māori, being Tangata Whenua, have a unique stakeholder role, and this should be explicitly considered in food-safety decision-making. More research is required to establish the nature of Māori participation in food-safety decision-making

Māori researchers consulted in this research reported a resurging desire within Māoridom to incorporate traditional knowledge, known as Matauranga Māori, into contemporary life, and for western science and Matauranga Māori to be acknowledged as equally valid, and officially incorporated within New Zealand decision-making frameworks. Recent years have seen the development by Māori, of quantitative methodologies, including MCDA methodologies, which utilise traditional Māori knowledge in decision-making (Morgan, 2006). Whether such models and ideas could, or should, be adapted for use in a variety of New Zealand wide and/or Māori specific food-safety decision-making contexts requires more investigation. This sounds like an interesting area. I was personally intrigued that MCDA methodologies, utilising indigenous knowledge, are being developed and applied by Māori researchers. These methodologies have a strong holistic focus on health and sustainability, and there may be potential for wider New Zealand to learn and benefit from the ideas presented by these methodologies. Any such research would need to be conducted in a culturally appropriate manner.

This thesis has tested three of the most popular MCDA models. There are however, a much larger number of MCDA models available to decision-makers, with new methods being developed. For stakeholders looking to employ MCDA on a large scale or on an ongoing basis, it may be worthwhile testing some other MCDA models before determining the most appropriate model for their requirements. It would also be a useful activity to test the three models in this thesis under real conditions, using real data elicited from a range of stakeholders. Understanding how different stakeholders feel about the different types of model and model configurations would be helpful, especially regarding specific concepts like indifference and preference thresholds, pairwise comparisons, direct rating, global and local scales, full or partial compensation.

This thesis sought input from stakeholders regarding many aspects of the MCDA process, especially the conceptualisation of the problem, identification of stakeholders, determination of criteria and their measurement, and the relative importance of these criteria. For the *Campylobacter* case study however, the majority of stakeholders consulted did not learn about, or use the three MCDA models, nor were they used to formally elicit model inputs and analyse outputs. Time restrictions and participant availability prevented this. This would have been an interesting piece of work and would have allowed participants themselves to test out the three different types of model and provide feedback regarding the practical aspects of their use, as well their intrinsic properties and the assumptions behind their algorithms. I believe it is important for stakeholders to remain engaged throughout an analysis in order for MCDA to remain an effective decision-making process. If users are losing interest because they feel a particular MCDA model is irrelevant, time consuming or overly complex, then this may impact the quality of the inputs they are providing and the quality of the process as a whole. Field testing these models is perhaps the only way to get an idea of how stakeholders are likely to respond to these methods.

There exists also a range of MCDA software, utilising different types of elicitation and aggregation procedures. Much of this software claims to make the MCDA process easy and straightforward, even when there are a large number of geographically separated stakeholders involved. Investigating and trialling some different MCDA software[26] may be a worthwhile

[26] A list of MCDA software can be found on the MCDM Society website: http://www.mcdmsociety.org/soft.html

activity for those looking to employ MCDA methods, especially on a large scale or on an ongoing basis. Such software has the potential to overcome some of the specific model difficulties identified in this thesis, such as achieving consistency in AHP pairwise comparison matrices.

MCDA software may also be able to assist analyses through the use of charts and other visual tools. Based on limited feedback from stakeholders in this thesis, visual tools have the potential to not only assist an analysis through the effective communication of data, but also through more effectively engaging stakeholders during both the input elicitation and output analysis stages of MCDA. The *Campylobacter* case study conducted in this thesis used a range of charts to present data, including tornado plots, bar charts and box-and-whisker plots. Some of these charts were interactive, allowing for a manual and visual sensitivity analysis. I think there is potential to develop or employ even more visual tools to help with the elicitation and communication of MCDA data. Using these tools made many aspects of the case study a more engaging experience for me personally, and allowed me to better communicate MCDA concepts and results to the few experts that I presented these models to. It would be interesting to investigate whether other stakeholders felt the same way as I regarding the value of these tools, and whether such tools were able to improve decision-making in practice.

As a whole, this thesis is intended to act as a useful source of information for those contemplating the use of MCDA for the evaluation and selection of food safety interventions. It details the strengths and weaknesses of three of the most common MCDA models and describes many of the intricacies of each in the context of food safety intervention evaluation. Although this thesis does not conduct an exhaustive evaluation of every MCDA method, it does detail a process that others could adopt as a guide to assess the appropriateness of other MCDA models.

The trend towards the use of MCDA for this type of decision-making looks set to continue. New Zealand, being at the forefront of food safety research and development, is positioned well to take a leading role in the widespread adoption and further development of MCDA methodologies as part of its risk prioritisation and management frameworks. As a country with limited resources available for food safety regulation, developing and applying suitable MCDA methodologies, or a set of methodological guidelines, could reduce the time and resources required for decision-making. MCDA appears to be well suited to New Zealand's focus on transparent, inclusive and, most importantly, evidence based, food-safety decision-making.

References

Abelson, P. (2003). The Value of Life and Health for Public Policy. *Economic Record, 79*, 2-13.

Adams, C. (2013). Firms fear Pure NZ ridicule, 8 August 2013. from http://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=10909601

AHRQ. (2002). Agency for Healthcare Research and Quality: Rating the strength of scientific research findings. Retrieved 11 July, 2012, from http://www.ahrq.gov/clinic/epcsums/strenfact.htm.

Alexander, M. (2012). Decision-Making using the Analytic Hierarchy Process (AHP) and SAS/IML. Retrieved from Institute for Advanced Analytics website: http://analytics.ncsu.edu/sesug/2012/SD-04.pdf

Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality Progress, 40*(7), 64-65.

Anand, S., & Hanson, K. (1997). Disability-adjusted life years: a critical review. *Journal of Health Economics, 16*(6), 685-702.

Anders, S., & Schmidt, C. (2011). The international quest for an integrated approach to microbial food-borne risk prioritization: where do we stand? *Journal of Risk Research, 14*(2), 215-239.

Anders, S. M., & Caswell, J. A. (2009). Standards as barriers versus standards as catalysts: Assessing the impact of HACCP implementation on US seafood imports. *American journal of agricultural economics, 91*(2), 310-321.

Antle, J. M. (1999). Benefits and costs of food safety regulation. *Food Policy, 24*(6), 605-623.

Antle, J. M. (2000). No such thing as a free safe lunch: the cost of food safety regulation in the meat industry. *American Journal of Agricultural Economics, 82*(2), 310-322.

Awatere, S., Pauling, C., Hoskin, R., et al. (2008). Tū Whare Ora: Building capacity for Māori driven design in sustainable settlement development: Research Report, Landcare Research, Hamilton, New Zealand.

Baggaley, A. R., & Hull, A. L. (1983). The effect of nonlinear transformations on a Likert scale. *Evaluation & the Health Professions, 6*(4), 483-491.

Baker, M. (2005). Findings and methodological lessons from a small case control study into campylobacteriosis in Wellington. *New Zealand Medical Journal, 118*(1220).

Baker, M., Wilson, N., & Edwards, R. (2007). Campylobacter infection and chicken: an update on New Zealand's largest 'common source outbreak'. *New Zealand Medical Journal, 120*(1261).

Baker, M., Wilson, N., Rosemary, I., et al., (2006). Regulation of chicken contamination urgently needed to control New Zealand's serious campylobacteriosis epidemic. *New Zealand Medical Journal, 119*(1243).

Baker, V., Cronin, K., Jenca, C. (2009). Coming to the Table: Literature Review and Annotated Bibliography of recent New Zealand research on consumer and public responses to emerging food technologies *Coming to the Table*: Institute of Environmental Science and Research Ltd.

Baltussen, R. (2013). *Third Plenary Session - Should Multi-Criteria Decision Analysis (MCDA) Replace Cost-Effectiveness Analysis (CEA) for Evaluation of Health Care Coverage Decisions?* Paper presented at the AN ISPOR 16th Annual European Congress, Dublin, Ireland.

Baltussen, R., & Niessen, L. (2006). Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Effectiveness and Resource Allocation, 4*(1), 14.

Banville, C., Landry, M., Martel, J. M., et al. (1998). A stakeholder approach to MCDA. *Systems Research and Behavioral Science, 15*(1), 15-32.

Baptista, F. M., Halasa, T., Alban, L., et al. (2006). Modelling food safety and economic consequences of surveillance and control strategies for Salmonella in pigs and pork. *Epidemiology and Infection*, 1-11.

Barcus, A., & Montibeller, G. (2008). Supporting the allocation of software development work in distributed teams with multi-criteria decision analysis. *Omega, 36*(3), 464-475.

Barendregt, J., Bonneux, L., & Van der Maas, P. (1996). DALYs: the age-weights on balance. *Bulletin of the World Health Organization, 74*(4), 439-444.

Bartolini, F., & Viaggi, D. (2010). Recent developments in multi-criteria evaluation of regulations. *Quality Assurance and Safety of Crops & Foods, 2*(4), 182-196.

Barzilai, J., & Golany, B. (1994). AHP rank reversal, normalization and aggregation rules. *INFOR-Information Systems and Operational Research, 32*(2), 57-64.

Behzadian, M., Kazemzadeh, R. B., Albadvi, A., et al. (2010). PROMETHEE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research, 200*(1), 198-215.

Belton, V., & Gear, T. (1983). On a short-coming of Saaty's method of analytic hierarchies. *Omega, 11*(3), 228-230.

Belton, V., & Pictet, J. (1997). A framework for group decision using a MCDA model: sharing, aggregating or comparing individual information? *Journal of Decision Systems, 6*(3), 283-303.

Belton, V., & Stewart, T. (2001). *Multiple criteria decision analysis: an integrated approach*: Springer.

Bertsch, V., Geldermann, J., Rentz, O., et al. (2006). Multi-criteria decision support and stakeholder involvement in emergency management. *International Journal of Emergency Management, 3*(2), 114-130.

Bonsel, G., Janssen, M., & Birnie, E. (2003). Mild diseases and ailments study (MiDAS). *Ziektelastbepaling van, 27*.

Brans, J.-P., & Vincke, P. (1985). A Preference Ranking Organisation Method: (The PROMETHEE Method for Multiple Criteria Decision-Making). *Management Science, 31*(6), 647-656.

Brown, J. D. (2011). Likert items and scales of measurement, *JALT Testing & Evaluation SIG Newsletter: University of Hawai'i at Manoa,* pp. 10-14. .

Bruggemann, R., & Carlsen, L. (2012). Multi-criteria decision analyses. Viewing MCDA in terms of both process and aggregation methods: Some thoughts, motivated by the paper of Huang, Keisler and Linkov. *Science of the Total Environment, 425*, 293-295.

Bruhn, C. M. (1998). Consumer acceptance of irradiated food: theory and reality. *Radiation Physics and Chemistry, 52*(1), 129-133.

Bruhn, C. M. (2007). Enhancing consumer acceptance of new processing technologies. *Innovative Food Science & Emerging Technologies, 8*(4), 555-558.

Buzby, J. C. (2003). *International trade and food safety: economic theory and case studies*: US Department of Agriculture, Economic Research Service Washington, DC.

Buzby, J. C., & Roberts, T. (2009). The economics of enteric infections: human foodborne disease costs. *Gastroenterology, 136*(6), 1851-1862.

Byanyima, M., Opiyo, G., Tumuboine, E., et al. (2013). Establishing Priorities for SPS Capacity-Building in Uganda Using Multi Criteria Decision Analysis. Uganda: Common Market for Eastern and Southern Africa, World Health Organization.

CAC. (1997). Hazard Analysis and Critical Control Point (HACCP) - System and Guidelines for its Application: Annex to CAC/RCP 1-1969, Rev. 3: Codex Alimentarius Commission.

CAC. (2005). Enhancing participation in Codex activities: Food and Agriculture Organization - World Health Organization. Geneva, Switzerland.

CAC. (2007a). Guidelines on the Application of General Principles of Food Hygiene to the Control of Listeria Monocytogenes in Foods: Codex Alimentarius Commission.

CAC. (2007b). Principles and Guidelines for the Conduct of Microbiological Risk Management (MRM): Codex Alimentarius Commission.

CAC. (2011). Joint FAO/WHO Food Standards Programme: Report of the Sixty-Fifth Session of the Executive Committee of the Codex Alimentarius Commission

Cao, K., Maurer, O., & Scrimgeour, F. (2003, February 12-14). *Estimating the Cost of Food Safety Regulation to the New Zealand Seafood Industry.* Paper presented at the Australian Agricultural and Resource Economics Society (47th), Freemantle, Australia.

Caswell, J. A. (2008). Expanding the focus of cost–benefit analysis for food safety: a multi-factorial risk prioritization approach. *Innovation: The European Journal of Social Science Research, 21*(2), 165-169.

CCNET. (2012). Cochrane Consumer Network: Levels of Evidence Retrieved 20 February, 2013, from http://consumers.cochrane.org/levels-evidence

Chang, C.-W., Wu, C.-R., Lin, C.-T., et al. (2007). An application of AHP and sensitivity analysis for selecting the best slicing machine. *Computers & Industrial Engineering, 52*(2), 296-307.

Checkland, P. B. (2001). Soft Systems Methodology. In J. Rosenhead & J. Mingers (Eds.), *Rational Analysis for a Problematic World Revisited* (Second ed.): John Wiley & Sons, Ltd, .

Chen, C., Yang, J., & Findlay, C. (2008). Measuring the effect of food safety standards on China's agricultural exports. *Review of World Economics, 144*(1), 83-106.

Claxton, K. (2013). Should Multi-Criteria Decisions Analysis (MCDA) Replace Cost Effectiveness Analysis (CEA) for Evaluation of Health Care Coverage Decisions? ISPOR 16th Annual European Congress, Dublin http://www.ispor.org.

Clinton, W. J. (1993). Executive Order 12866: Regulatory planning and review. *Federal Register, 58*(190), 51735-51744.

Codex. (2014). The Codex System: The Codex Alimentarius Commission and how it works Retrieved 21 June, 2014, from http://www.fao.org/docrep/008/y7867e/y7867e05.htm

Cox, D., & Evans, G. (2008). Construction and validation of a psychometric scale to measure consumers' fears of novel food technologies: The food technology neophobia scale. *Food quality and preference, 19*(8), 704-710.

Cressey, P. (2012). Risk Ranking: Updated Estimates of the Burden of Foodborne Disease for New Zealand in 2011 - MPI Technical Paper No: 2012/12. Christchurch, New Zealand: Environmental Science and Research.

Cressey, P., Lake, R. (2004). Ranking Food Safety Risks: A Prototype Methodology. Christchurch, New Zealand: Institute of Environmental Science & Research Limited.

Crutchfield, S. R., Buzby, J. C., Roberts, T., et al. (1997). Economic Assessment of Food Safety Regulations: The New Approach to Meat and Poultry Inspection: United States Department of Agriculture, Economic Research Service.

DAFF. (2008). Australias's Appropriate Level of Protection and Import Risk Analysis. Australia: Department of Agriculture Fisheries and Forestry.

DCLG. (2009). Multi-criteria analysis: a manual: Department for Communities and Local Government: London, United Kingdom.

de Montis, A., de Toro, P., Droste-Franke, B., et al. (2005). Assessing the quality of different MCDA methods. In Routledge (Ed.), *Alternatives for Environmental Evaluation*. New York: Routledge.

Donoso, C. (2008). *Application of Multicriteria Decision Analysis Tools to the Prioritisation of Microbial Hazards in Food Systems.* Thesis: Master of Science, The University of Guelph, Guelph, Canada.

Draper, A., & Green, J. (2002). Food safety and consumers: constructions of choice and risk. *Social Policy & Administration, 36*(6), 610-625.

Dunphy, S., & Herbig, P. A. (1995). Acceptance of innovations: The customer is the key! *The Journal of High Technology Management Research, 6*(2), 193-209.

Edwards, W. (1977). How to use multiattribute utility measurement for social decisionmaking. *Systems, Man and Cybernetics, IEEE Transactions on, 7*(5), 326-340.

Edwards, W., & Barron, F. H. (1994). SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement. *Organizational Behavior and Human Decision Processes, 60*(3), 306-325.

ESR. (2007). Food Safety Risk Profile for Campylobacter species in broiler (young) chickens: Institute of Environmental Science & Research Limited.

ESR. (2008). Consumer knowledge, attitudes and beliefs with respect to Campylobacter, Campylobacteriosis and Poultry: Institute of Environmental Science & Research Limited.

ESR. (2010). Clostridium Botulinum: The organism/toxin: Institute of Environmental Science & Research Limited.

ESR. (2014a). Environmental Science and Research, EpiSurv Retrieved 20 May, 2014, from https://surv.esr.cri.nz/episurv/index.php

ESR. (2014b). Environmental Science and Research, LabSurv Retrieved 20 May, 2014, from https://surv.esr.cri.nz/LabSurv/index.php

EU. (2000). The Precautionary Principle. *Europa: Summaries of EU legislation* Retrieved 23 February, 2013, from http://europa.eu/legislation_summaries/consumers/consumer_safety/l32042_en.htmFAO. (2003). The Codex System: The Codex Alimentarius Commission and how it works Retrieved 12 February, 2013, from http://www.fao.org/docrep/008/y7867e/y7867e05.htm

FAO. (2007a). Working Principles for Risk Analysis for Application by National Governments. Available from: ftp://ftp.fao.org/codex/Publications/Booklets/Risk/Risk_EN_FR_ES.pdf.

FAO. (2007b). Food Safety Risk Analysis: A guide for national food safety authorities: Food and Agriculture Organization.

FAO/WHO. (1995a). Application of Risk Analysis to Food Standards Issues. Geneva: Food and Agriculture Organization & World Health Organzsation.

FAO/WHO. (1995b). Application of Risk Analysis to Food Standards Issues, a Joint FAO/WHO Expert Consultation. Geneva, Switzerland: Food and Agriculture Organization & World Health Organization.

FAO/WHO. (1997). Risk Management and Food Safety. Rome: Food and Agriculture Organization & World Health Organization.

Fazil, A., Rajic, A., Sanchez, J., et al. (2008). Choices Choices, The application of MCDA to a food safety decision making problem. *Journal of Food Protection, 71*(11).

Finan, J. S., & Hurley, W. J. (1997). The analytic hierarchy process: does adjusting a pairwise comparison matrix to improve the consistency ratio help? *Computers & Operations Research, 24*(8), 749-755.

Finnie, L., Graham, H. (2004). Public Health Unit Profile: Public Health South.

Fischer, D. (2009). Multiple criteria decisions: opening the black box. *Journal of Real Estate Practice and Education, 12*(1), 43-55.

Flynn, C. (1999). *The Optimal Regulatory Model: the application of risk management to government regulation* Paper presented at the World Congress on Food Safety, Sydney, Australia.

Food Standards Australia New Zealand Act, C2011C00117, 118 of 1991 Stat. (Office of Legislative Drafting and Publishing, Attorney General's Department, Canberra 1991).

Frenzen, P. D., Majchrowicz, T. A., Buzby, J. C., et al. (2000). Consumer acceptance of irradiated meat and poultry products. *Agricultural Information Bulletins*: US Dept. of Agriculture, Economic Research Service.

Frewer, L. J., Arnout, R. H., Nidhi, G. (2009). Coming to the Table: Key issues regarding societal acceptance of future foods. An interpretation and extrapolation of research findings: Wageningen University.

FSANZ. (2013). Fact Sheets: Codex Retrieved 24 March, 2013, from http://www.foodstandards.gov.au/scienceandeducation/factsheets/factsheets/codexnovember2012.cfm

Gadiel, D., & Abelson, P. (2010). The economic cost of foodborne disease in New Zealand, from http://www.foodsafety.govt.nz/elibrary/industry/economic-cost-foodborne-disease/foodborne-disease.pdf

Gartlehner, G., Hansen, R. A., Nissman, D., et al. (2006). Criteria for distinguishing effectiveness from efficacy trials in systematic reviews. *Technical Review No. 12*: Agency for Healthcare Research and Quality (US).

Geldermann, J., Bertsch, V., Treitz, M., et al. (2009). Multi-criteria decision support and evaluation of strategies for nuclear remediation management. *Omega, 37*(1), 238-251.

Gilbert, S., Lake, R., Hudson, A. H., et al. (2006). Shiga-Toxin Producing Escherichia Coli in Leafy Vegetables. Christchurch, New Zealand: Institute of Environmental Science & Research Limited.

Glaser, B. G., & Strauss, A. L. (1967). The Discovery of Grounded Research: Strategies for Qualitative Research. New York: Aldine De Gruyter.

Gold, M., Stevenson, D., & Fryback, D. (2002). HALY S AND QALY S AND DALY S, OHMY: Similarities and Differences in Summary Measures of Population Health. *Annual Review of Public Health, 23*(1), 115-134.

Goodman, L. A. (1961). Snowball sampling. *The Annals of Mathematical Statistics, 32*(1), 148-170.

Greig, J. D., Todd, E. C., Bartleson, C. A., et al. (2007). Outbreaks where food workers have been implicated in the spread of foodborne disease. Part 1. Description of the problem, methods, and agents involved. *Journal of Food Protection, 70*(7), 1752-1761.

Griffin, R. (2012). Plant Pest Risk Analysis: Concepts and Applications (pp. 257): CAB International.

Guitouni, A., & Martel, J.-M. (1998). Tentative guidelines to help choosing an appropriate MCDA method. *European Journal of Operational Research, 109*(2), 501-521. doi: 10.1016/s0377-2217(98)00073-3

Haagsma, J. A. (2010). *Disability Adjusted Life Years And Acute Onset Disorders.* Doctor of Philosophy, Erasmus University, Rotterdam, Netherlands.

Haagsma, J. A., Havelaar, A. H., Janssen, B., et al. (2008). Disability Adjusted Life Years and minimal disease: application of a preference-based relevance criterion to rank enteric pathogens. *Population health metrics, 6*(1), 7.

Halbrendt, C. K., Pesek, J. D., Parsons, A., et al. (1994). Consumer Preference For Pst-Supplement Pork. *Australian Journal of Agricultural and Resource Economics, 38*(2), 189-202.

Hansen, P., & Ombler, F. (2014). 1000 Minds MCDA software Retrieved 25 August 2014, from http://www.1000minds.com/about/how

Harris, R., et al. (2001). Current methods of the U.S. Preventive Services Task Force: a review of the process. *American Journal of Preventive Medicine*, 21-35.

Havelaar, A., Bräunig, J., Christiansen, K., et al. (2007). Towards an integrated approach in supporting microbiological food safety decisions. *Zoonoses and public health, 54*(3 4), 103-117.

Havelaar, A., & Melse, J. (2003). Quantifying public health risk in the WHO Guidelines for drinking-water quality: A burden of disease approach.

Henson, S., & Caswell, J. (1999). Food safety regulation: an overview of contemporary issues. *Food Policy, 24*(6), 589-603.

Henson, S., Caswell, J. A., Cranfield, J. A. L., et al. (2007). A multi-factorial risk prioritization framework for food-borne pathogens: University of Massachusetts Department of Economics Working Paper No. 2007-8

Henson, S., & Masakure, O. (2011). Establishing Priorities for SPS Capacity-building: A Guide to Multi-Criteria Decision-Making. Guelph, Canada: Standards and Trade Development Facility.

Hikuroa, D., Morgan, T., Gravley, D., et al. (2010). *Integrating indigenous values in geothermal development.* Paper presented at the 4th International Traditional Knowledge Conference 2010.

Ho, W. (2008). Integrated analytic hierarchy process and its applications–a literature review. *European Journal of Operational Research, 186*(1), 211-228.

Hoffmann, S. (2010). Food safety policy and economics: a review of the literature: RFF Press, Washington DC

Houghton, J. R., van Kleef, E., Rowe, G., et al. (2006). Consumer perceptions of the effectiveness of food risk management practices: A cross-cultural study. *Health, Risk & Society, 8*(2), 165-183.

Huang, C. L., Wolfe, K., & McKissick, J. (2006). *Willingness to Pay for Irradiated Meat Products: A Comparison between Poultry and Pork*.

Hudson, A., Cressey, P., & Lake, R. (2008). On-Farm Factors for Campylobacter Contamination of Broilers: Literature Review and Overview of Broiler Farming in New Zealand. Christchurch, New Zealand: Institute of Environmental Science & Research Limited.

Hunter, R., & Fewtrell, L. (2001). Acceptable Risk *Water Quality: Guidelines, Standards and Health* London, UK: World Health Organization (WHO).

INAHTA. (2014). International Network of Agencies for Health Technology Assessment - HTA Resources Retrieved 3 June, 2014, from http://www.inahta.org/HTA/

IOM (Institute of Medicine). (2006). Valuing Health for Regulatory Cost-Effectiveness Analysis. Washington, DC: National Academies Press.

Jensen, H. H., Unnevehr, L. J., Gomez, M. I., et al. (1998). Costs of improving food safety in the meat sector. *Journal of Agricultural and Applied Economics, 30*, 83-94.

Johnson, R., Hanrahan, C. (2010). The U.S.-EU Beef Hormone Dispute: Congressional Research Service.

Kalafatelia, E., Fryer, K. (2011). Awareness and Knowledge of Folate and Folic Acid: A survey of New Zealand women of child-bearing age: Ministry of Agriculture and Forestry.

Karvetski, C. W., Lambert, J. H., & Linkov, I. (2011). Scenario and multiple criteria decision analysis for energy and environmental security of military and industrial installations. *Integrated environmental assessment and management, 7*(2), 228-236.

Keeney, R. (1992). Value-Focused Thinking: A Path to Creative Decisionmaking (Cambridge, MA: Harvard University).

Kepa Brian Morgan, T. K., Sardelic, D. N., & Waretini, A. F. (2012). The Three Gorges Project: How sustainable? *Journal of Hydrology, 460*, 1-12.

Kim, S., Lee, H. P. K., & Jung, C. (2007). *MAUT Approach for Selecting a Proper Decommissioning Scenario.* Paper presented at the WM Symposia, , 1628 E. Southern Avenue, Suite 9-332, Tempe, AZ 85282 (United States).

Koen, R. (2008). *Aspects of MCDA classification and sorting methods.* Thesis: Master of Science, University of South Africa (UNISA).

König, A., Kuiper, H. A., Marvin, H. J. P., et al. (2010). The SAFE FOODS framework for improved risk analysis of foods. *Food Control*.

Kraus, N. (2013). Indigenous Maori Values in Kawerau, Bay of Plenty, New Zealand: Assessing Environmental, Cultural, Social, and Economic Impacts of the Te Ahi O Maui Geothermal Project Using the Mauri Model. *MAI Journal (Submitted)*.

Kreiger, S., Schiefer, G., da Silva, C. (2007). Costs and benefits in food quality systems:

concepts and a multi-criteria evaluation approach. Rome: Food and Agriculture Organization.

Kumar, N. V., & Ganesh, L. (1996). A simulation-based evaluation of the approximate and the exact eigenvector methods employed in AHP. *European Journal of Operational Research, 95*(3), 656-662.

Lake, R. (2006). Risk Ranking: Development of a Single Metric for Risk Ranking by the NZFSA. Christchurch, New Zealand: Institute of Environmental Science & Research Limited.

Lake, R., Cressy, P., Hudson, A., et al. (2007). Quantitative model: Campylobacter spp in the poultry food chain Retrieved August 2, 2013, from http://www.foodsafety.govt.nz/elibrary/industry/Quantitative_Risk-Science_Research.pdf

Lake, R., Horn, B., & Ball, A. (2011). Campylobacter in food and the environment: examining the link with public health. Christchurch, New Zealand: Ministry of Agriculture and Forestry.

Lake, R., Horn, B., Dunn, A., et al. (2013). Cost-effectiveness analysis of potential interventions to control Campylobacter in the New Zealand poultry meat food supply. *Journal of Food Protection, 7*(July 2013), 1161-1167.

Lake, R., Hudson, A., & Cressey, P. (2002). Shiga Toxin-Producing Escherichia Coli in Red Meat and Meat Products. Christchurch, New Zealand: Institute of Environmental Science & Research Limited.

Lake, R., Hudson, A. H., Cressey, P., et al. (2007). Campylobacter Jejuni/Coli in Poultry (Whole and Pieces). Christchurch, New Zealand: Institute of Environmental Science & Research Limited.

Lake, R., van der Logt, P., McBride, G., et al. (2007). Campylobacteriosis in New Zealand: modelling as a way forward. *New Zealand Science Review, 64*, 37.

Lake, R. J., Adlam, S. B., Perera, S., et al. (2010). The disease pyramid for acute gastrointestinal illness in New Zealand. *Epidemiology & Infection, 138*(10), 1468-1471. doi: doi:10.1017/S0950268810000397

Lake, R. J., Cressey, P. J., Campbell, D. M., et al. (2010). Risk ranking for foodborne microbial hazards in New Zealand: burden of disease estimates. *Risk Analysis, 30*(5), 743-752.

Lawson, L., Jensen, J., Christiansen, P., et al. (2009). Cost-effectiveness of Salmonella reduction in Danish abattoirs. *International Journal of Food Microbiology, 134*(1-2), 126-132.

Lawson, L., Lund, M., & Jensen, J. (2007). The Costs of Food Safety - A Methodological Review, from http://www.foi.life.ku.dk/English/Publications/FOI%20Series/~/media/Foi/docs/Publikationer/Working%20Papers/2007/Working%20Paper%20no21%202007.ashx

Lee, J. (2012). Standardisation of Hazard Analysis and Critical Control Point (HACCP): MPI Technical Paper No: 2012/30: Ministry for Primary Industries: Standards Branch.

Li, Y., & Beghin, J. C. (2012). A meta-analysis of estimates of the impact of technical barriers to trade. *Journal of Policy Modeling, 34*(3), 497-511.

Lim, E., Lopez, L., Borman, A., et al. (2012). Annual Report Concerning Foodborne Disease in New Zealand 2011. Christchurch, New Zealand: Institute of Environmental Science & Research Limited.

Lim, E., Lopez, L., Cressey, P., et al. (2011). Annual Report Concerning Foodborne Disease in New Zealand 2010. Christchurch, New Zealand: Institute of Environmental Science & Research Limited.

Lim, E., Tisch, C., Cressey, P., et al. (2010). Annual Report Concerning Foodborne Disease in New Zealand 2009. Chirstchurch, New Zealand: Institute of Environmental Science & Research Limited.

Linkov, I., Satterstrom, F., Kiker, G., et al. (2006). From comparative risk assessment to multi-criteria decision analysis and adaptive management: Recent developments and applications. *Environment International, 32*(8), 1072-1093.

Linkov, I., Satterstrom, F. K., Steevens, J., et al. (2007). Multi-criteria decision analysis and environmental risk assessment for nanomaterials. *Journal of Nanoparticle Research, 9*(4), 543-554.

Lopert, R. (2009). Evidence-Based Decision-Making Within Australia's Pharmaceutical Benefits Scheme. ACT, Australia: Theraputic Goods Administration.

Mabin, V., Beattie, M. (2006). A Practical Guide to Multi-Criteria Decision Analysis. Victoria University of Wellington.

Macharis, C., Springael, J., De Brucker, K., et al. (2004). PROMETHEE and AHP: The design of operational synergies in multicriteria analysis.: Strengthening PROMETHEE with ideas of AHP. *European Journal of Operational Research, 153*(2), 307-317.

Mangen, M., Batz, M., Käsbohrer, A., et al. (2010). Integrated Approaches for the Public Health Prioritization of Foodborne and Zoonotic Pathogens. *Risk Analysis, 30*(5), 782-797.

Mangen, M. J. J., de Wit, G. A., & Havelaar, A. H. (2007). Economic analysis of Campylobacter control in the dutch broiler meat chain. *Agribusiness, 23*(2), 173-192.

Mangen, M. J. J., Havelaar, A. H., Poppe, K. P., et al. (2007). Cost-Utility Analysis to Control Campylobacter on Chicken Meat—Dealing with Data Limitations. *Risk Analysis, 27*(4), 815-830.

Mareschal, B., De Smet, Y., & Nemery, P. (2008). *Rank reversal in the PROMETHEE II method: some new results.* Paper presented at the Industrial Engineering and Engineering Management, 2008. IEEM 2008. IEEE International Conference on.

Mason, R. O., & Mitroff, I. I. (1981). *Challenging strategic planning assumptions: Theory, cases and techniques: Wiley- New York,* .

Maurer, T. J., & Pierce, H. R. (1998). A comparison of Likert scale and traditional measures of self-efficacy. *Journal of Applied Psychology, 83*(2), 324.

Mayon-White, W., & Wisudha, A. (2010). The use of MCDA in strategy and change management (pp. 105). London, United Kingdom: José-Rodrigo Córdoba School of Management, Royal Holloway University of London.

Mazzocchi, M., Ragona, M., & Zanoli, A. (2013). A fuzzy multi-criteria approach for the ex-ante impact assessment of food safety policies. *Food Policy, 38*, 177-189.

Mehta, M. D. (2002). Public Perceptions of Food Safety: Assessing the Risks Posed by Genetic Modification, Irradiation, Pesticides, Microbiological Contamination and High Fat/High Calorie Foods. *Pierce Law Review, 1*(1), 69-84.

Melse, J. M., Essink-Bot, M.-L., Kramers, P., et al. (2000). A national burden of disease calculation: Dutch disability-adjusted life-years. Dutch Burden of Disease Group. *American Journal of Public Health, 90*(8), 1241.

MercoPress. (2008). , WTO rules against EU import ban of beef with hormones. (Monday 31st March). Retrieved from http://en.mercopress.com/2008/03/31/wto-rules-against-eu-import-ban-of-beef-with-hormones

Metcalfe, S., Rodgers, A., Werner, R., et al. (2012). PHARMAC has no cost-effectiveness threshold. *The New Zealand Medical Journal, 125* (1350).

MfE. (1996). Hazardous Substances and New Organisms Act 1996. Wellington, New Zealand: Ministry for the Environment.

MOH. (2014). Ministry of Health - Publications Retrieved 20 May, 2014, from http://www.health.govt.nz/publications/cost%20effectiveness

MOH. (2015). Notifiable Diseases, Retrieved 20 March, 2015, from http://www.health.govt.nz/our-work/diseases-and-conditions/notifiable-diseases

Montibeller, G., Belton, V., Ackermann, F., et al. (2008). Reasoning maps for decision aid: an integrated approach for problem-structuring and multi-criteria evaluation. *Journal of the Operational Research Society, 59*(5), 575-589.

Morgan, T. (2006). Decision-support tools and the indigenous paradigm. *Proceedings of the ICE-Engineering Sustainability, 159*(4), 169-177.

Morgan, T. K. K. B. (2008). *Pacific Rim Indigeneity and Sustainable Municipalities.* Paper presented at the World Environmental and Water Resources Congress 2008, sAhupua'A.

Morgan, T. K. K. B. (2011). Waiora and Cultural Identity: Water quality assessment using the Mauri Model. *AlterNative: An International Journal of Indigenous Peoples, 3*(1).

Mørkbak, M., Christensen, T., & Gyrd-Hansen, D. (2010). Consumers' willingness to pay for safer meat depends on the risk reduction methods-a Danish case study on Salmonella risk in minced pork. *Food Control*.

MPI. (2008). Position Statement on the Application of Precaution in Managing Biosecurity Risks Associated with the Importation of Risk Goods. Wellington, New Zealand: Ministry for Primary Industries.

MPI. (2010). Food Safety and the New Zealand public: MPI Food Safety.

MPI. (2010/2011). *NZ Participation in Codex Working Groups* Wellington: MPI Food Safety Retrieved from http://www.foodsafety.govt.nz/policy-law/codex/nz/particpation.htm.

MPI. (2013a). About food recalls Retrieved March 15, 2013, from http://www.foodsafety.govt.nz/recalls-warnings/about-recalls/

MPI. (2013b). Campylobacter Risk Management Strategy 2013-2014. Wellington, New Zealand: Ministry for Primary Industries.

MPI. (2013c). Consultation: Ministry for Primary Industries Food Safety Retrieved March 20, 2013, from http://www.foodsafety.govt.nz/consultation/

MPI. (2013d). Listeria Risk Management Strategy 2013-2014. Wellingon, New Zealand: Ministry for Primary Industries.

MPI. (2013e). MAF's Precautionary Approach to Managing Food Safety Risks Retrieved 16 March, 2013, from http://www.foodsafety.govt.nz/elibrary/industry/Precautionary_Approach-Background_Paper.htm

MPI. (2013f). Market Access (OMARs/FYIs) Retrieved 28 February, 2013, from http://www.foodsafety.govt.nz/industry/exporting/market-access/

MPI. (2013g). Risk Profiles, Ministry for Primary Industries Retrieved March 15, 2013, from http://www.foodsafety.govt.nz/science-risk/risk-assessment/risk-profiles/

MPI. (2013h). Salmonella Risk Management Strategy 2013-2014. Wellington, New Zealand: Ministry for Primary Industries.

MPI. (2014a). Food has big chemical safety margins Retrieved June 19, 2014, from http://www.foodsafety.govt.nz/elibrary/industry/2010-11-food-tech-mrl-setting.htm

MPI. (2014b). Overview - HACCP Retrieved 3 September, 2014, from http://www.foodsafety.govt.nz/industry/general/haccp/overview.htm

MPI. (2014c). Risk Profiles Retrieved 23 August, 2014, from http://www.foodsafety.govt.nz/science-risk/risk-assessment/risk-profiles/

MPI. (2014d). Which microbiological limits do i use? Retrieved 2 August, 2014, from http://www.foodsafety.govt.nz/elibrary/industry/Which_Microbiological-Outlines_Four.htm

MPI. (2014e). Hormonal Growth Promotants Retrieved 26 December, 2014, from http://www.foodsmart.govt.nz/whats-in-our-food/chemicals-nutrients-additives-toxins/agricultural-production/hormonal-growth-promotants/

Munda, G. (1995). *Multicriteria Evaluation in a Fuzzy Environment: Theory and Applications in Ecological Economics*: Physical Verlag, Heidelberg

Müllner, P. (2009). *Estimating the contribution of different sources to the burden of human campylobacteriosis and salmonellosis.* Thesis: Doctor of Philosophy, Massey University, Palmerston North, New Zealand.

Müllner, P., Jones, G., Noble, A., et al. (2009). Source Attribution of Food-Borne Zoonoses in New Zealand: A Modified Hald Model. *Risk Analysis, 29*(7), 970-984.

Murray, C. J., & Lopez, A. D. (1996). *Global burden of disease* (Vol. 1): Harvard University Press Cambridge, MA.

NHC. (2013). Decision-making criteria Retrieved 27 May, 2014, from http://nhc.health.govt.nz/about-us/decision-making-criteria

NHC. (2014). National Health Committee, Our Work Retrieved 27 May, 2014, from http://nhc.health.govt.nz/our-work

Nord, E., Pinto, J. L., Richardson, J., Menzel, P., & Ubel, P. (1999). Incorporating societal concerns for fairness in numerical valuations of health programmes. *Health economics, 8*(1), 25-39.

NZFSA. (2003). Maximum Residue Limits/MRLs: New Zealand Food Safety Authority: Agricultural Compounds and Veterinary Medicines Group: New Zealand Food Safety Authority

NZFSA. (2006). Food Sector Risk Ranking and Prioritisaion Models: The Methods: New Zealand Food Safety Authority.

NZFSA. (2007). Code of Practice - Processing of Poultry: New Zealand Food Safety Authority.

NZFSA. (2008a). New rules for poultry processors. *Food Focus,* 6-7: New Zealand Food Safety Authority.

NZFSA. (2008b). The Regulatory Model: An Explanation: New Zealand Food Safety Authority.

NZFSA. (2009a). A guide to the Animal Products Act 1999, the Food Act 1981 and the Food Hygiene Regulations 1974: New Zealand Food Safety Authority.

NZFSA. (2009b). New Zealand's Strategic Objectives in Codex: New Zealand Food Safety Authority.

NZFSA. (2009c). Report of NZFSA Academy Annual Meeting.: New Zealand Food Safety Authority.

NZFSA. (2010a). Campylobacter Risk Management Strategy 2010-2013: New Zealand Food Safety Authority.

NZFSA. (2010b). *New Zealand's Food Safety Risk Management Framework*. Retrieved January 29 2013, from http://www.foodsafety.govt.nz/elibrary/industry/RMF_full_document_-_11604_NZFSA_Risk_Management_Framework_3.1.pdf.

NZFSA. (2010c). NZFSA update on BPA in babies' bottles Retrieved February 24, 2014, from http://www.foodsafety.govt.nz/elibrary/industry/Nzfsa_Update-Safety_Common.htm

NZFSA. (2010d). Prioritisation Process for Market Access Activites: New Zealand Food Safety Authority.

NZFSA. (2011). An Evidence-Based Approach to Reviewing the Science on the Safety of Chemicals in Foods: New Zealand Food Safety Authority.

NZFSA/FSANZ. (2008). The New Zealand Food Regulatory Environment: New Zealand Food Safety Authority and Food Standards Australia New Zealand.

Ollinger, M., Moore, D., & Chandran, R. (2004). Meat and poultry plants' food safety investments: Survey findings. *Agricultural Economics Report*(7893), 48.

Önal, H., Unnevehr, L., & Bekric, A. (2000). Regional shifts in pork production: implications for competition and food safety. *American Journal of Agricultural Economics, 82*(4), 968-978.

Orden, D., Josling, T., & Roberts, D. (2002). Product Differentiation, Sanitary Barriers, and Arbitrage in World Poultry Markets *Global Food Trade and Consumer Demand for Quality* (pp. 147-164): Springer.

Otsuki, T., Wilson, J. S., & Sewadeh, M. (2001). Saving two in a billion:: quantifying the trade effect of European food safety standards on African exports. *Food Policy, 26*(5), 495-514.

Oxford. (2008). *A Dictionary of Chemistry* (6th ed.): Oxford University Press

Papamichail, K. N., & French, S. (2013). 25 Years of MCDA in nuclear emergency management. *IMA Journal of Management Mathematics, 24*(4), 481-503.

Peacock, B. C. (2011). *A GIS-Based Habitat Restoration and Preservation Prioritization Tool Geospatially Integrating Hydrological, Ecological, Pollution, Economic, Social and Cultural Considerations*. Paper presented at the National Conference On Undergraduate Research (NCUR) March 21 - April 2, 2011, New York.

Peacock, B. C., Hikuroa, D., & Morgan, T. K. K. B. (2012). Watershed-scale prioritization of habitat restoration sites for non-point source pollution management. *Ecological Engineering, 42*, 174-182.

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology and Community Health, 57*(7), 527-529.

PHARMAC. (2012). *Prescription for Pharmacoeconomic Analysis: Methods for cost-utility analysis - Version 2.1* - Wellington, New Zealand: Parmaceutical Management Agency.

PHARMAC. (2014). Decision criteria consultation Retrieved June 21, 2014, from http://www.pharmac.health.nz/about/operating-policies-and-procedures/decision-criteria-consultation

Phillips, L. (2008). *Multi-criteria decision analysis for appraising benefit-risk.* Paper presented at the Ministerial Industry Strategy Group Forum, 3rd September 2008, United Kingdom.

Phillips, L., & Stock, A. (2003). Use of multi-criteria analysis in air quality policy: United Kingdom Department for Environment Food & Rural Affairs.

Platia, O. (2012). Using the Mauri Model to Assess the Impact of the Rena Grounding on the Mauri of the Bay of Plenty, New Zealand: Lafayette College, Dept of Civil and Environmental Engineering. Easton, PA, United States & Frontiers Abroad, The University of Auckland. Auckland, New Zealand.

Ragona, M., & Mazzocchi, M. (2008a). Food safety regulation, economic impact assessment and quantitative methods. *Innovation: The European Journal of Social Science Research, 21*(2), 145-158.

Ragona, M., & Mazzocchi, M. (2008b). *Impact evaluation of food safety regulations: A review of quantitative methods.* Paper presented at the European Association of Agricultural Economists 110th Seminar, February 18-22, 2008, Innsbruck-Igls, Austria.

Ragona, M., Mazzocchi, M., & Alldrick, A. (2012). *Multi-criteria analysis for the impact assessment of food safety policies: The case of EU regulation on dietary arsenic.* Paper presented at the Italian Association of Agricultural and Applied Economics (AIEAA).

Ragona, M., Mazzocchi, M., & Rose, M. (2012). *Regulatory impact assessment of food safety policies: A preliminary study on alternative EU interventions on dioxins.* Paper presented at the 86th Annual Conference, April 16-18, 2012, Warwick University, Coventry, UK.

Ragona, M., Mazzocchi, M., Zanoli, A., et al. (2011). Testing a toolbox for impact assessment of food safety regulations: maximum levels for T 2 and HT 2 toxins in the European Union. *Quality Assurance and Safety of Crops & Foods, 3*(2), 12-23.

Raiffa, H., & Keeney, R. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*: Wiley.

Redpath, S., Arroyo, B., Leckie, F., et al. (2004). Using decision modeling with stakeholders to reduce human–wildlife conflict: a raptor–grouse case study. *Conservation Biology, 18*(2), 350-359.

Rimal, A., Fletcher, S. M., McWatters, K. H., et al. (2001). Perception of food safety and changes in food consumption habits: a consumer analysis. *International Journal of Consumer Studies, 25*(1), 43-52.

Robberstad, B. (2005). QALYs vs DALYs vs LYs gained: What are the differences, and what difference do they make for health care priority setting? *Norsk Epidemiologi, 15*(2).

Ropkins, K., & Beck, A. J. (2000). Evaluation of worldwide approaches to the use of HACCP to control food safety. *Trends in Food Science & Technology, 11*(1), 10-21.

Rose, J. (2011). Sample size, statistical power and discrete choice experiments: How much is enough, University of Technology, Sydney Australia, from http://www.newton.ac.uk/programmes/DAE/seminars/2011081711001.html

Ross, T., & Sumner, J. (2002). A simple, spreadsheet-based, food safety risk assessment tool. *International journal of food microbiology, 77*(1-2), 39-53.

Roy, B., & Vanderpooten, D. (1996). The European school of MCDA: Emergence, basic features and current works. *Journal of Multi-Criteria Decision Analysis, 5*(1), 22-38.

Ruzante, J. M., Davidson, V. J., Caswell, J., et al. (2010). A multifactorial risk prioritization framework for foodborne pathogens. *Risk Analysis, 30*(5), 724-742.

Ryan, M., & Farrar, S. (2000). Using conjoint analysis to elicit preferences for health care. *BMJ: British Medical Journal, 320*(7248), 1530.

Saaty, T. L. (1980). The analytic hierarchy process: planning, priority setting, resources allocation. *M cGraw-Hill*.

Saaty, T. L. (2001). Deriving the AHP 1-9 scale from first principles. *ISAHP 2001 proceedings, Bern, Switzerland*.

Saaty, T. L. (2004). Decision making—the analytic hierarchy and network processes (AHP/ANP). *Journal of Systems Science and Systems Engineering, 13*(1), 1-35.

Saaty, T. L., & Vargas, L. G. (2012). The Seven Pillars of the Analytic Hierarchy Process *Models, Methods, Concepts and Applications of the Analytic Hierarchy Process* (Vol. XIII): Springer Science & Business Media New York.

Salo, A., & Hämäläinen, R. P. (2010). Multicriteria decision analysis in group decision processes *Handbook of group decision and negotiation* (pp. 269-283): Springer.

Sauian, M. S. (2010). *MCDM: A Practical Approach in Making Meaningful Decisions.* Paper presented at the Regional Conference on Statistical Sciences 2010 (RCSS'10) 13-14 June 2010, Kota Bharu, Kelantan.

Savage, G. T., Nix, T. W., Whitehead, C. J., et al. (1991). Strategies for assessing and managing organizational stakeholders. *The Executive, 5*(2), 61-75.

Savage, L. J. (1954). The Foundations of Statistics: New York: Wiley.

Sears, A., Baker, MG., Wilson, N., et al., (2011). Marked Campylobacteriosis Decline after Interventions Aimed at Poultry, New Zealand. *Emerging Infectious Diseases, 17*(6).

Sekaran, U. (1992). *Research Methods for Business: A Skill Building Approach*: John Wiley & Sons, Inc.

Shaw, I. (2013). *Food safety: the science of keeping food safe*: Wiley-Blackwell.

Siegrist, M. (2008). Factors influencing public acceptance of innovative food technologies and products. *Trends in Food Science & Technology, 19*(11), 603-608.

Slorach, S. (2008). Food safety risk management in New Zealand: A review of the New Zealand Food Safety Authority's risk management framework and its application. Available from http://www.foodsafety.govt.nz/elibrary/industry/Food_Safety-Nzfsa_Been.pdf.

Spackman, M., Dodgson, J., Pearman, A., et al. (2000). DTLR Multi-Criteria Analysis Manual. London, UK: NERA (National Economic Research Associates).

StatsNZ. (2014). Life Expectancy. Retrieved 19 June, 2014, from http://www.stats.govt.nz/browse_for_stats/health/life_expectancy.aspx

Stewart, T. P., & Johanson, D. S. (1998). SPS Agreement of the World Trade Organization and International Organizations: The Roles of the Codex Alimentarius Commission, the International Plant Protection Convention, and the International Office of Epizootics, The. *Syracuse J. Int'l L. & Com., 26*, 27.

Stouthard, M. E., Essink-Bot, M., Bonsel, G., et al. (1997). *Disability weights for diseases in the Netherlands*: Inst. Sociale Geneeskunde.

Sullivan, T. (2012). *Using MCDA (Multi-Criteria Decision Analysis) to prioritise publicly-funded health care.* Thesis: Doctor of Philosophy, University of Otago, New Zealand.

Tappero, J. W., Schuchat, A., Deaver, K. A., et al. (1995). Reduction in the incidence of human listeriosis in the United States. *JAMA: the journal of the American Medical Association, 273*(14), 1118-1122.

Tong, T. (2014). Tutorial: Scoring and Swing Weighting in MCDA, ScHAAR, University of Sheffield, from http://prezi.com/uwxki_u314to/tutorial-scoring-and-swing-weighting-in-mcda/

Tony, M., Goetghebeur, M., Wagner, M., et al. (2010). *Multicriteria decision analysis (MCDA) for drug coverage decision by a public health plan: case study of tramadol for chronic non-cancer pain in Canada (CNCP) in Canada.* Paper presented at the ISPOR, Atlanta.

Tony, M., Wagner, M., Khoury, H., et al. (2011). Bridging health technology assessment (HTA) with multicriteria decision analyses (MCDA): field testing of the EVIDEM framework for coverage decisions by a public payer in Canada. *BMC health services research, 11*(1), 329.

Treasury. (2005). New Zealand Treasury, Business Analysis Team: Cost-Benefit Anaysis Primer Retrieved July 12, 2013, from http://www.treasury.govt.nz/publications/guidance/planning/costbenefitanalysis/primer/cba-primer-v12.pdf

Treasury. (2013). New Zealand Treasury, Regulatory Impact Analysis Handbook. Available from: http://www.treasury.govt.nz/publications/guidance/regulatory/impactanalysis/ria-handbk-jul13.pdf. Wellington, New Zealand.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, S251-S278.

Unnevehr, L. J., Gómez, M. I., & Garcia, P. (1998). The incidence of producer welfare losses from food safety regulation in the meat industry. *Review of Agricultural Economics, 20*(1), 186-201.

USDHHS. (2013). Shigella Retrieved 23 November, 2013, from http://www.foodsafety.gov/poisoning/causes/bacteriaviruses/shigella/

U.S. OMB (Office of Management and Budget) (2003). Regulatory Analysis. Circular A-4. Washington, DC: U.S. OMB.

Vaidya, O. S., & Kumar, S. (2006). Analytic hierarchy process: An overview of applications. *European Journal of Operational Research, 169*(1), 1-29.

Vannoort, R. W., & Thomson, B. M. (2009). 2009 New Zealand Total Diet Study: Agricultural compound residues, selected contaminant and nutrient elements. Wellington, New Zealand: Ministry for Primary Industries.

Verbeke, W., Frewer, L. J., Scholderer, J., et al. (2007). Why consumers behave as they do with respect to food safety and risk information. *Analytica Chimica Acta, 586*(1), 2-7.

Vickers, A. J. (1999). Comparison of an ordinal and a continuous outcome measure of muscle soreness. *International Journal of Technology Assessment in Health Care, 15*(4), 709-716.

Von Neumann, J., & Morgenstern, O. (1947). *The theory of games and economic behavior*: Princeton University Press.

Von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research* (Vol. 604): Cambridge University Press.

Vos, T., Carter, R., Barendregt, J., et al. (2010). ACE–Prevention Team. Assessing Cost-Effectiveness in Prevention (ACE–Prevention): Final Report. University of Queensland, Brisbane and Deakin University, Melbourne.

Vose, D. (2007). PERT distribution. *ModelRisk STANDARD Edition: Distributions,* Retrieved 15 April, 2014, from http://www.vosesoftware.com/ModelRiskHelp/index.htm#Distributions/Continuous_distributions/PERT_distribution.htm

Vroom, V. (1974). A new look at managerial decision-making. In D. A. Kolb (Ed.), *Organizational Psychology: A Book of Readings* (pp. 63-80). Prentice-Hall, Englewood Cliffs, NJ.

WHO. (2009). Global health risks: mortality and burden of disease attributable to selected major risks: World Health Organization.

WHO/FAO. (2002). Review of the Working Procedures of the Join FAO/WHO Meeting on Pesticide Residues (JMPR): Food and Agriculture Organization & World Health Organization.

WHO/FAO. (2014). Multicriteria-based ranking for risk management of food-borne parasites - Report of a Joint FAO/WHO Expert Meeting, 3–7 September 2012, FAO Headquarters, Rome, Italy *Microbiological Risk Assessment Series* (Vol. 23). Rome, Italy: Food and Agriculture Organization & World Health Organization.

Winnard, D., Simmons, G. (2008). Food Safety Campaign Evaluation 2007-2008: New Zealand Foodsafe Partnership.

Wright, D. W. (2013). A Historical Overview by Former MIRINZ Director Doug Wright. Retrieved 20 March, 2013, from http://www.mirinz.org.nz/history.asp

WTO. (2014). SPS Agreement Training Module: Chapter 2 - The Key Provisions of the Agreement Retrieved 22 May, 2014, from http://www.wto.org/english/tratop_e/sps_e/sps_agreement_cbt_e/c2s1p1_e.htm

Yeung, R. M. W., & Morris, J. (2001). Food safety risk: Consumer perception and purchase behaviour. *British Food Journal, 103*(3), 170-187.

Young, S. M., McDonald, S., Edwards, H. M., et al. (2001). *Quality and people in the development of situationally specific methods.* Paper presented at the Second Asia-Pacific Conference on Quality Software 2001.

Appendix A: Letter to interviewees

Food Safety MCDA participant info-sheet

Dear potential participant

My name is Alex Dunn, a PhD student from the University of Canterbury's Management department. I am conducting my thesis in collaboration with Environmental Science and Research (ESR), studying methods of evaluating potential food safety interventions. In particular I am interested in Multi-Criteria Decision Analysis (MCDA) frameworks and I am looking at assembling a high level MCDA approach which includes all relevant criteria from a range of perspectives. My primary research question for this thesis is:

Key Research Question: How can food safety interventions be systematically evaluated using a framework which incorporates both quantitative data and stakeholder input, and which provides a basis for decision-making?

This research aims to construct a practical MCDA framework that includes criteria which are relevant from a range of stakeholder perspectives (government, industry, consumers, Māori etc.). Two crucial steps in this project will be to a) decide on a suitable set of criteria that can capture the range of potential impacts associated with a potential intervention, and b) decide on how to best measure (or estimate) the performance of interventions on each criterion. This research will be conducted over the next 12-18 months and the framework(s) will be tested on one or two food-pathogen case studies (probably norovirus in shellfish and/or *Campylobacter* in poultry at this stage).

You have been identified as a food safety expert and I am writing to ask whether you would be willing to participate in this research and provide your expert opinion regarding the formulation of some aspects of this MCDA framework. As indicated, the two main research questions I seek advice on answering have to do with the generation and measurement of the criteria to be used in the MCDA framework. However, additional to this, I am seeking to gain a clear understanding of current decision-making processes. I have included a brief list of questions at the end of this document designed to help me further understand this topic and to help shape the development of the MCDA framework. I thank you for taking the time to read this information sheet and for considering participating in this research. If you have any questions regarding any aspects of this research please feel free to contact myself or my supervisors.

Appendix B: Outline interview questions

MCDA framework: Outline interview questions

When evaluating and selecting potential food safety interventions/strategies, the direct financial cost and effectiveness of an intervention are usually two important criteria considered, but there may be other criteria involved in the decision-making process also.

1. What criteria would you consider relevant when evaluating potential food safety interventions?
2. How would you go about 'measuring' the performance of an intervention on each criterion? (I.e. what key data would you require for each criterion?)
3. Are there any additional criteria which should be included in these evaluations but are often excluded?
   a. Why are they excluded?
   b. How could/should these extra criteria be measured or estimated?
4. Which criteria (if any) are the most important in your opinion? (I.e. which should be given the most weight in the final decision?)
5. Are there any practical/feasible ways to improve current decision-making processes (from your perspective)?
6. Whose responsibility is food safety in NZ?
7. Who are the major stakeholders when it comes to food safety decision-making in NZ? (i.e. who should be included/consulted in the decision-making process?)
8. In your opinion, what is the most crucial part(s) of any intervention evaluation?

9. Is there anyone else within or outside of your organisation/industry who you think I should also talk to?

Thank you very much for your participation!

Appendix C: Key search terms used in literature search

The following key search terms were used during database searches for relevant MCDA, Cost-Effectiveness Analysis, food safety and stakeholder literature

AHP + Food Safety / Health Technology Assessment

Campylobacter + Cost-Effectiveness

Decision-making + Stakeholder

Food Safety + Decision Analysis

Food Safety + Decision Analysis + Stakeholder

Food Safety + Decision-making

Food Safety + Decision-making + Stakeholder

Food Safety + Evaluation

Food Safety + Evaluation + Stakeholder

Food Safety + Health Technology Assessment

Food Safety + Intervention + Assessment

Food Safety + Intervention + Evaluation

Food Safety + Stakeholder

Linear Additive Model + Food Safety / Health Technology Assessment

MAUT + Food Safety / Health Technology Assessment

MCDA / MCDM + Food Safety + Criteria

MCDA / MCDM + Health + Criteria

MCDA / MCDM / Cost-Effectiveness + Food Safety

MCDA / MCDM / Cost-Effectiveness + Guidelines

MCDA / MCDM / Cost-Effectiveness + Health

MCDA / MCDM / Cost-Effectiveness + Review

MCDA / MCDM / Cost-Effectiveness + Stakeholder

MCDA / MCDM / Cost-Effectiveness + Stakeholder Identification

Multi-Criteria Decision + Food Safety

Multi-Criteria Decision + Food Safety + Criteria

Multi-Criteria Decision + Food Safety + Intervention

PROMETHEE + Food Safety

Stakeholder Analysis

Stakeholder Analysis + Food Safety

Stakeholder Classification

Stakeholder Identification

Weighted Sum Model + Food Safety / Health Technology Assessment

Appendix D: Campylobacter Cost-Effectiveness paper

Publication removed due to copyright restrictions

Lake, Robin J., et al. "Cost-effectiveness of interventions to control Campylobacter in the New Zealand poultry meat food supply." *Journal of Food Protection* 76.7 (2013): 1161-1167. doi:10.4315/0362-028X.JFP-12-481