The Concept of Evidence in Health Technology Assessment (HTA)

by

Anna Stoklosa

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Institute for the History and Philosophy of Science and Technology

University of Toronto

© Copyright by Anna Stoklosa 2013
Abstract

This dissertation aims to make four contributions to knowledge. First, unlike other fields (e.g. science, medicine, etc.), the field of health technology assessment (HTA), has not thus far greatly benefited from a conceptually rigorous, philosophical analysis. Thus, the overall, and first, contribution of this dissertation is to begin to fill in some of this gap.

Second, because cross-jurisdictional comparisons and analyses are uncommon in the HTA literature, the dissertation devises a heuristically useful scheme for this purpose, and subsequently utilizes that scheme to compare Canada’s, Australia’s and UK’s approaches to HTA at the national level (focusing on CADTH, MSAC and NIHR, respectively). This analysis reveals that although HTA is claimed to be an evidence-based endeavour, the agencies’ assessments rely on a variety of probative inputs, raising the question of how, exactly, these bodies understand the concept of ‘evidence’?

An analysis of the HTA literature, as well as of the publications of the HTA agencies themselves, reveals that HTA, unlike other evidence based fields (such as evidence-based medicine (EBM) and evidence-based health policy (EBHP)), has thus far failed to flesh out an account of evidence. Thus, as a third contribution to knowledge, an account of evidence that is suitable for HTA purposes is proposed here. Evidence is argued to be: available data that is
relevant to the issue being addressed, question asked, decision made, etc. It is further argued that this definition enjoys numerous benefits, whilst simultaneously bypassing some of the problems around the definitions of evidence proposed in the EBM and EBHP literature.

The tenability and utility of the proposed definition is tested twice. First, the definition is tested against CADTH’s, MSAC’s and NIHR’s evaluative practices vis-à-vis health technologies in general, and second, the definition is tested against the agencies’ evaluative practices vis-à-vis genetic tests in particular. Thus, as a fourth contribution, the dissertation addresses genetic tests as an HTA issue, rather than as a bioethics issue, as it is predominantly done in the literature. In case of both the assessments of health technologies generally speaking, and assessments of genetic tests in particular, it is argued that the proposed definition of evidence is both tenable and useful, as it captures all and only the probative inputs that are required by the health technological assessments.
Acknowledgements

I owe an enormous debt of gratitude to David Castle, who has generously supervised this dissertation, and for many years now has – equally generously – also been my mentor, lending his moral and academic support, offering his time and wisdom, and being the kind of academic I aspire to be. Thanks are also due to the other members of the ‘Triumvirate’ – Ross Upshur and Paul Thompson – whose thoughts have helped to improve the dissertation, and to Mita Giacomini who served as my External Examiner and provided invaluable suggestions both on the dissertation’s contents and for future research.

I wrote the dissertation whilst a Visiting Postgraduate Student at the Sydney Centre for the Foundations of Science and living at the Women’s College, both at the University of Sydney. At SCFS, I would particularly like to thank Paul Griffiths and Karola Stotz for their kindness, support, and academic inspiration. At the Women’s College, I am particularly grateful to Jane Williamson and Bob Cope, Tiffany Donnelly, Mary McGuirk, Mary Assunta, and Tiffanny Junee, for their ‘typical Aussie’ friendliness and hospitality, wisdom on things academic and otherwise, encouragement, and many lovely meals and conversations. Special thank you is due Jane and Bob for their periodic ‘loans’ of Tigger for dissertation de-stressing purposes, and their seemingly endless supply of chocolate and wine, which also helped in that regard.

Thanks are also due to my friends and colleagues at the Systematic Reviews and Health Technology Assessment Unit, University of Sydney, and the Medical Services Advisory Committee, Department of Health and Ageing, Australia. In particular, gratitude is owed to Sally Wortley, Melina Willson, Toby Gould, Samara Lewis and Sally Lord at the University of Sydney, and Sally Gosewinckel at the Dept. of Health and Ageing, from whom I have learnt an enormous amount about systematic reviews, health technology assessment, and health policy.

Finally, for their unwavering moral support, encouragement, patience and friendship, I would like to thank Melanie Coughlin and Ola Stoklosa.
Table of Contents
The Concept of Evidence in Health Technology Assessment (HTA) ........................................ ii
Abstract .................................................................................................................................... ii
Acknowledgements .................................................................................................................. iv
Table of Contents ................................................................................................................... v
List of tables ............................................................................................................................. viii
CHAPTER 1: Health Technology Assessment: An Introduction ............................................... 1
1.1. Conceptual background ..................................................................................................... 2
  1.1.1. ‘Technology’ and ‘health technology’........................................................................ 2
  1.1.2. ‘Technology Assessment’ and ‘Health Technology Assessment’............................... 4
1.2. HTA: The Past and the Present ...................................................................................... 6
  1.2.1. Factors precipitating the development of HTA .......................................................... 6
  1.2.2. Key milestones in the development of the field of HTA ............................................. 7
  1.2.3. HTA at present .......................................................................................................... 10
  1.2.4. HTA methodology and evaluata .............................................................................. 12
1.3. Conclusion ...................................................................................................................... 14
CHAPTER 2: Canada, Australia and the United Kingdom: Health Systems and Approaches to HTA .................................................................................................................. 15
2.1. Why Canada, Australia and the United Kingdom? ........................................................ 16
2.2. Canada ............................................................................................................................ 16
  2.2.1. Canada’s Health System ......................................................................................... 16
  2.2.2. Health Technology Assessment in Canada .............................................................. 21
2.3. Australia .......................................................................................................................... 26
  2.3.1. Australia’s Health System ....................................................................................... 26
  2.3.2. Health Technology Assessment in Australia ............................................................ 30
2.4. United Kingdom ............................................................................................................ 34
  2.4.1. United Kingdom’s Health System ............................................................................ 34
  2.4.2. Health Technology Assessment in the United Kingdom ......................................... 38
2.5. Conclusion ...................................................................................................................... 43
CHAPTER 3: Comparing the 3 Jurisdictions’ Approaches to HTA ........................................... 44
3.1. Why a taxonomy? ............................................................................................................ 45
  3.1.1. Comparing HTA approaches across jurisdictions? .................................................... 45
  3.1.2. Why the field of HTA needs a taxonomy of approaches .......................................... 46
3.2. A general taxonomy of approaches to HTA ................................................................. 48
3.3. Comparing CADTH, MSAC and NIHR ........................................................................ 52
3.4. Detailed comparison of CADTH, MSAC and NIHR ..................................................... 53
  3.4.1. Comparison of Criterion 1: Centralisation ................................................................. 53
  3.4.2. Comparison of Criterion 2: Funding model ............................................................... 54
  3.4.3. Comparison of Criterion 3: Level ............................................................................ 55
  3.4.4. Comparison of Criterion 4: Scope ........................................................................... 56
  3.4.5. Comparison of Criterion 5: Method ......................................................................... 57
  3.4.6. Comparison of Criterion 6: Evaluata ...................................................................... 57
  3.4.7. Comparison of Criterion 7: Policy Advice ............................................................... 59
  3.4.8. Comparison of Criterion 8: Influence on funding .................................................... 60
3.5. Conclusions .................................................................................................................... 61
CHAPTER 4: Conceptualising ‘Evidence’ in HTA ................................................................. 62
CHAPTER 5: Testing the Proposed Concept of Evidence – Part I ........................................... 101
5.1. Canada: CADTH’s evaluation of health technologies .................................................. 102
  5.1.1. CADTH’s health technology assessment process .................................................. 102
  5.1.2. Method for establishing CADTH’s understanding of ‘evidence’ .......................... 102
  5.1.3. How CADTH understands ‘evidence’ ................................................................. 104
5.2. Australia: MSAC’s evaluation of health technologies .................................................. 107
  5.2.1. MSAC’s health technology assessment process .................................................. 107
  5.2.2. Method for establishing MSAC’s understanding of ‘evidence’ ............................ 107
  5.2.3. How MSAC understands ‘evidence’ ................................................................. 109
5.3. UK: NIHR’s evaluation of health technologies ............................................................ 111
  5.3.1. NIHR’s health technology assessment process .................................................. 111
  5.3.2. Method for establishing NIHR’s understanding of ‘evidence’ ............................. 112
  5.3.3. How NIHR understands ‘evidence’ ................................................................. 114
5.4. ‘Evidence’ in HTA: theory vs. practice ....................................................................... 115
  5.4.1. Comparison of the three HTA programmes’ practices ....................................... 116
  5.4.2. How do the 3 agencies conceptualise ‘evidence’? .............................................. 120
  5.4.3. Advantages of adopting the proposed conceptualisation of evidence ............... 125
  5.4.4. Mapping the agencies’ evidentiary practices onto the proposed conceptualisation of evidence .............................................................................................................................................................................. 125
5.5. Conclusion .................................................................................................................. 128

CHAPTER 6: The Road to Genetic Testing ............................................................................. 130
6.1. History of modern genetics .......................................................................................... 131
  6.1.1. Gregor Mendel ....................................................................................................... 131
  6.1.2. The Classical Period in Modern Genetics ............................................................ 135
  6.1.3. Molecular Genetics Period .................................................................................. 137
6.2. The Human Genome Project ....................................................................................... 141
  6.2.1. The Project Timeline ............................................................................................ 141
  6.2.2. The formal goals and the hopes for the Project .................................................. 143
List of tables

Table 1: Comparing CADTH, MSAC and NIHR ................................................................. 52
Table 2: Scan of the International Journal of Technology Assessment in Health Care .......... 66
Table 3: Scan of CADTH’s 10 most recent HTA reports for definitions of the term ‘evidence.’ ................................................................. 69
Table 4: Scan of MSAC’s 10 most recent HTA reports for definitions of the term ‘evidence.’ ................................................................. 69
Table 5: Scan of NIHR’s 10 most recent HTA reports for definitions of the term ‘evidence.’ 70
Table 6: CADTH’s 10 most recent HTA reports: evaluata .................................................. 103
Table 7: MSAC’s 10 most recent HTA reports: evaluata ..................................................... 108
Table 8: NIHR’s 10 most recent HTA reports: evaluata ...................................................... 113
Table 9: HTAs of genetic tests carried out by CADTH .......................................................... 165
Table 10: CADTH: evaluata considered in HTAs of genetic tests (ACCE nomenclature) .... 168
Table 11: CADTH: evaluata considered in HTAs of genetic tests (standard HTA nomenclature) ................................................................. 170
Table 12: HTAs of genetic tests carried out by MSAC .......................................................... 172
Table 13: MSAC: evaluata considered in HTAs of genetic tests (ACCE nomenclature) ...... 175
Table 14: MSAC: evaluata considered in HTAs of genetic tests (standard HTA nomenclature) ................................................................. 177
Table 15: HTAs of genetic tests carried out by UK GTN ........................................................ 181
Table 16: UK GTN: evaluata considered in HTAs of genetic tests (ACCE nomenclature) .... 184
Table 17: The agencies’ understanding of ‘evidence’ in HTAs of genetic tests ...................... 189
Table 18: Agencies’ conceptualisation of evidence: in terms of general HTA categories ...... 193
Table 19: Agencies’ conceptualisation of evidence: in terms of ACCE nomenclature .......... 194
CHAPTER 1: Health Technology Assessment: An Introduction

Summary: This chapter provides an introduction to the field of Health Technology Assessment (HTA). The chapter is divided into two parts: Part I reviews the HTA nomenclature, whilst Part II builds on this nomenclature to describe the factors which spurred the development of HTA, the field’s brief history, its aim, methodology, and evaluative practices. The primary aim of this chapter is to provide the necessary background for chapter two’s discussion of the HTA programmes in Canada, Australia and the United Kingdom.
1.1. Conceptual background

1.1.1. ‘Technology’ and ‘health technology’

The term ‘technology’ has its roots in the Greek word, ‘technologia’, which is a compound of ‘techne’ (a craft) and ‘logia’ (a saying) (Banta, 2009). Although a variety of definitions of ‘technology’ have been put forth in the literature, at the most general level, technology can be understood to refer to “[the] use and knowledge of tools and crafts and how [their] use affects the ability to control and adapt to the social and physical environment” (Banta, 2009, 34). Beyond the term’s general use, ‘technology’ can also be applied more specifically to particular areas, such as telecommunications technologies, information technologies, construction technologies, health technologies, and so forth. It is around the last type of technologies – health technologies – that the field of health technology assessment has grown.¹

Definitions of ‘health technologies’ put forth in both the academic and grey literature fall on a spectrum that ranges from fairly narrow to rather quite broad (Velasco Garrido et al 2010). Understood most narrowly, ‘health technologies’ are “drugs, devices and procedures that can be provided within the health care system² as it delivers health services.” Somewhat more broadly, health technologies may additionally encompass interventions that are applied to the health systems themselves – for example, changes that aim to increase access, modifications to health care provider payment procedures, the nature of the particular service delivery model adopted by a health system framework, and so on. Finally, at their broadest, ‘health technologies’ may also include health-promoting interventions that take place outside of the health care system, such as educational services or social services (Velasco Garrido et al, 2010).

No consensus currently exists on where, precisely, to locate the conceptual boundaries of ‘health technologies’ for the purpose of health technology assessment (henceforth, HTA) –

¹ Prior to the 1990s, the term ‘health technology’ (and its cognates), was used interchangeably with ‘medical technology’ (and its cognates). After the 1990s, however, the ‘health technology’ discourse became dominant (Banta, 2009). Consequently, the terms ‘health technology’ and its cognates will be preferred here.
² Various definitions of ‘health systems’ have been put forth. The one that is relied on here is one stipulated by Velasco Garrido et al, who define health systems as “the arrangements, individuals, and institutions through which personal health services are provided, organised, and controlled” (Velasco Garrido et al, 2010, 197)
HTA programmes in different jurisdictions adopt varying definitions of ‘health technologies.’ For example, the United States’ Office of Technology Assessment, which is credited with formulating one of the first definitions of health technologies, defined health technologies3 in 1976 as:

[T]he set of techniques, drugs, equipment, and procedures used by health-care professionals in delivering medical care to individuals and the systems within which such care is delivered (Office of Technology Assessment, 1976).

This definition subsequently underwent slight modifications and, in 1978, ‘health technologies’ became: “the drugs, devices and medical and surgical procedures used in medical care, and the organisational and supportive systems within which such care is provided” (Office of Technology Assessment, 1978). OTA’s 1978 definition of health technologies had gained a lot of currency, and several HTA programmes continue to use this definition even today. These include, for example, the United Kingdom’s National Institutes for Health Research HTA programme (National Institute for Health Research (NIHR), 2009a), EUR-ASSESS, which is a project aiming to coordinate HTA activities in Europe (EUR-ASSESS Steering Committee, 2009), and the Ontario’s Ministry of Health and Long-Term Care (Ministry of Health and Long-Term Care (Ontario), 2009), which is one of the provincial-level HTA programmes in Canada.

Canada’s federal agency for HTA – Canadian Agency for Drugs and Technologies in Health (CADTH) – on the other hand, recently defined ‘health technologies’ more elaborately to mean:

any method or intervention that is used to promote health; prevent, diagnose, or treat disease; or improve rehabilitation and long-term care. Technologies include drugs, devices, diagnostic agents, equipment, and medical and surgical procedures. The definition also includes organisational and service systems that provide health care, such as tele-health. (Canadian Agency for Drugs and Technologies in Health, 2011)

More concisely, yet also more broadly, the Australian government, finally, recently defined health technologies as “all innovations in the provision and arrangement of health care” (Government of Australia, 2009a, 34). The breadth of the definitions of health technologies adopted by the various HTA programmes is therefore quite varied – although none of the

---

3 The definition cited here was given for ‘medical technologies’ not ‘health technologies’ (see fn 1).
definitions appear to construe ‘health technologies’ in what Velasco Garrido et al consider to be their broadest sense.

1.1.2. ‘Technology Assessment’ and ‘Health Technology Assessment’

The focus of the field of health technology assessment, or HTA, is on evaluating health technologies. Before ‘health technology assessment’ was born, however, its slightly older sibling – ‘technology assessment’ – was first conceptualised. The United States’ Office of Technology Assessment (OTA) is credited with offering one of the first formal definitions of technology assessment, when, in 1976, it defined technology assessment as “a comprehensive form of policy research that examines the short- and long-term social consequences of the application or use of technology” (Office of Technology Assessment, 1976).

Nearly four decades later, OTA’s definition of technology assessment continues to enjoy some popularity among health technology assessments programmes; its vestiges are evident in the definitions of health technology assessment adopted by EUR-ASSESS, which stipulates that HTA is “a form of policy research that systematically examines short- and long-term consequences of the application of a health technology, a set of related technologies, or an issue related to technology” (EUR-ASSESS Steering Committee, 2009, 10). Canadian Agency for Drugs and Technologies in Health, on the other hand, defined health technology assessment more elaborately as an “evaluation of the clinical effectiveness, cost-effectiveness, and broader impact of drugs, medical technologies, and health systems, both on patient health and the health care system” (Canadian Agency for Drugs and Technologies in Health, 2011). Government of Australia construes HTA as a “field of policy analysis studying the medical, economic, social and ethical implications of development, diffusion and use of health technology” (Government of Australia, 2009a).4

It is evident even from this very brief scan of HTA definitions currently in use by the various jurisdictions, that OTA definition’s focus on the consequences or impacts of adopting the technologies has been retained. The one modification that is readily apparent in the more recent definitions, however, is the greater specificity with regard to which particular types of

---

4 United Kingdom’s HTA programme draws a very interesting – and rarely utilised elsewhere – distinction between health technology assessment and health technology appraisal; it will be discussed in greater detail in chapter 2.
consequences are evaluated. Whereas OTA stipulated as its evaluata the social consequences of technologies – a very broad and rather vague category – the contemporary definitions tend to be more precise, emphasising specifically the economic consequences or medical consequences, for example. The categorisation of HTA as a field of policy research⁵ seems also to have been largely retained from the OTA’s definition – it is certainly manifest in both the EUR-ASSESS and the Australian Government’s definitions of HTA. Although CADTH’s definition of HTA does not emphasise the identity of HTA as a field of policy analysis, CADTH itself does identify policy input as one of HTA’s functions. In short, the core of OTA’s definition – the emphasis on HTA as a field of policy research focused on consequences of adoption of health technologies – seems largely to have been retained.

How broadly – or narrowly – a health technology assessment programme construes ‘health technologies’ should correlate with the evaluative focus of a particular health technology assessment programme. However, although the theoretical definitions of health technologies adopted by the various HTA programmes are moderately broad, in practice, the focus of HTA programmes tends to be fairly narrow (Battista and Hodge, 2009). The first HTA programme – OTA’s – for example, adhered much more closely to the narrower end of the spectrum of understanding what health technologies are, than to the moderate one: its first object of a health technology assessment was a CT scanner, and this assessment was then followed by an evaluation of several other pieces of large and expensive medical equipment (Banta, 2003a).⁶

There is some evidence that the spectrum of technologies that are evaluated by HTA programmes is broadening somewhat. Banta, for example, reports that several programmes have more recently evaluated nursing care, physiotherapy programmes, and mental health programmes (Banta, 2003a). Nevertheless, those continue to be isolated cases; ‘health technologies’ – understood in the broadest sense – generally remain an infrequent subject of HTAs (Velasco Garrido et al, 2010). Contemporary HTA programmes continue to predominantly focus their assessment efforts on new, expensive equipment, pharmaceuticals

---

⁵ The relationship between the fields of HTA and evidence-based health policy will be discussed in more detail in chapter 4.

⁶ Although a narrow construal of the evaluata of HTA is consistent with HTA’s task – to evaluate health technologies – a broader construal may be preferable here, since, as Oliver et al argue, clinical and medical services have less influence on health than other determinants of health, such as nutritional, environmental and biological factors (Oliver et al, 2004).
and clinical procedures (Oliver et al, 2004; Walley, 2007; Velasco Garrido et al, 2010). This is particularly evident in the recent overview of the Health Technology Process in Australia, authored by the Australian Government, which, although defining ‘health technologies’ quite broadly – as was noted above – emphasises the evaluation of diagnostic and therapeutic goods and services, such as prostheses, diagnostic tests, medical procedures and devices (Government of Australia, 2009a). Thus, although ‘health technologies’ may be – and often are – conceptualised moderately broadly, in practice, HTA programmes tend to focus their attention closer to the narrow end of the definitional spectrum.

1.2. HTA: The Past and the Present

1.2.1. Factors precipitating the development of HTA

What, however, precipitated all of this effort toward conceptualising ‘health technologies’ and ‘health technology assessments’, in the first place? Historically-speaking, health technology assessment programmes are a relatively new phenomenon. The field’s relative youth is not surprising, given the relative youth of effective health technologies themselves. As Banta notes, while health technologies as such enjoy a long history, effective health technologies are a fairly new development (Banta, 2003a). As recently as 1927, only about three percent of medical remedies and therapies constituted effective treatment or prevention (Beeson, 1980). Less than a century later, health technologies have come to constitute an effective – and indispensable – component of health systems worldwide, functioning in areas ranging from prevention, diagnosis, and treatment, to disability alleviation (World Health Organisation, 2010). Australia’s Productivity Commission – an independent advisory body to the Australian Government – for example, recently estimated that approximately one-half of the increase in length and improvement in the quality of life may be ascribed to medical innovations (Productivity Commission, 2005).

It can be – and has been – argued that the field of health technology assessment arose as part of the more general climate of criticism towards sciences and technologies which dominated the discourse in the 1960s and the 1970s (ten Have, 1995; see also Florman, 1996). This view, however, is not a common one among the scholars of HTA; the majority credit factors closely
Health technologies do contribute quite heavily to the ever-increasing costs of health care systems. The aforementioned Productivity Commission, for example, estimates that health technologies account for approximately 1/3 of the increases in health expenditures between 1992 and 2002 (Government of Australia, 2009a). It is precisely the concerns about this issue – the fiscal impact of health technologies on health systems – that are commonly credited with spurring the development of the field of Health Technology Assessment (Office of Technology Assessment, 1980; Banta, 2003a; Oliver et al, 2004; McGregor and Brophy, 2005; Petherick et al, 2007; Walley, 2007).

However, while concerns about the fiscal impact of health technologies on health systems are the dominant factor credited here, it was, by no means, the only factor contributing to the field’s birth and development. Other factors that played a role in the birth and ongoing development of the field include: the visibility of new technologies (Banta and Luce, 1993), concerns about clinical effectiveness about the technologies already integrated into health systems and being used by health care providers (Oliver et al, 2004; Stevens and Milne, 2004), a desire to de-politicise the decisions about allocation of health care budgets (Battista and Hodge, 2009), and a commitment to evidence-based practices (Walley, 2007).

1.2.2. Key milestones in the development of the field of HTA

As these factors began to receive ever-increasing attention from governments and policymakers, HTA programmes gradually began to come into being (Banta, 2003a; Oliver et al, 2004; Walley, 2007; Battista and Hodge, 2009). Banta divides the – admittedly, brief – history of HTA into three key periods. The early period begins with the inception of HTA in the 1970s (Saarni et al, 2008; Banta and Jonsson, 2009). This period in the field’s development is generally agreed to have begun in the United States, although there is no consensus on the field’s precise birth date. Some date the beginnings of HTA to the publication of reports on the implications of several health technologies in 1973 by the United States’ Academy of Science and the National Institutes of Health; others place HTA’s beginnings either at the launch of the
United States’ Office of Technology Assessment in the early 1970s, or in 1976 – the date of OTA’s first formal report (Banta, 2003a).

Lack of consensus about the precise beginnings notwithstanding, the early period of the field’s development is characterised by its focus: predominantly, on large and expensive machine-based technologies. The primary evaluata of health technological assessments at that time were the health technologies’ cost-effectiveness and efficacy (Banta, 2003a), which is consistent with the claim that economic considerations were one of the main factors behind the development of the field. Although it is the United States that is said to be the birthplace of HTA (Banta, 2003a), the international spread of HTA began shortly after its inception; the HTA activity at the federal level in Australia, for example, began with the establishment of the National Health Technology Advisory Panel in 1982 (Hailey, 2009).

The second period in HTA’s history – what Banta labels the ‘middle period’ – began in 1985. During this time, the HTA began to acquire more prominence in the United States, and continued to spread internationally. It was during this period, for example, that the World Health Organisation’s European Office published a document urging its member states to establish “a formal mechanism to systematically assess the appropriate use of health technologies and to verify that they respond to the national health program needs” (World Health Organisation, 1985). It is also during the second period that the first Canadian HTA agencies were established: the first one, at the provincial level in Quebec in 1988, and shortly thereafter, the federal HTA agency was established in 1989 (Roehrig and Kargus, 2003; Battista et al, 2009; Menon and Stafinski, 2009).

The third period in the development of HTA began in the late 1990s and continues to present day. This period is characterised by an emphasis on increasing both the general importance of HTA bodies and their activities, and enhancing their level of influence vis-à-vis policy decisions (Banta, 2003a). It is worth noting that although the spread of HTA internationally is continuing, not all nations have an HTA programme as of yet – and where HTA activity is evident, its level is uneven across jurisdictions. For example, while nearly all members of the European Union now have an HTA programme, the HTA activity is significantly more pronounced among the Western European Members of the EU than among its Eastern and
Central European Members (Banta, 2003a). Some HTA activity has recently began to take root in Latin America, Asia and the Middle East (Banta and Jonsson, 2009), but not all members of these jurisdictions have an HTA programme at present; HTA activity in Africa remains minimal to non-existent (Sivalal, 2009).

Moreover, while overall more and more HTA activity is taking place – at least in some jurisdictions – it is often recognised that it is unclear whether – and if so, to what extent – these HTA activities actually do affect policy decisions (Stevens and Milne, 2004; McGregor and Brophy, 2005). Several scholars and policy-makers have recently concluded that the available evidence about HTA’s influence is rather discouraging (Romanow, 2002; Oliver et al, 2004; McGregor and Brophy, 2005). This conclusion also seems to underlie Banta and Jonsson’s recent claim that the levels of influence and impact of HTA remain among the major challenges for the field’s future (Banta and Jonsson, 2009). Although one may suspect that political factors play a dominant role in diminishing the levels of HTA’s influence, the low influence is at least partly due to the finite fiscal and temporal resources under which HTA agencies operate. The United Kingdom’s HTA programme, for example, evaluates only approximately 40-50 health technologies annually (NIHR, undated). Time and money have been – and will likely continue to be – an impediment to HTA’s acquiring substantial policy influence (Oliver et al, 2004).

Nevertheless, at least in several specific areas, HTA’s influence appears to be fairly substantial. For example, health technological assessments are acknowledged to have a significant impact on decisions about funding and reimbursement for services and products in several jurisdictions (Banta, 2003a; Velasco Garrido et al, 2010). This is because in those jurisdictions, the use of HTA is mandated by law – this is the case in both France and Germany, for example (Velasco Garrido et al, 2010). Similarly, health technology assessments are now required to be used by Australia’s major federal-level HTA programmes (namely, the Pharmaceutical Benefits Advisory Committee and the Medical Services Advisory Committee), as well as by the Health Technology Programme of the United Kingdom’s National Health Service (Gallego et al, 2009).
1.2.3. HTA at present

As this suggests, the main aim of HTA is to influence health policy. This is, in fact, quite consistently recognised in the literature as the key function of health technology assessments. Thus, for example, Banta states that the primary aim of HTA is to affect decisions in various health policy areas (Banta, 2003a; see also Banta, 2009) and McGregor, similarly, notes that the purpose of HTA is “to inform health policy decisions” (McGregor and Brophy, 2005, p. 263). This claim is frequently echoed in the academic literature (see, e.g., ten Have, 1995; McGregor and Brophy, 2005; Walley, 2007; Saarni et al, 2008; Velasco Garrido et al, 2010), and those involved with the HTA process are in agreement with the scholars on this point. That the purpose of HTA is to provide input into health policy decisions was, for example, the stated aim of the United States’ Office of Technology Assessment in the early 1970s (Banta and Perry, 1997). EUR-ASSESS, also, has espoused this view as far back as 1997 (EUR-ASSESS, 1997a) and has recently reiterated its commitment thereto (EUR-ASSESS Steering Committee, 2009). United Kingdom’s HTA programme (Walley, 2007), Canada’s CADTH (Canadian Agency for Drugs and Technologies in Health, 2011), and the Australian Government’s HTA (Government of Australia, 2009a) have also endorsed it.

Consequently, a very broad range of health policy decisions can be – and has been – affected at least to some extent by health technology assessments. Some of these decisions have included, for example: decisions about whether to continue or de-fund a particular programme, decisions about changes to clinical-level guidelines, technology investment and disinvestment decisions, decisions about the types and levels of health human resources required by an adoption of a new technology, decisions about provision or rescission of coverage for a particular health technology, and cost-containment decisions (Banta, 2003a; Roehrig and Kargus, 2003), to name but a few.

In light of HTA’s purpose – to provide input into health policy decisions – it is unsurprising that the key audience for HTAs are the health policy-makers at various levels of government, including federal, provincial (or its equivalent in other jurisdictions) and local levels, as well as the regional health authorities and health care institutions (Roehrig and Kargus, 2003; Hailey,
One might also expect that, since the function of HTA is to produce health policy input, the preponderance of HTA-producing activity takes place at a variety of governmental organisations – bodies either constitutive of or affiliated with the national ministries of health (Banta and Jonsson, 2009). Indeed, a lot of HTAs are produced by entities that match this description. Among these, one may include, for example: the United Kingdom’s National Institutes for Health Research HTA programme (Walley, 2007), as well as Australia’s three federal HTA bodies – Pharmaceutical Benefits Advisory Committee, Medical Services Advisory Committee and the Prosthesis and Devices Committee (Government of Australia, 2009a). While not a part of Canada’s federal health department – that is, Health Canada – the Canadian Agency for Drugs and Technologies in Health may arguably also be counted here, insofar as it is funded by the federal and provincial governments of Canada (Menon and Stafinski, 2009).

The production of HTAs, however, is not circumscribed only to those types of entities – in other words, health technology assessment activity also takes place at what can be collectively called the ‘sub-national level’ bodies. Several provincial-level HTA bodies exist in Canada, including, for example, Quebec’s L’Agence d’Évaluation des Technologies et des Modes d’Intervention en Santé and Ontario’s Medical Advisory Secretariat (Roehrig and Kargus, 2003; Hailey, 2007). Sub-national level of HTA activity is not particularly widespread in Australia – in part due to the health system’s centralised structure – but inchoate efforts include the Policy Advisory Committee on Clinical Practice and Technology in the state of Victoria (Bulfone et al, 2009). Some HTA activity is also evident at the level of regional health authorities, hospital associations or even individual hospitals in Canada, Australia and several other jurisdictions (McGregor and Brophy, 2005; Gallego et al, 2009). A variety of other sub-
national entities also conduct HTA activities. These include: funding agencies, universities, manufacturers and industry, aid agencies, insurance providers and professional organisations (such as the American Medical Association) (Banta, 2003a).

1.2.4. HTA methodology and evaluata

As was noted above, the function of HTA is to provide input into health policy. HTA, however, is not just any form of policy input – it is an evidence-based form of policy input.10 The Australian Government, for example, in its recent evaluation of the state of HTA activities in Australia, noted that its HTA programme “conducts evidence-based assessment” (Government of Australia, 2009a, emphasis added) and Canada’s CADTH similarly described its primary purpose as providing “decision makers with credible, impartial, and evidence-based information on health technologies” (Canadian Agency for Drugs and Technologies in Health, 2011, emphasis added). The United Kingdom’s HTA programme was created specifically “because of a commitment by the [National Health Service] to evidence-based practice” and, as Walley points out, its aim is to “produce robust scientific assessments” (Walley, 2007, 283, emphases added). EUR-ASSESS, too, holds that HTA must be “based on scientific evidence” (EUR-ASSESS Steering Committee, 2009, 10, emphasis added). Scholars often echo this point (see, e.g., ten Have, 1995; Saarni et al, 2008; Banta and Jonsson, 2009).

The evaluata11 assessed in the health technology assessment process can, in theory, span a fairly broad spectrum, including some or all of the following: economic considerations, scientific data (often in the form of randomised control trials), social and legal aspects, ethical considerations, organisational considerations (Saarni et al, 2008), as well as expert opinion, political judgements, and stakeholder and patient views (Velasco Garrido et al, 2010). In practice, however, HTA programmes are frequently charged with focusing on an unduly narrow sub-set of evaluative considerations (ten Have, 1995; DeJean et al, 2009; Banta,

---

10 Occasionally, HTA is even said to be a science or like a science although what is meant by the claim usually is that HTA is an evidence-based process (See, e.g., Roehrig and Kargus, 2003; Walley 2007). The concept of evidence underlying this claim will be fleshed out in chapter 4.

11 The term ‘evaluata’ is preferred here to the term ‘evidence’ so as not to beg the question. The issues around establishing an appropriate definition of the term ‘evidence’ in context of HTA will be discussed in chapter 4.
The charges seem well merited, as the documents and reports authored by the HTA programmes themselves suggest. Australian Government’s review of its HTA programmes, for example, emphasises safety, clinical effectiveness and cost-effectiveness (Government of Australia, 2009a), as does the recent EUR-ASSESS position paper (EUR-ASSESS Steering Committee, 2009). This narrow approach may in part reflect the priorities of the publicly funded health system model adopted by some jurisdictions, and need not be problematic with respect to all health technologies. Nevertheless, at least with respect to new biotechnologies – such as genetic testing or IVF – the omission of ethical and social considerations, for example, has been the subject of frequent criticism (see, e.g., ten Have, 1995; Banta, 2003a; Roehrig and Kargus 2003; Oliver et al, 2004; Banta, 2009).

On the other hand, a significant constraint on the breadth of the evaluata that can be considered by the various HTA programmes is imposed by the methodologies that they adopt. Because HTA programmes in different jurisdictions develop partly in isolation from each other (Banta, 2003a) and partly in response to local contexts (Dobrow, 2004; Hofmann, 2008), different methodological approaches have been adopted by in HTA programmes vis-à-vis treatment and procurement of the evaluata for the assessments. Typically, one of two approaches is adopted: research synthesis approach or primary research approach.

The research synthesis method (sometimes also referred to as a ‘systematic review’) tends to be the most commonly adopted methodology (Banta, 2003a; Hofmann, 2005; EUR-ASSESS Steering Committee, 2009; Banta and Jonsson, 2009; DeJean et al, 2009; Velasco Garrido et al, 2010). This approach involves the systematic review by the HTA programme of the already existing and available evidence, and generally involves database searches, literature reviews, reviews of the results of randomised controlled trials and meta-analyses, etc. (Roehrig and Kargus, 2003). While the research synthesis approach is quite amenable to assessing technologies about which voluminous data exists, not all health technologies meet this

---

12 The types of evaluata that are (or are not) considered may be contingent at least in part on the health system model within which the HTA programme is embedded, and on whose views are represented at the decision table. Banta, for example, notes that policy-makers and insurers tend to prioritise economic considerations, while health providers focus primarily on evidence of quality, industry members on efficacy and cost-effectiveness, and the general public on access and quality. (Banta, 2003a)

13 Although a common methodological approach may seem to be desirable, the different methodological approaches are often driven by the structure and the goals of particular HTA programmes and the particular health system frameworks within which they operate. More will be said about this in Chapter 2.
criterion – new biotechnologies such as nutrigenomics and genetic testing are likely to be especially problematic here.

An alternate methodological approach has therefore been adopted by several HTA programmes: the conduct of new empirical research. This approach typically involves the identification of existing knowledge gaps vis-à-vis the technology that is being assessed and collection and analysis of data in order to close those gaps. (Velasco Garrido et al, 2010, p.198). While this methodological approach is not commonly adopted by the HTA programmes (Walley, 2007), programmes in the United Kingdom, Netherlands, and Denmark, for example, have adopted it, with some good results (Velasco Garrido et al, 2010).

1.3. Conclusion

To summarise, the primary aim of this chapter was to offer an introductory overview of the field of Health Technology Assessment (HTA) and to establish the main elements of the conceptual framework required for the discussion of three specific HTA programmes in chapter 2. From the brief overview presented here, it is evident that although some commonalities are evident – with regard to the HTA’s purpose and its evidence-basedness, for example – there seems to be no consensus internationally with regard to many of the key elements of HTA. Jurisdictions conceptualise the field and its scope differently, and differences are evident with regard to HTAs’ methodological and evidentiary approaches. This is partly due to the various health system frameworks which those HTA programmes serve and within which they operate. In the next chapter, this will be highlighted via a discussion of the health system frameworks and approaches to health technology assessments at the national level in three jurisdictions: Canada, Australia and the United Kingdom.
Summary: This chapter provides an overview of the health systems, and the health technology assessment programmes that operate within those health systems, for three jurisdictions: Canada, Australia and the United Kingdom. The primary aim of this chapter is to describe in some detail the key elements of each jurisdiction’s approach to HTA. The secondary aim of this chapter is to lay groundwork for drawing out the similarities and contrasts between the three jurisdictions’ approaches to HTA at the national level.
2.1. Why Canada, Australia and the United Kingdom?

The jurisdictions chosen for comparison in this dissertation are: Canada, Australia and the United Kingdom.¹⁴ Three sets of reasons ground this selection, which can be categorised as: jurisdiction-specific, health system-specific and HTA approach-specific reasons.

Under the rubric of jurisdiction-specific considerations, the reason for the selection of these three countries in particular is largely pragmatic: all three jurisdictions are developed and English-speaking countries. This translates into, respectively, a wealth of data (about their respective health systems, approaches to HTA, etc.), and the accessibility of said data.

Health system-specific reasons for focusing on these jurisdictions primarily centre on their sharing the single-payer, universal approach to health care provision. Nonetheless, differences between the three are also evident. United Kingdom has a highly centralised public system (although a small parallel private health system is also in existence). Canada has a largely decentralised, public system – and private provision of health services is largely prohibited by law. Finally, Australia has a largely centralised, publicly funded health system, with a well-developed and substantial parallel private health system in operation.

Finally, HTA-specific reasons for selecting these jurisdictions are two-fold. First, all three jurisdictions have well-established HTA programmes. Second, these jurisdictions evince a fair degree of variety in their approaches to health technological assessment and therefore will yield interesting comparisons in chapter 3.

2.2. Canada

2.2.1. Canada’s Health System

2.2.1.1. System of Governance

Canada is a constitutional monarchy and a parliamentary democracy, with two orders of government: federal and provincial (or territorial) (Marchildon 2005). The legislative branch

¹⁴ It needs to be emphasised here that arguments made regarding these jurisdictions specifically are not intended to generalise to all jurisdictions (unless otherwise noted).
consists of the House of Commons (whose members are elected by popular vote) and the Senate (whose members are appointed by the Prime Minister). The Constitution consists of the Constitution Act of 1867, and the Constitution Act of 1982 (which added the Canadian Charter of Rights and Freedoms) (CIA 2011a). The Constitution documents outline how the responsibilities are divided between the federal and the 10 provincial and 3 territorial governments (Romanow, 2002).

2.2.1.2. A brief history of Canada’s health system

Prior to the 1950s, Canadian patients paid for medical and hospital services out of pocket, which resulted in increasing incidence of patients failing to seek care. The beginnings of the contemporary Canadian health care system can therefore be traced to 1954, when the Canadian Government addressed this trend by agreeing to reimburse part of the provinces’ costs of providing hospital care to residents under the Hospital Insurance and Diagnostic Services Act (HIDSA). The subsequent 1966 Medical Care Act (MCA) expanded the hospitalisation coverage to also include basic physician services, thereby resulting in a full coverage of required hospital and physician services under the single-payer publicly funded model. This model was further entrenched in the 1984 Canada Health Act, which superseded both the 1954 HISDA and 1966 MCA (Romanow 2002; Marchildon 2005)

The Canada Health Act, which was enacted unanimously by the Canadian Parliament in 1984, remains in place to the present (Kirby 2002). The Act elucidates the five key principles that define the Canadian healthcare system; these principles have now also become conditions that must be met by provinces and territories in order to receive the federal funds (Romanow 2002). The five principles are as follows:

1. **Universality**: public health care insurance must be provided to all Canadians and eligible Canadian residents
2. **Comprehensiveness**: medically necessary hospital and doctor services are included under the public health care insurance
3. **Accessibility**: financial barriers (such as, for example, co-pays) to accessing publicly funded health services are discouraged, so as to render health services available to all Canadians and eligible Canadian residents

---

15 At that point, several provinces (e.g. Saskatchewan, British Columbia) had already instituted hospital insurance (Marchildon 2005). The point here, however, is about the federal involvement in care.
4. **Portability**: all Canadians and eligible Canadian residents are covered by the public health insurance, regardless of whether they are traveling within Canada, internationally, or move between provinces/territories.

5. **Public administration**: provincial and territorial health care insurance plans are to be managed publicly on a non-profit basis (Canada Health Act 1984, Kirby 2002, Marchildon 2005).

### 2.2.1.3. Health System Funding

The system that operates within the constraints of these five principles is a publicly-funded, universal one. Public financing of the system enjoys strong support from the Canadian public – this was the case even during the period of government budget cutbacks in the mid-1990s when accessibility and quality of health services were impaired (Romanow 2002; see also Kirby 2002, Marchildon 2008). The public financing approach also has strong support among the policy-makers (Kirby 2002). The influential Romanow Commission Report, for example, stated that:

> The Commission is strongly of the view that a properly funded public system can continue to provide the high quality services to which Canadians have become accustomed. Rather than subsidise private facilities with public dollars, governments should choose to ensure that the public system has sufficient capacity and is universally accessible. (Romanow, 2002, emphases added)

The support for the publicly funded approach stems from the view that awarding financial responsibility for the system to a single funder: permits a more efficient administration of the system (than a multiple funder approach), minimises costs associated with advertising and billing, distributes risk across a large pool, increases Canada’s industrial competitiveness, and allows all Canadians – independently of their economic status – to be covered (Kirby 2002).

Canada’s publicly funded universal health system – commonly referred to as Medicare – covers all services that are medically necessary (Romanow, 2002). Canada Health Act of 1984 prohibits purchasing private health insurance for services defined as medically necessary under the Act – that is, hospital and physician services provided by the publicly funded system (Dhalla, 2007).16 While forbidden to duplicate services provided under the Canada Health Act,

---

16 It is worth noting, however, that in *Chaoulli v. Quebec*, 2005, the Supreme Court of Quebec struck down the Quebec law which prohibits purchase of private insurance for publicly funded hospital and physician services. It was concluded that – given the unreasonably long waiting lists – banning private health insurance was at odds with the rights granted to Quebeckers under the provincial Charter of Rights and Freedoms (Dickens, 2005).
private health insurance is widely held for health care needs such as dental care, pharmaceuticals and optometry services, as well as ancillary services such as physiotherapy, etc. (OECD 2005a, Marchildon 2008). However, in contrast to the ‘voluntary’ private health insurance system in other countries, private health insurance in Canada is generally provided as part of a benefit package subsidised (to a varying degree) by employers (Marchildon 2008). Consequently, the overall role of private health insurance in Canada is often described as fairly limited (Flood and Archibald 2001), even though approximately 65% of the Canadian population has private health insurance (Dhalla 2007).

2.2.1.4. Health Services Delivery

In Canada, the authorities and responsibilities are divided between the aforementioned two levels of government – the federal government and the provincial (or territorial) governments – as is outlined in the Constitution (Romanow 2002, Leeson 2002). However, as was pointed out in Schneider v. The Queen,

> Health is not a subject specifically dealt with in the Constitution Act either in 1867 or by way of subsequent amendment. It is by the Constitution not assigned either to the federal or provincial legislative authority (Romanow, 2002, 3; emphases added).

Nevertheless, the judicial interpretation of various provisions in the Constitution has established that the delivery and organisation of health services is primarily the domain of Canadian provinces and territories (Romanow, 2002). The federal government, on the other hand, is tasked with: public health, health data collection (through Statistics Canada), funding of health research, as well as provision of a research and, for the most part, the regulatory infrastructure (Marchildon 2005, Marchildon 2008). The most important role of the federal government, however, is its transfer of funds to provinces and territories for financing of the delivery and organisation of health services – the so-called ‘federal spending power.’ Because the receipt of funds by provinces and territories is contingent on meeting the five conditions

---

17 While the provinces and territories are generally tasked with health care delivery, the federal government is responsible for health care delivery to several specific groups, which include: First Nations people, Inuit people, current members and veterans of the Canadian Armed Forces, members of the RCMP and persons who are incarcerated (Romanow 2002).

18 Regulation, however, is not exclusively the domain of the federal government. For example, while the physician licensing is regulated federally, province-specific licensing is regulated provincially.
enumerated by the Canada Health Act, the federal government – despite adopting a hands-off approach to the *delivery* of services – plays an active role in the setting of the national *standards* in healthcare (Marchildon 2008).

Canada’s approach to division of responsibility with regard to health yields a fairly decentralised system, which, although underpinned by a set of common values elucidated in the Canada Health Act, consists of 13 separate (provincial and territorial) health insurance plans (Menon and Stafinski, 2009; Marchildon 2005). This translates into regional differences with regard to health care services. For example, differences can be observed in the modes of service delivery that are adopted across the country (Marchildon 2005). Provinces and territories rely on various combinations of regional health authorities, hospitals, physicians’ practices, and health clinics (Romanow 2002). More importantly for this dissertation, provinces and territories also differ with regard to which particular services – in addition to those provided in hospital and by physicians – are offered, and the extent to which they are funded (Romanow 2002). With regard to health technologies specifically, the provinces and territories decide individually which health technologies to cover – the role of the federal government here is largely limited to pre-market approval (Menon and Stafinski, 2009).

**2.2.1.5. Canada’s Health System by the Numbers**

The funding for the Canadian health system is generated largely through taxation at both federal and provincial levels: in the recent years, approximately 70% of the total dollars for Canada’s health care have come from taxes (Romanow 2002; Marchildon 2008).\(^{19}\) Because most developed countries fund between 70-80% of health care expenditures via taxation (Mossialos et al, 2002; OECD 2009a), Canada is on the lower end of that trend.

Canada’s health expenditure/GDP ratio, however, has consistently exceeded the OECD average health expenditure/GDP ratio for nearly 15 years. In 1997, for example, this ratio was 8.8% for Canada and 8.1% for OECD (OECD 2010a; Huber 1999). In 2002, these numbers

---

\(^{19}\) The remaining 30% of health expenditures is incurred in the private sector, and is funded either through (employer-based) private health insurance or out of pocket. These expenditures include dental care, vision care, physiotherapy, pharmaceuticals and home care (Marchildon 2005, OECD 2005a). In 2005, of the total expenditure on health of $142 billion, this translated into a public financing of $99 billion worth of health expenditures, and private sector expenditure of $43 billion (Marchildon 2008).
were: 9.6% for Canada and 8.6% for OECD (OECD 2010a, OECD 2005b). In 2007, these numbers were: 10.1% for Canada and 9.0% for OECD (OECD 2010a, OECD 2010b). Thus, in addition to consistently exceeding the OECD average health expenditure to GDP ratio, the gap between the OECD average and the Canadian average has grown from 0.7% in 1997 to 1.1% in 2007.

2.2.2. Health Technology Assessment in Canada

2.2.2.1. A brief history of the Canadian HTA

Prior to 1988, HTA activities in Canada at both a national and provincial level took place predominantly on an ad-hoc basis. A combination of factors favouring the establishment of permanent HTA agencies, however, finally converged in the late 1980s. Those included: the growth of evidence-based medicine movement, lack of Canadian-based health technology producers, and the identification of health technology assessment as one of the key priorities of the federal and provincial or territorial ministries of health (McDaid, 2003; Menon and Stafinski, 2009; Battista et al, 2009). Although what is generally recognised as the dominant factor credited with the rise of HTA internationally – namely, the costs associated with new health technologies – is not cited explicitly here, it may well be implied by the identification of HTA as one of the key priorities of the ministries of health. Indeed, in a publicly funded system, it would be a surprise were this not the case.

The first (non-ad-hoc) Health Technology Assessment programme in Canada was established well over a decade after United States established the OTA in the 1970s. The first Canadian HTA agency was established in Quebec in 1988 – it was the Conseil d'évaluation des technologies de la santé (CETS), later renamed the Agence d'évaluation des technologies et des modes intervention en santé (AETMIS) (Hailey, 2007; Menon and Stafinski, 2009; Battista et al 2009). The agency’s purpose was – and is – to produce health technology assessments, to advise Quebec’s Minister of Health, and to disseminate its assessments to key stakeholders in the provincial health system of Quebec (Hailey 2007; Battista et al 2009).

The first national level agency in Canada – the Canadian Coordinating Office for Health Technology Assessment (CCOHTA) – quickly followed on the heels of Quebec’s CETS, in
1989 (Battista et al, 2009). In 2006, CCOHTA was renamed the Canadian Agency for Drugs and Technologies in Health (CADTH) (Battista et al, 2009), which persists to this day.

**2.2.2.2. Sub-national level HTA activity in Canada**

Currently, HTA activity takes place at a variety of both governmental and non-governmental institutions in Canada. Amongst the latter, HTA activities are carried out, for example, by regional health authorities, and individual hospitals, as well as at university-based groups and by the health technology manufacturers (Roehrig and Kargus, 2003; Menon and Stafinski 2009; Hailey 2007; McGregor and Brophy, 2005). But while HTA activity *does* take place among non-government bodies in Canada, its preponderance is carried out by bodies that are affiliated with either a provincial or the federal government (Hailey, 2007).

Because the decentralised nature of the health system in Canada leaves provinces and territories largely to make their own decisions about which health technologies to fund, the vigorousness of the HTA activity at the provincial level in Canada is unsurprising. Provincial HTA bodies currently exist (or have recently existed) in the provinces of: British Columbia, Alberta, Ontario, and Quebec.

British Columbia Office of Health Technology Assessment (BCOHTA) was a BC provincial-government funded body. It was established in 1990 and ceased to exist in 2002 (McDaid 2003; Roehrig and Kargus, 2003). Its aim was to promote the use of HTA by the government of British Columbia, as well as to promote it to healthcare executives and providers (Battista et al, 2009). Its methodology consisted of secondary data collection – that is, systematic reviews and meta-analyses (McDaid, 2003). BCOHTA’s scope was quite broad, encompassing: drugs, medical devices, surgical procedures, medical procedures, medical tests, and clinical guidelines, as well as health administration, and programme delivery and planning (Roehrig and Kargus, 2003). However, BCOHTA’s task was circumscribed only to the assessments of particular health technologies – BCOHTA had *no* involvement in policy decisions (Roehrig and Kargus, 2003).

---

20 Although provincial level HTA activity is discussed here for completeness, the main focus of the dissertation is on national level HTA activity. The statements and arguments made in later chapters intended to apply only to national level HTA agencies and HTA activities, unless otherwise indicated.
Alberta’s provincial HTA body was established in 1993 within Alberta’s Department of Health; this body exists to this day, although it is now located at the Institute of Health Economics (Battista et al, 2009). It is a provincial government-funded entity, and its aim is the production of HTAs for Alberta’s policymakers (Menon and Stafinski, 2009; Hailey 2007). Just like BCOHTA, the methodological emphasis is on the secondary data collection (McDaid, 2003). The scope of the agency’s assessment activities is very broad albeit rather vague – its focus is on “topics relevant to policy development and decision making in Alberta’s health care system” (Roehrig and Kargus, 2003). The programme carries out assessments but does not provide policy advice. Reports on the impact of the programme’s activities are encouraging – with some suggesting that approximately 70% of its HTA reports exert some level influence on the decisions made (Hailey et al, 2000), and that requesting bodies have high confidence in the quality, relevance and objectivity of the information provided (Roehrig and Kargus, 2003).

The Medical Advisory Secretariat (located within the Department of Health and Long-Term Care) is Ontario’s provincial HTA body. It was established in 2001, and its function is to “conduct evidence-based analyses to help stakeholders make policy and funding decisions about health technologies in Ontario” (MOHLTC, 2010). MAS does not conduct primary research – its methodological approach is circumscribed to conducting systematic reviews (Roehrig and Kargus, 2003). Its scope appears to be quite broad, in that it conducts assessment of individual technologies, groups of integrated technologies related to particular diseases, emerging technologies, and existing technologies. Similarly to the other Canadian provincial HTA bodies, MAS does not produce policy recommendations; this task is carried out by the Ontario Health Technology Advisory Committee (MOHLTC 2010).

The function of the aforementioned AETMIS – a descendant of CETS, or the first HTA agency in Canada – is to produce HTAs for Quebec’s health policy-makers (Menon and Stafinski, 2009). AETMIS is an independent agency, albeit one that is funded by the provincial government of Quebec. It reports to Quebec’s Minister of Health and Social Services (McDaid, 2003). Like Canada’s other provincial HTA bodies, AETMIS conducts systematic reviews (AEMTIS 2011). The health technologies that are assessed by AETMIS span a very broad spectrum, encompassing a variety of devices, screening technologies, pharmaceuticals, vaccines, substances, instruments, medical and surgical procedures and professional practices.
(AETMIS 2011). In a notable divergence from other provincial bodies, AETMIS provides policy advice to the Minister of Health (Roehrig and Kargus, 2003; AETMIS 2011).

### 2.2.2.3. National level HTA activity in Canada

Unsurprisingly, the provincial level HTA agencies are (or were) instituted among the largest Canadian provinces. The remaining provinces and territories lack formally established HTA bodies at present. In those jurisdictions, the federal Canadian HTA agency is therefore the *de facto* provincial HTA agency (McDaid, 2003).

As previously noted, the first Canadian federal agency for HTA was the Canadian Coordinating Office for Health Technology Assessment (CCOHTA), which was established in 1989. CCOHTA was initially set up on a trial basis for 3 years, however, upon a favourable review of the agency, it was established as a permanent body in 1993 (Roehrig and Kargus, 2003). In 2006, CCOHTA was renamed the Canadian Agency for Drugs and Technologies in Health (CADTH). CADTH’s funding is shared between the two levels of government: the federal and the provincial (and territorial) governments (excepting the government of the Province of Quebec) (Menon and Stafinski, 2009). CADTH is currently the largest producer of health technology assessments in Canada (Menon and Stafinski, 2009), and its aim is the provision of “credible, impartial advice and evidence-based information about the effectiveness of drugs and other health technologies to Canadian health care decision makers” (CADTH 2009a).

Because CADTH’s level of funding is relatively low, its methodological approach is primarily to rely on secondary research rather than to conduct primary research (McDaid, 2003), like most other HTA programmes. The research is either carried out by CADTH itself or contracted to external research organisations (Menon and Stafinski, 2009).\(^{21}\)

According to CADTH, its evaluative considerations include: clinical effectiveness, cost-effectiveness and the broader impact of health technologies (CADTH 2009a). Although the concept of ‘broader impact’ is rather quite vague, one might suspect that it includes social and ethical considerations around health technologies. Nevertheless, DeJean et al recently found

---

\(^{21}\) This process differs slightly for new drugs (i.e., those recently approved for sale in Canada but not yet reimbursed or covered through one of the publicly funded drug benefit plans) – see, e.g., Menon and Stafinski (2009).
that social and ethical issues do not engender much consideration at CADTH.\textsuperscript{22} Between January 1997 and December 2006, over 600 Canadian-produced HTA reports were examined. These reports typically assessed: effectiveness (92%), economic issues (57%), and safety (25%); only 17% of the reports assessed issues that can be categorised as socio-ethical issues (DeJean et al, 2009).\textsuperscript{23}

At the most general level, CADTH’s HTA programme claims to evaluate pharmaceuticals, medical technologies and health systems (CADTH 2009a).\textsuperscript{24} More specifically, CADTH defines as health technologies:

any method or intervention that is used to promote health; prevent, diagnose, or treat disease; or improve rehabilitation and long-term care. Technologies include drugs, devices, diagnostic agents, equipment, and medical and surgical procedures. The definition also includes organisational and service systems that provide health care, such as tele-health. (CADTH 2009a)

Characterised against Velasco Garrido et al’s spectrum of the definitions of health technologies (see chapter 1), this definition suggests a moderately broad (rather than narrow) understanding of health technologies by CADTH. However, a 2002 review by Menon of Canadian HTAs shows that a vast majority of Canadian HTAs – over two-thirds – focus predominantly on therapeutic technologies – that is, technologies that are used to treat diseases (Menon 2000).\textsuperscript{25} This finding is not surprising, given that the finite amount of funding available to CADTH limits the possible assessments of health technologies to only a small percentage of the many health technologies that constitute the Canadian health care system (Menon and Stafinski, 2009).

Upon the completion of the evidence assessment relevant to a particular health technology, CADTH translates its findings into a report, which it makes available publicly (CADTH

\textsuperscript{22} A substantial drawback of DeJean et al’s analysis is that it amalgamated CADTH’s HTA reports with those produced by several provincial and university-based HTAs. This limits to some extent the reliability of the conclusion for CADTH specifically. Nevertheless, the failure of HTA programmes to adequately consider socio-ethical issues is well documented in the HTA literature (see, e.g., Hoffman (2008), Lehoux and Williams-Jones (2007) and ten Have (1995)).

\textsuperscript{23} This is intended only to give a general idea of the evaluative concerns of Canada’s national level HTA programme. More will be said about this issue in Canada (as well as Australia and the UK) in chapters 4, 5 and 7.

\textsuperscript{24} CADTH is divided into two sub-committees: the Devices and Systems Advisory Committee, and the Advisory Committee on Pharmaceuticals (Menon and Stafinski, 2009)

\textsuperscript{25} For a more recent analysis of CADTH’s evaluative foci, see discussion in chapters 5 (vis-à-vis health technologies, generally) and chapter 7 (vis-à-vis genetic testing in particular).
2009a). CADTH itself does not provide policy advice in the report (unlike some of its provincial counterparts). This is likely a reflection of the largely decentralised structure of the Canadian health system, in which the role of the federal government is rather limited with regard to controlling which health technologies are (or are not) funded by the provinces and territories.

2.3. Australia

2.3.1. Australia’s Health System

2.3.1.1. System of Governance

Just like Canada, Australia is a parliamentary democracy and a constitutional monarchy. Australia’s Constitution establishes the existence of the federal government – the Commonwealth of Australia – and the state (and territorial) governments. The ‘governing bodies’ are the Senate (whose members are elected by popular vote) and the House of Representatives (whose members are also elected by popular vote) (CIA 2011b).

Australian Constitution enumerates the powers of the federal government; those powers not identified in the Constitution as federal are the jurisdiction of the state and territorial governments. Each of Australia’s six states and two territories has its own parliament (Hailey, 2009), and each is also further sub-divided into local governments, which include municipal governments and shire councils (CDHAC 2000).

2.3.1.2. A brief history of Australia’s health system

Three key stages characterise the history of the health system that operates within this political framework: pre-1975, 1975-1984, and 1984 to present. Prior to 1975, Australia lacked a public health system; its approach to health care can be most succinctly summarised as ‘voluntary private health insurance model’ which was regulated by the 1953 National Health Act (OECD 2003).

In 1975, Medibank – a universal, tax-financed public health insurance whose aim was to provide equitable and efficient health care coverage for Australians – was introduced. The programme proved very popular with Australians – within 9 months of the programme’s
inception, over 90% of Australians were issued health cards, and the demand for public medical services (doctors, hospitals) rose, while at the same time, the percentage of coverage by private insurance dropped (Scotton, 2000).

Nevertheless, the Fraser Government closed Medibank in 1981, as a result of budgetary considerations and its commitment to policies that were more amenable to private health care providers and insurers. The health care system reverted to the pre-Medibank arrangements; publicly funded, free at the point of service care was provided only to the pensioners and the indigent, while the general population had to purchase private health insurance (Scotton, 2000).

This led to widespread instances of hardship, and consequently, with the 1982 election of the Labour Government under Bob Hawke, the tide changed back in favour of establishing publicly-funded, universal coverage. The passage of the 1983 Health Legislation Amendment Act amended the existing laws governing health care (the Health Insurance Act of 1973 and National Health Act of 1953), and, in 1984, the universal public health insurance was re-introduced. Although the newly re-introduced publicly funded health system quite closely resembled the original Medibank model, it was renamed Medicare. This system remains in place until the present day (Access Economics 2002, OECD 2003).

2.3.1.3. Health System Funding

The aim of Medicare was – and is – to “provide eligible Australian residents with affordable, accessible and high-quality health care” (Medicare Australia, 2010). That is, Medicare provides access to free treatment to patients in public hospitals, as well as either free or subsidised access to the services provided by physicians, specialists, and other allied health services including optometrists and dentists (Medicare Australia 2010; Hailey 2009, Bulfone et al 2009). Currently, Medicare policy development falls under the jurisdiction of the federal Department of Health and Ageing (Medicare Australia, 2009) and its administration falls to the Health Insurance Commission (AIHW 2004). As of 2005, over 20.5 million people were registered for Medicare benefits, and during the fiscal period 2004-2005, in excess of 236 million services were provided (Medicare Australia 2010).
Medicare is funded through both taxation and an income-dependent Medicare levy (Medicare Australia 2010, Swerissen 2004, AIHW 2010). Medicare levy is presently set at 1.5% of taxable income, although low-income earners are exempt from the levy. As of 1997, an additional levy at the level of 1% of taxable income is imposed on high-income individuals who fail to purchase private health insurance (CDHAC 2000, AIHW 2004). Medicare levy, however, only provides approximately 27% of the Commonwealth’s funding for Medicare – the majority of the funding comes from income tax and general sales tax (CDHAC 2000).

The introduction of Medicare, however, did not mark the end of the evolution of the Australian health system; in the late 1990s, several reforms were implemented with intent to expand and strengthen the parallel private health system. These reforms reflected the Howard Government’s belief that involvement of the private sector in the provision and financing of health services was crucial to the continued viability of Australia’s public health system (CDHAC 2000), insofar as the parallel private system both improved access to care and relieved cost pressures on public hospitals (OECD 2003).

The promotion of private health insurance uptake has not been without impact. Until 1984 – that is, prior to the introduction of Medicare – approximately 50% of the Australian population held private insurance coverage. However, the numbers of privately insured decreased after 1984 – only approximately 32% of the population had private health insurance in 1997 (that is, around the time of the introduction of the first private health insurance incentive, PHIIS). In 2001 – i.e., approximately one year after implementation of private health insurance incentives – 44.3% of Australians were found to have private health insurance (AIHW 2004). This number has held relatively steady since then; in December 2009, 44.7% of the Australian population held private health insurance (AIHW 2010).

Private health insurance is offered by private health insurance funds, which must be registered with the Commonwealth of Australia (that is, the federal government) and are prohibited from

---

26 These included: the Private Health Insurance Incentive Scheme (PHIIS) enacted in 1997 (a health insurance premium subsidy and the aforementioned 1% additional Medicare levy imposed on high-income earners who failed to purchase private health insurance), 1998 amendments to PHIIS that expanded the premium subsidy scheme, and 2000 Lifetime Health Cover arrangements, which support maintaining continuous private health insurance coverage by individuals (OECD 2003, AIHW 2002).
conducting other business (Armstrong et al 2007, CDHAC 2000). The private health insurance sector in Australia is heavily regulated in order to: ensure appropriate interaction with the public coverage system; promote private insurance funds’ fiscal stability; ensure equity of private health insurance coverage; ensure affordability of coverage; protect consumers and their confidence in the private system; protect insurers against adverse selection; and ensure fair competitive environment (Armstrong et al, 2007; OECD 2003). 43 private health insurance funds are currently registered, however, the market is fairly concentrated – top three insurance funds combined insure over 57% of the privately insured population (OECD 2003). Membership in the funds is for the most part open (although some are restricted by profession or union-membership) and all but 6 funds are non-profit (OECD 2003, CDHAC 2000).

Private health insurance offers coverage for approved services in both private and public hospitals, as well as for ancillary services (such as physiotherapy, dental services, optical services, etc.). A variety of insurance plans, with or without front-end deductibles, as well as including or excluding particular services (such as obstetrics), are available (AIHW 2002).

2.3.1.4. Health Services Delivery

Although – unlike in Canada – a parallel private/public health system operates in Australia, the Australian split of powers with regard to health care oversight and delivery mimics the Canadian one.27 Initially, the Australian Constitution only granted to the Commonwealth government the power with respect to quarantine matters (CDHAC 2000). In the aftermath of several Constitutional amendments, however, the main role of the Commonwealth Government currently is to provide funding for health expenditures to the territories and the states (AIHW 2002). The Commonwealth also assumes the leadership of issues such public health policy, regulation of private health insurance industry, research and information management, as well as oversight of the safety and quality of health technologies (CDHAC 2002, Hailey 2009).

The responsibility for the actual provision of health services falls to the State and Territory governments (CDHAC 2000). Although state health authorities are responsible for delivery of

27 Predictably, the Commonwealth/State government split of responsibilities with regard to the health care system is criticised as one of the “single most significant problem in [Australian] health system design” (Dwyer 2004; see also Rix et al, 2005).
health services, sometimes the provision of services is delegated to local governments or community service organisations (AIHW 2002). Generally, however, the states’ responsibilities encompass delivery of services in the public hospitals (including both acute care and mental health care), as well as in infant health centres; dental, health promotion and ambulation services also fall under the remit of state governments, as does the regulation and inspection of the facilities, the institutions and the personnel (AIHW 2002).

2.3.1.5. Australia’s Health System by the Numbers

The funding for these services comes from a variety of sources, which include: the Commonwealth Government, State Governments, Local Governments and Non-Governmental sources (including: individuals, health insurance funds, motor vehicle third party insurance, and workers’ compensation insurance) (AIHW 2004). Approximately 70% of the total expenditure on health (both in the private and public system) is funded by governments,28 which puts Australia – together with Canada – on the lower end of most developed countries, which fund between 70 and 80% of health expenditures through taxation (OECD 2009).

Nevertheless, unlike in the case of Canada, Australia’s health expenditure/GDP ratio is consistently below the OECD average. In 1997, the health expenditure/GDP ratio in Australia was 7.5%, in 2002 it was 8.4% and in 2007 it was 8.5% (OECD 2010a). In comparison, in 1997, the health expenditure/GDP ratio among OECD jurisdictions was 8.1% (Huber, 1999), in 2002 it was 8.6% (OECD 2005b), and in 2007 it was 9.0% (OECD 2010b).

2.3.2. Health Technology Assessment in Australia

2.3.2.1. A brief history of the Australian HTA

In the late 1970s, the increasing costs of health services in Australia and a simultaneous desire for policy-makers to justify their resource allocation decisions led to the creation of the Committee on Applications and Costs of Modern Technology in Medical Practice (also known as the Sax Committee). It was one of that Committee’s recommendations that a permanent national expert panel be struck in order to provide the advice on the funding of new health technologies (Hailey, 2009; Bulfone et al 2009). It is as a result of that recommendation that

28 Of that 70%, the Commonwealth Government contributes approximately two-thirds, and state/territorial and local governments contribute the remainder (Hailey, 2009; Bulfone et al, 2009)
Australia became one of the first jurisdictions outside of the United States to establish an HTA programme (Gallego et al, 2009; Petherick et al, 2007).29

This programme was the National Health Technology Advisory Panel (NHTAP), established in 1982 – that is, predating by 6 years the first government-funded Canadian HTA programme. During its existence, the NHTAP produced in excess of 40 technology reports. In 1990, however, the Panel was subsumed by the Australian Health Technology Advisory Committee. Several years later, in 1998, the AHTAC was replaced by MSAC (Hailey, 2009) – the Medical Services Advisory Committee – which continues operation to the present day.

2.3.2.2. Sub-national level HTA activity in Australia

Just like in Canada, some HTA activity does presently take place at the local and institutional levels in Australia. Gallego et al, for example, report that regular HTA activities take place at an Area Health Service in North Sydney (Gallego et al, 2009). Moreover, several hospitals in the States of Queensland, Western Australia, Victoria and South Australia are reported to have established their own HTA programmes (Jackson, 2007). In Victoria, for example, these include Bayside Health and Southern Health (Productivity Commission, 2005). These examples notwithstanding, overall, not a lot of information is generally available about HTA activity among Australian entities not associated with either the federal or state governments (Gallego et al, 2009).

Unlike in Canada (where a reasonable amount of government-funded HTA activity takes place at the provincial level), relatively little HTA activity takes place at the State (or Territorial) level in Australia. One notable exception to this trend is the Victorian Policy Advisory Committee on Clinical Practice and Technology (VPACT) established in 2004 in the state of Victoria (Jackson, 2007; Gallego et al 2009; Bulfone et al, 2009). VPACT’s function is rather broadly identified as addressing the issues surrounding the introduction of health technologies as they arise specifically within the Victorian context (DHVA 2011). What methodology VPACT adopts in fulfilling its mandate is unclear, although the agency does stress that it considers both the cost-effectiveness and clinical effectiveness of health technologies. The

29 Unlike Canada, then, Australia explicitly recognised the economics of new health technologies as one of the driving factors for the establishment of its HTA programme.
spectrum of technologies that VPACT assesses includes: clinical and treatment interventions, prosthetic and implantable devices, diagnostic tests, pharmaceuticals, and medical and surgical procedures. VPACT both carries out the assessments of evidence relating to these technologies, and produces policy recommendations to the Victorian Department of Human Services (DHVA 2011).

2.3.2.3. National level HTA activity in Australia

Although some HTA activity is therefore evident at both State and local levels in Australia, the preponderance of Australia’s HTA work is carried out at the national level. Unlike in Canada, where a single agency – CADTH – performs the HTA assessments at the national level, Australia divides the HTA load among three agencies: Prostheses and Devices Committee (PDC), Pharmaceutical Benefits Advisory Committee (PBAC) and Medical Services Advisory Committee (MSAC) (Government of Australia 2009b, Hailey 2009).

Australian government construes prostheses and prosthetic devices to include devices such as cardiac defibrillators, pacemakers and stents, joint replacements (including knee and hip replacements), as well as a variety of human tissues (for example: heart valves, bones, corneas and muscle tissues) (Government of Australia, 2009b). From a methodological standpoint, the Prostheses and Devices Committee (PDC) reviews safety and clinical effectiveness evidence that is provided by the manufacturer as part of the application, rather than conducting its own systematic reviews (Government of Australia, 2009b). PDC’s function is two-fold: to carry out the health technology assessments of these devices (in particular focusing on their safety and clinical effectiveness), and to provide advice to the Minister of Health and Ageing regarding which devices ought to be covered.

The Pharmaceutical Benefits Advisory Committee (PBAC) evaluates pharmaceuticals prior to their inclusion on the Pharmaceutical Benefits Scheme (PBS). PBAC’s remit includes pharmaceuticals and vaccines (Government of Australia, 2009a; Bulfone et al 2009). The Committee does not conduct either primary research (RCTs, etc.) or secondary research.

---

30 All three bodies will be described, although the focus here and subsequently will predominantly be on MSAC, since it is the closest national level analogue to Canada’s CADTH and UK’s NIHR (discussed shortly)
31 Pharmaceutical Benefits Scheme (PBS) is Australia’s list of publicly subsidised pharmaceuticals. For more on PBS, please see: http://www.medicareaustralia.gov.au/about/whatwedo/pbs.jsp
(evidence synthesis); rather, the ‘burden of proof’ lies with the applicant or manufacturer to submit a complete HTA, which is then critiqued by PBAC (Bulfone et al, 2009). PBAC’s key evaluative considerations are the drug’s safety, effectiveness and cost-effectiveness (Government of Australia, 2009b; Jackson 2007). In addition to critiquing the HTAs supplied to it by applicants, PBAC also provides recommendations to the Minister of Health and Ageing regarding which pharmaceuticals and vaccines ought to be listed on the PBS (i.e., publicly subsidised) (Bulfone et al, 2009).

The third Australian national level HTA entity is MSAC – the Medical Services Advisory Committee. MSAC is the third incarnation of the National Health Technology Advisory Panel (1982-1990), which was subsequently replaced by the Australian Health Technology Advisory Committee (1990-1998), before ultimately metamorphosing into MSAC in 1998.

MSAC focuses on assessments of new medical technologies and procedures which do not fall under the remit of either PDC or PBAC (Government of Australia, 2009b). Although one might expect that – given the division of the total HTA load among three agencies – the number of health technological assessments completed in Australia annually is substantial, this is not actually the case. Between 1998 and 2003, for example, MSAC received 129 applications and completed only 75 reviews (Hailey, 2009).32

MSAC permits (but does not require) that applicants submit a complete HTA. Consequently, applicant submissions are rare; it is more typical that MSAC carries out its own review. Like in the case of many other HTA agencies – including Canada’s CADTH – MSAC’s assessments consist of a review of the existing data (Petherick et al, 2007). In 2007, Jackson observed that MSAC lacks the capacity to commission or carry out new research where such research is not available (Jackson, 2007). More recently, the situation remains unchanged (Bulfone et al, 2009).

32 Hailey (2009) notes that the situation is similar with regard to PBAC; the number of pharmaceuticals assessed annually is very low, and only the new pharmaceuticals go through the assessment process. For a more detailed overview of MSAC’s evaluative output more recently, see chapter 5.
The evaluata prioritised by MSAC pertain to the safety, clinical effectiveness, and cost-effectiveness\textsuperscript{33} of health technologies (Government of Australia, 2009b; Hailey, 2009; Petherick et al, 2007; Jackson 2007). In fact, Australia is credited with being the first jurisdiction internationally to require a cost-effectiveness assessment as part of the HTAs; Australia introduced this requirement into its HTA process in 1993 (Jackson 2007). The cost-effectiveness data is required to be specific to the Australian context, although in practice Australian HTA bodies have frequently adopted and adapted the economic evaluations carried out in other jurisdictions (Bulfone et al, 2009). In cases where evidence – whether economic or otherwise – is insufficient or inconclusive, but does suggest that the assessed technology may be more cost-effective, more effective or safer than the currently used comparable technology, MSAC has the ability to recommend interim funding for the health technology. The interim funding is intended to permit further data collection and subsequent further assessment of the health technology (Bulfone et al, 2009). As this suggests, MSAC’s methodological approach normally involves comparing the new technology (or treatment) against currently funded and most commonly used technology (or treatment) (Government of Australia, 2009b).

Beyond assessing particular health technologies, MSAC also provides advice to the Minister on Health and Ageing on whether the technology should be funded (Hailey, 2009; Bulfone et al, 2009; Petherick et al, 2007; Government of Australia, 2009b). Because in Australia – unlike in Canada – funding decisions for health technologies are made at the national level, and, moreover, these decisions are directly linked to the assessment carried out by MSAC, MSAC therefore plays a much more influential role in the national health policy in Australia than, mutatis mutandis, does CADTH.

\textbf{2.4. United Kingdom}

\textbf{2.4.1. United Kingdom’s Health System}

\textbf{2.4.1.1. System of Governance}

Just like Canada and Australia, the United Kingdom is a constitutional monarchy. The Legislative power rests with the government and two houses of Parliament: the House of Lords,\textsuperscript{33} In 2009, MSAC’s Terms of reference were broadened to also include an evaluation of the health technology’s impact on the budget.
and the House of Commons. The majority of the members of the House of Lords are unelected (elections take place only as vacancies arise), while the members of the House of Commons are elected every five years (unless the House is dissolved prior to that) (EOHCS 1999). The UK is divided into four countries: England, Northern Ireland, Wales and Scotland. Partial devolution of power exists in Scotland (Scottish Parliament), Wales (Welsh Assembly) and Northern Ireland (Northern Ireland Assembly) (EOHCS 1999).

2.4.1.2. A brief history of United Kingdom’s health system

During the 19th and early 20th century, the UK lacked a centrally organised health system. Health care was provided by voluntary and municipal (government-run) hospitals, and physician care was provided by community doctors. Some mutual insurance funds did exist, but health insurance was not widely held; the costs of health care were normally borne either by the individuals themselves, or by charitable donations. Throughout the 1920s and 1930s, however, it became increasingly recognised that this situation resulted in substantial financial barriers to health care access – especially for women and the indigent (Boyle 2011).

Thus, in the aftermath of 1942’s Beveridge Report and World War II, the Labour Government passed the National Health Service Act in 1946 (Webster 2002). As a result of the passage of this Act, the National Health Service (NHS) was implemented in 1948 (EOHCS 1999) – several decades before the federal governments of either Canada or Australia became involved in establishing a universal health system for their own citizens.

At its launching in 1948, the original objectives of the NHS included: (1) that it meet the needs of everyone; (2) that it be free at the point of delivery; and (3) that it be based on clinical need, not ability to pay (NHS 2009). Thus, the implementation of the NHS eliminated direct charges for health care, thereby providing access to health services to groups that were previously unable to access care (Webster 2002; EOHCS 1999). Over 60 years later, the NHS continues to provide universal coverage for primary and hospital care to all those ordinarily resident in the UK, and these three principles remain at the heart of the NHS34 (NHS 2009; Banta 2003a).

34 Although a review in 2000 resulted in an addition of several further objectives, which included, for example, emphasis on taking advantage of new health technologies.
2.4.1.3. Health System Funding

Much like both Canada’s Medicare and its Australian namesake, UK’s NHS is a single-payer, publicly funded system. NHS is, in fact, currently the world’s largest publicly funded health care system (NHS 2009). The funding for the system comes predominantly from general taxation (Roehrig and Kargus 2003), although some funding is generated from private medical insurance or PMI (discussed below) and via user charges for services (although services are generally provided free of charge). The funding for the system is a result of negotiations between the Department of Health and the Treasury Department; the Department of Health is in charge of allocating the negotiated resources to the NHS (Boyle 2011).

Private health sector does exist in parallel with the publicly funded system, although it is rather small – only approximately 13% of the population carries private medical insurance (Boyle 2011), as compared to Australia’s 45%. Moreover, the private health sector in the UK is not as extensively developed as in Australia – it focuses predominantly on providing elective acute care. The private sector’s funding comes from private insurance premiums, as well as from patients’ out-of-pocket payments, and payments by the Department of Health (Boyle 2011). The sector is overseen by the Care Quality Commission – one of several arm’s length, national level bodies that assist the Department of Health. Care Quality Commission’s role is to regulate service providers within the framework of the Care Standards Act of 2000, as well as to oversee the quality of the services provided (Care Quality Commission, undated).

2.4.1.4. Health Service Delivery

The Department of Health is the national body charged with responsibility for health care in the United Kingdom. At the most general level, the Department’s function is “to support the government in improving the population’s health” (Boyle 2011). More specifically, the functions of the department include setting of national policy on public health, as well as on mental health, clinical quality, health improvement, and so forth. The Department also serves as the national headquarters of the NHS (Boyle 2011). Until 1999, the Department of Health was responsible for the National Health Service in all four countries, however, following a

35 Several arm’s length bodies (including the National Institute for Health and Clinical Excellence, or NICE, and the aforementioned Care Quality Commission) assist the Department in setting and monitoring standards, as well as regulating the health system (Boyle 2011). See the subsequent section on the UK’s approach to HTA for a more expanded description of the function of NICE.)
partial devolution, this responsibility was taken on in Scotland by the Department of Health of the Scottish Office, while in Wales the Welsh Office performs this task, and in Northern Ireland, it became the function of the Department of Health and Social Services.\textsuperscript{36}

Although health policy is determined centrally at the national level by the Department of Health, the oversight at the \textit{regional} level is carried out by 10 Strategic Health Authorities or SHAs. The SHAs’ mandate is to manage the NHS at the local level (by developing plans for local health services), as well as to oversee the quality of the local health services and the integration of national priorities into the local service provision (NHS, 2010).

While SHAs engage in oversight, the \textit{provision} of care at the local level is organised by organisations known as PCTs – Primary Care Trusts. The PCTs manage the local individual health provider offices (doctors, dentists, opticians, mental health service providers, etc.), as well as the NHS walk-in centres, pharmacies, hospitals, and the NHS direct telephone centres (NHS 2010). The funds for the Primary Care Trusts are allocated by the Department of Health; there are currently 151 PCTs, each of which covers, on average, approximately 340,000 people (Boyle 2011).

\textbf{2.4.1.5. United Kingdom’s Health System by the Numbers}

The health system is financed by the government through taxation. United Kingdom’s provision of 82\% (in 2008) of the health system’s funding via taxation (Boyle 2011) thus places it above the average range of 70-80\% taxation-based funding that is provided by other countries. This percentage is also a substantial 12\% higher than either Canada or Australia.

Despite this, United Kingdom’s expenditure on the health system (expressed as a percentage of the GDP) has – like Australia – for nearly the past decade and a half consistently been below the OECD average. In 1997, for example, UK’s health expenditure/GDP ratio was 6.6\% while OECD’s ratio was 8.1\% (OECD 2010a; Huber, 1999). Five years later, in 2002, United Kingdom’s ratio was 7.6\% while OECD’s was 8.6\% (OECD 2010a, OECD 2005b). Finally, in 2007, these ratios were 8.4\% for the United Kingdom and 9.0\% for the OECD (OECD 2009b, OECD 2010b). While the UK ratio remains under the OECD average, the gap between the UK

\textsuperscript{36} For more information, please see EOHCS 1999, especially appendix 1.
ratio and the OECD average has noticeably narrowed in the recent years (from 1.5% in 1997 to 0.6% in 2007).

2.4.2. Health Technology Assessment in the United Kingdom

2.4.2.1. A brief history of HTA in the United Kingdom

Archie Cochrane’s 1972 book “Effectiveness and Efficiency” is credited with spurring the efforts to establish a formal HTA programme in the United Kingdom (Stevens and Milne, 2004). In his book, Cochrane stressed that the National Health Service’s modus operandi essentially constituted a “giving [of] a blank cheque both to the demands of patients and the wishes of doctors”, and consequently, the NHS faced ever-increasing technology costs (cited in Drummond and Banta, 2009, 178). Around the same time, concerns about both costs and effectiveness of health technologies were also being raised in other influential quarters, including the Health Committee of the House of Commons (Woolf and Henshall, 2000). This ultimately resulted in several HTA efforts in the 1980s by the Medical Research Council, Department of Health, industry, university and medical centres. For the most part, however, these efforts remained sporadic and uncoordinated (Drummond and Banta 2009).

The first sustained and official HTA effort in the United Kingdom came rather later than parallel efforts in either Australia or Canada. UK’s first HTA efforts can be traced to the appointment, in the early 1990s, of Michael Peckham as the Director of Research and Development for the NHS (Stevens and Milne, 2004). Peckham established Health Technology Assessment as one of the priorities for the Research and Development Programme (R&D Programme), and, under his direction, several evidence-based programs targeting clinical practice and clinical guidelines were both developed and promoted (Drummond & Banta, 2009). Peckham’s efforts were further supported and entrenched by the 1992 Department of Health’s report which called for health technological assessments that would address the issue of effectiveness of health technologies in the National Health Service (DOHUK, 1992). As Stevens and Milne (2004) note, by mid-1990s, this yielded an official recognition of the value of health technological assessments and a robust level of officially supported HTA activities.
In 1993, in what was largely a re-branding exercise, the R&D HTA Programme was relocated to the National Institute for Health Research (NIHR) and renamed the NIHR HTA programme.\(^{37}\) It is worth emphasising that although the HTA efforts in the United Kingdom came about as a result of concerns about the costs and effectiveness of health technologies, the establishment of the HTA programme was framed as a result of the NHS’s commitment to evidence based practice, rather than strictly out of a desire to ‘contain’ new technologies (Walley, 2007).

2.4.2.2. Sub-national level HTA activity in the United Kingdom

Just like in Canada and Australia, there is some evidence of HTA activity across a variety of levels and locations in the United Kingdom. Some health technology manufacturers conduct self-funded HTAs (Roehrig and Kargus, 2003). Several universities – which include the University College of London, the University of Oxford, the University of York and the University of Sheffield – also have HTA research bodies\(^{38}\) (Roehrig and Kargus, 2003). Nonetheless, similarly to the arrangements in place in Australia, in the United Kingdom, the preponderance of HTA activity takes place at the national level.

2.4.2.3. National level HTA activity in the United Kingdom

In context of its national level HTA efforts, The United Kingdom draws an important distinction not often drawn by the national level HTA programmes in other jurisdictions – certainly, it is not drawn by either Canada’s CADTH or Australia’s MSAC – namely, a distinction between health technological assessment and health technological appraisal (Walley, 2007; Oliver et al, 2004, Stevens and Milne, 2004).

The assessment stage of the HTA process is sometimes referred to as the ‘science’ of HTA (Oliver et al, 2004) or the ‘scientific’ stage of the HTA process (Walley, 2007). United Kingdom’s National Institutes of Clinical Excellence (NICE) defines the assessment stage of the HTA as “a systematic evaluation of the relevant evidence available on a technology” (NICE 2008) and in the literature, this stage has been defined as “the analytical process of

---

\(^{37}\) Jill Weeden, Senior Programme Manager, NIHR Evaluation, Trials and Studies Coordinating Centre (NETSCC). Personal Communication, April 15, 2011.

\(^{38}\) These bodies are often contracted by the NIHR (discussed below) to carry out systematic reviews.
gathering and summarising information about health technologies” (Stevens and Milne, 2004). The assessment is regarded as ‘objective’ (NICE 2008), and transferable internationally, subject to adjustments for economic and cultural circumstances (Stevens and Milne, 2004).

The health technological appraisal, by contrast, is the domain of the health policy-makers (Walley 2007). Health technology appraisal has been defined as “the process by which the [HTA] science is considered at the policy-making level” (Oliver et al, 2004) or alternately as “the political process of making a decision about health technologies, taking into account assessment information but also values and other factors” (Stevens and Milne, 2004). NICE itself views the appraisal as a stage wherein the assessment is considered together with a range of other factors (which include additional information supplied by stakeholders) (NICE 2008). The appraisal stage of the HTA process is regarded as more context-specific than the assessment stage – and the output of the appraisal process is a policy decision rather than an evidence assessment report (Stevens and Milne, 2004).

The aim of this division of the HTA process into the assessment and the appraisal stages is to protect – as much as is possible – the ‘scientific’ assessment stage from the pressures and influences of the various stakeholder groups (manufacturers, patient groups, etc.) (Walley, 2007).

Assessment: NIHR

The current, national level HTA programme was established in 1993, and it is both funded and managed by the National Health System (NHS) (Boyle, 2011; Walley, 2007). Institutionally-speaking, the HTA programme is currently situated within the National Institute for Health Research (NIHR). Annually, the NIHR HTA programme produces on average 40-50 assessment reports, which are referred to as TARs or Technology Assessment Reports (NIHR, undated).

Reflecting the assessment/appraisal distinction adopted in the United Kingdom, the function of the NIHR HTA programme is to provide health technology assessments only (Walley 2007). This is achieved by commissioning primary research and assessment of the information (e.g. economic evaluation or review of manufacturers’ submissions) pertaining to a particular health
technology (Boyle, 2011; Walley 2007). Thus, unlike the majority of national level HTA programmes (Walley 2007) – including MSAC and CADTH – which are limited only to systematic reviews and meta-analyses, the NIHR HTA programme also engages in primary research projects (Banta, 2003a; Boyle, 2011). The programme’s evaluative considerations span cost-effectiveness, clinical effectiveness and safety of the health technologies (Walley 2007; NIHR undated).

The spectrum of technologies assessed by the NIHR HTA programme evinces what Velasco Garrido et al (2010) would characterise as a moderately or perhaps even very broad understanding of ‘health technologies.’ Between 1993 and 2008, the Programme performed HTAs for: drugs (e.g. antibiotics), devices (e.g. dialysis machines), mental care, surgical techniques, screening programmes (e.g. for STIs), counselling, and so forth (NIHR, 2009b). Once complete, the assessment of these technologies – in the form of a Technology Assessment Report (TAR) – is submitted to NICE in order for it to carry out the appraisal part of the HTA process. The TAR itself does not contain any policy recommendations to the NICE (Walley, 2007).

**Appraisal: NICE**

Prior to the establishment of the National Institute for Clinical Excellence or NICE, health technology appraisal was carried out by a wide range of bodies across the UK, at the local, regional and national level (Stevens and Milne, 2004). With the creation of the NICE in 1999, however, the task of health technology appraisal was centralised at the national level, and assigned to NICE (Stevens and Milne 2004). Within NICE, the Centre for Health Technology Evaluation (CHTE) is the body that performs the health technology appraisals (Drummond and Banta, 2009).

NICE’s mandate includes ensuring equal access to new technologies (Pearson and Rawlins, 2005; Walley 2007; Oliver et al 2004). NICE is therefore commonly – and not entirely

---

39 Initially, the Institute’s name was the National Institute for Clinical Excellence, hence the acronym NICE. In 2005, the Institute was amalgamated with the Health Development Agency, and renamed the National Institute for Health and Clinical Excellence, but the acronym NICE remained.

40 NICE’s formal remit covers only England and Wales; nevertheless, its informal influence extends to Northern Ireland and Scotland, as well (Stevens and Milne, 2004).
surprisingly – regarded as the United Kingdom’s HTA agency (Walley 2007). In light of the distinction between the health technology assessment and health technology appraisal that the UK system draws, however, it is worth keeping in mind that NICE’s role in the HTA process is that of a health technology appraiser rather than a health technology assessor.

The TARs produced by the NIHR HTA programme are one of the evidentiary bases on which NICE’s CHTE relies to make its decisions, but by no means is it the only consideration that factors into NICE’s decision. Other considerations include comments and submissions from: health professionals, health organisations, patient groups, and manufacturers of the technology under appraisal (Stevens and Milne, 2004). Generally speaking, NICE’s appraisal focuses on the balance that a particular health technology achieves between costs (such as quality adjusted life-years) and benefits. More specifically, NICE’s appraisal decisions are based on the following six criteria: (1) the clinical needs of the patients, (2) the priorities of the NHS, (3) the balance between the costs and benefits, (4) the impact on NHS’s other resources, (5) encouragement of innovation, and (6) the guidance from the Ministry of Health regarding the available resources (Rawlins, 2001). Importantly, NICE does not consider the budget impact of the appraised health technology into account (Pearson and Rawlins, 2005). 41

On the basis of the TAR, as well as the other considerations, NICE produces an FAD – a Final Appraisal Determination – which is subsequently translated into a guidance document for the evaluated technology. In light of UK’s separation – into the assessment and appraisal stages – of what is elsewhere a single HTA process, it is worth asking whether this approach translates into frequent divergence between the conclusions suggested by the assessment stage (bearing in mind that no actual recommendations are offered by the NIHR HTA) and the decisions made by the NICE at the appraisal stage. Research shows, however, that generally, the NICE decisions are aligned with the conclusions suggested by the assessment stage; divergence is rare (Walley 2007). 42

---

41 Affordability is not the responsibility of NICE; the government remains accountable for the overall NHS budget and must therefore judge a particular intervention unaffordable for the NHS even if NICE might have judged it cost effective. In practice, however, such a situation has not yet arisen. (Pearson and Rawlins, 2005)

42 Barnett (2001) describes a case of an exception to this trend: the assessment of riluzole (a pharmaceutical used to treat motor neuron disease). In that case, the assessment showed a cost-effectiveness that exceeded ratios normally approved by NICE. Nevertheless, as a result of the evidence presented by patient groups, NICE approved the drug.
Because the NICE’s guidance is binding on the National Health Service (that is, the guidance must be implemented within 3 months of issue) (Stevens and Milne, 2004), the UK’s health technology assessment system (understood as including both the assessment and the appraisal stages) appears to enjoy the level of influence on national health policy that is more akin to that of Australia’s MSAC, than that of Canada’s CADTH.

2.5. Conclusion

To briefly summarise, the health systems in the three jurisdictions considered here – Canada, Australia and the UK – all share the single payer approach to provision of services. Nevertheless, some differences are also evident – in particular with regard to the centralisation or decentralisation of the health systems.

These differences translate into disparities in these jurisdictions’ respective approaches to HTA. In the case of Australia and the United Kingdom, the centralised nature of their health systems limits the vigorousness of HTA activity at the sub-national level. Conversely, in the decentralised Canada, HTA activity among provincial governments is comparatively high. Moreover, the centralisation (or lack thereof) also impacts the level of influence that the national level HTA programme has on health policy. In both UK and Australia, the health technology funding decisions are closely linked to the evaluations conducted by HTA bodies (MSAC in the case of Australia, and NIHR HTA in tandem with NICE, in the case of the UK). This results in a high degree of influence of these bodies. In Canada, conversely, where the provinces make their own funding decisions, the impact of the national HTA programme – CADTH – is far less pronounced.

In addition to the differences that are linked to the minutiae of the health system frameworks within which they are embedded, a variety of other differences (e.g., with regard to scope, methodology, etc.) is also observed between Canada’s, Australia’s and UK’s national level HTA programmes. The next chapter will build on the findings in this and the previous chapter in order to first, establish a set of criteria on which the various HTA approaches can be compared, and second, to compare the three jurisdictions’ HTA programmes thereon.
CHAPTER 3: Comparing the 3 Jurisdictions’ Approaches to HTA

*Summary:* This chapter begins by making a case for generating a scheme for comparing different HTA approaches, and subsequently, offers such a scheme. It then places the 3 national level HTA programmes described in Chapter 2 – Canadian CADTH, Australian MSAC and UK’s NIHR – along that scheme. Finally, the chapter offers a systematic, criterion-by-criterion comparison of the 3 HTA programmes, revealing that although (as was established in chapters 1 and 2) all three programmes claim to be evidence-based, differences are evident with regard to the evaluata they consider.
3.1. Why a taxonomy?

3.1.1. Comparing HTA approaches across jurisdictions?

In chapter 2, it became quite clear that the three jurisdictions under consideration here – Canada, Australia and the United Kingdom – share numerous characteristics. Similarities are evident both with regard to their approach to health system funding (all three jurisdictions publicly fund their health services), as well as with regard to their approaches to HTA (all three jurisdictions claim that their HTAs are evidence-based and that their function is to provide input into health policy).

On the other hand, numerous differences also emerged. Some of these included: the evaluative scope of the particular agencies (i.e. which technologies are evaluated by the particular bodies), the vigorousness of HTA activity at levels other than national or federal, the methodologies adapted by the particular HTA bodies, whether the HTA process is directly linked to funding decisions, and so forth.

It is possible – perhaps even plausible – that the differences among the various HTA bodies are likely to translate into differences, for example, in what is evaluated by these bodies – an HTA body adopting the research synthesis (i.e. secondary) approach to methodology may be more limited in the range of inputs that it is able to consider, than an HTA that adopts primary methodology, for example. Differences in whether an HTA programme is limited by its terms of reference to assessment only, or whether it additionally produces policy advice, may also impact its willingness to consider evidence other than the standardly encountered safety, cost-effectiveness and clinical effectiveness (e.g., patient views, socio-ethical issues, etc.). An ability to systematically compare the similarities and differences between these programmes would therefore be quite useful.

Nevertheless, systematic comparisons of different HTA programmes are quite difficult, as no heuristically useful taxonomy of HTA approaches exists at present. This may be due to a lack of interest by HTA scholars in this type of work. The preponderance of current HTA literature focuses on describing methodological issues, the process and outcomes of particular health
technological assessments, establishing the need for the integration of ethical issues into the HTA process, or on describing the various jurisdictions’ programmes. Comparative or cross-jurisdictional theoretical HTA work, on the other hand, is practically non-existent.

The only two HTA articles that most closely approach a comparison of different jurisdictions’ approaches to HTA were located in a 2007 edition of *The Medical Journal of Australia*. In “Health technology assessment in England: assessment and appraisal,” Tom Walley describes in detail the English approach to HTA, devoting a brief section to a comparison of HTA in England and Australia. In this section, he notes that Australia divides its national level HTA work among three agencies, which leads to a confusion over responsibility for combined technologies, whereas England is “getting better at acting as a corporate whole” (Walley, 2007, 285). He additionally points out that the English HTA programme is more transparent and amenable to public involvement than its Australian counterpart. In “Health technology assessment in Canada: diversity and evolution,” David Hailey devotes the majority of the article to describing Canada’s HTA programmes, like Walley, also devoting a brief section to a comparison – between Canada and Australia. In this section, he notes the relatively higher incidence of provincial-level HTA activity than the state-level activity in Australia, and notes that Australia’s programme is more focused “on providing input to national health insurance programmes” than the Canadian one (Hailey, 2007, 287).

The primary shortcomings of these comparisons are first, their brevity, and second, their failure to establish a systematic classification scheme of the key elements of each HTA approach for the purposes of the comparison. There are, however, both general and HTA-specific reasons why generating such a taxonomy of HTA approaches would be useful. First, however, some background about taxonomies or classificatory schemes.43

### 3.1.2. Why the field of HTA needs a taxonomy of approaches.

Taxonomies enjoy a long history, and evidence of their use dates back at least to the ancient Greeks – Theophrastus’ (372 BC – 287 BC) *Historia Planetarum* is among one of the earliest

43 ‘Taxonomy’ and ‘classification’ will be used interchangeably here, to mean grouping of objects on the bases of their relationships to each other, although see e.g., Sokal (1974) for a discussion of formal differences between these terms.
surviving works that attempts to group plants (into trees, shrubs and herbs) according to their physical characteristics (Keaney, 1968; Weiher et al, 1999). Since Theophrastus’ time, classificatory systems have been applied in areas ranging from biology (animals, plants, soils) to diseases (both physical and psychiatric), languages, manuscripts, and so forth, and, as Sokal points out, presently, “classification is an important aspect of most sciences” (Sokal, 1974, 1115).

Taxonomies wouldn’t enjoy such long history and wide range of application if they weren’t useful. The main benefits of the use of taxonomies encountered in the literature fall into three main categories: communicative, economy of memory and relationship analysis. First, taxonomies facilitate communication by allowing us to readily summarise information and to affix convenient labels to that information (Raven, 1971). Second, they also effect the economy of memory, insofar as they group numerous individual entities or objects into a group, subsuming all objects contained in that group under a single description. Finally, by collecting objects into groups, they allow us to describe the relationships between the members of the group internally, as well as the relationships between that and other groups (Sokal, 1974).

As these benefits are generic to any taxonomy, they will also apply to any potential taxonomy of HTA approaches. First, a taxonomy would permit us to summarise the information about the main elements of the HTA approaches in a systematic and readily accessible manner, facilitating communication. Second, although ‘HTA approaches’ certainly do not form a class as large and disparate as, say, ‘animals,’ some economy of memory is achieved insofar as a fair degree of variety is evinced with by the various HTA programmes (e.g., their level, types of evidence they evaluate, the types of technologies they evaluate, and so on). Finally, it is the third potential benefit – the relationship analysis – that is anticipated to constitute the greatest benefit here. The ability to engage in meaningful discourse about HTA approaches adopted by various jurisdictions, and to speculate about the causes and effects of the differences and similarities between them, requires the ability to identify these differences and similarities in a systematic manner in the first place. This is currently difficult; generating a taxonomical system of HTA approaches would help to resolve this difficulty.
3.2. A general taxonomy of approaches to HTA

Classification systems are generally one of two types: monothetic and polythetic. In monothetic classification systems, each member of a class shares a particular property (call it A), which differs it from members of other classes, which lack property A. In polythetic classification systems, on the other hand, classes share numerous properties but it is not necessarily the case that they all share any one particular property (e.g., A). In other words, neither a particular combination of characteristics nor a single uniform property defines a particular group or class (Sokal, 1974).

The classificatory system constructed here will be of the latter type – that is, polythetic. The HTA programmes will have some characteristics or properties in common, but it is not the case that any one property will be shared by all of them (with the exception that they all are HTA programmes). The HTA programmes will be characterised according to two sets of criteria: the contextual criteria and the HTA programme-specific criteria.

The first category – that is, the contextual criteria – encompasses two criteria that describe the health system within which the HTA programme is situated. That a relationship between the structure of the HTA programme and the health system structure within which it operates does exist is often asserted or implied in the literature (see, e.g., Walley, 2007; Hailey, 2007; Stevens and Milne, 2004). What the nature of this relationship is, however, is unclear. The category is included here in order to shed some light on the nature of this relationship.

**Criterion 1** thus centres on the first of the two key elements of a health system within which the HTA programme is operating, namely, the centralisation of the health system (or lack thereof). This particular criterion is of interest for two reasons. First, the vigorousness of HTA activity at the sub-national level will likely be affected by the level of centralisation of a health system – in health systems that are more heavily centralised, it is likely that the HTA activity, too, will be more concentrated at the national level. Second, the influence of the national level HTA programme is also likely linked to the level of centralisation of the health system; a national level HTA programme in a heavily centralised health system is likely to have a greater influence on funding decisions than a national level HTA programme in a heavily
decentralised system. The health systems will be categorised under this criterion fairly generally, as either centralised or decentralised.

**Criterion 2** describes the health system funding model that is in place in a particular jurisdiction: that is, it describes whether the health system is publicly funded, privately funded, or a mixture of both. It is possible – perhaps even plausible – that an HTA process operating within a predominantly privately funded health system would differ from one operating within a predominantly publicly funded health system (e.g. with respect to the types of evaluata prioritised). This criterion is less crucial than criterion 1 for this particular dissertation – as the health systems of all three jurisdictions of interest are predominantly publicly funded ones. Nevertheless, because the taxonomical system described here is a general one, and intended to be applicable to *all* HTA programmes (which, after all, operate in variously funded health systems), the criterion is included here for completeness of the taxonomical system.

The second category of criteria in this taxonomy of HTA approaches – that is, the category of HTA-specific factors – encompasses six criteria that describe the structure and operation of the HTA programme itself. The criteria include: the jurisdictional level at which the HTA programme operates, the scope of its assessment activities (that is, what type of health technologies it assesses), the method it utilises in its assessments, the evaluata that it considers in conducting its assessments, whether it provides policy advice, and finally, what role it plays with regard to health technology funding decisions. These particular criteria were selected because they evince the greatest degree of variety among the particular HTA programmes described in chapter 2, as well as among other programmes, considered more generally in chapter 1.

**Criterion 3** describes the level at which the HTA programme is situated. As discussed in chapters 1 and 2, HTA activity is evident at a variety of levels, including both national and sub-national levels. The sub-national levels can be further differentiated into provincial/territorial/state (or jurisdictional equivalent) level, regional level and institutional level. Where a precise ascription of a level may be difficult, the deciding factor will be the programme’s self-identification. Thus, if an HTA programme self-identifies (e.g. on its website) or is generally recognised in the academic literature as a national HTA programme, it
will be classified as such. *Mutatis mutandis* if an HTA programme self-identifies as a provincial (or state, or jurisdictional equivalent) one, regional one, and so on.  

**Criterion 4** notes the scope of assessment – i.e. it considers the range of health technologies that fall under the remit of a particular HTA programme. Many HTA bodies – particularly at the national or provincial/territorial/state levels – do *not* restrict their evaluative scope, allocating all HTA efforts to a single agency. On the other hand, some jurisdictions *do* divide their HTA workload among several agencies, drawing the division by health technology type. One often encountered division here is between pharmaceutical and non-pharmaceutical health technologies. This category will clarify which health technologies are assessed by a particular HTA programme.

**Criterion 5** specifies the method adopted by a particular HTA body. HTA programmes usually adopt one of the following methodological approaches: the primary research approach or the secondary approach (or a mix of the two). The primary approach involves conduct of original research and generation of new evidence by the HTA programme itself. The secondary approach involves a systematic review of the *existing information* regarding a particular technology; no new studies or research are generated by the HTA programme itself.

**Criterion 6** outlines the types of evaluata that are assessed by a particular HTA body. Both academic literature on health technology assessment and the HTA programmes themselves heavily emphasise the *evidence-based* nature of HTA programmes. Nevertheless, what is assessed differs by programme, and sometimes even by technology (that is, a particular HTA programme’s assessment of technology A may evaluate X and Y, while its assessment of technology B may evaluate X and Z). The category here lists the evaluata that the HTA programme *claims* to assess.  

---

44 The focus of this dissertation is on the national level HTA bodies, however, as previously discussed, HTA bodies operate at a variety of levels, both national and sub-national. The criterion is therefore included here for completeness.

45 I will consider the issue of what is *actually assessed* by these programmes, and the issues around conceptualising evidence, here, in subsequent chapters.
Criterion 7 considers whether the HTA programme puts forth policy advice together with its assessment report. As became evident in chapter two’s discussion of several of the provincial HTA programmes in Canada (as well as in the discussion of the UK’s approach to HTA), some HTA programmes deliberately separate out their assessment process from the policy advice or policy formulation process. On the other hand, remaining programmes collapse the two elements of the process. This criterion will therefore categorise HTA programmes according to this distinction. Where an HTA programme issues policy advice, it will be labelled as a prescriptive one; where it does not do so, it will be labelled a descriptive one.

Criterion 8, finally, describes the influence of the HTA programme. Although both the academic literature and the HTA programmes themselves consistently aver that the function of HTA programmes is to influence health policy (see Chapter 1), the concept of ‘influence’ is not usually clearly fleshed out. Consequently, it is often unclear whether the HTA programme does successfully carry out its function. To avoid this problem, ‘influence’ here will be construed strictly to mean whether the HTA programme’s assessment has any impact specifically on funding decisions. Programmes will be categorised as one of the following types: direct influence, indirect influence, limited/no influence. Programmes will be categorised as ‘directly influencing’ where their assessment feeds directly into the funding decisions made regarding a particular technology; they will be categorised as ‘indirectly influencing’ where their assessment forms indirect input into the funding decision, and as limited/no influence where the assessment does not link to the funding decisions in a readily discernible manner.
3.3. Comparing CADTH, MSAC and NIHR

Drawing on the descriptions in chapter 2 of the national level HTA programmes in Canada, Australia and the UK – CADTH, MSAC and NIHR, respectively – these programmes are categorised in Table 1, below.

Table 1: Comparing CADTH, MSAC and NIHR

<table>
<thead>
<tr>
<th>Comparators</th>
<th>Jurisdiction</th>
<th>Canada</th>
<th>Australia</th>
<th>United Kingdom</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTA programme</td>
<td>CADTH</td>
<td>MSAC</td>
<td>NIHR</td>
<td></td>
</tr>
<tr>
<td><strong>Context: Health System</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Criterion 1:</strong> Centralisation</td>
<td>Decentralised</td>
<td>Centralised</td>
<td>Centralised</td>
<td></td>
</tr>
<tr>
<td><strong>Criterion 2:</strong> Funding model</td>
<td>Public</td>
<td>Public and (well-developed) private</td>
<td>Public and (minimal) private</td>
<td></td>
</tr>
<tr>
<td><strong>HTA Programme</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Criterion 3:</strong> Level</td>
<td>National</td>
<td>National</td>
<td>National</td>
<td></td>
</tr>
<tr>
<td><strong>Criterion 4:</strong> Scope</td>
<td>Drugs, devices, systems and services</td>
<td>New medical techs and procedures (outside of PBAC and PDC remit)</td>
<td>Drugs, devices, mental care, techniques, screening programmes</td>
<td></td>
</tr>
<tr>
<td><strong>Criterion 5:</strong> Method</td>
<td>Secondary</td>
<td>Secondary</td>
<td>Primary and Secondary</td>
<td></td>
</tr>
<tr>
<td><strong>Criterion 6:</strong> Evaluata</td>
<td>Clinical effectiveness, cost-effectiveness and broader impact of health technologies</td>
<td>Safety, clinical effectiveness, cost-effectiveness</td>
<td>Safety, clinical effectiveness and cost-effectiveness</td>
<td></td>
</tr>
<tr>
<td><strong>Criterion 7:</strong> Policy Advice</td>
<td>No; descriptive</td>
<td>Yes; prescriptive</td>
<td>No (delegated to NICE); descriptive</td>
<td></td>
</tr>
<tr>
<td><strong>Criterion 8:</strong> Influence on funding decisions</td>
<td>Little/None</td>
<td>Yes, direct</td>
<td>Yes, indirect</td>
<td></td>
</tr>
</tbody>
</table>

Categorising these HTA programmes according to this scheme permits us to conduct a systematic side-by-side comparison on a number of key criteria, and consequently, to discern some of the relationships between them. Let us therefore consider the three HTA programmes against each criterion, in turn.
3.4. Detailed comparison of CADTH, MSAC and NIHR

3.4.1. Comparison of Criterion 1: Centralisation

Criterion 1 describes the first of two key characteristics of the health system within which the HTA programme is located, namely, the centralisation (or lack thereof) of the health system.

Side-by-side comparison reveals that two of the three comparators – Australia and the United Kingdom – have centralised health systems, while Canada’s system is decentralised. Centralisation indicates that the decisions regarding the funding of particular health technologies by the health system are made at the national level (as is the case in UK and Australia); decentralisation, conversely, suggests that those decisions are made by sub-national (provincial or territorial) governments (as is the case in Canada). This has two inter-related consequences.

The first consequence of centralisation arises vis-à-vis the vigorousness of HTA activity at the sub-national levels. In jurisdictions with centralised health systems, the decisions regarding the funding of particular health technologies are made at the national level. Consequently, the national level HTA is more likely to be connected to the policy-makers making these funding decisions. In jurisdictions with decentralised health systems, on the other hand, the national level HTA cannot be connected to funding decisions made at the national level simply because funding decisions are not made at the national level; instead, those decisions are made by the state, provincial, or territorial governments. Consequently, those provinces, states or territories are likely to establish their own HTA agencies, for the purpose of conducting health technology assessments that reflect those jurisdictions’ particular demographic or budgetary needs, for example. This trend seems to be borne out by all 3 jurisdictions. As discussed in chapter 2, both centralised jurisdictions – England and Australia – evince fairly limited HTA activity at sub-national level. Conversely, Canada – the only decentralised jurisdiction examined here – enjoys a fairly vigorous level of HTA activity at provincial level.

The second consequence follows from the first, and pertains to the influence of the national level HTA programmes. Recall, ‘influence’ is construed here as the impact that an HTA
programme has on funding decisions. In a decentralised health system, a national level HTA programme’s influence is likely to be lower than its counterpart’s in a centralised system. This is because – unless the national level HTA agency is the only (or perhaps the main) HTA agency in the country, the provinces, territories or states in charge of making their own funding decisions are likely to establish their own HTA agencies – the second consequence of decentralisation. The provincial-, state- or territorial-level HTA agencies – by virtue of their proximity to the local policy-makers – are thus likely to dilute the national level agency’s influence.

3.4.2. Comparison of Criterion 2: Funding model

**Criterion 2** describes the second of the two main characteristics of the health system within which the HTA body is situated: that is, the jurisdiction’s approach to funding of its health care system. The main categorisations possible here are: public funding, private funding, or a combination of the two.

One of the main reasons for selecting Canada, Australia and the UK as the comparators for this dissertation was an interest in comparing how *publicly funded* health care systems approach the HTA process at the national level. Consequently, all three jurisdictions examined here share the public approach to funding of health system. The differences with respect to this criterion therefore arise only with regard to the role of the private health services and expenditure. While Canada’s legislation prohibits private provision of medically necessary services, thus eliminating the possibility of establishing a parallel private system, a parallel private system is operational and well-developed in Australia, and a nascent parallel private system is in existence in the UK.

It is possible that an HTA process operating at the national level in a wholly publicly funded system would differ from an HTA process operating at the national level in a wholly privately funded one. The most plausible difference might arise with regard to the evaluative behaviour of the HTA approaches; it is possible that cost-effectiveness and budget impact considerations would play a diminished role, and safety and effectiveness (and perhaps other, additional considerations) would play a larger one. However, because all three jurisdictions examined here are (deliberately chosen) to be instances of publicly funded health systems, it is
impossible in this dissertation to make this particular comparison; at most, a comparison may be made between an HTA operating in a wholly publicly funded system and an HTA operating in a mixed public/private system. However, as not all health systems in the world are publicly funded ones – and those health systems where purchase of private health insurance is required by law, such as The Netherlands or Switzerland (Dhalla, 2007) do have their own HTA agencies – completeness necessitated the inclusion of this criterion in the taxonomy.

3.4.3. Comparison of Criterion 3: Level

Criterion 3 elaborates the level at which the HTA programme under examination is located. The main categorisations possible here are national and sub-national (with the latter possibly further differentiated into provincial/state/territorial, regional or institutional)

Because the comparator HTA programmes focused on here were deliberately selected to be the national level programmes, all HTA programmes under consideration here share this characteristic. Nevertheless, the categorisation is not entirely straightforward as subtle differences arise with regard to these three programmes. Australia’s MSAC is quite unproblematically categorised as a national level agency insofar as it is both situated within the federal Department of Health and Ageing and it is often explicitly identified as such both by the government reports and in the academic literature (see, e.g., Government of Australia 2009a, Government of Australia 2009b, Walley, 2007). As discussed in chapter 2, UK’s NIHR HTA programme is both funded and managed by the National Health Service. Because, moreover, NIHR’s HTAs form the backbone of the evaluation carried out by NICE (which in turn is binding on the National Health Service), it, too, is fairly unproblematically classified as a national level HTA. A somewhat challenging categorisation here is Canada’s CADTH. This is because CADTH is not located within Canada’s federal Department of Health (i.e. Health Canada). Nevertheless, CADTH is funded by both federal, and provincial and territorial governments, and, more importantly, it self-identifies as a national body, it is categorised here as a national level programme (CADTH 2011b).

The information about the level at which a particular HTA programme operates is important, because HTA processes at national level are likely to differ from HTA processes operating at sub-national level. It has been argued that HTA processes operating at the local (hospital) level,
for example, are more likely to be sensitive to the local technological needs and more reflective of community values (McGregor and Brophy, 2005). This translates into evaluative differences with respect to inputs assessed (the local programmes are more likely to likely to consider community or ethical values) and the scope of technologies assessed (narrower and more targeted to local needs), for example. Conversely, national level agencies may be more likely to have access to larger budgets, which, again, may translate into evaluative differences (e.g. the ability to assess a greater range of evaluata) and methodological differences (e.g. the ability to conduct both primary research and systematic reviews). Although comparisons between national and sub-national level HTA programmes are not possible on account of the comparators chosen here, nevertheless, interesting differences arise between HTA programmes operating at the national level, too, and it is those that are the focus here.

3.4.4. Comparison of Criterion 4: Scope

Criterion 4 describes the range of health technologies that are assessed by each health technology assessment programme.

An examination of the three comparator jurisdictions shows that Canada and the United Kingdom’s national level HTA programmes (CADTH and NIHR, respectively) assess a similar – reasonably broad – range of health technologies, which include drugs, devices, systems and services. Australia’s MSAC stipulates that it assesses ‘new medical technologies and procedures’ which prima facie appears to cast a very wide net, but one must bear in mind that – as outlined in chapter two – Australia divides its HTA workload among three national level HTA agencies: PBAC (whose remit covers pharmaceuticals and vaccinations), PDC (whose remit covers prostheses and prosthetic devices) and MSAC (whose remit covers the remaining health technologies).

What is a benefit of Canada and UK’s single-agency approach is simultaneously a drawback for Australia’s tri-agency approach, and vice versa. The benefit of dividing the entire HTA workload among three agencies consists in that this approach may lead to an increased total HTA output. On the other hand, problems can – and do – arise with respect to hybrid and co-dependent technologies – where it is uncertain under which agency’s remit a particular co-dependent or hybrid technology falls, a technology goes through an assessment process at both
agencies, resulting in a duplication of labour and a delay. The single-agency approach sidesteps this issue, but this benefit is purchased at a cost of greater workload for that agency, and, consequently, a lower HTA output by that agency.

3.4.5. Comparison of Criterion 5: Method

Criterion 5 specifies the method adopted by a particular HTA agency. The following classifications are possible here: primary methodology (conducts own research and generates own evidence), secondary methodology (systematic review of existing evidence), or a mix of primary and secondary methodology.

As was noted in chapters 1 and 2, HTA programmes generally tend to adopt the secondary – that is, the systematic review – methodology. This is partly on account of budgetary constraints and partly due to temporal constraints. In this regard, Canada and Australia are exemplars of the general trend, both adopting the secondary (systematic review) methodology. United Kingdom’s NIHR on the other hand, rather unusually among HTA programmes, relies on both the primary and secondary methodological approaches.

Adopting the primary methodology – as UK does – allows an HTA programme to fill in the knowledge gaps it identifies in its evaluative process. The ability to do so is particularly useful in context of new health technologies – including new biotechnologies, such as genetic testing or nutrigenomics – where existing evidence may be quite scant. On the other hand, adopting the primary methodology approach is both financially and temporally costly, and in refraining from doing so, Canada and Australia avoid those costs.

3.4.6. Comparison of Criterion 6: Evaluata

Criterion 6 lists the types of evaluata that a particular HTA programme assesses. As noted in chapter 1, some of the many and varied inputs considered by HTA programmes may include some or all of the following: clinical effectiveness, safety data, economic considerations (including QALY, cost-effectiveness and budget impact), socio-ethical issues, legal issues,

---

46 Recent examples of this at MSAC are Application 1221 (transurethral injection of botox into bladder wall for urinary incontinence due to neurogenic detrusor overactivity), and Application 1168, (injection of botulinum toxin for prophylaxis of headaches in adults with chronic migraine).
stakeholder views, and so on. In practice, however, the evaluative considerations of the HTA programmes’ typically tend to be limited to issues of economics, safety and clinical effectiveness.

Among the national level HTA programmes in the jurisdictions of interest, Australia’s and United Kingdom’s programmes assess identical evaluata: safety, clinical effectiveness and cost-effectiveness. Both jurisdictions are therefore in keeping with the general trend among HTA programmes – to consider a fairly narrow range of evaluata.47 Canada shares with Australia and UK their fairly narrow evaluative approaches, however, three points need to be highlighted about CADTH, here. First, just like Australia and UK, Canada’s HTA programme assesses clinical effectiveness and cost-effectiveness. Second, unlike Australia or the UK, Canada includes among its evaluative considerations the category of ‘broader impact of health technologies.’ The category is fairly broad and vague, and therefore not particularly informative. Finally, Canada’s national HTA programme does not list ‘safety’ as one of its evaluative priorities, which is rather surprising. Although ‘safety’ could arguably fall under the criterion of ‘broader impact of health technologies’, an analysis of several recent HTAs completed by CADTH reveals that this category comprises a variety of considerations, which include, for example, budget impact and impact of a health technology on patients. Safety is sometimes considered, but not particularly frequently – only in approximately 30% of HTAs.48

It must be emphasised here, however, that at this stage, these observations apply only at the most general level – that is, they apply only to the evaluata that the national level HTA programmes examined here state they consider. It remains to be seen whether these differences extend to assessments of particular technologies, which will be discussed in further detail in chapters 5 and 7 of this dissertation.

47 It is interesting to note here that although MSAC and NIHR do not share a methodological approach (the former carries out systematic reviews, whilst the latter also conducts original research), their evaluata are identical. I am not sure how to account for this.

48 A detailed examination of the evaluative contents of CADTH’s 10 most recent (as at June 2011) HTA reports is provided in chapter 5.
3.4.7. Comparison of Criterion 7: Policy Advice

Criterion 7 describes whether the HTA programme provides policy advice together with its assessment, or whether it refrains from doing so, limiting itself only to conducting assessments. Those HTA programmes that provide policy advice are categorised as prescriptive, while those that do not issue policy advice are categorised as descriptive.

On this criterion, Canada and United Kingdom share their descriptiveness, while Australia’s MSAC is categorised as an instance of a prescriptive type of an HTA body.

In the case of United Kingdom, the HTA programme is deliberately structured to separate the assessment (the process of evaluation of the evidence) from the appraisal (the process of making a decision about health technologies) stage. The intent behind this division is to protect the evidence evaluation stage from undue influence by patients, policy-makers or manufacturers. In Canada’s case, the reasons why CADTH refrains from issuing policy advice are not entirely clear, although they may be due both to a similar desire to protect the ‘objectiveness’ of its assessments and the fact that CADTH’s evaluations does not affect funding decisions (see discussion of criterion 8, below). In light of the generally received view (discussed in chapter 1) that the function of HTA programmes is to provide input into health policy, Canada’s and UK’s approaches may appear surprising. However, failure to issue policy advice need not preclude either CADTH or NIHR from ‘providing input’ into the policy process – their evaluations of the current state of evidence regarding a health technology may well feature in policy deliberations regarding the particular health technology under considerations (and, at least in UK’s case, certainly do so, during the appraisal stage conducted by NICE).

Unlike in the case of HTAs conducted in Canada or the UK, each of MSAC’s HTA reports concludes with a section containing its policy advice. The section contains a recommendation (that a particular health technology be funded, be funded on an interim basis, or not be funded), together with a reasoning behind that recommendation. The advantage of this approach is that – unlike in cases of Canada or UK – it becomes reasonably straightforward to assess the
relationship between the HTA assessment and the decision that is ultimately made by the Department of Health and Ageing.

3.4.8. Comparison of Criterion 8: Influence on funding

Finally, criterion 8 considers the influence of an HTA programme. ‘Influence’ is here construed to mean whether an HTA programme has an impact on the funding decisions regarding health technologies. Possible categorisations here are: direct influence, indirect influence, little/no influence.

On this categorisation, the entire range of options is evinced by the three HTA programmes of interest here. Canada’s CADTH is categorised as evincing little/no influence, UK’s NIHR enjoys influence, albeit an indirect one, and finally, Australia’s MSAC enjoys direct influence.

The categorisation of the influence of Canada’s CADTH is a result of that jurisdiction’s decentralised approach to health system structuring. The decisions regarding funding (or not funding) of particular health technologies rest with the provincial (and territorial) governments, many of which have their own HTA agencies, thus limiting CADTH’s influence. (There may be some indirect influence here, as those provinces that lack their own provincial HTA bodies may use the national agency as their de facto provincial agency (see chapter 2) but this is a tenuous link at best). Consequently, CADTH’s assessments are categorised here as having little/no impact on health technology funding decisions.

UK’s NIHR is categorised as indirectly influencing that jurisdiction’s funding decisions due to the afore-discussed split between the assessment stage (which is conducted by the NIHR’s HTA programme) and the appraisal or policy formulation stage (which is the remit of the NICE). The HTAs produced by NIHR are one of the key inputs into NICE’s policy decisions. Moreover, as the decisions made by NICE and the reports produced by NIHR often point towards the same conclusions, NIHR’s HTA programme is here judged to influence policy regarding the funding of health technologies, albeit indirectly.

Finally, Australia’s MSAC’s assessments are classified as directly linked to the funding decisions that are ultimately made. This is because MSAC’s terms of reference specify that its
function is to advise the Minister for Health and Ageing (i.e., the federal minister of health) on which technologies ought to be funded and under what circumstances. Indeed, MSAC’s direct influence on funding decisions is often recognised in the literature (see, e.g., Hailey, 2009, Bulfone et al, 2009; Jackson 2007).

3.5. Conclusions

To summarise, then, a comparison of the three HTA bodies of interest here clearly demonstrates that although all three HTA programmes operate at the national level, and do so within (a wholly or predominantly) publicly funded health system context, a fair degree of variety is evident with regards to: their respective scopes, methodology adopted, whether policy advice is issued, and the degree of influence on funding decisions.

The comparison here also demonstrates both the similarities and differences with regards to the *evaluata* all three bodies claim to consider in their assessments. Whilst the evaluata claimed to be considered by NIHR and MSAC are identical, CADTH both omits ‘safety’ from its evaluata and adds the ‘broader impact’ category. Yet, as was established in chapters 1 and 2, all three bodies claim that their assessments are *evidence*-based. How, then, to account for these evaluative discrepancies? The remainder of this dissertation will aim to shed some light on this issue. Towards this end, the next chapter – chapter 4 – will focus on establishing how the field of HTA understands the term ‘evidence’ in the first place.
CHAPTER 4: Conceptualising ‘Evidence’ in HTA

Summary: This chapter will offer a definition of evidence that is specific to HTA. Toward that end, it will proceed by, first, establishing the dearth of conceptual work on this issue in HTA. The chapter will then consider the current conceptualisations of ‘evidence’ in two cognate fields – evidence-based medicine (in section 2) and evidence-based health policy (in section 3). These fields offer both narrow and broader conceptualisations of evidence, and it will be argued here that a broader conceptualisation is more appropriate for HTA purposes. Specifically, the conception proposed here is that evidence is: available data relevant to the issue being addressed, question asked, decision made, etc. Main benefits and drawbacks of this approach will also be discussed.
4.1. A gap in HTA

4.1.1. What does it mean for HTA to be ‘evidence-based’?

As discussed in chapter 1, definitions of Health Technology Assessment vary somewhat from one jurisdiction to another, but the gist of the definitions is (quite consistently) that HTA is a field whose purpose is to inform health policy decisions. This view is evident both in the original definition of HTA outlined by the United States’ Office of Technology Assessment in the 1970s, and in the more recent definitions put forth by UK’s national level HTA programme, Canada’s CADTH, and the Australian government’s 2009 review of its three national level HTA agencies – as well as in the numerous statements made by various other HTA agencies.

The policy input that HTA provides is, moreover, evidence-based. This has been asserted both by and about the national level HTA programmes of all three of the jurisdictions under examination here. Thus, all three of the Australian national level HTA programmes (viz. PBAC, PDC and MSAC) are said to carry out evidence-based assessments (Government of Australia, 2009a). CADTH, similarly, asserts that its purpose is to provide decisions makers with “evidence-based information on health technologies” (Canadian Agency for Drugs and Technologies in Health, 2011), and the birth of UK’s HTA programme is specifically credited to the National Health Service’s commitment to evidence-based practice (Walley 2007).

In generating this ‘evidence-based’ ‘policy-input’, HTA programmes can potentially consider a wide range of data in making their decisions. These may include data on cost-effectiveness, safety, effectiveness, social aspects, patient and other stakeholder preferences, as well as expert opinion, ethical issues, organisational constraints, and so forth (Saarni et al 2008; Velasco Garrido et al 2010). In practice, the spectrum of data evaluated by the HTA bodies, as discussed in chapter 3, is much narrower. Usually, it includes: safety, effectiveness and cost-effectiveness, occasionally also extending to include data on patient preferences or overall impact on the health system budget. If HTA is an evidence-based process, then, presumably, at

---

49 The term ‘data’ is used here in lieu of the term ‘evidence’ (in order not to beg any questions) or the term ‘research.’ The term ‘research’ (as will be discussed shortly) is frequently used in the literature in an unclear manner. ‘Data’ is used here in the sense defined by the Oxford English Dictionary as: something known; an assumption or premise from which inferences are drawn. In pl. Facts, esp. numerical facts, collected together for reference or information (Oxford English Dictionary 2012).
least some of these inputs (or perhaps even all) count as ‘evidence.’ The question that arises here, however, is which ones – and why?

Even the narrower range of inputs considered in HTA practice involves multiple disciplines and a variety of methodological approaches. For example, effectiveness data may be obtained from a randomised controlled trial, or an observational study (e.g. a historical control study); data on cost-effectiveness may come from a cost-utility analysis (which measures the outcomes in quality adjusted life years) or a cost-benefit analysis (which measures both the benefits and the costs in dollar terms); and so on. If a yet broader range of inputs is considered, the disciplines and methodologies multiply further still. In short, a concept of ‘evidence’ that is grounded either in a particular methodology or in a particular disciplinary origin (e.g. that the data originates from economics or from clinical studies, etc.) seems like a prima facie non-starter in a multi-disciplinary field like HTA. But if not this, then what?

4.1.2. Methodology

In order to establish what needs to be done vis-à-vis adequately conceptualising ‘evidence’ in HTA, one must first establish what has already been done. Towards this aim, the process adopted here consisted of a scan of (both academic and grey) literature judged to be most likely to provide a definition of ‘evidence’ in HTA. The following sources were examined:

2. Most recent reviews of the national level HTA programme and/or national HTA strategy undertaken by each jurisdiction
3. CADTH’s, MSAC’s and NIHR’s websites
4. Glossaries of 10 most recent HTA reports produced by CADTH, MSAC and NIHR
5. International Network of Agencies in HTA (INAHTA) website’s glossary list

A more detailed justification for the selection of each of the above sources, and a discussion of the search results, is provided below.

---

50 In other words, the analysis is limited to the information contained in publicly available documents; no formal interviews were conducted with policy-makers nor confidential documents accessed.
4.1.3. The current understanding of ‘evidence’ in HTA?

4.1.3.1. International Journal of Technology Assessment in Health Care

The *International Journal of Technology Assessment in Health Care* is HTA’s main journal. It is published by the field’s chief international organisation, the International Network of Agencies for Health Technology Assessment (INAHTA). The *Journal* normally produces 4 issues per year (occasionally additional, supplementary issues are produced), and each issue typically contains 15-20 articles (with more articles frequently included in ‘theme’ issues). Issues usually contain the following sections: assessments (e.g., an evaluation of the cost-effectiveness of a particular treatment, a systematic review of a particular screening programme, etc.), discussions of methods (e.g. a description of a model for an adequate assessment of particular technologies such as tele-medicine, methods for integrating patient views into HTA, etc.), and policy discussions (e.g. on the impact of HTA on decisions made in various jurisdictions). Less frequently, editorials, letters to the editor and book reviews are also included.

A hand search was undertaken of the journal archives in January and February 2012. Full journal archives are available online at: [http://journals.cambridge.org/action/displayJournal?jid=THC](http://journals.cambridge.org/action/displayJournal?jid=THC). The hand search consisted of a scan of article titles for the terms like ‘definition,’ ‘concept,’ ‘evidence,’ and their cognates.\(^{51}\) The period searched was 1990-present (first issue of 2012), in order to overlap with the period of development of Evidence-Based Medicine (and its own conceptual advances). Articles that seemed promising were retrieved for examination. In total, 93 issues were searched, and because each issue averages 15-20 articles, the search encompassed at minimum 1400 articles. The search results are reproduced in Table 2, below.

\(^{51}\) Although further insights may have been obtained from extending the analysis to terms such as ‘decision-making,’ ‘criteria,’ and ‘frameworks,’ the overarching emphasis in the field of HTA is on its evidence-based nature, hence the focus on this particular concept here.
Table 2: Scan of the International Journal of Technology Assessment in Health Care

<table>
<thead>
<tr>
<th>Year</th>
<th>Issues</th>
<th># articles retrieved for closer examination</th>
<th># articles addressing the concept of evidence explicitly</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>1 (as at Feb 10, 2012)</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>2011</td>
<td>4</td>
<td>1</td>
<td>--</td>
</tr>
<tr>
<td>2010</td>
<td>4</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>2009</td>
<td>6 (4 regular, 2 supplements)</td>
<td>1</td>
<td>--</td>
</tr>
<tr>
<td>2008</td>
<td>4</td>
<td>1</td>
<td>--</td>
</tr>
<tr>
<td>2007</td>
<td>4</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>2006</td>
<td>4</td>
<td>2</td>
<td>--</td>
</tr>
<tr>
<td>2005</td>
<td>4</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>2004</td>
<td>4</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>2003</td>
<td>4</td>
<td>1</td>
<td>1 (Banta, 2003b)</td>
</tr>
<tr>
<td>2002</td>
<td>4</td>
<td>1</td>
<td>1 (Weatherly et al, 2002)</td>
</tr>
<tr>
<td>2001</td>
<td>4</td>
<td>1</td>
<td>--</td>
</tr>
<tr>
<td>2000</td>
<td>4</td>
<td>1</td>
<td>--</td>
</tr>
<tr>
<td>1999</td>
<td>4</td>
<td>1</td>
<td>1 (Granados, 1999)</td>
</tr>
<tr>
<td>1998</td>
<td>4</td>
<td>1</td>
<td>--</td>
</tr>
<tr>
<td>1997</td>
<td>4</td>
<td>1</td>
<td>--</td>
</tr>
<tr>
<td>1996</td>
<td>4</td>
<td>2</td>
<td>--</td>
</tr>
<tr>
<td>1995</td>
<td>4</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>1994</td>
<td>4</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>1993</td>
<td>4</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>1992</td>
<td>5 (4 regular, 1 supplement)</td>
<td>2</td>
<td>--</td>
</tr>
<tr>
<td>1991</td>
<td>5 (4 regular, 1 supplement)</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>1990</td>
<td>4</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>Total</td>
<td>87</td>
<td>16</td>
<td>--</td>
</tr>
</tbody>
</table>

The search located three sources\textsuperscript{52} that explicitly discussed the concept of evidence.

In 1999, Granados published a paper titled “Health Technology Assessment and Clinical Decision Making: Which is the Best Evidence?” (Granados, 1999). The article offers a definition of evidence both for EBM and for HTA. The definitions, however, are specifically of ‘best evidence’ rather than evidence \textit{simpliciter}. These definitions will be set aside here, and discussed in more detail in Sections 3 and 4 of this chapter, respectively.

\textsuperscript{52} A fourth source is worth mentioning here, although it was discarded as it did not contain an explicit definition of evidence or any of its cognate terms. One of the Journal’s 1997 issues was devoted to discussing the EUR-ASSESS project (previously mentioned in chapter 1). One of the articles in that issue consisted of the entirety of EUR-ASSESS’s glossary (EUR-ASSESS, 1997b). The glossary contained no entries on ‘evidence,’ or cognate terms, however.
In 2002’s paper “Using Evidence in the Development of Local Health Policies: Some Evidence from the United Kingdom,” Weatherly et al drew a distinction between internal (experiential) evidence and external (empirical) evidence. The former was defined as “based on professional opinion and tacit knowledge” whilst the latter was defined as “based on research from primary or secondary studies, published papers or guidelines” (Weatherly et al, 2002). The definitions seem to have been intended more for EBM than for HTA (although this is not completely clear), and are incomplete (insofar as they fail to indicate what is based on professional opinion, etc.). Evidence of definition tout court was not provided. Consequently, this article was discarded for the purposes of the present discussion.

Finally, in 2003, David Banta published a conceptual paper on ‘evidence’ as it pertained specifically to the field of public health policy (Banta, 2003b). In “Considerations in defining evidence for public health,” he discusses several potentially viable approaches for defining the concept. The conceptualisations he considers and the issues he raises will be discussed in more detail in Section 3, which focuses on evidence-based health policy.

4.1.3.2. Reviews of national level HTA programmes

Where available, each jurisdiction’s most recent review of its national level HTA programme and/or national HTA strategy were located and consulted, in order to establish whether any of the jurisdictions of interest here provided their own understanding of ‘evidence.’

No such document was located for UK’s NIHR. It is not clear whether this is because no such review was recently undertaken, or if it had been undertaken but is not readily locatable. Canada most recently reviewed its health technology strategy in 2004. The document itself is fairly brief (17 pages in length) and predominantly focuses on providing general policy recommendations (e.g. it describes deficiencies in ‘traditional HTA’, how to move beyond them, and so on) (Health Technology Assessment Group, 2004). No glossary is provided, nor is the definition of ‘evidence’ contained in the text of the document.

Of the three jurisdictions considered here, Australia has recently undertaken the most extensive review of its national level HTA process. This review culminated in a substantive report (338 pages) released in 2009 (Government of Australia, 2009a). The report contains a
moderately-sized and reasonably comprehensive glossary, although here, too, neither the term evidence nor any of its cognates were defined.

4.1.3.3. CADTH, MSAC and NIHR websites

The national agencies under consideration here generally failed to provide glossaries of their terms on their websites (although the terms ‘health technology assessment’ and ‘health technologies’ were generally defined, as noted in the previous chapters). Neither Canada’s CADTH nor Australia’s MSAC provided a glossary on their websites. UK’s NIHR does provide a glossary of its terms (available at: http://www.nihr.ac.uk/Pages/Glossary.aspx), although the glossary is more akin to a list of acronyms than a list of definitions. Neither ‘evidence’ nor cognate terms (e.g. evidence-based practice, evidence-based medicine, etc.) were defined.

4.1.3.4. Ten (10) most recent reports produced by CADTH, MSAC and NIHR

In June and July of 2011, 10 most recent full HTA reports from each agency were obtained (other types of publications – such as horizon scans, etc., were excluded; also excluded were reports pertaining to genetic testing, as these will be discussed separately in subsequent chapters). Each jurisdiction’s reports were examined for inclusion of a glossary, and, if present, the glossary was examined for the inclusion of the term ‘evidence’ or cognate terms. The research findings are reproduced in Table 3, Table 4, and Table 5, below.
Table 3: Scan of CADTH’s 10 most recent HTA reports for definitions of the term ‘evidence.’

<table>
<thead>
<tr>
<th>No.</th>
<th>Report Title</th>
<th>Glossary included?</th>
<th>Definition of ‘evidence’ or cognate terms?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Robot-Assisted Surgery Compared with Open Surgery and Laparoscopic Surgery: Clinical Effectiveness and Economic Analyses</td>
<td>No</td>
<td>--</td>
</tr>
<tr>
<td>2</td>
<td>Octaplas Compared with Fresh Frozen Plasma to reduce the risk of transmitting lipid-enveloped viruses</td>
<td>No</td>
<td>--</td>
</tr>
<tr>
<td>3</td>
<td>Vancomycin or Metronidazole for Treatment of Clostridium difficile Infection: Clinical and Economic Analyses</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Bariatric Surgery for Severe Obesity: Systematic Review and Economic Evaluation</td>
<td>No</td>
<td>--</td>
</tr>
<tr>
<td>5</td>
<td>Ablation Procedures for Rhythm Control in Patients With Atrial Fibrillation: Clinical and Cost-Effectiveness Analyses</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>Triple Therapy for Moderate-to-Severe Chronic Obstructive Pulmonary Disease</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Pulmonary rehabilitation for chronic obstructive pulmonary disease</td>
<td>No</td>
<td>--</td>
</tr>
<tr>
<td>8</td>
<td>Recombinant Activated Factor VII for Prevention of Bleeding Unrelated to Hemophilia: Clinical and Economic Systematic Review</td>
<td>No</td>
<td>--</td>
</tr>
<tr>
<td>9</td>
<td>Clopidogrel versus Other Antiplatelet Agents in the Secondary Prevention of Vascular Events in Adults with Cerebrovascular Disease: Clinical and Cost-Effectiveness Analyses</td>
<td>No</td>
<td>--</td>
</tr>
<tr>
<td>10</td>
<td>Long-Acting Beta2-Agonist and Inhaled Corticosteroid Combination Therapy for Adult Persistent Asthma: Systematic Review of Clinical Outcomes and Economic Evaluation</td>
<td>No</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 4: Scan of MSAC’s 10 most recent HTA reports for definitions of the term ‘evidence.’

<table>
<thead>
<tr>
<th>No.</th>
<th>Report Title</th>
<th>Glossary included?</th>
<th>Definition of ‘evidence’ or cognate terms?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Matrix Induced Autologous Chondrocyte Implantation and Autologous Chondrocyte Implantation</td>
<td>No</td>
<td>--</td>
</tr>
<tr>
<td>2</td>
<td>Radiofrequency Ablation in Barrett’s Oesophagus with Dysplasia</td>
<td>No</td>
<td>--</td>
</tr>
<tr>
<td>3</td>
<td>Review of Interim Funded Service: Brachytherapy for the Treatment of Prostate Cancer</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Middle Ear Implant for Sensorineural, Conductive and Mixed Hearing Losses</td>
<td>No</td>
<td>--</td>
</tr>
<tr>
<td>5</td>
<td>Assessment of Liver Iron by R2-MRI data analysis</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>Unattended Sleep Studies in the Diagnosis of Obstructive Sleep Apnoea</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Second Generation Contrast Agents for Use in Patients with Suboptimal Echocardiograms</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>Cryotherapy for Recurrent Prostate Cancer and Renal Cancer</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>Computer-Navigated Total Knee Arthroscopy (CATKA)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>Vagus Nerve Stimulation for Epilepsy</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
Table 5: Scan of NIHR’s 10 most recent HTA reports for definitions of the term ‘evidence.’

<table>
<thead>
<tr>
<th>Agency: NIHR</th>
<th>Report Title</th>
<th>Glossary included?</th>
<th>Definition of ‘evidence’ or cognate terms?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a tumour necrosis factor inhibitor: a systematic review and economic evaluation (Malottki).</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>Clinical effectiveness and cost-effectiveness of continuous subcutaneous insulin infusion for diabetes: systematic review and economic evaluation (Cummins).</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>Clinical and cost-effectiveness of epoprostenol, iloprost, bosentan, sitaxentan and sildenafil for the treatment of pulmonary arterial hypertension within their licensed indications: a systematic review and economic evaluation (Chen).</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Endovascular stents for abdominal aortic aneurysms: a systematic review and economic model (Chambers).</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>The effectiveness and cost-effectiveness of methods of storing donated kidneys from deceased donors: a systematic review and economic evaluation (Bond).</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>Spinal cord stimulation for chronic pain of neuropathic or ischaemic origin: systematic review and economic evaluation</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Routine antenatal anti-D prophylaxis for RhD-negative women: a systematic review and economic evaluation (Squires).</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>Structural neuroimaging in psychosis: a systematic review and economic evaluation (Albon).</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>The clinical effectiveness and cost-effectiveness of cardiac resynchronisation (biventricular pacing) for heart failure: systematic review and economic model (Bond).</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a tumour necrosis factor inhibitor: a systematic review and economic evaluation (Malottki).</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Of the 10 most recent HTA reports by CADTH, only 3 provided a glossary. None of the three glossary-containing HTAs provided a definition of ‘evidence’ or any of the cognate terms. Glossaries were more frequently included in the HTAs produced by MSAC (7 of the 10 reports included them). However, none of these 7 HTAs defined the terms of interest here. Finally, all of the HTA reports produced by NIHR examined here included glossaries (10 of 10 reports) – however, as in the previous cases, none defined the relevant terms.

4.1.3.5. International Network of Agencies in HTA (INAHTA)

The International Network of Agencies in Health Technology Assessment (INAHTA) is the largest international body in the field of HTA. INAHTA was established in 1993, and is both the publisher of the *International Journal of Technology Assessment in Health Care* and the organiser of the field of HTA’s main annual meeting, Health Technology Assessment.
International. Its membership consists of organisations from 30 countries. All of the member organisations are engaged in production of HTAs, and are generally linked to either national or sub-national governments.

On its website, INAHTA hosts a glossary whose explicit aim is to provide the HTA community with a common nomenclature (most recently updated in July 2006). The glossary itself is quite comprehensive – its pdf version spans over 50 pages. Three evidence-related concepts are defined in the glossary: evidence-based decision making, evidence-based health care, and evidence-based medicine.\textsuperscript{53} The term ‘evidence’ itself, however, is neither listed nor defined.

4.1.3.6. Cochrane Collaboration’s Handbook for Systematic Reviews

Cochrane Collaboration was established in 1993. It is a network of health care practitioners, policy-makers, patient advocates, and so forth, who prepare and disseminate systematic reviews on a wide variety of topics, ranging from breast cancer, to effective practice of organisation and care, ear/nose/throat disorders, skin diseases, etc. Over 4600 systematic reviews have already been placed in the Cochrane Library, and the \textit{Cochrane Handbook for Systematic Reviews of Interventions} (current edition: March 2011) is one of the most comprehensive documents on systematic reviews’ methodology in existence.

Due to its prominent status in the field of HTA, it was hoped that the Handbook would contain a definition of the concept of evidence; consequently, the \textit{Handbook} was searched for definitions of terms of interest. Like INAHTA’s, Cochrane Review’s glossary is very extensive (pdf version spanning over 50 pages). Nevertheless, neither the definition of

\begin{itemize}
\item \textbf{Evidence-based decision making}: “the consideration of evidence in the process of making health care decisions. Also known as ‘evidence-informed decision making.’”
\item \textbf{Evidence-based health care}: “An extension of the application of the principles of evidence-based medicine to all professions associated with health care, including purchasing and management.”
\item \textbf{Evidence-based medicine}: “The use of current best evidence from scientific and medical research to make decisions about the care of individual patients. It involves formulating questions relevant to the care of particular patients, systematically searching the scientific and medical literature, identifying and critically appraising relevant research results, and applying the findings to patients.” (INAHTA 2006). Insofar as these definitions themselves invoke the term ‘evidence,’ they are not particularly enlightening \textit{vis-à-vis} that concept (although the definition of EBM \textit{does} suggest that ‘evidence’ in that field can be understood as data coming from scientific and medical research – more discussion on this shortly)
\end{itemize}

\textsuperscript{53} The definitions are as follows: Evidence-based decision making: “the consideration of evidence in the process of making health care decisions. Also known as ‘evidence-informed decision making.’” Evidence-based health care: “An extension of the application of the principles of evidence-based medicine to all professions associated with health care, including purchasing and management.” Evidence-based medicine: “The use of current best evidence from scientific and medical research to make decisions about the care of individual patients. It involves formulating questions relevant to the care of particular patients, systematically searching the scientific and medical literature, identifying and critically appraising relevant research results, and applying the findings to patients.” (INAHTA 2006). Insofar as these definitions themselves invoke the term ‘evidence,’ they are not particularly enlightening \textit{vis-à-vis} that concept (although the definition of EBM \textit{does} suggest that ‘evidence’ in that field can be understood as data coming from scientific and medical research – more discussion on this shortly)
‘evidence’ nor any definitions of related terms were provided (Cochrane Collaboration, 2005/2010).

4.1.4. Is lack of a definition of evidence a problem for HTA?

In short, the current situation vis-à-vis the concept of evidence, as it emerges from an examination of the key sources judged the most likely to furnish such a definition, is fairly bleak. Examination of over 20 years’ worth of contents in HTA’s key journal produced 3 definitions – of which, one is specific to the field of public health policy (Banta, 2003b), one described external and internal evidence, although the definitions were not particularly illuminating (Weatherly et al 2002), and one offered an HTA-specific definition which describes ‘best evidence’ (rather than evidence simpliciter) (Granados, 1999; to be discussed shortly). Scan of the websites of the agencies of interest here (namely, CADTH, MSAC and NIHR) and of recent national level reviews of the HTA programmes, did not yield a definition of evidence; nor did a scan of each agency’s 10 most recent HTA reports. The comprehensive glossaries authored by INAHTA and Cochrane Collaboration likewise failed to define the term. In other words, although the field of HTA regards itself as an evidence-based one, it is not presently clear how the term ‘evidence’ is actually understood by those in the field.

However, to establish the gap exists is not yet to establish that the gap is in need of filling. One could argue here, for example, that HTA is a methodological rather than a conceptual field. Battista, for example, has explicitly acknowledged that the field of Health Technology Assessment has been – and continues to be – predominantly focused on the methodology for conducting assessments (Battista, 2006). Indeed, the scan documented above did convey the impression that the field of HTA overwhelmingly focuses on methodological issues. (For example, each issue of the Journal contains a ‘methodology’ section, usually comprising several articles). However, the importance of methodological issues in a field does not establish the unimportance of conceptual work in that field. Nor does an ‘is’ establish an ‘ought.’ Whilst the field may have historically focused on methodological issues to the exclusion of the conceptual issues, it need not continue to do so.

One could nevertheless further argue that the field of HTA has now been in existence for nearly 40 years, and has thus far managed to carry on without a definition of ‘evidence.’ Why,
therefore, fix something that isn’t broken? The answer to that argument, of course, is that ‘something’ is broken. The development of evidence-based fields (including evidence-based medicine and evidence-based policy) was driven by the desire for a more rational and defensible approach to decision-making (Lin, 2003; Upshur and Colak 2003). We do not know if the evidence-based field of HTA is succeeding in its aim (of being evidence-based) unless we have a metric against which to compare its current practices. Generating a definition of the concept of ‘evidence’ would constitute one of the steps necessary towards establishing such a metric, and, consequently, towards assessing whether HTA is successfully fulfilling its goal.

4.1.5. Next steps

Battista’s aforecited observation – that the field of HTA has focused predominantly on methodological issues – has led him to conclude that “it is high time for HTA to bring together aspects of conceptual and theoretical works from other fields to strengthen the foundation of HTA” (Battista, 2006; emphasis added). In this, I think, he is exactly right.

Because of the dearth of conceptual work in the field of HTA vis-à-vis the concept of evidence, the methodology that Battista suggests was adopted here. Unlike HTA, two other health-centred, evidence-based fields have expanded a fair degree of effort to clarifying the definitions of evidence – evidence-based medicine (EBM) and evidence-based-health-policy (EBHP). Section 2 will therefore offer a brief introduction to EBM and devote some space to considering the definitions of evidence that have been offered in that field. Subsequently, Section 3 will take a parallel approach vis-à-vis EBHP. Section 4, finally, will consider some of the similarities and differences between HTA and EBM/EBHP and, taking those into consideration, will adopt and adapt the definitions in EBM/EBHP for the purposes of HTA. In addition to filling in a conceptual gap clearly in need of filling, establishing the definition of ‘evidence’ for HTA will lay the groundwork for chapter 5’s and 7’s testing of the theoretical concept of evidence against HTA’s actual practices. More specifically, chapter 5 will test the conceptualisation of evidence against HTA practices more generally, and Chapter 7 will test the conceptualisation of evidence against HTA practices vis-à-vis a particular type of health technology – namely, genetic testing.
4.2. EBM and its concept of evidence

4.2.1. Background to Evidence-Based Medicine

Although the *beginnings* of the debate about the function and proper conceptualisation of evidence in medicine can be traced to the mid-1970s (Rycroft-Malone et al 2004), it was not until the early 1990s that these issues truly entered the mainstream of medicine. The term ‘evidence-based medicine’ (hereafter, EBM) itself was coined in 1992 by Sackett and his colleagues at McMaster University and rapidly spread (Lambert, 2006; Banta 2003b; Haynes 2002). By 2001, the New York Times Magazine’s ‘Year in Review’ proclaimed EBM to be one of the most influential ideas that year.

The original definition of EBM, put forth in 1992 by the Evidence-Based Medicine Working Group (which included Sackett and his McMaster colleagues) described EBM in the following way:

> Evidence-based medicine deemphasises intuition, unsystematic clinical experience and pathophysiologic rationale as sufficient grounds for clinical decision making and stresses the examination of evidence from clinical research. Evidence-based medicine requires new skills of the physician, including efficient literature searching and the application of formal rules of evidence evaluating the clinical literature (Evidence-Based Medicine Working Group, 1992).

The target of the EBM advocates was the then-dominant paradigm of medicine, which prioritised the knowledge of the basic mechanisms of diseases and the expertise of the clinical expert him- or herself (or, alternately, the collective expert opinion of professional societies) (Haynes 2000). EBM devalued those elements – as well as patient values and other ‘unsystematic’ clinical evidence – and in their place placed the emphasis on clinical research.

4.2.2. Why evidence and ‘evidence’ are important in EBM

Since its original conceptualisation, the definition of EBM has undergone some transformations, however. Its central emphasis remained the same – the emphasis on the “explicit, judicious, and conscientious use of current best evidence from healthcare research in decisions about the care of individuals and populations” (Sackett et al, 2000; also Lambert 2006, Dobrow et al 2004). However, EBM now *also* acknowledges the importance of the clinical expertise. As Sackett et al concede, clinical research cannot displace clinical expertise...
– it is, after all, the clinical expert who decides whether the clinical research is applicable to the individual patient’s care, and how that research ought to be integrated into care. Thus, Sackett et al have subsequently accepted that the practice of EBM requires an integration of the clinical research with the clinical expertise (Sackett et al, 1996; also Haynes 2002, Dobrow et al 2004, Banta 2003b).

Numerous characterisations of ‘clinical expertise’ can be found in the literature. (‘Clinical expertise’ also sometimes referred to as ‘craft knowledge’, ‘practical know-how’ or ‘practical knowledge’). Clinical expertise has been defined as, for example, the health practitioner’s possession of the information relevant to patient’s care, and the ability to apply this information (Pearson et al 2007; Banta 2003b). Sackett et al themselves have defined clinical expertise as the proficiency that clinicians acquire through clinical practice (Sackett et al 1996). Elsewhere, clinical expertise has been similarly defined as the experientially-acquired skills of clinical practice (Lambert 2006). Although these definitions evince some slight differences, they all share the focus on the professional judgement of the health care providers, and usually identify previous practice as its source.

The other key conceptual constituent of EBM, namely clinical research, likewise has benefited from close attention in the literature. The definitions of ‘clinical research’, however, tend to be more uniform than those of ‘clinical expertise.’ Generally, they follow the line of the definition offered by Sackett et al, in which clinical research is construed as originating “from the basic sciences of medicine” (Sackett et al 1996). Where further elaborations are offered, they tend to specify the types of questions that clinical research attempts to answer. Among those frequently mentioned are the efficacy and safety of therapeutics, and accuracy and precision of diagnostic modalities54 (Sackett et al 1996; Haynes 2002)

Although EBM has tended to focus on clinical research and clinical expertise, these two categories fail to exhaust the spectrum of possible probative considerations in EBM. One might also consider salient to the practice of EBM the views, values and preferences of the patients themselves, the cost-effectiveness of the medical care, and organisational factors (both equipment and health human resource constraints), for example.

54 Some issues around these definitions will be discussed shortly.
The latter two categories generally garner little, if any, attention in the EBM literature. Rycroft-Malone et al are among the few who consider the value of the information about organisational factors; they include organisational factors under the ‘local context’ category (which they list as one of the main sources of input in EBM) (Rycroft-Malone et al 2004). The issue of cost-effectiveness, on the other hand, tends to be raised primarily as a means of criticising EBM – EBM, the critics’ argument goes, will serve as the means to cost-cutting ends. In response to this, Sackett et al argue that EBM – insofar as it requires the application of the most effective interventions – may actually increase rather than decrease the care costs (Sackett et al 1996).

In contrast to cost-effectiveness and organisational issues, patient values and preferences have lately been gaining a fair bit of recognition in EBM as valuable evidentiary considerations. Neither EBM’s initial definition (Evidence-Based Medicine Working Group, 1992), nor the subsequent article clarifying “What Evidence-Based Medicine Is and Isn’t” (Sackett et al 1996) mention patient preferences. Our methods for determining patient values and preferences, moreover, continue to be – as Haynes puts it – primitive in comparison to our methods for determining the efficacy of therapeutics (Haynes 2002). Nevertheless, EBM scholars have lately argued for the inclusion of patient values and preferences as the third evidentiary pillar of EBM (see, e.g., Lambert 2006; Upshur 2000, discussed below).

As its name indicates, the field of EBM regards itself as evidence-based. Consequently, clarifying the function of evidence in the field of EBM is of vital importance. The function of evidence is justificatory – evidence supports the courses of actions chosen, positions adopted or recommendations made. Relying on evidence in EBM also serves to “reduce unnecessary variations in practice, lessen arbitrariness in the use of prescription medication and diagnostic testing, and eliminate the influence of values in decision making.” (Upshur and Colak, 2003; see also Dobrow et al 2004). These considerations are undoubtedly important – but perhaps especially so in publicly funded health systems, like the ones under consideration here. This is
because publicly-funded systems are typically underpinned by the principle of equal access to care.\textsuperscript{55} The reliance on evidence thus helps these systems to live up to that principle.

What is at stake in clearly conceptualising ‘evidence’, then? Clearly delineating the conceptual boundaries of ‘evidence’ allows us to identify data that constitute ‘evidence’ and those data which are not ‘evidence’ – and consequently, to assess whether a decision, course of action, position or recommendation are, indeed, evidence-based and hence, justified and non-arbitrary. In order to be able to do so, the concept of ‘evidence’ must be clear. Not only must the concept be clear – it must also, as Upshur et al point out, be sufficiently robust to resonate with those involved in the field of EBM (Upshur et al 2000).

\textbf{4.2.3. Conceptualising ‘evidence’ in EBM}

Although thus clearly recognised as important to the field of EBM, the concept of ‘evidence’ poses a considerable challenge. Rycroft-Malone et al (2004), for example, emphasise that clearly conceptualising ‘evidence’ in EBM is a problem, and Haynes (2002), in a similar vein, identifies that consensus on the definition of evidence as a pressing matter in EBM. Upshur, likewise, points out that “the concept of evidence has yet to be analysed systematically” (Upshur 2000). Two sets of issues underlie the debate in the literature: one, whether to adopt a broad or a narrow conceptualisation of ‘evidence,’ and two, the precise nature of the definition of evidence that ought to be adopted.

\textbf{4.2.3.1. Narrow conceptualisations of evidence in EBM}

Narrower conceptualisations of evidence in EBM are the ‘dominant orthodoxy of EBM’ (Pearson et al 2007). Emphasis is on the experimental, quantitative, and/or the scientific, and ‘evidence’ is typically identified with the findings from ‘research.’

Among those definitions that circumscribe ‘evidence’ to findings from ‘scientific research’ one may count the definition given by Pearson et al: “only the result of objective, scientific research, where validity is determined by the degree to which bias is minimised, constitutes evidence” (Pearson et al 2007, 86; emphasis added). Dobrow et al, similarly, regard as one of

\textsuperscript{55} See, e.g., chapter 2’s discussion of the key principles of Canada’s Health Act, the aim of the Australian Medicare, or NHS’s underlying objectives.
the most dominant EBM conceptualisations to be “a scientific conception of evidence – evidence developed through systematic and methodologically rigorous clinical research, emphasising the use of science” (Dobrow et al 2004; emphasis added).56 In this conceptualisation, Dobrow et al argue, the emphasis is predominantly on the rigour of the method (its validity, reliability). High quality evidence is usually held to consist in randomised controlled trials or systematic reviews thereof (Dobrow et al 2004). Somewhat more broadly, albeit still keeping to the narrower end of the spectrum of definitions of evidence, Scott-Findlay and Pollock “urge the restriction of the term evidence to research findings” – the importance of other inputs into the clinical decision-making process is recognised, however, they are not considered to be evidence (Scott-Findlay and Pollock, 2004).

There are several problems with adopting the narrow conceptualisations of evidence, however. I will elaborate here on the three main ones. First, the narrow conceptualisations of evidence fail to reflect the newer, broader definition of EBM. Second, the narrow conceptualisations rest on a common (albeit mistaken) view of science as an ‘objective’ and ‘value-free’ enterprise. Third, and finally, a proper account of ‘research’ (as well as ‘scientific research’ or ‘clinical research’) would need to be fleshed out in order to adequately ground the definitions considered above. Consider each problem in turn.

First, the narrow definitions of evidence are congruent only with the original conceptualisation of EBM, offered in 1992. Recall, that definition simultaneously emphasised the importance of clinical research and de-emphasised clinical judgement, patient values, and so forth. Subsequently, however, EBM itself has transformed, moving away from an exclusive emphasis on clinical research, and acknowledging the salience of the clinical judgement (and lately, also, patient values). This broadening of what is recognised (by the discipline itself) as salient inputs into the evidence-based medicine should also be reflected in its conceptualisation of ‘evidence.’ Indeed, although typically framed as an issue of failing to recognise reality (rather than a failure to be faithful to the field’s definition), one of the most frequent criticisms of the narrow conceptualisation of ‘evidence’ in EBM has been that it fails to consider a wide range of other inputs salient to the practice of evidence-based medicine (Dobrow et al 2004; Upshur et al 2000, Rycroft-Malone et al 2004).

56 Dobrow et al discuss this definition, but themselves do not subscribe to it.
One may, of course, concede here that clinical research, clinical expertise and patient values can be – and perhaps even should be – recognised as key inputs into evidence-based medicine. But, one could argue, only clinical research merits the label ‘evidence’ – other inputs are something other than evidence (call them ‘influences,’ or ‘inputs,’ for example). (Rycroft-Malone et al 2004, for example, advocate something like this). For this move to be legitimate, however, a principled reason is required for privileging clinical research over clinical expertise and patient values. The hint of the reasoning behind this move is suggested in the rationale offered for the shift from ‘old medicine’ to EBM – i.e., the desire to do away with the subjective, in favour of the objective. Pearson et al (2007) make the point more explicitly – EBM’s emphasis on the experimental, or scientific, or quantitative is underpinned by the desire for the ‘objective’. This perception of the experimental, scientific, or quantitative as ‘objective,’ however, is mistaken.

This brings us to the second problem with narrow conceptualisations of evidence. On a commonly held view, science is an objective enterprise (Douglas 2007; McMullin 1982). Its goal is the discovery of facts and there is nothing personal about it; water just is H2O, regardless of the scientist’s emotions, or religious or political beliefs. General population tends to subscribe to this view, regarding scientists as discoverers of ‘objective truths’; this perception is also common amongst scientists themselves, many of whom see their own work as beyond the influence of cultural or social factors (Lederman and Bartsch, 2001). Policy-makers, too, are not immune to the seductiveness of this view. The view is evident, for example, in the European Commission report which notes that science that underpins policy is one that is “isolated from contamination by social values or interests” (European Commission, 2007, 34). And, as Brian Wynne points out, this view is also evident in the United States’ National Council’s Red Book, which divides the risk governance process into the risk assessment and risk management stages. Risk assessment is “an exclusively scientific process of objective factual discovery,” whilst risk management is the stage at which the ethical, economical and subjective social values enter into consideration (European Commission, 2007, 32; emphases added). Insofar as the Red Book forms a widely-used framework for the
governance of risk, its expectation that the scientific is that which is objective and value-free pervades the regulatory and policy milieux.\footnote{Wynne identifies cases of adherence to this framework in European regulations spanning the fields of nuclear safety (CEC 1996), chemical safety (CEC 2001a), control of medicines (EMEA 2006), planning (CEC 1997) waste (CEC 2002a) and water management (CEC 2000a), as well as the regulation of food safety (CEC 2002b), and biotechnology (CEC 2001b) (European Commission, 2007). Similar understanding seems to underpin the process of risk analysis employed by Australia’s Office of the Gene Technology Regulator (see, e.g., OGTR 2002), as well as by Health Canada (see, e.g., Health Canada 2000).}

The view of science as an objective and value-free enterprise, however, has long since been shown to be untenable by the philosophers of science. This is because values – both epistemic and non-epistemic ones – have been shown to pervade both the process and concepts of science (see, e.g., Kuhn, 1977; McMullin, 1982; Brown, 1989; Longino 1990). These values enter, for example, by way of selecting which questions are asked, which questions are not asked, which effects are measured, which data is discarded, where the thresholds are set (e.g. for inclusion, exclusion, significance, etc.), the interpretation of results (and unconscious biases in that process), the relative weighing of the significance of apparently contradictory results, and so on. In short, one cannot elevate scientific (or clinical) research to a privileged probative position on the grounds that this type of research is objective or value-free – because it is neither. But if not this, then what else could possibly underpin such privileging? Absent a proper account of a property (or set of properties) that would ground this, the privileging of the scientific (or clinical) research over other EBM inputs is unmerited.

Finally, the third problem to raise vis-à-vis the narrow definitions of evidence in EBM, centres on another concept they all deploy: namely, ‘research.’ Pearson et al identify evidence with the results of “objective, scientific research.” Dobrow et al identify evidence as that which is developed through methodologically rigorous clinical research, and finally, Scott-Findlay and Pollock also restrict evidence to research findings. Whether ‘research’ is considered more generally (as in the case of Scott-Findlay and Pollock) or more specifically (e.g. as ‘scientific’ or clinical research’ as in the cases of Pearson et al and Dobrow et al, respectively), all of these definitions invoke the concept of research. In order for these definitions to be clear, however, an adequate account of ‘research’ is required. Neither Dobrow et al, nor Pearson et al, offer such an account. Scott-Findlay and Pollock seem to limit ‘research’ only to ‘scientific research’ but do not clarify further.
There are several possibilities for what is meant by ‘research’ here. One could, for example, identify ‘research’ with activities undertaken in particular domains – typically, the basic sciences or clinical science (as Sackett et al’s definition does, see Section 4.2.2.). But this is too vague – for it raises a further question – namely, which activities in these domains are included? Scientific methodologies are very diverse. Even amongst the basic sciences, some disciplines are more reliant on observation and experimentation, whilst others depend more heavily on mathematical proofs. As Edmond and Mercer correctly point out, it is more plausible to talk about “scientific methods or heuristics – plural – where only particular subsets of these apply to any one scientific specialisation” (Edmond and Mercer, 1998). Subsets – not subset – must be emphasised here.

Alternately, one may identify research strictly with randomised controlled trials and systematic reviews thereof (although these are not methodologies widely used in basic sciences, they are widely used in clinical research). Banta, for example, defines research in a manner suggestive of this view. For him, research is an “unbiased and objective process of enquiry that produces knowledge.” He further differentiates between primary research (systematic empirical enquiry) and secondary research (accumulation of research findings into a body of knowledge; typically via a systematic review) (Banta 2003b). Setting aside the aforenoted problems with the issues around objectivity, this conceptualisation of research is problematic – for several reasons. First, randomised controlled trials (and a fortiori systematic reviews thereof) are most appropriate for assessing therapeutic interventions – however, they are not particularly useful for assessing diagnostic health technologies, surgical interventions or public health interventions. Second, RCTs are not able to assess some interventions for ethical reasons (e.g. one cannot deliberately expose people to toxic substances in order to assess the effectiveness of an antidote, for example). Finally, RCTs are frequently unable to properly assess safety outcomes – their duration is often insufficient to do so adequately, and their patient numbers too limited to adequately identify rare complications (Rychetnik et al, 2002; Hynes et al 2002; Upshur 2000; Head 2010).

In short, if ‘research’ is construed as RCTs or systematic reviews of RCTs, it would be impossible to generate ‘research’ on a number of areas falling under the aegis of EBM. If, on the other hand, ‘research’ is an activity circumscribed to particular domains and/or activities,
we are owed an account of which domains, and which activities are included, given the plurality of both the domains and the activities in existence. Absent such an account, grounding the concept of ‘evidence’ in another concept that is itself unclear (viz. ‘research’), does not constitute a viable approach.

4.2.3.2. Broad conceptualisations of evidence in EBM

In light of some of the problems around the narrow conceptions of evidence, arguments have been put forth that such conceptions ought to be replaced (see, e.g., Upshur and Colak 2003; Upshur, 2000; Rycroft-Malone et al 2004). Broad conceptualisations of evidence have therefore been proposed in their place. Although the broad conceptualisations tend to retain the ‘scientific’ in some way or another, they also allow other inputs to carry probative weight. Four examples of broader definitions will be considered here: Rycroft-Malone et al’s, Granados’, and Pearson et al’s and Upshur’s (the last two of which are grouped together due to some similarities underlying the definitions offered).

a. Rycroft-Malone et al’s conceptualisation of ‘evidence’ in EBM

A definition of evidence recently offered in context of evidence-based care, stipulates that evidence is “knowledge derived from a variety of sources that has been subjected to testing and has been found to be credible” (Rycroft-Malone et al 2004; also cited in Scott-Findlay and Pollock, 2004). The ‘variety of sources,’ here, is not as broad as it would prima facie appear, because subsequently, four specific sources are identified: knowledge from research evidence, knowledge from clinical experience, knowledge from patients, clients and carers, and knowledge from local context.

This account suffers from drawbacks both vis-à-vis the definition itself, and the domains identified as the sources of the evidence. With respect to the definition, it is not ‘knowledge’ that poses a problem here, as this is clearly defined as “an awareness or familiarity gained by experience, a person’s range of information” (Rycroft-Malone et al 2004). Instead, the problematic aspects of the definition centre on the other concepts crucial to it – namely the terms ‘testing’ and ‘credible.’ It is not clear what, exactly, would constitute satisfactory ‘testing’ – or how ‘credible’ does ‘credible’ have to be – and to whom? These issues are left unaddressed, however.
It is also not clear how several of these domains would meet the testing and credibility criteria. The most intractable here are likely to be: knowledge from clinical experience, knowledge from patients, clients, and carers; and knowledge from local context. Whilst potential methods could be devised for testing and credibility assessment of clinical experience, as well as for testing and credibility assessment of knowledge from the local context, it is by no means clear that such methods currently exist. The knowledge from patients, clients and carers, moreover, is by definition subjective and thus it is difficult to see how credibility could be assessed or testing carried out. In short, although correct in encompassing ‘a variety of sources,’ this particular definition is a non-starter.

b. Granados’ conceptualisation of ‘evidence’ in EBM

Let us turn, then, to Granados’ definition. This definition – while also a definition of the ‘broad’ conceptualisation flavour – takes a different approach to Rycroft-Malone et al.’s. Granados considers “the best evidence as that which gives the most valid and objective answer to the different types of questions arising both in clinical practice and in health care resources management, and coming from the synthesis and integration of the results of scientific studies from different knowledge fields aside from health sciences—such as social, economic, and political sciences” (Granados, 1999, 586). There are several benefits and several drawbacks to this definition, so I will address each in turn.

Grounding the definition of evidence in data’s relevance to the question asked is, I think, the right approach. Consider the following (as far as I know, fictitious) bit of data: 10% of women with family history of breast cancer undergo genetic testing for BRCA1 and 2 mutations. This data is relevant to answering the question of the costs likely to be incurred by a health system considering whether to publicly fund BRCA1 and 2 genetic testing. In other words, the data would count as evidence in this case. The same data, however, would not count as evidence when, say, the health system is considering the funding of prostate specific antigen testing for men over 50. It is the same bit of data, yet its status as ‘evidence’ changes from one case to the next. The difference, of course, consists in the data’s relevance to the question asked. In other words, Granados is right in linking the definition of evidence to the issue under consideration, decision to be made, or the question asked.
The definition potentially treads on quicksand, however, insofar as it invokes ‘objectivity’ and ‘validity.’ What is meant by validity is not clear, and is left unexplained. The definition sidesteps the issues raised previously about the objectivity by implicitly acknowledging that objectivity is a matter of degree (via highlighting that data that is salient is one that is most objective rather than objective tout court). However, if something like patient preferences or values, as well as clinical judgement, are to be captured by the definition of evidence in EBM, it is not clear that the objectivity standard could constitute a part of the definition. In fairness, Granados mentions only the health sciences and social sciences (social, economic and political) as potential sources of evidence, so perhaps she means to exclude both patient values and clinical judgement. It would be, however, rather peculiar for a broad definition of evidence in context of EBM to exclude both of these elements.

c. Pearson et al’s and Upshur’s conceptualisation of ‘evidence’ in EBM

Let us consider, then, another broad type of definition identified in the literature. This type of definition links the concept of evidence to its function. Two exemplars were identified in the literature. First, Pearson et al consider the following definition of evidence: “as the available facts and circumstances supporting, or otherwise, a belief or proposition, or indicating whether a thing is true or valid” (Pearson et al 2007, 86; definition cited is originally from Pearsall and Trumble, 1995). Second, a definition in a similar vein was offered by Upshur et al; evidence was defined as “an observation, fact or organised body of information, offered to support or justify inferences or beliefs in the demonstration of some proposition or matter at issue.”

Upshur et al further differentiate four types of evidence: qualitative/personal (which centres on particular individuals, e.g. narrative); qualitative/general (evidence is primarily social); quantitative/general (evidence is primarily statistical); and quantitative personal (e.g. quality of life studies, psychology studies) (Upshur et al 2000).

Although differently phrased, the Pearson et al, and Upshur et al, definitions of evidence share with Granados’ definition the grounding of the definition in the function of evidence – to answer the question asked, issue raised, etc. The added qualification ‘support or otherwise’ (in Pearson’s case) is important because not all of the relevant information will offer a supportive or affirmative answer. (For example, a study showing, say, that vaccination rates are actually decreasing would, of course, be relevant to answering the question “Are vaccination rates
higher in publicly funded health systems than in private health systems?”, even though it does not offer an affirmative answer to the question).

Pearson et al’s stipulation that evidence establishes whether a thing is ‘true or valid’ is problematic. This is because, as Upshur notes elsewhere, in medicine, at best we have access to probabilities rather than certainties; data is inherently defeasible and subject to being superseded (Upshur, 2000).

Pearson et al do not stipulate particular domains, and thus, sidestep aforementioned worries about which methodologies within particular domains would count and which would not. Upshur et al, however, remain vulnerable to this objection. Their move here is an interesting one – rather than identifying particular domains, they specify general characteristics of the domains which would furnish the evidence. All possible domains are covered, insofar as all of the logical combinations of any two characteristics are included (that is, P and Q, P and not-Q, and so on). The worry here, however, is therefore two-fold. First, the aforementioned issue regarding which of any particular domain’s methodologies count and which do not, remains salient here. Second, the worry here is that the approach captures more (qua evidence) than it needs to, and in so doing, comes closer to a conceptualisation of evidence that is more appropriate to evidence-based health policy than to EBM.

Of all the definitions considered so far, however, Upshur et al’s definition is the most promising one. How it could be adapted and adopted for HTA purposes will therefore be considered in the final section of this chapter. Before addressing this, however, first we must take a detour into the field of evidence-based health policy, to see if further definitions – and conceptualisations – potentially adoptable and adaptable for HTA purposes could be identified there.

4.3. EBHP and its concept of evidence

4.3.1. Background to Evidence-Based Health Policy

In the last 20 years or so, the growing influence of the principles of EBM has increasingly led to the belief that other fields, too, should be founded on evidence. A number of other evidence-
based fields have therefore arisen, including: evidence-based practice, evidence-based nursing, evidence-based decision-making, evidence-based policy (hereafter, EBP), and so forth (Rycroft-Malone et al 2004). More recently, the field of evidence-based health policy (hereafter, EBHP) also emerged (Lin 2003). ‘Health policy’ has been defined as “those actions of governments and other actors in society aimed at achieving the health of the population” (Niessen et al, 2000). ‘Evidence-based health policy,’ therefore, entails carrying out these actions whilst relying on evidence.

4.3.2. Why evidence and ‘evidence’ are important in EBHP

The typical inputs into evidence-based health policy span a wider range than those typically associated with EBM. Just like in the case of EBM, research evidence is considered important; on the other hand, patient values (whilst occasionally mentioned) seem to be considered less frequently. EBHP’s counterpart to EBM’s clinical judgement – that is, the professional or ‘field’ knowledge of those delivering the programmes – has often been dismissed in the past. Lately, however, it has garnered some attention as a potentially valuable probative input into EBHP (Head 2010). Beyond those, one might also include among salient EBHP inputs: the domains of the political judgement (of politicians or other organised groups), cultural/ethical values of individuals, organisations and society as a whole (e.g. social justice, accessibility, cultural appropriateness), the experiential knowledge of service users, as well as the organisational resources (human resources, system capacity, etc.) (Head, 2010; Lin 2003; Head 2008; Banta 2003b)

The reasons offered for the importance of evidence in EBHP echo those put forth in EBM (mutatis mutandis) – that decisions ought to be made on bases other than solely the authority and the views belonging to the decision-makers. For example, Brian Head points out that the aim of evidence-based policy (including evidence-based health policy), is to generate “legitimate forms of decision-making that are alternatives to ideological or faith-based policy-making” (Head 2010). Evidence, that is, serves as a justification for the decisions made, and the expectation is that sounder health policy follows from the use of evidence (Head 2010, Lin 2003, UK Cabinet Office 1999).
4.3.3. Conceptualising ‘evidence’ in EBHP

Very much like in the case of EBM, the emphasis on the evidence-basedness in EBHP has manifested itself in the focus on the concept of ‘evidence’ in the EBHP literature. Also very much like in the case of EBM, the concept remains the subject of some scholarly disagreement, with the main disagreement echoing those previously encountered in context of EBM – whether to adopt a broad or a narrow conceptualisation of ‘evidence,’ and the precise nature of the definition of evidence that ought to be adopted.

However, before engaging in a discussion of the particular concepts of evidence proffered, as well as the general types they fall under (i.e. narrow or broad), one challenge first needs to be dismissed. It has been argued that it is not important how evidence should be defined in EBHP. Dobrow et al (2004), for example, argued that how evidence is utilised is more important than how it is defined. Banta (2003), similarly, argued that the precise nature of this concept is less important than it is to accept that evidence ought to be used in decision-making in health policy. (Banta argues specifically for a conclusion relating to the field of public health policy, but the argument seems to be equally applicable to other sub-fields of health policy).

The function of evidence is undoubtedly important. However, as was previously noted, this does not by itself show that the definition of evidence is unimportant. Moreover, the two can readily be united – the particular tradition of grounding an entity’s definition in that entity’s function enjoys a long philosophical tradition, dating all the way back to Aristotle. Far from having fallen into disuse – as many other ancient practices have – this particular tradition continues to flourish. In fact, it continues to flourish in relatively new epistemic domains – as became evident in the previous section, some definitions of evidence in EBM ground their conceptualisations of evidence in its very function.

Aside from those – admittedly, rather infrequent – invitation to dismiss the import of the concept of evidence in EBHP, the debate in the literature flourishes. As in the case of EBM, the definitions offered tend to fall either on the narrower or the broader end of the spectrum. The former tend to construe evidence as findings from research (where ‘research’ is variously fleshed out, or fails to be fleshed out). The broader definitions, on the other hand, tend to
refrain from linking the definition of ‘evidence’ to a particular domain in the definition itself, although the accounts offered often do specify salient domains. Consider each in turn.

4.3.3.1. Narrow conceptualisations of evidence in EBHP

Arguments for a narrow conceptualisation of evidence in EBHP are common but less frequent than in EBM. Generally, they tend to share the understanding of evidence with the understanding manifest in EBM – namely, that a proper understanding of ‘evidence’ requires us to restrict that concept to ‘research’. It is, at times, acknowledged that elements other than ‘research’ factor into decisions made in EBHP, but those other factors are generally demoted to labels like ‘information’ or ‘knowledge’ or ‘additional influences.’ (see, e.g., Scott-Findlay and Pollock, 2004). ‘Research’ here is variously understood, although the precise definition is typically left unclear.

Among the exemplars of such definitions one may count that offered by Miles et al (1999; see also Miles et al 2000 and Banta 2003b), who hold that evidence is the findings of research studies. Somewhat more circumscribed are the definitions of ‘evidence’ which stipulate that evidence in EBHP is the findings of scientific or clinical research. This view is evident, for example, in the writings of Weiss, who conceptualises evidence as information derived from scientific research (Weiss, 1983). Other definitions also include findings from both clinical research and social scientific research. Rychetnik et al, for example, have defined evidence as “compris[ing] the interpretation of empirical data derived from formal research or systematic investigations, using any type of science or social science methods.” (Rychetnik et al, 2002). Finally, the concept of evidence that underpins the evidence-based policy (including health policy) has also at times been grounded in ‘applied research.’ Brian Head, for example, adopts this view of evidence, defining the latter as “the knowledge generated by applied research” (Head 2008). Under ‘applied research,’ Head includes “general evidence about broad trends and explanations of social and organisational phenomena, as well as specific evidence generated through performance indicators and program evaluations”(Head 2008). This is not a particularly common understanding of applied research as it seems unduly narrow; it is not clear, for example, why Head does not seem to permit ‘applied research’ to take place within the natural sciences.
Many of the problems with the narrow definitions of evidence in context of EBHP mirror those identified with regard to the narrow definitions considered in context of EBM, so they will only be recapitulated briefly, here. First, three of these definitions rely on the concept of ‘research’ which itself is potentially problematic, yet typically is left undefined. The problems centre primarily around the fact that ‘research’ potentially captures a lot of activities in a variety of epistemic domains, so it is not clear what exactly is meant here.

Second, the specification that that ‘research’ is scientific or social scientific or clinical, muddies, rather than clarifies, the waters. This is because what counts as research within these domains, again, encompasses many different activities. In clinical area alone, ‘conducting research’ could be any one of the following activities: performing a systematic review, a randomised controlled study, a cohort study, case-control study, interrupted time series, historical control study, and so forth. Head’s approach, to grounding his definition in the concept of ‘applied research’ seems prima facie to sidestep the problem, but the particular definition he adopts for his purposes is rather unusual, and, in the end, fails to bypass the problems around relying on the term ‘research.’

The final problem with the narrow definitions of evidence in context of EBHP is that they are incongruent with the practices of the field. As discussed above, in EBHP, the spectrum of inputs into the decisions is broad – even more so than in EBM. Failing to reflect that breadth of inputs requires an account of why some types of inputs are elevated to the status of ‘evidence’ while others are not. This account has not been forthcoming, although in all likelihood, the same reasoning underpins this as in the case of EBM – the perception of ‘scientific’ as objective and therefore ‘better.’ This view, albeit pervasive, is mistaken, as was outlined in Section 2.

**4.3.3.2. Broad conceptualisations of evidence in EBHP**

In short, the overall problem with the narrow definitions of evidence is that – despite their narrowness – these definitions are quite vague. As in the case of EBM, this has led some in the field of EBHP to argue that a proper conceptualisation of ‘evidence’ in EBHP is a broad one (see, e.g., Banta 2003b, Lin 2003, Miles et al 1999). This is generally argued on the grounds
that evidence based health policy decisions are a complex amalgam of multiple inputs, and a broader concept of ‘evidence’ is necessary in order to reflect this (Lin 2003, Head 2008; Banta 2003b; Velasco Garrido et al 2010; Lavis et al 2002). Several examples of broad conceptualisations of evidence in EBHP are discussed below; what these definitions have in common is their focus on the function of evidence.

**a. Banta’s conceptualisation of ‘evidence’ in EBHP**

For example, Banta defines evidence as “any useful information that serves as a basis for making decisions” (Banta, 2003b, 562). Elsewhere, and in a somewhat similar vein, Dobrow et al fail to offer a definition of evidence – as noted before, they think that the function of evidence takes precedence over its definition – but stress that the key aspect of evidence is its applicability to a specific purpose (Dobrow et al 2004). As discussed previously, grounding the concept of ‘evidence’ in its function (answering a particular question, addressing a specific issue, offering justification for a particular decision, and so forth) is, I think, the most promising one. It accounts for the fact that the same data may vis-à-vis one issue, and not-evidence vis-à-vis another.\(^{58}\) I will return to this in Section 4.

**b. Lin’s conceptualisation of ‘evidence’ in EBHP**

Vivian Lin puts forth a definition that is quite similar to Banta’s and Dobrow et al’s definitions, considered above; she defines evidence in the context of policy-making as “anything that establishes a fact or gives a reason for believing something” (Lin, 2003). Unlike Banta and Dobrow et al, however, she further specifies the ‘three rationalities’ that are the key determinants of health policy. These include: cultural rationality, political rationality, and technical rationality. Cultural rationality reflects ethics, values, and social opinions regarding health policy; political rationality focuses on issues like transparency of the process and accountability of the policy-makers; finally, technical rationality is the ‘research domain.’ Presumably, then, Lin’s definition is not ‘anything’ that establishes a fact, but rather, ‘anything drawn from one of these domains.’ Identifying specific domains as sources of evidence is potentially problematic, however – the issue previously raised, regarding which methodologies count and which do not, remains salient. In fairness to Lin, given her definition, it seems that

\(^{58}\) See the discussion of Granados’ definition in the EBM section
she wishes to include ‘anything’ only from these three domains. Nevertheless, specifying salient domains before identifying a particular policy question, issue, or problem, is risky business as different questions (or issues or problems) will require evidence drawn from different domains. Moreover, it is not entirely clear how some of the key inputs in health policy (e.g. economic inputs, organisational issues, or professional ‘field’ knowledge) fit into these ‘3 rationalities.’

c. Head’s conceptualisation of ‘evidence’ in EBHP

Brian Head takes a similar approach to Lin, insofar as he identifies the potential domains from which evidence may originate. He does not stipulate a definition of ‘evidence’ but because he also emphasises that ‘evidence’ is what helps us to answer ‘what works’ and ‘what happens if we change these settings’ – that is, he emphasises the functional understanding of ‘evidence’ – his conceptualisation of evidence is included under the ‘broad’ category. He argues that there is not one evidence base (typically construed as research or science) but rather, there are several bases (which he refers to as the ‘three lenses’ for understanding policy inputs). The three lenses are: (1) political know-how; (2) scientific and technical analysis; and (3) practical and professional field experience (Head 2008).

Lens (1), the political knowledge lens, is the knowledge of the political actors (he includes here the individual politicians and political parties, as well as other organised groups and public media). Lens (2) is the scientific (research based) knowledge. For Head, ‘scientific’ knowledge is not limited to natural sciences; instead, it can originate from a large number of fields, which include natural sciences, social sciences, law, and public administration. He acknowledges the methodological variety of these domains, but stresses that the unifying characteristic is that all of these disciplines and methodologies they deploy are predominantly focused on the quality and the consistency of the data they produce. Finally, lens (3), is the professional or practical ‘field knowledge.’ This is the knowledge of those implementing the health policy programmes – those working ‘on the ground’ so to speak. The ‘knowledge’ here is their practical ‘know-how,’ or, in other words, the knowledge of issues around implementation of health policies in the local context (Head 2008).
The problem with Head’s conceptualisation is one that is shared with the other definitions relying on the concept ‘research’ and thus need not be reiterated here. It is worth highlighting here, however, that (unlike all of the previously considered conceptualisations of evidence) Head acknowledges the methodological variety of the domains he identifies as sources of evidence. Moreover, Head does capture the domains which seemed to have been left out under Lin’s approach (viz., the organisational issues and professional ‘field’ knowledge – the economic is less clearly captured, although possibly so, insofar as Head includes social sciences among his domains). On the other hand, he shares with Lin the general problem around the stipulation of particular domains – as noted, it is at least peculiar to specify sources of evidence before a particular question is asked, an issue raised, or a decision put forth, as each of those will have different evidentiary needs. As previously noted, the emphasis on the function of evidence – that is, its role in answering questions such as ‘what works’ is, however, a promising one.

4.3.4. In summary

To summarise then, although far being problem-free, the broader approach to conceptualising ‘evidence’ is more promising than the narrow one. As became apparent in the discussion of the definitions put forth in EBM and EBHP, however, it is preferable to refrain from invoking intuitively seductive but conceptually problematic terms (such as ‘objective,’ ‘scientific’ or ‘research’). It is also better to avoid stipulating particular domains as sources of evidence, because different issues are likely to require evidence drawn from different domains. On the other hand, an adequate definition of evidence must capture the domains recognised as salient in the field (e.g., economics, etc.). The next section will therefore offer and defend a broad conceptualisation of evidence that both meets these criteria and is appropriate for the HTA purposes.

4.4. Conceptualising ‘evidence’ in HTA

4.4.1. Contrasting HTA with EBM and EBHP

As became clear in Section 1, the concept of ‘evidence’ thus far has not benefitted from much attention in the field of HTA. Therefore, the state of conceptual work on ‘evidence’ in the
fields of EBM and EBHP was assessed. Those two fields, in particular, were judged as most promising for the purposes here, because both are evidence-based and both are centred around health – just like HTA.

There are, however, some important differences between these fields. The most salient ones are, first, those around both the number and the particular types of inputs into the decision-making process, and second, the levels at which decisions apply (as well as the degree of certainty about these decisions).

First, in EBM, the range of inputs has been historically quite narrow, limited only to clinical research, although lately this has broadened to also include patient values and clinical judgement. In EBHP, the range of inputs is quite a bit broader, encompassing ‘research,’ professional ‘field’ knowledge, political judgement, cultural and ethical values, knowledge of service users, etc. HTA falls somewhere in between EBM and EBHP. In theory, HTA can consider data on: cost-effectiveness, safety, effectiveness, social aspects, patient and other stakeholder preferences, as well as expert opinion, ethical issues, organisational constraints, etc. In practice, however, it is typical for HTA bodies to focus predominantly on safety, effectiveness and cost-effectiveness (as was discussed in chapters 1 and 2).

Beyond the number of inputs typically considered, differences between EBM, EBHP and HTA, also arise with regard to the particular domains from which these inputs are drawn. The most notable difference here is the salience of political considerations (relevant in EBHP, not so in EBM, and arguably very little in HTA insofar as the latter constitutes an input into policy rather than itself being a policy field). The second key difference here centres on the salience of economic input: EBM seems to pay it little heed, while both HTA and EBHP regard it as quite salient. (In fact, some jurisdictions go as far as to legally mandate a cost-effectiveness assessment as part of the HTA process (Niessen et al 2000; see also the discussion in chapter 1).

The second key distinction between EBM, EBHP and HTA concerns the level at which the decisions apply. EBM operates on the level of the individual patient – it is, to borrow from Harari, “a science of individuals” (Harari, 2001; see also Lambert 2006, Haynes 2002, Dobrow
et al 2004). EBHP, on the other hand, operates on the population level; decisions made affect the entirety (or large portions) of the population and the health system (Morgan, 2010, Dobrow et al 2004). In this regard, HTA is more similar to EBHP than to EBM – it assesses issues at system rather than individual level (although it is worth reiterating here that HTA is an *input* into health policy; it is not itself a policy-making field). The implication of the ‘level’ at which the field operates is this: as we move from individual to population level, the issues increase in complexity (Sindall 2003; Dobrow et al 2004). What follows from this is that uncertainties multiply.59

### 4.4.2. Proposed definition of evidence

#### 4.4.2.1. Existing definitions of evidence

Before proposing a definition of evidence that is appropriate to the field of HTA, a discussion of the two HTA-specific definitions of evidence currently in existence, is in order.

The first of these is the aforementioned (see 4.1.3.1) definition proposed by Granados. Her definition reads as follows: “In the HTA framework the best evidence should actually be the one able to answer the uncertainties of decision makers” (Granados, 1999, 586). This definition is not particularly illuminating, as it defines ‘best evidence’ rather than evidence simpliciter. However, this definition does two things right: it implicitly emphasises the importance of differentiating between better quality and lesser quality evidence, and it emphasises that ‘evidence’ is frequently uncertain. As noted above, this issue becomes more salient as the decisions shift from the individual patient to the population level, so Granados is right to emphasise this in context of HTA.

The second definition of evidence in the HTA literature was identified not through literature searches, but rather through personal contact with the authors of MSAC’s *Guidelines for the assessment of diagnostic technologies*. That document contains a glossary, which, unusually, provides a definition of ‘evidence.’ The definition stipulates that evidence is “data about the effectiveness of a new treatment or intervention derived from studies comparing it with an appropriate alternative. The evidence is preferably derived from a good-quality randomised

---

59 It needs to be flagged here, that uncertainty is a problem even for EBM, which operates at the individual patient level. This is because medical knowledge is inherently fallible and provisional in nature (Upshur and Colak 2003; Upshur 2000).
controlled trial, but it may not be” (MSAC 2005). The definition is problematic in that it limits ‘evidence’ only to data about effectiveness; even on the narrow range of inputs that is considered by HTA bodies in practice, the concept of ‘evidence’ should also encompass safety and cost-effectiveness. On the other hand, the definition rightly recognises that RCTs are not particularly well suited as sources of data about diagnostic technologies; the flexibility to account for the fact that different types of health technologies require different types of evidence is necessary in an HTA-specific definition of evidence.

4.4.2.2. Proposed definition of evidence

Bearing all of this in mind, I think the appropriate definition of ‘evidence’ for HTA purposes should stipulate that evidence is: available data relevant to the issue being addressed, question asked, decision made, etc. Let us parse each constituent of the definition in turn.

First, how are we to understand ‘data’ here? The understanding is in keeping with that indicated in the Oxford English Dictionary: something known, a premise from which inferences are drawn (Oxford English Dictionary 2012). The term ‘research’ is deliberately being avoided here in light of the problems identified in both the EBM and EBHP sections.

Why stipulate that data be ‘available’? This is meant to draw out a distinction between available data and conceivable data. This is because ‘best conceivable data’ is a standard that is not realistically attainable. This standard is particularly problematic in areas such as complex public health interventions, environmental health interventions, toxicology, etc. But it is also difficult to attain in many other areas due to temporal, fiscal and other practical limitations. The stipulation ‘available,’ therefore, allows us to make evidence-based decision under real (rather than ideal) conditions.

The stipulation of data’s relevance is meant to invoke what is generally regarded as the function of evidence. As noted previously, the function of evidence is to support (or otherwise) a decision, or a matter at hand. Not all data supports a decision – some tells against a particular decision. The phrasing ‘relevant’ is meant to capture both possibilities whilst maintaining the emphasis on the role of evidence.
Finally, the phrasing ‘issue being addressed, question asked, decision made, etc.’ is meant to be interpreted broadly. This is because HTA assessments are undertaken for a variety of reasons and purposes: some are meant to generally assess the state of the evidence at a particular point in time, some are meant to address a general policy question (e.g. is particular technology X funded by other jurisdictions?), while others still are conducted to answer a jurisdiction-specific question (should we – jurisdiction A – fund technology Z?) and so on. The broad phrasing allows these types of assessments to be captured as evidence-based activities.

4.4.2.2. Benefits of the proposed approach

There are several benefits to conceptualising evidence as available data relevant to the issue being addressed, question asked, etc.

First, this conceptualisation recognises that evaluations of different technologies will require different types of data. As noted previously, with respect to the clinical domain, assessments of therapeutic health technologies typically require RCT data, whilst assessments of diagnostic technologies and public health interventions draw on observational studies. Similarly, assessments of some health technologies will require data on QALYs, while others will require a cost/benefit analysis. In invoking the relevance criterion, the proposed definition is flexible enough to capture that.

Second, this conceptualisation accounts for the fact that technologies that undergo HTA assessments are at various stages in their development cycle – and consequently, that varying amounts of data about them exist. Few studies will have been done and little data will be available if a technology is a new one, and, conversely, a lot more data will be available for a technology later in its development cycle. The stipulation that data be available permits HTA bodies to conduct an assessment of both new and more entrenched health technologies, regardless of where they are in their development cycle, whilst keeping to the (reasonable) expectation that decisions ought to be evidence-based.
Third, as discussed in chapter 1, definitions of ‘health technologies’ adopted by the various HTA models range from fairly narrow to rather quite broad. As Velasco Garrido et al (2010) note, at their narrowest, ‘health technologies’ typically include pharmaceuticals, devices and procedures; somewhat more broadly, they may also include service delivery models, whilst at their broadest, they may also encompass health-promotion interventions outside of the health system (e.g. educational campaigns). The phrasing ‘issue being addressed, question asked, or decision made’ is sufficiently broad to permit carrying out evidence-based assessments of the entire range of ‘health technologies’ – from narrowest to broadest.

Fourth, this conceptualisation also takes into consideration the previously raised issue that the same data is not ‘evidence’ in all circumstances. To reiterate, data on the rates of penetrance of a particular genetic disease (Huntington’s, say) will count as ‘evidence’ in an assessment of the likely cost of publicly funding the genetic counselling services; the same data, however, will not be ‘evidence’ vis-à-vis an evaluation of the costs incurred by the system for vaccination of infants. The grounding of the concept of evidence in the data’s relevance to the issue under consideration, decision being made, etc., takes this into account.

Fifth, and finally, the stipulation of particular domains as sources of data is deliberately avoided here. This is because, as noted in context of both EBM and EBHP, different decisions (issues, questions) will have different evidentiary needs, and stipulating particular domains as salient sources of data prior to the particular question being asked (issue raised, etc.) runs the risk of encompassing too many domains, too few, or ones that are entirely inapplicable. (While it is likely that some domains will recur in many assessments, they certainly will not in all recur in all assessments – and the proper definition of evidence ought to capture all and only the relevant domains.) Thus, the choice here was made to forego the stipulation of specific domains, in favour of the ‘relevance’ criterion.

4.4.2.3. Some challenges for the proposed approach

Whilst clearly enjoying some advantages, this account is not without potential challenges. Two particularly pressing ones need to be addressed here.
The first challenge centres on the issue of relevance. More specifically, the question is, how is ‘relevance’ established – and by whom? I think the answer here requires an approach parallel to that adopted by EBM (where an individual physician decides whether clinical research is relevant to a particular patient’s case) and EBHP (where those who implement programs assess the relevance of a particular intervention to their local circumstances). That is, it is the members of the HTA community who should be the ones to resolve the issue of the relevance of data to a particular assessment – on a case by case basis. These members include: the members of the HTA bodies, policy-makers who make decisions on the basis of the HTA input, scholars in the field, and so on. Decisions about relevance may be made in light of existing guidance documents from the policy-makers (to whom the HTA reports are submitted), in light of the assessors’ previous experience, other background knowledge, familiarity with the constraints of the HTA model they operate under (e.g. its level – national or sub-national, the health system funding model, etc.), and so on.

The second challenge here centres around how to differentiate between high quality and low quality evidence. Unarguably, some evidence is of better quality than other evidence, but the definition proposed here is itself unable to clearly ground this distinction. I think, however, that this problem is not as intractable as it prima facie appears. This is because it is not the function of the definition of evidence to provide the means of drawing out this distinction. Notice, for example, that none of the definitions considered either in the context of EBM or in the context of EBHP offered a way of cashing this distinction out. It is, typically, the hierarchies of evidence that are deployed for this purpose.60 In short, although it certainly is the case that a means of differentiating between higher and lower quality of evidence is required, this is not a task that must be accomplished by the definition.

4.4.3. A brief note about the function of evidence in HTA

As noted above, the EBM and EBHP literature – unlike the HTA literature – have devoted a fair amount of attention to the definition of evidence. The case is similar vis-à-vis the function

---

60 Hierarchies of evidence are not without their problems, however. With regard to HTA specifically, existing hierarchies target primarily the studies assessing a health technology’s effectiveness; they are not adequate for assessing safety, economic, and other inputs typically considered in HTA. Some of the other criticisms targeted at the hierarchies more generally centre on: the absence of evidence for the hierarchies themselves (Dobrow et al 2006); hierarchies’ exclusion of normative/values considerations (Lambert 2006); and the multitude of hierarchies in existence and their inconsistence with each other (Upshur 2003).
of evidence. Thus, in context of EBM, Upshur and Colak, for example, note that “using the best available research evidence would place clinical decision-making on a more objective basis” (Upshur and Colak, 2003, 284; emphasis added). Similar views are apparent in the EBHP literature. Dobrow et al, for example, point out that evidence “is applied to support or justify a decision” (Dobrow et al 2004, 214; see also Dobrow et al 2006), and Head, in a similar vein, emphasises that the primary goal of the evidence-based policy is to “improve the reliability of advice concerning […] policy settings and possible alternatives” (Head, 2010, 82).

In short, the view that emerges from the EBM and EBHP literature is that the role of evidence is justificatory – evidence underpins objective (or non-arbitrary) decisions, conclusions, guidelines, recommendations, and so forth.

Unlike in the case of EBM and EBHP, the picture that emerges from the HTA literature is far less clear. Fortunately, the HTA agencies make reasonably clear what they take the function of evidence to be. CADTH’s website, for example, notes that its evidence-based reviews are key to making informed decisions in health care (CADTH 2009a), and its most recent 2010/11 Annual Report reiterated that CADTH’s dissemination of evidence “promotes the optimal use of drugs and other health technologies in Canada” (CADTH 2010). MSAC makes similar claims. It states that its evidence-based advice informs Australian Government’s decisions about health technologies (MSAC 2011), and, moreover that its “evaluation of evidence associated with medical services has been an integral part of the process for the listing of new medical technologies and services on the Medicare Benefits Schedule” (MSAC 2011, emphasis added). In a similar vein, NIHR’s most recent, 2010/11 Annual Report notes that its Technology Assessment Reports, produced for NICE, inform the latter’s guidelines issued for the NHS, and “provide decision-makers with the best possible information about the effects of tests, treatments and other interventions used in health and social care” (NIHR 2011). In short, the HTA agencies’ view – similarly to the views propounded in the EBM and EBHP literature – is that the function of evidence in their evidence-based assessments is to ground informed decisions – or, in other words, to play a justificatory role.
4.5. Conclusion

To briefly summarise, then, there is currently a gap in the HTA literature regarding the conceptualisation of ‘evidence.’ As this gap is gradually being filled in both EBM and EBHP, the literature in those fields was examined. The ‘state of play’ in both fields is remarkably similar – the debate around ‘evidence’ centres, in both fields, around whether to adopt a narrower or a broader definition, and the precise nature of the definition that ought to be adopted. Narrower definitions in both fields tend to share the two main problems: they typically rely on a problematic (yet usually left unclear) concept of ‘research’ to ground them, and they are at odds with the practices in both fields. The broader approach to conceptualising ‘evidence’ was therefore adopted here, as the more promising one for the purposes of HTA. The particular definition proposed here construes evidence to be: available data relevant to the issue being addressed, question asked, decision made, etc. This definition seems to bypass some of the problems regarding the broad definitions put forth in EBM in EBHP, identified in sections 2 and 3, whilst simultaneously enjoying numerous advantages. Thus, the definition seems prima facie promising. However, to more firmly establish its tenability, it needs to be tested against the actual evaluative practices in the field of HTA – and the next chapter will focus on precisely this task.
Summary: Chapter 4 offered a definition of evidence that is specific to HTA purposes. It proposed that, in HTA, ‘evidence’ should be understood as: available data that is relevant to the issue being addressed, question asked, decision made, etc. The aim of the present chapter is to test the tenability of this definition. The chapter will therefore consist of two parts. In the first part, I will establish how each of the three HTA bodies under examination here understands the concept of ‘evidence’ (both in theory and in practice). Because, as was established in chapter 4, the HTA agencies’ understanding is not made explicit, a contextual approach will be adopted here. In the second part, each programme’s construal of evidence will be tested for coherence with the definition proposed in chapter 4. I will argue that the agencies’ theoretical understanding of ‘evidence’ is insufficiently clear to assess its coherence with the proposed conceptualisation of evidence, whilst the agencies’ practical understanding of ‘evidence’ is captured to a fair degree by the proposed conceptualisation of evidence. The focus in this chapter will be on health technologies generally; the issues around a particular type of health technology – namely, genetic testing – will be addressed in chapter 7.
5.1. Canada: CADTH’s evaluation of health technologies

5.1.1. CADTH’s health technology assessment process

CADTH’s standard health technology assessment process consists of the following steps:

1. A research team is formed, consisting of individuals with expertise in the topic being evaluated. Experts typically include: physicians, allied health professionals, epidemiologists, health economists, pharmacologists and bioethicists.
2. A protocol describing search strategies for clinical and economic literature, and its evaluation, is established.
3. Data is gathered, analysed and interpreted.
4. A report which summarises the findings is written.
5. The report undergoes peer-review by external experts (clinicians, economists, methodological experts, etc.).
6. The report is modified in light of received comments, if required.
7. The report is disseminated via CADTH website and other knowledge exchange strategies. (CADTH 2012)

5.1.2. Method for establishing CADTH’s understanding of ‘evidence’

In order to establish how CADTH conceptualises ‘evidence’, CADTH’s 10 most recent, sequential, full health technology assessments were identified in June and July 2011. Care was taken to exclude evaluations of genetic testing, as those will be considered separately in chapter 7. Reports that were included in the present analysis are listed below, in Table 6. The health technologies evaluated by each report are categorised according to CADTH’s stated evaluative remit, which includes: drugs, devices, diagnostic agents, equipment, and medical and surgical procedures (hereafter, M&S procedures). The date that each report was published by CADTH is also noted in the table.

Because, as was noted in chapter 4, neither CADTH itself nor any of its reports explicitly identify what they mean by evidence, a context-based approach was adopted here. To this end, the ‘Objectives’ section of each report was identified and analysed, as it generally contained an implicit suggestion of the report’s construal of ‘evidence.’ Typically, this construal was identified from statements like “this report assessed evidence on clinical and cost-effectiveness of technology X” or others that utilised similar phrasing. The construal of ‘evidence’ that was

---

61 Reports other than full health technology assessments (e.g. rapid reviews, horizon scans, etc.) were excluded. CADTH search was updated in Dec 2011, in order to replace an erroneously included ‘rapid review’ with a full health technology assessment.
suggested by the report’s Objectives section was in each case confirmed with the contents of the body of the report.

Table 6: CADTH’s 10 most recent HTA reports: evaluata

<table>
<thead>
<tr>
<th>No.</th>
<th>Report Title</th>
<th>Type of health technology</th>
<th>Date</th>
<th>Focus of assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Robot-Assisted Surgery Compared with Open Surgery and Laparoscopic Surgery: Clinical Effectiveness and Economic Analyses</td>
<td>Medical and surgical (M&amp;S) procedure</td>
<td>2011</td>
<td>clinical effectiveness, cost-effectiveness</td>
</tr>
<tr>
<td>2</td>
<td>Octaplas Compared with Fresh Frozen Plasma to Reduce the Risk of Transmitting Lipid-Enveloped Viruses</td>
<td>Unclear</td>
<td>2011</td>
<td>cost-effectiveness, budget impact</td>
</tr>
<tr>
<td>3</td>
<td>Vancomycin or Metronidazole for Treatment of Clostridium difficile Infection: Clinical and Economic Analyses</td>
<td>Drug</td>
<td>2011</td>
<td>clinical effectiveness, cost-effectiveness, budget impact</td>
</tr>
<tr>
<td>4</td>
<td>Bariatric Surgery for Severe Obesity: Systematic Review and Economic Evaluation</td>
<td>M&amp;S Procedure</td>
<td>2010</td>
<td>clinical effectiveness, safety, economic implications</td>
</tr>
<tr>
<td>5</td>
<td>Ablation Procedures for Rhythm Control in Patients With Atrial Fibrillation: Clinical and Cost-Effectiveness Analyses</td>
<td>M&amp;S Procedure</td>
<td>2010</td>
<td>clinical effectiveness, cost-effectiveness, impact (of procedure) on patients</td>
</tr>
<tr>
<td>6</td>
<td>Triple Therapy for Moderate-to-Severe Chronic Obstructive Pulmonary Disease</td>
<td>M&amp;S procedure</td>
<td>2010</td>
<td>clinical effectiveness, cost-effectiveness, impact on health services</td>
</tr>
<tr>
<td>7</td>
<td>Pulmonary Rehabilitation for Chronic Obstructive Pulmonary Disease</td>
<td>M&amp;S procedure</td>
<td>2010</td>
<td>clinical effectiveness; cost-effectiveness; health services impact</td>
</tr>
<tr>
<td>8</td>
<td>Recombinant Activated Factor VII for Prevention of Bleeding Unrelated to Hemophilia: Clinical and Economic Systematic Review</td>
<td>Drug</td>
<td>2010</td>
<td>clinical effectiveness; safety; efficacy; cost-effectiveness</td>
</tr>
<tr>
<td>9</td>
<td>Clopidogrel versus Other Antiplatelet Agents in the Secondary Prevention of Vascular Events in Adults with Cerebrovascular Disease: Clinical and Cost-Effectiveness Analyses</td>
<td>Drug</td>
<td>2009</td>
<td>clinical effectiveness, cost-effectiveness; optimal duration of treatment [safety?]</td>
</tr>
<tr>
<td>10</td>
<td>Long-Acting Beta2-Agonist and Inhaled Corticosteroid Combination Therapy for Adult Persistent Asthma: Systematic Review of Clinical Outcomes and Economic Evaluation</td>
<td>M&amp;S procedure</td>
<td>2009</td>
<td>clinical efficacy, safety, and cost-effectiveness</td>
</tr>
</tbody>
</table>
5.1.3. How CADTH understands ‘evidence’

On the basis of the above-listed 10 reports, the following conclusions can be drawn.

5.1.3.1. Health technologies evaluated

Of the 10 reports, the majority (6/10) evaluated medical and surgical procedures. A substantial minority evaluated pharmaceuticals (3/10), and the categorisation of one health technology assessment (report 2, octoplas) is unclear.

CADTH’s evaluative mandate includes a wide range of health technologies, encompassing: pharmaceuticals, devices, diagnostic agents, equipment, and M&S procedures. Therefore, it is rather surprising that the agency’s evaluative efforts seem to focus predominantly on M&S procedures and pharmaceuticals. Although the sample under analysis (10 reports) is rather small, insofar as these 10 reports span 3 years of the agency’s work, the sample is likely to be representative of CADTH’s evaluative endeavours.

5.1.3.2. Possible evaluata versus stated evaluata

As was noted in chapter 2, the potential evaluata in health technology assessments may include: clinical research (e.g. effectiveness, safety), economic data, social aspects, legal aspects, ethical issues, organisational issues, stakeholder and patient views, political judgements, and expert opinion.

Against this broad range of possible evaluata, CADTH states that it considers health technologies’ clinical effectiveness, cost-effectiveness and broader impact (CADTH 2009a). Thus, the oft-levied charge by the scholars of HTA, that the HTA bodies’ focus is on an unduly narrow sub-set of relevant information, applies to CADTH. CADTH’s ‘broader impact’ category is conceptually vague, although, as established in chapter 2, it does not translate into a consideration of ethical and social issues; part of the aim of this chapter is to establish what, exactly, this category actually encompasses.

---

62 See, e.g., ten Have (1995); DeJean et al (2009); Banta (2009)
5.1.3.3. Stated evaluata versus actual evaluata

CADTH states that it assesses health technologies’ clinical effectiveness, cost-effectiveness and their broader impact (CADTH 2009a). But what actually happens in practice?

An analysis of the 10 reports, listed above, suggests the following. Ten (10) of the 10 reports assess clinical effectiveness (and/or clinical efficacy), and cost-effectiveness (and/or economic implications of the technology). In short, the first two stated evaluata match CADTH’s practice quite well. The ‘broader impact of health technologies,’ however, is a peculiar category. The category itself is not clearly defined by CADTH, so it could potentially encompass ‘impact’ on a variety of levels, including the individual patient level, community level, regional level, provincial level or the federal level. Alternately, it could capture a variety of input-types, including: financial, physical, quality of life, or yet another type of impact.

Regardless of what is intended to be captured by the ‘broader impact’ category, CADTH’s actual practices suggest that the category is plausibly assessed only 50% of the time. Among the 10 reports listed above, what may possibly fall under the ‘broader impact’ category was assessed in only 5 reports, which assessed: budget impact (2 of the 10 reports), impact of the health technology on health services (2 of the 10 reports), and impact of the procedure on the patients (1 of the 10 reports).

In short, a perfect consistency obtains between the inputs that CADTH claims to assess and its actual assessment practices vis-à-vis clinical effectiveness and cost-effectiveness. On the other hand, only a 50% consistency obtains between CADTH’s stated evaluata and its actual evaluative practices with regard to the ‘broader impact of health technology’ category.

5.1.3.4. Evidentiary considerations in addition to the stated ones?

What also needs to be flagged here regarding CADTH’s evaluative conduct is the peculiar case of ‘safety.’ Safety is quite often recognised as one of the key considerations in the HTA process (see chapter 1 and 2). It is therefore somewhat odd that CADTH does not list this domain among its stated evidentiary evaluata. Nevertheless, CADTH does, occasionally assess
safety in practice. Of the 10 reports above, 3 HTAs assessed the safety issues around the technology under evaluation (possibly 4, if the ‘optimal duration of treatment’ is categorised as a safety issue).

5.1.3.4. Consistency of evidentiary practices within a single health technology type?

One might suspect here that the discrepancy between the stated evaluata and the evaluata assessed in practice is due to the fact that a variety of health technology sub-types are being assessed by these 10 reports. After all, pharmaceuticals differ from diagnostics, procedures, and so forth – it is unreasonable to expect that the same inputs would be assessed for all of these health technology types. Perhaps then, if we were to consider only a single type of health technology, the discrepancies identified above might disappear?

To test this hypothesis, let us consider M&S procedures, insofar as they constitute the largest single assessed health technology type (6 of 10 reports). All 6 reports assessed clinical effectiveness (and/or efficacy) and cost-effectiveness (and/or economic implications of the health technology). The inconsistency with respect to the third assessed domain, viz. ‘broader impact of health technology’ remains, however. Of the 6 reports that assessed M&S procedures, 2 assessed impact on health services, and 1 assessed impact of the procedure on the patient. Thus, the ‘broader impact’ category remains considered only about one-half of the time. (Safety, again, is assessed but only infrequently – 2 of 6 reports on M&S procedures assessed it). In short, the inconsistency in CADTH’s practices cannot be explained by appeal to different health technology sub-types – discrepancies remain between stated and assessed inputs even when the assessed health technologies are limited to a single type.

5.1.3.5. In summary

To briefly sum up, then, it appears that CADTH construes ‘evidence’ as data originating from specific domains. CADTH states that those domains are: clinical effectiveness, cost-effectiveness, and broader impact of health technologies. (I.e. CADTH’s theoretical understanding of evidence is as of data on health technologies’ clinical effectiveness, cost-effectiveness and broader impact). Analysis of CADTH’s actual practices reveals that although clinical effectiveness and cost-effectiveness are consistently assessed, the ‘broader impact of
health technologies’ category is considered only 50% of the time. In addition, although not stated as an assessed domain, ‘safety’ is assessed by CADTH approximately one-third of the time. Thus, in practice, CADTH’s practical understanding of ‘evidence’ is as of data on effectiveness, cost-effectiveness, greater impact of health technology (some of the time) and safety (some of the time).

5.2. Australia: MSAC’s evaluation of health technologies

5.2.1. MSAC’s health technology assessment process

MSAC’s assessment procedure for the inclusion of any new health technology on the listing of publicly funded health technologies (Medical Benefits Schedule, or the MBS), is as follows:

1. An application is made to the Medical Services Advisory Committee (MSAC) for assessment of a health technology
2. MSAC appoints an expert group to conduct an extensive review of the available literature and other sources of evidence63
3. MSAC considers the report produced by the expert group, and makes a recommendation to the Minister for Health and Ageing regarding whether the item should be publicly funded
4. Minister decides whether or not to accept MSAC’s recommendation and, if accepted, arranges for the placement of the health technology on the MBS (RCPA 2008).

5.2.2. Method for establishing MSAC’s understanding of ‘evidence’

In order to establish how MSAC understands ‘evidence’ in its health technology assessments, 10 most recent health technology assessment reports available on the MSAC website in June and July 2011 were identified. The reports selected for inclusion were sequential; the only exception was an exclusion of any health technology assessments of genetic tests, as those will be considered separately in chapter 7. The reports are listed below, in Table 7. The table categorises health technologies according to the stated remit of Australia’s national level HTA programmes, that is: medicines; diagnostics, devices, equipment and supplies; medical and surgical procedures (hereafter, M&S procedures); support systems; and organisational and managerial systems (hereafter, O&M systems). (In light of Australia’s three-agency split of its

---

63 The group contracted to carry this out is usually (although not always) one of the following: University of Sydney’s NHMRC Clinical Trials Centre (which includes a Systematic Reviews and Health Technology Assessment team), Health Technology Assessment Group at Deakin University, Adelaide Health Technology Assessment (located at the University of Adelaide), or Royal Australasian College of Surgeons (more specifically, its Australian Safety and Efficacy Register of New Interventions and Procedures – Surgical group, i.e. ASERNIP-S)
national level HTA workload, all prostheses and prosthetic devices captured by this
description fall under PDC’s evaluative remit, whilst pharmaceuticals are the domain of
PBAC, as discussed in chapter 2). The date that each report was published by MSAC is also
given in the table.

As established in the previous chapter, MSAC does not clarify how it understands the concept
of evidence. However, each MSAC HTA report includes a Conclusion section which typically
includes the following statement: “On the strength of the available evidence of safety,
effectiveness and cost-effectiveness, MSAC does [or does not] support public funding for
[technology xyz]” (emphasis added). This was interpreted to mean that MSAC’s theoretical
understanding of ‘evidence’ is as of data on safety, effectiveness and cost-effectiveness. This
construal was subsequently verified against the contents of the full report in each case (i.e.
whether data on safety, effectiveness and cost-effectiveness, were indeed assessed).

Table 7: MSAC’s 10 most recent HTA reports: evaluata

<table>
<thead>
<tr>
<th>No.</th>
<th>Report Title</th>
<th>Type of health technology</th>
<th>Date</th>
<th>Focus of assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Matrix Induced Autologous Chondrocyte Implantation and Autologous Chondrocyte Implantation</td>
<td>M&amp;S Procedure</td>
<td>2010</td>
<td>safety, clinical effectiveness, cost-effectiveness, budgetary impact</td>
</tr>
<tr>
<td>2</td>
<td>Radiofrequency Ablation in Barrett’s Oesophagus with Dysplasia</td>
<td>M&amp;S Procedure</td>
<td>2010</td>
<td>safety, clinical effectiveness, cost-effectiveness, budgetary impact</td>
</tr>
<tr>
<td>3</td>
<td>Review of Interim Funded Service: Brachytherapy for the Treatment of Prostate Cancer</td>
<td>M&amp;S Procedure</td>
<td>2010</td>
<td>safety, clinical effectiveness, cost-effectiveness, budgetary impact</td>
</tr>
<tr>
<td>4</td>
<td>Middle Ear Implant for Sensorineural, Conductive and Mixed Hearing Losses</td>
<td>Device</td>
<td>2010</td>
<td>safety, clinical effectiveness, cost-effectiveness, budgetary impact</td>
</tr>
<tr>
<td>5</td>
<td>Assessment of Liver Iron by R2-MRI Data Analysis</td>
<td>M&amp;S Procedure</td>
<td>2010</td>
<td>safety, clinical effectiveness, cost-effectiveness, budgetary impact</td>
</tr>
<tr>
<td>6</td>
<td>Unattended Sleep Studies in the Diagnosis of Obstructive Sleep Apnoea</td>
<td>Diagnostics</td>
<td>2010</td>
<td>safety, clinical effectiveness, cost-effectiveness, budgetary impact</td>
</tr>
<tr>
<td>7</td>
<td>Second Generation Contrast Agents for Use in Patients with Suboptimal Echocardiograms</td>
<td>Unclear</td>
<td>2010</td>
<td>safety, clinical effectiveness, cost-effectiveness, budgetary impact</td>
</tr>
<tr>
<td>8</td>
<td>Cryotherapy for Recurrent Prostate Cancer and Renal Cancer</td>
<td>M&amp;S Procedure</td>
<td>2009</td>
<td>safety, clinical effectiveness, cost-effectiveness</td>
</tr>
<tr>
<td>10</td>
<td>Vagus Nerve Stimulation for Epilepsy</td>
<td>M&amp;S Procedure</td>
<td>2008</td>
<td>safety, clinical effectiveness, cost-effectiveness</td>
</tr>
</tbody>
</table>
5.2.3. How MSAC understands ‘evidence’

On the basis of the above-listed 10 reports, the following conclusions can be drawn.

5.2.3.1. Health technologies evaluated

The majority of MSAC’s evaluative work focuses on medical and surgical procedures (7 of 10 reports), with the remainder divided between devices (2/10) and diagnostics (1/10). On account of Australia’s split of its national level HTA work between three agencies (namely, PBAC, MSAC and PDC), one would not expect to see in Table 7 either pharmaceuticals (which fall under PBAC’s remit), or prostheses and prosthetic devices (which fall under PDC’s remit) – and, indeed, this is the case.

Nevertheless, the following comment about MSAC’s scope is in order. MSAC’s remit is, in theory, quite wide – it encompasses: diagnostics, devices, equipment and supplies; medical and surgical procedures (M&S procedures); support systems; and organisational and managerial systems (O&M systems). Insofar as the sample here spans 3 years of MSAC’s work (and so, can be taken to be representative of MSAC’s actual practices), it is therefore surprising to see this list so strongly dominated by the M&S procedures.

5.2.3.2. Possible evaluata versus stated evaluata

The potential inputs into HTA programmes may include: clinical research (e.g. effectiveness, safety), economic data, social aspects, legal aspects, ethical issues, organisational issues, stakeholder and patient views, political judgements, and expert opinion.

MSAC’s current terms of reference state that MSAC is to “Advise the Minister for Health and Ageing on [...] the strength of evidence in relation to the comparative safety, effectiveness, cost-effectiveness and total cost of the medical service” (MSAC 2010; emphasis added). (It needs to be noted here, that these terms of reference are effective as of 2010; previous terms of reference omitted the mention of the ‘total cost of the medical service’). Like many other HTA agencies, then, MSAC can also be charged with a fairly narrow evaluative focus.
5.2.3.3. Stated evaluata versus actual evaluata

MSAC thus *states* that it considers the following elements: safety, effectiveness, cost-effectiveness and the total cost of the medical service (the last as of 2010 only). But what actually happens in practice?

Analysis of the 10 reports reflects that MSAC’s practices exactly map onto its stated evaluata as indicated in the Terms of Reference. That is, pre-2010, all of the analysed reports assessed: safety, effectiveness and cost-effectiveness, in accordance with its then-current Terms of Reference. The subsequent reports (subject to the broadened Terms of Reference enacted in 2010), also consistently adhere to the evaluative inputs stipulated in the Terms by, additionally, assessing the impact of the health technology on the budget. It is also worth noting, here, that in order to maintain this consistency, MSAC provides an HTA report template to the groups that conduct evidence assessments on its behalf. Each template is pre-populated with a table of contents and section headings, together with a set of instructions for completing each section; the evaluative domains noted in MSAC’s Terms of Reference (viz., safety, effectiveness, cost-effectiveness, budget impact) are included in that template, and adherence thereto is expected by MSAC.

5.2.3.4. Evidentiary considerations in addition to the stated ones?

MSAC’s evaluative consistency is therefore unsurprising but still remarkable. All of its assessments are faithful to its terms of reference, focusing on – and only on – the safety, effectiveness, cost-effectiveness, and budgetary impact of the health technology under assessment.

5.2.3.5. Consistency of evidentiary practices within a single health technology type?

As apparent in Table 7, the consistency between stated inputs into the HTA and the inputs that are assessed in practice holds across various technology types – in other words, whether MSAC’s assessment is of an M&S procedure, a diagnostic, or yet another health technology, the evaluata remain the same.
5.2.3.6. In summary

In summary, MSAC construes ‘evidence’ as data originating from stipulated domains. MSAC states that those domains are: safety, effectiveness, cost-effectiveness and budgetary impact. (I.e. MSAC’s theoretical understanding of evidence is as of data on safety, effectiveness, cost-effectiveness and budgetary impact). An analysis of MSAC’s actual practices over 10 reports (spanning a period of 3 years), reveals a remarkable degree of consistency in this regard – no matter what the particular technology type (diagnostic, device, procedure, etc.), MSAC’s actual evaluata are the ones stated. Thus, MSAC’s practical construal of evidence matches its theoretical construal thereof.

5.3. UK: NIHR’s evaluation of health technologies

5.3.1. NIHR’s health technology assessment process

As noted in chapter 2, the evaluation process for health technologies in the UK is – rather unusually – split into two stages: the assessment stage (NIHR’s remit) and the appraisal stage (NICE’s remit).

This process proceeds as follows. The NIHR commissions a review (referred to as a Technology Assessment Report, or TAR) from one of several Technology Assessment Review Centres. The TARs produced by the review centres identify, assess and synthesise the research evidence regarding the health technology in question (NIHR 2012). The NIHR itself oversees the Review Centres’ work; the oversight consists primarily in the monitoring of deadlines and provision of peer-review and editorial-review process.

Once the TAR itself is completed, and the NIHR completes its review process, NIHR submits the TAR to the NICE. Whereas NIHR’s remit is the health technology assessment, NICE’s remit is the health technology appraisal. In addition to the TAR itself, NICE considers comments from a variety of organisations (including, e.g., professional organisations, health

---

64 These include: Aberdeen Health Technology Assessment Group, Liverpool Reviews and Implementation Group, Peninsula Technology Assessment Group, Sheffield School of Health and Related Research, Southampton Health Technology Assessment Centre, York Centre for Reviews and Dissemination (CRD) and Centre for Health Economics (CHE), Warwick Evidence Review Group, Kleijnen Systematic Reviews Ltd and BMJ Evidence Centre.
organisations, patient groups, and manufacturers). On those bases, NICE ultimately produces a guidance document regarding the integration of the evaluated technology into the NHS. The guidance document is binding on the NHS and has to be implemented within three months of its issue (Stevens and Milne, 2004).

In the interest of comparing health technology assessment programmes to health technology assessment programmes, the focus here will be on NIHR.

5.3.2. Method for establishing NIHR’s understanding of ‘evidence’

Given the lack of explicit indication of how NIHR conceptualises evidence, its understanding of ‘evidence’ was established in the following manner. Ten (10) most recent, sequential reports labelled ‘NICE Technology Assessment Report’ were identified on NIHR’s website in June and July 2011. As in the case of CADTH and MSAC, assessment reports pertaining to new biotechnologies (in particular, genetic testing) were excluded. The remaining identified reports are listed below, in Table 8, which categorises each health technology according to NIHR’s stated evaluative remit, which includes: drugs, devices, procedures, settings of care, and screening technologies. The date that each report was published by NIHR is also provided in the table.

An approach similar to the one adopted with respect to CADTH was utilised here. That is, the ‘objectives’ section of each Technology Assessment Report was identified and analysed for statements suggesting what the report construed as ‘evidence.’ The suggested construal was subsequently verified against the contents of the body of the report. No cases of discrepancies (between the objectives and the body of the report) were identified.
<table>
<thead>
<tr>
<th>No</th>
<th>Report</th>
<th>Type of health tech</th>
<th>Date</th>
<th>Focus of assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a tumour necrosis factor inhibitor: a systematic review and economic evaluation</td>
<td>Drug</td>
<td>2011</td>
<td>clinical effectiveness, cost-effectiveness</td>
</tr>
<tr>
<td>2</td>
<td>The clinical effectiveness and cost-effectiveness of topotecan for small cell lung cancer: a systematic review and economic evaluation</td>
<td>Drug</td>
<td>2010</td>
<td>clinical effectiveness, cost-effectiveness</td>
</tr>
<tr>
<td>3</td>
<td>Clinical effectiveness and cost-effectiveness of continuous subcutaneous insulin infusion for diabetes: systematic review and economic evaluation</td>
<td>Procedure</td>
<td>2010</td>
<td>clinical effectiveness, cost-effectiveness</td>
</tr>
<tr>
<td>4</td>
<td>Clinical and cost-effectiveness of epoprostenol, iloprost, bosentan, sitaxentan and sildenafil for the treatment of pulmonary arterial hypertension within their licensed indications: a systematic review and economic evaluation</td>
<td>Drug</td>
<td>2009</td>
<td>clinical effectiveness, cost-effectiveness</td>
</tr>
<tr>
<td>5</td>
<td>Endovascular stents for abdominal aortic aneurysms: a systematic review and economic model</td>
<td>Device</td>
<td>2009</td>
<td>clinical effectiveness, cost-effectiveness</td>
</tr>
<tr>
<td>6</td>
<td>The effectiveness and cost-effectiveness of methods of storing donated kidneys from deceased donors: a systematic review and economic evaluation</td>
<td>Procedure</td>
<td>2009</td>
<td>clinical effectiveness and cost-effectiveness</td>
</tr>
<tr>
<td>7</td>
<td>Spinal cord stimulation for chronic pain of neuropathic or ischaemic origin: systematic review and economic evaluation</td>
<td>Procedure</td>
<td>2009</td>
<td>clinical effectiveness, cost-effectiveness</td>
</tr>
<tr>
<td>8</td>
<td>Routine antenatal anti-D prophylaxis for RhD-negative women: a systematic review and economic evaluation</td>
<td>Procedure</td>
<td>2009</td>
<td>clinical effectiveness, cost-effectiveness</td>
</tr>
<tr>
<td>9</td>
<td>Structural neuroimaging in psychosis: a systematic review and economic evaluation</td>
<td>Procedure</td>
<td>2008</td>
<td>clinical effectiveness, cost-effectiveness</td>
</tr>
<tr>
<td>10</td>
<td>The clinical effectiveness and cost-effectiveness of cardiac resynchronisation (biventricular pacing) for heart failure: systematic review and economic model</td>
<td>Procedure</td>
<td>2007</td>
<td>clinical effectiveness, cost-effectiveness</td>
</tr>
</tbody>
</table>
5.3.3. How NIHR understands ‘evidence’

On the basis of the above-listed 10 reports, the following conclusions can be drawn.

5.3.3.1. Health technologies evaluated

As is readily apparent, the majority of NIHR’s assessments focus on procedures (6/10 reports), and a substantial minority of its assessments are of pharmaceuticals (3/10). One device was also assessed. Insofar as the sample here (10 reports) spans over 4 years of NIHR’s work, it may be taken to be representative of NIHR’s evaluative practices.

5.3.3.2. Possible evaluata versus stated evaluata

As previously noted, the potentially salient inputs into the HTA process, in theory, can span a fairly broad spectrum. They can include inputs such: clinical research (e.g. effectiveness, safety), economic data, social aspects, legal aspects, ethical issues, organisational issues, stakeholder and patient views, political judgements, and expert opinion.

Against these potential inputs into the HTA process, the range of inputs that NIHR claims to consider is rather narrow; NIHR states that its evaluative considerations include: cost-effectiveness, clinical effectiveness and safety of the health technologies (Walley 2007; NIHR undated).

5.3.3.3. Stated evaluata versus actual evaluata

How do NIHR’s stated evaluata (viz. safety, effectiveness and cost-effectiveness) map onto its evaluative practices, however? The 10 reports analysed here show that NIHR’s practices closely match its stated evaluata – with regard to two of these three domains. All 10 reports assess clinical effectiveness and cost-effectiveness. However, not one of the reports analysed actually assessed the safety of the health technology, despite NIHR’s claim that safety is included among the inputs it assesses.
5.3.3.4. Evidentiary considerations in addition to the stated ones?

No inputs in addition to the stated ones (viz., safety, effectiveness, cost-effectiveness) were considered across the 10 reports. (As noted above, however, in practice, NIHR assesses less than it claims to assess – evaluation of safety was omitted from all of the reports.)

5.3.3.5. Consistency of evidentiary practices within a single health technology type?

No variation is evident (with respect to the evaluata assessed) across the various types of health technologies that undergo NIHR’s assessment – in other words, whether a device, a pharmaceutical, or a procedure is assessed, NIHR’s actual evaluata remain: clinical effectiveness and cost-effectiveness.

5.3.3.6. In summary

NIHR, thus, quite clearly understands ‘evidence’ to be data on safety, effectiveness and cost-effectiveness of health technologies – at least in theory. A closer examination of NIHR’s actual assessment practices reveals that although the latter two categories are universally assessed, the first one (viz. safety) is not. Regardless of the health technology type under assessment, only its effectiveness and cost-effectiveness are assessed, whilst safety is not. Thus, NIHR’s practical understanding of evidence is narrower than its theoretical one, including only effectiveness and cost-effectiveness. The assessment, although limited to 10 reports, spans a period of 4 years, and can thus be thought to be representative of NIHR’s actual practices.

5.4. ‘Evidence’ in HTA: theory vs. practice

This section compares the practices of the three HTA programmes under examination, focusing in particular on their theoretical and practical understanding of the concept of evidence and comparison thereof to the conceptualisation of evidence proposed in chapter 4.
5.4.1. Comparison of the three HTA programmes’ practices

5.4.1.1. Health technologies evaluated

The analysis of 10 most recent HTA reports authored by the three national agencies under examination here covers a period of between 3-4 years in each agency’s history. Therefore, the results of the analysis can be assumed to be representative of the agencies’ actual practices.

Analysis showed that all three agencies focus predominantly on the assessment of procedures – specified as ‘medical and surgical procedures’ in CADTH and MSAC’s case, and more generically as ‘procedures’ in case of NIHR. Of 10 HTA reports, CADTH undertook an evaluation of M&S procedures in 6 cases, MSAC evaluated 7 M&S procedures, and NIHR evaluated 6 procedures.

In both CADTH and NIHR’s case, the majority of the remainder of the reports focused on pharmaceuticals. This is not – and cannot be – the case for MSAC due to Australia’s division of HTA labour at the national level between the three agencies (the evaluation of pharmaceuticals falls under PBAC’s remit). In their stead, the remainder of MSAC’s assessments focused on evaluation of devices and diagnostics.

5.4.1.2. Possible evaluata versus stated evaluata

As previously noted, salient to the HTA decisions may be data from (or about): clinical studies, economics, social aspects, legal aspects, ethics, organisational issues, stakeholder and patient views, expert opinion and political considerations. However, the range of inputs that the three agencies claim to consider is actually quite narrow.

NIHR and MSAC’s stated evaluata both include: safety, effectiveness and cost-effectiveness. As of 2010, MSAC additionally claims to evaluate the health technologies’ impact on the budget. Like NIHR and MSAC, CADTH considers clinical effectiveness and cost-effectiveness inputs. Unlike both NIHR and MSAC, CADTH does not claim to evaluate safety. Instead, it lists as its evaluata a rather vague ‘broader impact of health technologies’ category.
5.4.1.3. Stated evaluata versus actual evaluata

These, however, are only the inputs that agencies claim to assess. How well does what is claimed match what is actually evaluated?

Clinical effectiveness: All three agencies state that they assess evidence of clinical effectiveness (and/or clinical efficacy) in their HTAs. In practice, it was found that all 10 reports in each agency’s case actually assessed this domain. In other words, a 100% coherence obtained between the agencies’ stated evaluata and their actual evaluative practices with respect to clinical effectiveness.

Cost-effectiveness: All three agencies also list cost-effectiveness as an assessed domain. In all 10 reports, CADTH evaluated cost-effectiveness; all ten of MSAC’s and NIHR’s reports likewise assessed cost-effectiveness of the technologies under evaluation. In short, just like in the case of clinical effectiveness, a perfect consistence obtains between the agencies stated and actual evaluative practices vis-à-vis cost-effectiveness.

Safety: both MSAC and NIHR claim to assess safety; CADTH does not list ‘safety’ among the inputs it assesses. In practice, only MSAC is consistent between what is stated and what actually occurs – all 10 of its reports actually assessed safety of the health technology under evaluation. NIHR, despite claiming that safety is one of its evaluata, did not in actuality evaluate safety in any of the reports. CADTH, on the other hand, while not listing safety as its evaluata, did actually evaluate safety in 3 (possibly 4) reports.

One possible explanation for these discrepancies vis-à-vis safety may be that the HTA agencies do not wish to duplicate the evaluative work of agencies that assess health technologies for market entry, as those typically evaluate their safety.

But this explanation, at most, may account for NIHR’s evaluative behaviour. In the UK, Medicines and Healthcare products Regulatory Agency (MHRA; located within the Department of Health) assesses the safety of health technologies prior to entry onto the UK
market (MHRA, 2012). Thus, it may perhaps be reasonable to claim that NIHR’s failure to assess safety is an avoidance of duplicate evaluation. 65

In Canada’s case, pre-market-entry assessment is conducted by Health Canada (or, more specifically, its Health Products and Food Branch, or HPFB). HPFB carries out an assessment of safety and risks associated with the technology (Health Canada (HPFB), 2006). But the claim that CADTH does not evaluate safety so as not to duplicate HPFB’s work is untenable – for CADTH did evaluate safety in 3 (possibly 4) of its reports.

In Australia, the Therapeutic Goods Administration (or TGA) assesses health technologies’ safety prior to the market entry. Yet, all ten of MSAC’s reports evaluated safety. Thus, the claim that agencies fail to evaluate safety on account of the desire to avoid duplication of another agency’s work is even less tenable in Australia’s case than it is in Canada’s case.

Other: As of 2010, MSAC also evaluates the total cost of the health technology, which is typically framed as ‘budget impact of publicly funding the health technology’; this category has been consistently assessed by all of its reports since 2010. Somewhat similarly, CADTH also considers the input on the ‘broader impact of health technologies.’ The category, however, is not entirely clear. First, (as previously noted), it is not clear what, precisely, this category is meant to encompass. Second, even if the category is interpreted broadly (to encompass impact on a variety of levels and in a variety of ways – including financial, social, ethical, etc.), this domain is not consistently assessed by CADTH; only half (5/10) of CADTH’s reports assess what may reasonably be thought to be intended to be captured by it.

5.4.1.4. Evidentiary considerations in addition to the stated ones?

CADTH states that its evaluative considerations include: effectiveness, cost-effectiveness and the broader impact of health technologies. Its practices, however, are not entirely consistent with its statements, as safety is also occasionally assessed (3 or possibly 4 of 10 reports).

65 While this may account for NIHR’s actual evaluative conduct, it does not explain why safety is stated as one of NIHR’s assessed inputs, however.
NIHR does not *add* anything to its stated evaluata (safety, effectiveness and cost-effectiveness). (Although, as noted above, NIHR *subtracts* from the stated evaluata, in omitting an evaluation of safety in all 10 reports).

Unlike its Canadian and UK counterparts, MSAC’s stated evaluata match precisely its actual evaluata. Prior to 2010, MSAC’s terms of reference specified that it assesses safety, effectiveness, and cost-effectiveness of health technologies – and, as is clear from Table 7, above, this is precisely what its reports assessed. Post-2010, MSAC’s terms of reference added to its evaluata the budget impact of publicly funding the health technology. Since then, again, MSAC consistently assessed exactly what was listed under its terms of reference: safety, effectiveness, cost-effectiveness and the budgetary impact of publicly funding the health technology under assessment.

5.4.1.5. Consistency of evidentiary practices within a single health technology type?

It was thought that the inconsistencies in agencies’ evaluata could potentially be explained by the fact that the 10 reports analysed in each jurisdiction evaluated different health technology types. Upon closer analysis of CADTH’s evaluative practices vis-à-vis a single technology type (namely, M&S procedures), however, the inconsistencies held. In fact, the trend observed generally matched the trend observed with respect to M&S procedures: safety was assessed one-third of the time, and ‘greater impact’ was assessed half of the time; effectiveness and cost-effectiveness were assessed in all of CADTH’s reports.

Unlike CADTH, MSAC and NIHR are very consistent in their assessment practices assessment-to-assessment. As evident from Table 7 and Table 8, above, regardless of the health technology type assessed (whether it be a procedure, diagnostic, device, etc.), the assessed inputs were the same in each case – that is, safety, effectiveness, cost-effectiveness and budgetary impact (as of 2010) in MSAC’s case; effectiveness and cost-effectiveness in NIHR’s case.
5.4.2. How do the 3 agencies conceptualise ‘evidence’?

5.4.2.1. How do the 3 agencies conceptualise ‘evidence’ in theory and in practice?

What does the above suggest about how CADTH, MSAC, NIHR conceptualise ‘evidence’? To answer this question, first, a distinction needs to be drawn here between how the agencies understand ‘evidence’ in theory, and how they understand it in practice. By agencies’ understanding of evidence in theory, I mean what do the agencies’ statements indicate about the agencies’ understanding of the concept of evidence? On the other hand, by agencies’ understanding of evidence in practice, I mean what the agencies’ actual evaluative practices suggest about their understanding of ‘evidence.’ Consider each in turn.

As was shown in chapter 4, the agencies fail to offer an explicit definition of ‘evidence.’ Thus, their theoretical understanding of ‘evidence’ can only be derived from an analysis of their (rather limited) claims. The analysis suggests that the agencies theoretically construe ‘evidence’ to be data that originates from particular domains or areas (hereafter, the ‘stipulated domains’ approach). That is, in CADTH’s case, ‘evidence’ was theoretically construed as data on clinical effectiveness, cost-effectiveness and the broader impact of a health technology. In MSAC’s case, ‘evidence’ was theoretically understood as data on safety, effectiveness, cost-effectiveness, and budget impact. Finally, NIHR’s – like MSAC’s – theoretical construal of evidence is of data on safety, effectiveness, and cost-effectiveness of health technologies.

The agencies’ practical understanding of evidence, on the other hand, can be derived from their actual evaluative behaviour, as illustrated in Table 6, Table 7 and Table 8. So, how do CADTH, MSAC, NIHR understand ‘evidence’ in practice? The analysis shows that, just like in theory, the agencies’ practices suggest that their understanding of ‘evidence’ is as of data originating from specific domains. Thus, CADTH’s understanding of evidence in practice is as of data on effectiveness, cost-effectiveness, broader impact of health technologies (approximately one-half of the time) and safety (approximately one-third of the time) of health technologies. In MSAC’s case, the (current, i.e. post-2010 change in Terms of Reference) practical construal of evidence is as of data on safety, effectiveness, cost-effectiveness, and
budget impact of health technologies. Finally, in practice, NIHR understands ‘evidence’ to be data on effectiveness and cost-effectiveness of health technologies.

5.4.2.2. Some challenges for the ‘stipulated domains’ approach

The ‘stipulated domains’ approach, adopted by the agencies, does enjoy a not insignificant benefit. Recall that the spectrum of salient inputs into the HTAs is quite broad, potentially encompassing: clinical and economic data, as well as social, legal, ethical and organisational issues; stakeholder, patient and expert views; and political inputs. Identification and assessment of the existing data from all of these domains would undoubtedly be both time- and resource-intensive. In reality, however, both financial and temporal limitations constrain how much data can be located and assessed for any given HTA. Thus, by understanding evidence as, for example, data only on safety, effectiveness and cost-effectiveness, an agency is able to constrain those temporal and financial costs. (Temporal costs may be particularly salient in cases where policy decisions need to be made quickly, and so, the assessments need to be conducted rapidly.)

On the other hand, the ‘stipulated domains’ approach is not without its share of problems. Two main ones will be considered here. First, in construing evidence as data from stipulated domains, the agencies are vulnerable to the same ‘cart before the horse’ charge that arose in the discussion of this approach in context of EBM and EBHP, in chapter 4. Second, mismatches occur between the domains that are claimed to be assessed and the domains that are assessed in practice, which undermines the approach. Consider each in turn.

‘Cart before the horse’ problem

Adopting the ‘stipulated domains’ approach is hardly without precedent. As discussed in chapter 4, in both the EBM and the EBHP literature one encounters conceptualisations of evidence which stipulate particular domains as sources of evidence. On the EBM side, for example, Rycroft-Malone et al (2004) specified the following domains: research, clinical experience, knowledge from patients/clients/carers, and local context. Granados (1999) listed the following domains: scientific, social, economic and political. On the EBHP side, Vivian Lin (2003) included here: cultural rationality, political rationality, and technical rationality.
Brian Head (2008) counted: political know-how, scientific and technical analysis, professional and practical field experience.

However, a salient difference between these accounts’ construal of evidence and the agencies’ construal thereof needs to be emphasised here. That is, while the EBM and EBHP accounts coupled a definition of evidence to specified domains, the agencies’ theoretical understanding of ‘evidence’ omits the first half of that conjunction. While this entails that the agencies are not subject to the problem of accommodating the stipulated domains by the definition of evidence,66 they do remain vulnerable to a general problem underlying this approach.

More specifically, the problem is that specifying particular domains as sources of evidence prior to a particular assessment being undertaken, effectively amounts to putting the cart before the horse. Different questions (or issues, or decisions) will require different evidentiary inputs. Assessment of modalities involved in treating benign prostatic hyperplasia in men over 50 (e.g. open prostatectomy, or a holmium laser enucleation of the prostate with tissue morcellation), will require different evidence than will an assessment of a programme targeted at increasing Aboriginal participation in a childhood screening programme. While in both cases some of the same evaluata may be considered (e.g., effectiveness), the latter (but not the former) will likely also require a consideration of culturally-sensitive methodology of implementing the programme, a consideration of local organisational and health human resources constraints, and so forth. These inputs cannot be stipulated ahead of the particular assessment.

One could, of course, object here that it is open to the agencies to specify that several core domains (call them domains a, b, and c), which will be assessed in each HTA – and add a caveat that additional domains will be assessed as is required by particular HTAs. This would give the agencies some flexibility to work around this issue. However, as became evident earlier, agencies do not consistently adhere to their own stipulated domains for the most part (MSAC excepted67), rendering the feasibility of any domain-specified approach questionable.

66 See, e.g., the discussion of this problems in Rycroft-Malone et al’s construal of evidence in chapter 4 (section 4.2.3.2)

67 As will become apparent in chapter 7, however, not even MSAC is entirely immune to this problem.
Mismatches between stated domains and domains evaluated in practice

The problem with stipulating domains ahead of time became evident in consideration of the differences between the agencies’ construal of ‘evidence’ in practice and in theory. More specifically, when a particular assessment, issue, question, and so forth is before the agency, the actual evidentiary requirements (for addressing the issue, question, etc.), gave rise to mismatches between what was listed as assessed (‘stipulated domains’) and the domains assessed in practice. It needs to be noted, here, that Australia’s MSAC is an exception to this. MSAC states that its aim is to assess the safety, effectiveness, cost-effectiveness and budgetary impact of the adoption of a new health technology – and these are precisely the domains it assesses. (In other words, MSAC’s stated understanding of evidence (as data from these particular domains) matches its understanding of evidence in practice). The assessments conducted by the remaining agencies, however, evince three types of mismatches:

(a) A domain is not stated but is evaluated in practice
(b) A domain is stated but is not evaluated in practice
(c) A domain is stated and is evaluated in practice but only some of the time

Mismatch (a), that is, a domain not stated (as evaluated) but is evaluated in practice, is exemplified by CADTH’s approach to safety. Safety is evaluated approximately 1/3rd of the time, although it is not listed by CADTH as one of its evaluata.

Under (b), or the domain that is stated as evaluated (but is not actually evaluated in practice), one can categorise NIHR’s approach to evaluating safety. NIHR states that safety is one of the domains it assesses – however, in practice, safety was not assessed in any of the reports examined.

Mismatch (c), a domain that is stated and is evaluated in practice but only some of the time, can, again, be illustrated using CADTH. CADTH states that it assesses the ‘broader impact of health technologies,’ but an analysis of its actual evaluative practices shows that this happens only one-half of the time.

These mismatches are problematic for two reasons. First, they put into question the very benefit of the ‘stated domains’ approach. As noted above, the benefit of the ‘stipulated
domains’ approach consists in limiting the temporal and other resources associated with conducting an evidence-based assessment. However, insofar as the agencies’ evaluative behaviour is fairly ad-hoc (MSAC is excepted here) it is unclear whether this benefit is realised.

Second, the mismatches between stated domains and domains assessed in practice raise some troubling questions about the evidentiary status of the ‘mismatched’ domains. Is it the case that, in mismatch (a), data on safety constitutes ‘evidence’ in practice (but not in theory) for CADTH; in case (b), data on safety constitutes ‘evidence’ in theory (but not in practice) for NIHR; and in case (c), that data on the broader impact of health technologies constitutes evidence in theory all of the time but it constitutes evidence in practice only some of the time, for CADTH? How are we to account for the variable status (as evidence) of these inputs?

One could point out, here, that the variable status of data as evidence (or not-evidence) – from case to case – is something that has already been conceded. It was noted in chapter 4 that a proper understanding of evidence ought to permit that the same data (e.g. on the percentage of women who undergo BRCA1 and 2 testing) may constitute ‘evidence’ for the purposes of one assessment (e.g. vis-à-vis a genetic counselling programme) but not for the purposes of another assessment (e.g. vis-à-vis a particular diagnostic test for lung disease). The mismatches could therefore be defended, here, as simply an instance of this – already conceded to be permissible – characteristic.

The problem with this explanation, however, is that if one stipulates data from particular domains as ‘evidence,’ then the options to either omit what one has designated as evidence or to include data beyond ‘evidence’ (by including data from additional domains) are not viable. This is because it leaves one vulnerable to the charge of carrying out only a partially evidence-based assessments (in the first scenario) and raises questions about the evidentiary-status of the additional information considered (in the second scenario). In short, the problem with the ‘specification of domains’ approach is that it lacks the flexibility to decide what is evidence on

---

68 The subsequent question that arises in the second scenario, is that if the additional information is evidence, why was it not listed under the set of stipulated domains in the first place?
a case-by-case basis – the very flexibility that constitutes one of the main advantages of adopting, instead, the conceptualisation of evidence proposed in chapter 4.

5.4.3. Advantages of adopting the proposed conceptualisation of evidence

Adopting the proposed definition offers several additional advantages. The first advantage of adopting the proposed conceptualisation is generic – that is, it would be enjoyed by any definition of evidence that the agencies would adopt. More specifically, the benefit here would consist in actually having a definition of evidence. As noted previously, the importance of a clear definition of evidence in EBM and EBHP consists in that such a definition allows us to evaluate whether the decisions (or assessments or courses of action, etc.) in those fields really are ‘evidence-based.’ In other words, the definition would allow us (and the HTA agencies themselves) to assess whether the agencies are living up to their mandates.

The second advantage would be that the proposed conceptualisation would bypass the problems faced by the ‘stipulated domains’ approach. The proposed conceptualisation stipulates that the ‘evidence-ness’ of evidence rests on data’s relevance to the issue (or question, decision, etc.). By eliminating the ‘stipulated domains,’ the proposed conceptualisation is invulnerable to problems around ‘mismatches’ (for no domains which could potentially mismatch are stipulated), and allows the agencies’ practices to meet the evidentiary-based mandate no matter from what domains the data that is assessed originates – as long as the data assessed is relevant to the issue at hand.

5.4.4. Mapping the agencies’ evidentiary practices onto the proposed conceptualisation of evidence

However, as Upshur points out in context of EBM, any proposed concept of evidence must resonate with the practitioners within the field. The same ought to be true of the concept of ‘evidence’ in HTA. Does the proposed conceptualisation meet this criterion? One way to establish this is to consider whether – and if so, to what extent – the proposed conceptualisation maps on to the theoretical and practical understanding of evidence evinced by the 3 agencies. Consider each in turn.
5.4.4.1. Proposed conceptualisation vs. the agencies’ theoretical construal of evidence

As described above, the agencies’ theoretical understanding of ‘evidence’ is of data originating from specific domains. In CADTH’s case, ‘evidence’ is data on clinical effectiveness, cost-effectiveness and the broader impact of health technology. In MSAC’s case, ‘evidence’ is data on safety, effectiveness, cost-effectiveness, and impact on budget. In NIHR’s case, ‘evidence’ is data on safety, effectiveness, and cost-effectiveness. On the other hand, the proposed definition of ‘evidence’ for HTA purposes stipulated that ‘evidence’ is: available data that is relevant to the issue being addressed, question asked, decision made, etc. Are the proposed definition of evidence and the agencies’ theoretical understanding of ‘evidence’ consistent with each other?

The agencies’ theoretical understanding of ‘evidence’ certainly construes evidence as data. However, as noted above, the agencies’ theoretical understanding of ‘evidence’ is not ‘data simpliciter’ but rather ‘data from stipulated domains.’ This suggests that data from other domains is either not considered to be ‘evidence,’ or that their evidentiary status is in question. However, this cannot be inferred with a great degree of certainty, as the agencies’ conceptual understanding of evidence is insufficiently fleshed out and so, insufficiently clear. The conception is also too poorly fleshed out for assessing the consistency between the agencies’ theoretical understanding of ‘evidence’ and the remaining elements of the proposed definition of ‘evidence.’ It is not clear, for example, whether the data is (theoretically) construed as ‘available’ (rather than ‘ideal’), ‘relevant’, and pertinent to the issue, question or decision. The principle of charity would dictate that the answer is affirmative on all counts. However, given the agencies’ failure to flesh out their theoretical conception of evidence, and the reliance here on the derivation thereof from the agencies’ (very limited) claims on this issue, this cannot be concluded with certainty.

5.4.4.2. Proposed conceptualisation vs. the agencies’ practical construal of evidence

The agencies’ practical understanding of ‘evidence’ – much like their theoretical understanding thereof – is also qua data originating from specific domains. The domains assessed in practice, however, do not entirely match the ones indicated in theory (MSAC
excepted). Thus, in practice, CADTH understands evidence as data on: clinical effectiveness, cost-effectiveness, the broader impact of health technology (some of the time) and safety (some of the time). In MSAC’s case, ‘evidence’ is (consistently) data on safety, effectiveness, cost-effectiveness, and health technologies’ impact on budget. In NIHR’s case, ‘evidence’ is understood in practice as data on effectiveness and cost-effectiveness of health technologies.

The definition of ‘evidence’ for HTA purposes proposed in chapter 4, stipulated that ‘evidence’ is available data that is relevant to the issue being addressed, question asked, decision made, etc. Are the proposed definition of evidence and the agencies’ practical understanding of ‘evidence’ consistent, however?

The agencies’ practical understanding of ‘evidence’ does involve data. Again, as in their theoretical understanding of ‘evidence,’ the data here is ‘data from stipulated domains’ (identified above), rather than data simpliciter. Data that undergoes assessment is, a fortiori, available data (rather than ideal data).

Recall that ‘relevance’ was a characteristic of data that was resolved by the members of the HTA community (in parallel to EBM, where the individual physician decides whether clinical research is salient to the patient presenting before her). It is very clear that the agencies’ practices meet this criterion. As discussed earlier, CADTH either contracts or deploys its own team of experts to write the report (that is, to collect, synthesise and interpret the data) and sends the report out for peer review to other experts; MSAC and NIHR adopt a similar approach.69 Insofar as it is the members of the HTA community who decide what data is (or is not) salient for a given assessment, this criterion thus appears to have been met.

This leaves to be settled whether ‘issues, questions, or decisions’ are what is being, respectively, resolved, asked or made. As is evident from Table 6, Table 7 and Table 8, above, typically, the question that arises is around the effectiveness and economics of a particular technology; less frequently, decisions about extending the funding of an interim funded technology, or issues around comparing two specific health technologies are addressed in the HTA reports. In short, the agencies’ evaluative behaviour suggests that their construal of

69 See sections 5.1.1, 5.2.1, and 5.3.1, respectively.
‘evidence’ in practice is sufficiently broad to be consistent with this element of the proposed conceptualisation, as well.

5.4.4.3. How much change would be required of agencies, were they to adopt the proposed conceptualisation of evidence?

The agencies fail to explicitly flesh out a theoretical understanding of evidence, and their statements are too limited to derive an adequately detailed account of their conceptualisation thereof. The consistency between their theoretical understanding of evidence and the proposed definition of evidence is, therefore, difficult to assess; it is impossible to establish whether the proposed conceptualisation of evidence would resonate with the agencies’ theoretical understanding of ‘evidence,’ in other words.

With regards to the agencies’ evidentiary conduct in practice, a good degree of coherence obtains between the agencies’ practices and the definition of evidence proposed in chapter 4. The key difference is that the agencies’ understanding of ‘evidence’ is as of data from stipulated domains, rather than data simpliciter – which, as noted above, gives rise to several problems. On the other hand, the evidence that constitutes the input into the HTAs in practice does consist of data which is available and relevant to answering a variety of questions or addressing a variety of issues or making various decisions. Thus, the required change would be to eliminate the stipulated domains and simply adopt the proposed conceptualisation and adhere thereto. Insofar as the proposed conceptualisation of evidence would not require substantial behavioural (that is, evaluative) changes on the agencies’ parts, it is therefore likely to resonate with the field’s practitioners quite well.

5.5. Conclusion

In summary, the aim of this chapter was to test the definition of evidence proposed in chapter 4 against HTA practices. Towards this end, the agencies’ statements and evidentiary practices were examined, in order to establish their understanding of the concept of evidence (theoretical and practical, respectively). I argued that the agencies’ theoretical conceptualisation of evidence was too poorly fleshed out to test it for coherence with the proposed conceptualisation of evidence – on the other hand, the agencies’ practical understanding of evidence cohered quite well with the proposed conceptualisation. If – as Upshur suggests – an
adequate definition is characterised by its resonance among the practitioners of a field, it seems that the proposed conceptualisation meets this standard. In other words, the proposed conceptualisation coheres with HTAs’ evaluative practices – at least vis-à-vis ‘health technologies’ considered more generally.

I shall now turn to testing the proposed conceptualisation against a specific type of health technology – namely, genetic testing. Genetic testing was deliberately chosen here as it is a relatively new and widely recognised as an ethically-challenging health technology, which is therefore likely to raise difficulties for the HTA agencies and processes. Prior to considering genetic tests as a health technology assessment issue in more details, however, a brief detour into some background to genetics and genetic testing is first in order – this will be the focus of chapter 6.
CHAPTER 6: The Road to Genetic Testing

Summary: Chapter 6 consists of three parts. The first part traces some of the key events in the period of modern genetics – beginning with the work of Gregor Mendel. The second part discusses the Human Genome Project. Finally, the third part focuses on genetic testing, describing the science and the technology of genetic testing. The underlying aim of this chapter is to provide the necessary background to the discussion of genetic testing as a health technology assessment issue, in chapter 7.
6.1. History of modern genetics

6.1.1. Gregor Mendel

6.1.1.1. Mendel and the peas

Starting with the Ancient Greeks, the study of heredity tended to focus predominantly on *human* heredity.\(^{70}\) Ironically, it was necessary to return to the study of heredity in *plants* in order to solve the riddle of how heredity worked in humans. The protagonist of the story is Gregor Mendel (1822-1884). Often referred to as the ‘founding father of modern genetics,’ Mendel studied science at the University of Vienna between 1851-1853, where he took courses in plant physiology, experimental physics and mathematical physics (Sturtevant, 2001a; Lorentz et al 2002).

Mendel was a prolific scientist– in addition to experimenting with peas, he also experimented on *phaseolus*, maize, and a variety of other plants, as well as on bees and mice; he also conducted meteorological studies. Unfortunately, the totality of Mendel’s published scientific output is quite miniscule.\(^{71}\) The contents of the seminal paper – titled *Experiments in Plant Hybridisation* – were first made public in 1865. On February 8 and March 8 of that year, Mendel read the paper at the meeting of the Natural History Society of Bruun, and the paper was subsequently published in the Society’s *Proceedings* (Vorzimmer, 1968; Sturtevant, 2001a). The experiments described therein commenced in the summer of 1854 – Mendel initially began by growing 34 different strains of peas, and tested them all for consistency in 1855. The experiments began in earnest in 1856 (Sturtevant 2001a); they included 22 varieties of peas, and tested the behaviour of seven sets of characteristics: seed shape (round vs. wrinkled), cotyledon colour (yellow vs. green), seed coat colour (white vs. grey), pod shape (inflated vs. constricted), colour of unripe pod (green vs. yellow), position of the flowers (axial vs. terminal) and stem length (long vs. short) (Mendel, 1865).

\(^{70}\) For an overview of the history of genetics prior to the modern period, see Stubbe (1972) and Zirkle (1946)

\(^{71}\) Aside for his seminal 1865 paper describing the experiments with the peas, surviving are his 1870 *Übereinigetausländlicher Befruchtunggewonnene Hieracium-Bastarde* (On the fertilisation of Hieracium hybrids) and 1871’s *Die Windhose vom 13. Oktober 1870* (The tornado on October 13, 1870). Some of his letters (particularly to Nageli) also survive, and were published, in 1905 (Stubbe 1972)
Upon crossing peas displaying the different characteristics, Mendel found that some of these characteristics (the dominant ones) occurred in much greater number in the offspring than other characteristics (the recessive ones). The following characteristics were identified as dominant: round seed shape, yellow cotyledon colour, grey seed coat, inflated pod shape, green coloured unripe pod, axial position of the flowers, and long stem length (Mendel 1865). Mendel observed that each hybrid displayed either the dominant characteristics or the recessive characteristics, and the ratio of the dominant to recessive characteristics was 3:1.\(^{72}\) Importantly, the hybrids did not demonstrate any transitional (blended) characteristics (Mendel, 1865).\(^{73}\)

### 6.1.1.2. Oblivion and rediscovery

Not only did Mendel’s paper largely languish in oblivion during Mendel’s lifetime, but those few Mendel contemporaries who were familiar with its contents generally failed to grasp its significance. Only five individuals are known to have come in contact with Mendel’s work before its rediscovery in 1900: Anton Kerner von Marilaun (1831-1898), a professor at the University of Innsbruck and director of the Botanical Gardens in Vienna; Heinrich Hoffman (1819-1891), a Professor of Botany at Giessen; Wilhelm Focke (1834-1922), a German botanist and physician in Bremen; Carl Wilhelm von Nageli (1817-1891), a Swiss botanist, and a professor of botany at the University of Munich; and Ivan Fyodorovich Schmalhausen (1849-1894), a Russian botanist. Of those, von Marilaun’s reply to Mendel is lost, and Hoffman, Focke and Nageli did not appear to have grasped the significance of Mendel’s work. Schmalhausen seems to be the only one to have understood the import of Mendel’s discoveries, however, the sections of his dissertation that contained the discussion of Mendel were never translated into English (Vorzimmer, 1968; Stubbe 1972; Sturtevant 2001a; Mawer, 2006).

\(^{72}\) The ratios actually ranged from 3.15 to 1 (grey seed coats to white seed coats) to 2.82 to 1 (green unripe pods to yellow unripe pods); the ratio average across all 7 characteristics was 2.98 to 1 (Mendel, 1965)

\(^{73}\) In order to explain how an offspring inherits his or her parents’ traits, Darwin relied on blended inheritance (according to which, the offspring’s traits are an intermediate state between the parents’ traits). This was problematic, insofar as blending inheritance, over time, undermines natural selection. Had Darwin read the uncut copy of Mendel’s paper later found in his library (Henig, 2000), he would have known that blending inheritance does not work; instead, he would have learned that the transmission of traits from parents to offspring proceeds in discrete units. Alas, Darwin remained unfamiliar with Mendel’s paper his entire lifetime. In this, he was hardly alone; following its publication, Mendel’s paper largely fell stillborn from the press, as discussed below.
It is thus that Mendel’s work largely languished in oblivion for over 30 years\(^7\); its rediscovery had to wait until the year 1900 and the work of three individuals: Hugo de Vries, Erich von Tschermak and Carl Correns (Lorentz et al, 2002).

Hugo de Vries (1848-1935) was a Dutch botanist, and professor at the University of Amsterdam. He began his study of variation in plants in 1880 and conducted a series of experiments on hybridisation over the following decade. Those experiments confirmed Mendelian conclusions that heredity is non-blended but rather discrete, and that constant ratios are observed in the hybrid offspring (Stubbe 1972). In 1900, de Vries published three papers on his research. The first appeared in March 1900 in Paris Academy of Science’s *Comptes Rendus*. The second was also published in March 1900, by *Revue général de botanique*. The third was published in April 1900, by the German journal *Berichte der Deutschen Botanischen Gesellschaft* (Sturtevant, 2001a; Stubbe 1972). The second and the third papers both mention and credit Mendel with his discoveries; in the second (French) paper, de Vries conceded that:

> This law is not new. It was stated more than thirty years ago, for a particular case (the garden pea). Gregor Mendel formulated it in a memoir entitled ‘Versucheüüber Pflanzenhybriden’ in the *Proceedings of the Brünn Society*. Mendel has there shown the results not only for monohybrids but also for dihybrids. This memoir, very beautiful for its time, has been misunderstood and then forgotten. (Sturtevant, 2001a, 26)

The second re-discoverer of Mendel’s work was an Austrian botanist also interested in the effects of crossing and inbreeding of plants – Erich von Tschermak (1871-1962). In 1898, he engaged in a series of experiments crossing peas, observing the same 3:1 ratios that Mendel had observed 30 years prior (Stubbe 1972). He stated in one of his letters that he realised the implications of his results before becoming familiar with Mendel’s paper; he located the paper itself only after coming across its mention in a paper by Focke. Von Tschermak published the results of his experiments in 1900 in a paper titled “Concerning Artificial Crossing in *Pisumsativum*” and published – like de Vries’ – in *Berichte der Deutschen Botanischen Gesellschaft* Vol. XVIII; Mendel is mentioned several times in the paper (Stubbe, 1972).

---

\(^7\) Vorzimmer (1968) argues that the oblivion of Mendel’s work is at least partly due to the novelty of his approach; the union of statistics and biology was not standard methodology at the time. Sturtevant (2001a) argues that a further factor contributing here was that the attention of the scientific world was, at the time, monopolised by the issues raised in Darwin’s *The Origin of Species*, which was published 6 years before the publication of Mendel’s paper. Against that kind of competition, it is quite likely that a paper by a monk from Austria never stood a chance.
Finally, the third re-discoverer was Carl Correns (1864-1933), a German, and a botanist like the others. Correns’ work focused on fertilisation of maize and peas, and involved counting out the characteristics in resulting hybrids, again, much like Mendel had done. In the process, he discovered Mendelian ratios in the offspring generations. His discoveries were his own, so to speak; he did not become aware of Mendel’s essay until his experiments were nearly complete and the conclusions drawn, in 1899 (Stubbe, 1972; Sturtevant, 2001a).

Like de Vries and von Tschermak, Correns published his results with peas and maize in 1900 in *Berichte der Deutschen Botanischen Gesellschaft*. The paper was titled *G. Mendels Regel Über das Verhalten der Nachkommenschaft der Rassenbastarde* (G. Mendel’s law concerning the behaviour of progeny of varietal hybrids). There, he noted that his results were consistent with those of de Vries. Importantly, he also acknowledged that:

> When I discovered the regularity of the phenomena, and the explanation thereof — to which I shall return presently — the same thing happened to me which now seems to be happening to de Vries: I thought that I had found something new. But then I convinced myself that the Abbot Gregor Mendel in Brünn, had, during the sixties, not only obtained the same result through extensive experiments with peas, which lasted for many years, as did de Vries and I, but had also given exactly the same explanation, as far as that was possible in 1866. (Correns, 1900, 39; emphasis added)

Although Mendel’s work was rediscovered by the scientists on the continent, it may well have continued to languish in obscurity were it not for its enthusiastic promotion by William Bateson (1861-1926), a Cambridge zoologist interested in heredity and – by 1899 – a recognised name in the field of heredity (Sturtevant 2001a).

A fateful event took place during a train ride on May 8, 1900. On that date, Bateson was travelling from Cambridge to London, in order to read a paper on the problems of heredity at the meeting of the Royal Horticultural Society. During that trip, he is said to have read Mendel’s work; the work made such an impression on Bateson that he immediately incorporated its contents into his lecture (Sturtevant 2001a; Lorentz 2002). He subsequently translated Mendel’s paper into English and became one of its most enthusiastic advocates, both verbally and scientifically – striking together an active group of researchers at Cambridge who pursued Mendelian studies (Stubbe 1972). ‘Mendelian studies’ were not limited to England, however; groups devoted to the subject soon were struck in Germany (under Correns), Austria
(under von Tschermak), as well as France and the US, rapidly gaining in popularity and laying the foundations for the ‘classical’ period in the history of modern genetics (Sturtevant, 2001a).

6.1.2. The Classical Period in Modern Genetics

The classical period in modern genetics, or more simply, classical genetics, is the period of the first fifty or so years of the 20th century – roughly, from the rediscovery of Mendel’s work to just before the discovery of the doubly helical structure of the DNA. During that period, the mechanism of heredity became the subject of research focus (Allen, 2008). Although the period is rich in colourful characters and discoveries, the following will focus only on 3 key events, arguably contributing the most to the subsequent discovery of the structure of DNA.

6.1.2.1. Thomas Hunt Morgan

Thomas Hunt Morgan (1866-1945) was an American embryologist and zoologist. Following the receipt of his PhD in 1890, he initially taught at Bryn Mawr, before moving to Columbia University in 1904. There, in the aftermath of the rediscovery of Mendel’s work, he was conducting experiments involving evolution and heredity, particularly focusing on mutations in Drosophila melanogaster (Sturtevant, 2001b). Morgan himself, however, was no Mendelian; during his early days at Columbia he could best be characterised as an anti-Mendelian (Beadle, 1974; Sturtevant, 2001b).

It was the very work with Drosophila that was to cure him of his anti-Mendelianism, however. Around 1908, Morgan began to experiment with Drosophila – his interest at the time was in mutations. He came across his first significant mutation – a male Drosophila with white eyes (the normal colour is red) in 1910 (Morgan, 1910). To test the heredity of the white-eyed mutation, Morgan crossed the white-eyed male with the standard red-eyed female. The offspring displayed the Mendelian 3:1 red to white ratios, suggesting that redness is dominant while the whiteness is recessive. What was surprising, however, was that it was only the males that displayed the recessive trait for whiteness – no females displayed the trait (Morgan, 1910).

Three major findings resulting from these and subsequent experiments were published in two papers in Science, in 1910 and 1911. First, Morgan postulated that genes are carried on chromosomes. Second, genes are located at particular locations on the chromosomes. Finally,
genes for some traits – the aforementioned eye colour, as well as body colour and wing mutations – reside on sex chromosomes (Morgan, 1910; Morgan, 1911). Morgan thus both developed the chromosomal theory of heredity, and bestowed on genes their location – and thus, for the first time, a physical existence (Deichmann, 2004; Winchester, 2002).

6.1.2.2. Jean Brachet

Jean Brachet (1909-1988) was a Belgian biochemist and embryologist, whose academic interest was in the role of nucleic acids in the synthesis of protein (Thomas, 1992). While still undergoing his doctoral training – in the late 1920s – Brachet was offered a choice of two problems for his laboratory project: the role of mitogenetic rays in the development of embryos or the localisation and behaviour of DNA (then referred to as thymonucleic acid) in the growing oocytes. He elected to focus on the latter (Burian 1997; Brachet, 1957).

At the time that Brachet was beginning his work, it was generally believed that thymonucleic acid (DNA) was a low molecular weight substance which existed only in animal cells; it was believed that its function was to carry out only some kind of a buffering effect. The genes were thought to be proteins (Alexandre, 1992). Against this, Brachet’s experiments with fertilised eggs of sea urchin showed that DNA actually resided on chromosomes, that the amount of DNA remains constant across the different types of cells, and that DNA’s quantity increases proportionally to the number of cells. In short, his work suggested that DNA was somehow implicated in inheritance (Alexandre, 1992; Thomas, 1992).

6.1.2.3. Edward Tatum & George Beadie

About a decade later, Edward Lawrie Tatum (1909-1975), an American geneticist and biochemist, was researching biochemical, nutritional and genetic issues around Drosophila. His research partner was George Wells Beadle (1903-1989) – also an American geneticist, and likewise a researcher of Drosophila. Like many Drosophila researchers at the time, Beadle conducted some of his training – more specifically, his post-doctoral studies – under Thomas Hunt Morgan (Dronamraju, 1991; Beadle, 1974).

In the early 1940s, the answer to the question ‘What do genes do?’ was still somewhat of a mystery. Whilst sitting in on Tatum’s lecture, it occurred to Beadle that an appropriate way of
tackling this question would be to identify genes with known chemical reactions (Beadle, 1974). Where Beadle’s strength was genetics, Tatum’s strength was his biochemistry expertise; it was those strengths combined that were needed to answer the question (Singer and Berg, 2004).

Beadle and Tatum’s experiment involved Neurospora – the red bread mould – since both its cytogenetics had previously been worked out and the biochemistry of a closely related species was already known. They exposed Neurospora to mutation-causing X-rays and ultraviolet radiations; the idea was to test mutants for the loss of the ability to synthesise components required for a minimal growth medium (Beadle, 1974). 299 cultures and 5 months later, Beadle and Tatum found the first nutritional mutants, whose offspring grew only when vitamin B6 was added to the standard growth medium. From this, they were able to determine which single gene had been mutated (Beadle, 1974; Singer and Berg, 2004).

These experiments – published in a paper titled “Genetic Control of Biochemical Reactions in Neurospora” in 1941’s Proceedings of the National Academy of Sciences of the United States of America – suggested a direct relationship between genes and proteins. Their hypothesis was rapidly christened “one gene – one enzyme” hypothesis. Although the findings were revolutionary – in that they finally assigned to genes a function (viz., coding for proteins) – they were not readily received by the scientific community. This was because it was thought that coding for proteins would be much too simplistic a role to assign to genes (Beadle, 1974). Even a decade later – when Franklin, Watson and Crick were attempting to solve the riddle of the structure of DNA in the early 1950s – Beadle and Tatum’s findings were still under fire (Singer and Berg, 2004).

6.1.3. Molecular Genetics Period

6.1.3.1. Osvald Avery et al

Ultimately, it was the very work of Beadle and Tatum that laid the first foundations for what was to become the molecular genetics period in modern genetics (Singer and Berg, 2004; 75 Tatum and Beadle were not actually the first to discover the relationship; in 1902 Garrod was the first to discuss the gene/enzyme relationship, although never using the term gene (Dronamraju, 1991). Beadle acknowledged as much, both in print (Beadle 1974) and in his Nobel Prize acceptance speech (Singer and Berg, 2004).
Lederberg, 1994). However, while Beadle and Tatum laid the foundations, it was the work of Avery et al that erected some of the most important superstructure atop these foundations (Deichmann, 2004; Alexandre, 1992).

Although he was a Canadian-born medical researcher and physician, Osvald T. Avery (1877-1955) worked in the US. His research was devoted to the study of pneumococcal pneumonia (Steinman and Moberg, 1994). When Brachet was conducting his research with sea urchin in the 1930s, DNA was thought to be a low molecular weight, uninteresting substance; genes were thought to be proteins. Little had changed by the time Avery was conducting his research with pneumococcus in the 1930s and 40s – that genes were proteins was still a fairly widely held view amongst the scientific community (Deichmann, 2004; Steinman and Moberg, 1994). Similarly, DNA was still regarded as an unlikely candidate for biological specificity – partly on account of its relative chemical (and therefore structural) simplicity, and partly on account of the unavailability of homogenous samples of DNA which would have permitted a detailed chemical analysis (Lederberg, 1994; McGregor and Poon, 2003).

It is against this scientific backdrop that Avery – together with his colleagues, Colin MacLeod and Maclyn McCarty – published a paper in 1941’s Journal of Experimental Medicine, titled Studies on the Chemical Nature of the Substance-inducing Transformation of Pneumococcal Types. Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus type III. The paper reported on 15 years’ worth of experiments, whose aim was to determine the chemical properties of the substance which changes the heritable properties of pneumococcus. The key points put forth in the paper can be summarised as follows:

1. Pneumococci contains a biologically active ‘fraction’ that can be isolated
2. The methods for isolating and purifying this ‘fraction’ are described
3. The ‘fraction’ contains no protein, lipids, or reactive polysaccharide; it consists solely of deoxyribonucleic acid

The paper considers the possibility that it is something adsorbed to the deoxyribonucleic acid (or, alternately that it is some other substance, present in undetectably small amounts) that is the source of the transforming principle, but ultimately concludes that the evidence strongly suggests that deoxyribonucleic acid “is the fundamental unit of the transforming principle of Pneumococcus” (Avery et al, 1944).
In short, what Avery and colleagues showed that it was deoxyribonucleic acid – DNA – that transmitted hereditary properties, thereby upending the ‘gene qua protein’ dogma. Interestingly, while the response to similarly revolutionary work by Tatum and Beadle was very unfavourable, the work of Avery et al was received quite positively. Although the paper was published in a journal that was not commonly read among geneticists at the time, it was widely cited\(^{76}\) and its findings were quickly confirmed by other scientists (Steinman and Moberg, 1994).

6.1.3.2. Rosalind Franklin – and Watson and Crick

Thus, by the early 1950s, the chemical and biological characteristics of DNA were gradually emerging. A key piece of the puzzle was still missing, however – the structure of DNA. And, although it is conventional to cite Watson and Crick’s work here, one would be remiss in failing to acknowledge Rosalind Franklin. Franklin (1920-1958) was a UK molecular biologist and biophysicist, who received her PhD from Cambridge in 1945. It was the research she carried out on the DNA structure at King’s College, London, between 1951 and 1953, which was crucial to finally solving the puzzle (Delamont, 2003; Elkin, 2003).

When Franklin started at King’s College, some of the foremost scientific minds of the time were attempting to solve the problem. Among those were Linus Pauling’s group at Caltech, Watson and Crick at Cambridge, and finally Franklin and Maurice Wilkins\(^{77}\) at King’s College (Rapoport, 2002). Shortly after she began at King’s in January 1951, Franklin identified two distinct configurations of DNA – the A form and the B form. From those, she deduced two important pieces of the puzzle regarding DNA’s structure – that the DNA’s structure is helical, and that its backbone lies on the outside rather than the inside. She presented a lecture detailing these findings at an internal seminar at King’s College in November of 1951 (Elkin, 2003).\(^{78}\) By May 1952, Franklin had clear photographs of both the A and the B forms of DNA – the photograph of the B-form is the famous photograph #51 (see Franklin, 1953). Taken

\(^{76}\) E.g., Deichmann counted 239 citations between 1945-1954 (Deichmann, 2004) while Lederberg counted nearly 300 citations of the paper for the same period (Lederberg, 1994).

\(^{77}\) However, see, e.g. Hellman (2001) who argues that Franklin and Wilkins were not a team in any meaningful sense.

\(^{78}\) Watson attended her talk (see below).
looking ‘down’ the molecule, the photograph – in showing an X – provided some of the clearest evidence yet of the helical structure of DNA (Elkin, 2003; Watson, 2001).

While Franklin was working on DNA’s structure at King’s College, James Watson and Francis Crick were attempting to solve the same riddle at Cambridge. Watson (b. 1928), is an American; by training, he is a molecular biologist and a geneticist. His partner, Francis Crick (1916-2004), on the other hand, was an Englishman – a biophysicist like Franklin, as well as a molecular biologist (Watson, 2001).

Although Watson and Crick were initially studying the structure of proteins, Watson became bored with the subject, and turned his attention to DNA. When Franklin gave her lecture at King’s internal seminar in November of 1951, Watson attended. However, he took no notes and misremembered her key points. Consequently, the model that he and Crick built on the basis of that information erroneously placed the DNA’s backbone inside the molecule rather than on the outside (Watson, 2001). Their model, which was subsequently shown to Franklin and Wilkins, led to embarrassment, and they were forbidden by the Laboratory’s head – Sir Lawrence Bragg – from building any further DNA models and ordered to return to the study of proteins (de Chaderevian 2003). The model-building endeavours were subsequently suspended (Elkin, 2003).

Watson and Crick’s next breakthrough in unravelling the structure of DNA came only after another encounter with Rosalind Franklin’s work. In January of 1953, Watson came to visit Wilkins at King’s College; the latter showed him a (secretly duplicated) copy of Franklin’s photograph #51 (Watson, 2001; Selva, 2003). On Watson’s return to Cambridge, he and Crick began modelling the DNA structure anew – within a week, they had the correct model of DNA (Ropoport, 2002; Elkin, 2003). They submitted an article detailing that model to Nature very quickly – the editor’s acknowledgement of the receipt of the manuscript is dated March 18, 1953 (Elkin, 2003). The article proposed that the structure of DNA consists of two helical chains, both coiled around the same axis; that the helix is right-handed; and that the two chains
are held together by paired nucleobases, which always pair adenine with thymine and guanine with cytosine (Watson and Crick, 1953).\textsuperscript{79}

Franklin, Watson and Crick’s solution of the riddle of the structure of DNA was not only one of the key steps in the development of molecular genetics (Mcgregor and Poon, 2003) – it also helped to finally put out the flames of the controversy around the plausibility of Avery et al’s findings (Lederberg, 1994). Beyond that, the discovery was one of the last key puzzles to fall into place before the Human Genome Project could begin in earnest.

6.2. The Human Genome Project

6.2.1. The Project Timeline

The first meeting to explore the feasibility of the sequencing of the human genome was hosted in 1985 by Robert Sinsheimer, then the Chancellor of the University of California in Santa Cruz. The United States’ government came on-board fairly quickly thereafter, as in March of 1986 a conference on the topic was commissioned by Charles DeLisi, the Director of the Office of Health and Environmental Research at the Department of Energy, and the Department of Energy announced its Human Genome Initiative in 1986. Pilot projects were begun at the Department of Energy the same year (Roberts et al, 2001; Watson 1990).

In 1988, James Wyngaarden (then the Director of the National Institutes of Health) announced that NIH will also participate in the Human Genome Project, appointing James Watson – of the doubly-helical structure of DNA fame – as the head of its Office of Human Genome Research (later renamed National Centre for Human Genome Research). That same year (1988), the US Congress appropriated initial funding of $17.3 million to the NIH and $11.8 million to the DOE for the human genome project. National Institutes of Health signed a memorandum of agreement with the Department of Health, describing plans for the two agencies’ cooperation in genomic research in 1988 (Watson and Cook-Deegan, 1991).

\textsuperscript{79} In the article, Franklin’s role is explicitly recognised just once, although the acknowledgement itself is grossly understated. Watson and Crick note, at the end of the paper, that they “have also been stimulated by knowledge of the general nature of the unpublished experimental results and ideas of Dr. MHF Wilkins, Dr. RE Franklin and their co-workers at King’s College, London.” (Watson & Crick, 1953, 737). Neither Watson nor Crick mentioned Franklin’s name in their 1962 Noble Prize speeches; Wilkins mentioned Franklin twice, again, however, grossly understating her role.
Although the US governmental agencies came onboard fairly quickly, that the Project actually commenced was a triumph against numerous objections from the scientific community. Among those most-often levied was that the HGP constituted a substantial deviation from the way that biology had been practiced up until that point – consisting of small, investigator-led studies, conducted predominantly in individual labs. The Project’s price tag – estimated at 3 billion dollars at the start of the project – led to further concerns that its funding would deplete funding from other research. Remaining worries centred around the technical feasibility of the project – the technology in the late 80s and early 90s was insufficiently developed to facilitate the completion of the Project – and the Project’s scientific value – the latter being seen as quite limited in light of the view that 95% of the human DNA was the so-called ‘junk DNA’, i.e. DNA that did not code for genes (Roberts, 2001; Watson and Cook-Deegan, 1991; Watson 1990; Guyer and Collins, 1995). Objections notwithstanding, the 15-year project formally begun in the United States on October 1, 1990 and was named the Human Genome Project (HGP) (DOE 2011; Lorentz et al 2002).

The United States was not the only jurisdiction focused on sequencing the human genome; in the late 1980s, several other jurisdictions were testing the very same waters. For example, in 1987, Italy’s National Research Council involved fifteen Italian research groups in a pilot project on genomic research. United Kingdom began its own collaborative genome programme – joint effort between the Medical Research Council and the Imperial Cancer Research Fund – in February of 1989. That same year, USSR’s genome project was appropriated funds by the government. The following year, France’s government, too, announced its own genome programme (Watson and Cook-Deegan, 1991).

Subsequently, the Project turned into a collaborative, international endeavour. Toward that end, HUGO – the Human Genome Organisation – was established and funded. Its aim was to provide the international coordination of the genomic research for the project (IHGSC 2001). Not only was HUGO’s aim the international coordination, it was also internationally funded – for example, by 1990, HUGO had obtained funding commitments from both US (the Howard Hughes Medical Institute) and UK (Wellcome Trust) bodies (Watson and Cook-Deegan, 1991).
The international collaboration on the HGP was referred to as the International Human Genome Sequencing Consortium. As noted above, the coordination was facilitated by HUGO, however, it took place largely on a scientist-to-scientist level. Individual participants included over 2000 scientists, from over 20 institutions, in countries including: the US, UK, USSR, Canada, Australia, Italy, Japan, France, Germany, China (Watson 1990; IHGSC 2001; Government of Ontario 2002; Lorentz et al, 2002; Baltimore 2001). Among the many fruitful international collaborations on the Project, one may count: the sequencing of chromosome 22 (by a collaboration of the British, US and Japanese researchers), the sequencing of the C.elegans genome (a US-UK collaboration), a physical map of chromosome 16 (a US-Australia collaboration), the physical mapping of chromosome 21 (US-Japanese collaboration), and numerous others (Roberts et al, 2001; Collins and Galas, 1993).

6.2.2. The formal goals and the hopes for the Project

At the inception of the Project, the following key goals were laid out:

1. Development of the entire human genome map (both genetic and physical)
2. Determination of the human DNA sequence, as well as the sequence of several model organisms (including fruit fly, nematode, E. coli, yeast and mice)
3. Development of a means for storage, analysis and interpretation of the data
4. Development of the technologies required to accomplish these goals
5. Evaluation of the ethical, legal and social implications of the project (the ELSI programme) (Lorenz, 2002; Sawicki et al 1993; IHGSC 2001)

Although it was the Human Genome Project, the decision to analyse genomes of model organisms was quite crucial to its success. Model organisms were included within the scope of the Project for two reasons. First, a remarkable level of conservation of the genomic sequence obtains between the lower and higher level organisms (Ankeny, 2001). Thus, a study of the genetic characteristics of one organism – a model organism – will yield useful information about other organisms – including human organisms. Consequently, it was hoped that model organisms would help the scientists to interpret the information gained about the human sequence (DOE 1990). The second reason for the inclusion of model organisms in the Human Genome Project was that research on and with model organisms was an economical way of developing novel genetic technologies (such as gene sequencing and gene mapping, for example). These technologies – perfected on simpler organisms – could subsequently be applied to the study of a more complex, human organism (DOE 1990).
Although the Project was envisioned as a series of 5-year goals, and thus the first 5-year plan was originally set to last until 1995, by 1993 a revision to the initial set of goals was required. This was, partly, because it was becoming clearer what the Project needed to accomplish, and partly because the project was meeting its goals ahead of time (Collins and Galas, 1993).

Among the 5-year goals listed for the Project in 1993, were:

1. A 2-cM to 5-cM human genetic map (related to 15-year plan’s goal 1)
2. A physical map at resolution of 100kb (related to goal 1)
3. Increase of the DNA sequencing capacity through the development of new technologies (related to goal 4)
4. Development of methods for gene identification (related to goal 4)
5. Development of new and improved technologies required for the completion of the HGP (related to goal 4)
6. Continued sequencing of model organisms (related to goal 2)
7. Continued development of informatics tools (databases, software, etc.) required for the HGP (related to goal 3)
8. Continued research in the ELSI areas, and development of policy options to address issues that are identified (related to goal 5)
9. Training of personnel in genome-related interdisciplinary sciences (goal addition)
10. Technology transfer (goal addition)
11. Dissemination of the products of HGP research to the wider community (goal addition) (Collins and Galas, 1993).

However, as soon as 1995, it was becoming apparent that the 1993 plan will itself need to be superseded by a new 5-year plan prior to its planned end-date (Guyer and Collins, 1995). Although the next 5-year plan was ultimately not released until 1998, it is noteworthy that it accelerated the completion of the Project to 2003 (DOE 2011).\textsuperscript{80} The 1998 plan listed the following goals:

1. Human DNA sequence: working draft to be completed by 2001, complete sequence by 2003 (related to 15-year plan’s goal 1)
2. Reiterated the need for improvements to the efficiency of the DNA sequencing technology (related to goal 4)
3. Emphasised the need for basic information about the types, frequency and distribution of variation in the human genome and the function of the known genes (related to goal 2)

\textsuperscript{80} This was in response to the actions of Celera Genomics. Celera Genomics was formed in May of 1998, with Craig Venter at its head. That same year, Venter announced that they will sequence the human genome in 3 years at the cost of $300 million, beating the publicly funded HGP by 4 years. In response to the announcement, the publicly funded Human Genome Project advanced its timeline to obtain the first draft of the complete human genome sequence by 2001, and the completion date for the final draft was moved from 2005 to 2003 (DOE 2011; Roberts et al, 2001; Collins et al, 1998).
4. Underscored the value of continued study of genomes of model organisms as a source of insight into human gene structure and function (related to goal 2)
5. Reiterated the 1993 plan’s recognition of the import of the study of the ELSI implications of genome research (related to goal 5)
6. Emphasised the import of bioinformatics and computational biology to the implementation of the genome project (related to goal 3)
7. Underscored the need for specialists with skills at the intersection of biology and other disciplines (including other natural sciences, engineering, computing, etc.) (goal addition to the 15-year plan, although previously listed as goal 9 in 1993) (Collins et al 1998)

The Human Genome Project had enormous symbolic significance to the participating scientists, and to the non-participating scientists, the public, policy-makers and the media alike; according to some observers, the HGP was perceived as on the same scale as humanity’s first landing on the moon, or detonation of the first atomic bomb (Paabo, 2001). Therefore, in addition to the Project’s formal goals, a lot of hopes were also expressed with respect to the benefits anticipated from its completion.

In 1990, that is, shortly after the official start of the Project, James Watson expressed the belief that the Project will “explain, at the chemical level, the role of genetic factors in a multitude of diseases, such as cancer, Alzheimer's disease, and schizophrenia, that diminish the individual lives of so many millions of people” (Watson, 1990, 44). In 1993, Sawicki et al noted that the knowledge of the sequence will make it possible to intervene at the genetic level, envisioning treatment of genetic diseases and tumours with gene therapy becoming de rigueur. Beyond that, the HGP was hoped to provide the tools required for treatment of both monogenetic and multifactorial genetic diseases (Sawicki et al, 1993). Putting forth what turned out to be the Project’s final 5 year plan, Collins et al noted in a similar vein that the Project will facilitate the understanding of the relationship between genetic mutations and diseases (Collins et al 1998).

Much the same hopes were still being expressed by observers near or just after the completion of the Project. In 1999, Gostin and Hodge were hoping that the following medical benefits would follow from the Project:

(1) Availability of predictive information about individuals’ health status
(2) Identification of aetiology and physiology of diseases
(3) Advances in treatment of genetic diseases
(4) Improvement in public health protection through tracking of patterns of genetic carrier states and diseases (Gostin and Hodge, 1999)

Health-care policy makers were anticipating similar benefits. For example, the Government of Ontario hoped that the medical advances stemming from the HGP would fundamentally transform healthcare for Canadians, in areas ranging from diagnostics, to disease management and treatment (Government of Ontario, 2002).

6.2.3. The Project’s major outcomes

In December of 2000, the announcement of the completion of the working draft of the DNA sequence of the entire human genome was made; the initial working draft of the publicly funded project was published in Nature’s February 15, 2001 edition.81 Human Genome Project was officially declared to be complete two years later, in April 2003 (DOE 2011) – two years ahead of the original 15-year plan. But how did the outcomes of the project map onto the both the formal goals set out for the Project, and the hopes associated with it?

6.2.3.1. Achievement of the formal goals?

Goal 1: Development of the entire human genome map (both genetic and physical)

The physical map (that is, a map of distances between DNA sequences) of the human genome was published by the HGP in a nearly completed form on February 12, 2001. The map covered more than 96% of the genome, and guided the assembly of the draft human genome sequence (HGP 2001). The genetic map (also sometimes referred to as a linkage map), on the other hand, represents the linear order of the genes on the chromosome. The goal for 1993-1998 plan was to produce a genetic map with resolution of 2 to 5 cM; a 1 cM map was published ahead of schedule, in September 1994 (Collins et al, 1998)

Goal 2: Determination of the human DNA sequence, as well as that of several model organisms

By the end of 2001, 98.5% of the sequence was deposited in the public database (GenBank) in either a draft or finished form, with approximately 50% being deposited in the finished form (defined as less than 1 error per 10,000 base pairs). Genomic sequences of model organisms –

---

81 Celera Genomics’ draft sequence was published the next day, in Science’s February 16, 2001 edition.
including Mycobacterium tuberculosis, E.coli, S. cerevisiae (yeast), C.elegans (nematode worm), D. melanogaster (fruit fly) – were also published prior to or by that date. Mouse genome was also nearly completed at the time that the draft sequence was released (Lorentz et al, 2002).

Goal 3: Development of a means for storage, analysis and interpretation of the data

The main bioinformatic needs of the HGP fell into two categories: databases and analytical tools. Databases were required in order to store the data ensuing from the project (mapping, sequencing, etc.). The analytical tools (in particular, software) were required to process the data – view it, correlate it, analyse it, and so forth (Collins et al, 1998). Among the many databases that were developed for the Project were: GenBank (the public database where the DNA sequences were deposited), and EGAD (the expressed gene anatomy database). Among software developed for the project, one may mention: BLAST (which can be used to identify similar sequences, e.g. between different organisms), FLAST (another rapid sequence comparison program), CAP 3 (which assembles sequences from ESTs or Expressed Sequence Tags – partial sequences from either end of a cDNA clone), and many others (Liang et al, 2000).

Goal 4: Development of the technology required to accomplish HGP’s goals

The project – as planned – gave rise to numerous technological advances. For example, new sequencing methods were developed, including the ‘shot-gun’ approach to large-scale sequencing. In this approach, the genome is essentially sliced into small fragments, and a computer reassembles the entire sequence by identifying overlapping regions (Guyer and Collins, 1995; Roberts 2001). Other notable technological developments included: the development of new types of genetic markers, improved vectors for the cloning of large DNA fragments, and improved technology and automation for DNA sequencing (Collins and Galas 1993).

Goal 5: Evaluation of the ethical, legal and social implications of the project (ELSI)

In the United States, ELSI research was focused predominantly on issues around the clinical integration of new genetic technologies, privacy of genetic information, as well as the education of professionals and the public around these issues. Some of the voluminous research carried out in these areas included: a project on the genetic counselling of patients
considering genetic testing of cystic fibrosis patients, and a similar project on counselling and testing of patients for several cancers (including breast, ovarian and colon) (Guyer and Collins, 1995). It is worth noting that ELSI research is still ongoing both in the United States and in its various international permutations (e.g. GE3LS – Genomics, Ethics, Environmental, Economic, Legal and Social Aspects – project, in Canada).

6.2.3.2. … and what about the hopes?

Recall, the hopes associated with the project, centred predominantly on anticipated medical or clinical benefits – including the understanding of the diseases, prediction of diseases, treatment of diseases and new medical technologies. How much of that actually materialised?

With respect to the understanding of disease, as the PHG Foundation82 succinctly put it in its 2010 report, “there is little doubt about the [HGP’s] potential to provide a better understanding of disease mechanisms” (PHG Foundation 2010). The publication of the draft sequence in 2001 has allowed the identification of multiple gene/disease links, including those for a variety of cancers (including breast, skin and colon), Huntington’s disease, Alzheimer’s disease, fragile X syndrome, cardiovascular disease, diabetes, arthritis and muscle disease (Government of Ontario, 2002; PHG Foundation 2010). On the other hand, the vast majority of disease-gene associations identified have been for single-gene (rather than multifactorial) disorders, with notable exceptions including cystic fibrosis, Duchenne muscular dystrophy and Bruton’s disease (Bentley, 2000). Moreover, although the function of approximately 15,000 genes was known at the completion of the HGP, the function of the remaining genes – whether linked to diseases or otherwise – remained unknown (Lorentz et al, 2002).

Understanding the aetiology of disease, however, does not necessarily translate into identification of treatment for that disease. Although many genes linked with particular diseases have been identified, treatment for the vast majority does not yet exist. Some genetic diseases, admittedly, do have treatment. For example, in the case of multiple endocrine neoplasia type 2, undergoing thyroidectomy greatly reduces the likelihood of dying from the medullary thyroid carcinoma (that is almost certain to develop in those who are carriers)

82 Its full name is the Foundation for Genomics and Population Health; PHG is an independent, Cambridge (UK)-based organisation with emphasis on evidence-based medicine and health policy.
(Evans et al, 2001). Many other genetic diseases, however, still lack treatment. This is the case, for example, with both Huntington’s Chorea and Alzheimer’s disease (CRS 2007).

For the most part, the hopes associated with HGP vis-à-vis prediction of diseases have also been overestimated (PHG Foundation 2010). Some predictive tests are currently available – including those for Alzheimer’s disease, Huntington’s Chorea, breast and ovarian cancer (BRCA1 and 2 tests), multiple endocrine neoplasia type 2, haemochromatosis, colorectal cancer (Sudell, 2001; Evans et al 2001). However, the utility of those tests is largely questionable. This is because in many cases, only a percentage of carriers will go on to develop the disease – the estimate for breast cancer mutations, for example, is that 15-60% of carriers will never go on to develop the disease. In some cases, admittedly, predictive tests will identify carriers who are certain or nearly certain to develop the disease – this is the case in Huntington’s and multiple endocrine neoplasia type 2. However, those tests are unable to predict either the precise time of symptom onset or their severity, and, as noted above, no treatment for these diseases yet exists. At most, the value of these tests depends on the individual patient; it may be circumscribed to their allowing the patient to make long-term planning decisions (including reproductive, career, education, life insurance, and similar decisions) (Hildt, 1999; Rothstein 2007; Sherwin and Simpson, 1999, Evans et al 2001; CRS 2007).

Some unfulfilled – or, perhaps more accurately, as yet unfulfilled – hopes notwithstanding, two important medical technologies have materialised as a result of the discoveries of the HGP: pharmacogenomics applications and genetic tests. Pharmacogenomics studies how genes affect individual’s response to drugs; TRUEGENE HIV-1 test (which provides information on an individual’s responsiveness and resistance to particular medications used in treatment of patients with HIV) is just one example of such application.

Genetic tests are another important medical applications stemming from the HGP. Before the HGP was officially concluded, over 600 genetic tests were available for use, including tests for diagnosis of sickle cell, Down’s syndrome, Cystic Fibrosis, Haemochromatosis, Breast Cancer and Colon Cancer (Government of Ontario, 2002; Collins and Galas, 1993; PHG Foundation 2010). As of December 2011, genetic tests exist for nearly 2500 diseases (NCBI 2012).
6.3. Genetic Testing

6.3.1. The Science...

Before considering genetic tests in more detail, however, a brief detour into their scientific background is needed. With the exception of mature red blood cells, all of human cells possess nuclei, which in turn contain chromosomes. Normally, human cells have 23 pairs of chromosome – one member of each pair is inherited from each parent. Those 23 pairs include: 22 pairs of autosomal chromosomes (numbered 1-22) and one pair of sex chromosomes (chromosome X and chromosome Y). Chromosomes are composed of deoxyribonucleic acid (DNA), whose doubly-helical structure, as described previously, was first published by Watson and Crick in 1953. DNA itself is composed of a sequence of nucleotides: adenine (A), thymine (T), cytosine (C), and guanine (G). The entirety of DNA in the human genome consists of 3 billion nucleotides (and because a human cell possesses 2 copies of the genome – one from each parent – each cell contains 6 billion nucleotides).\(^{83}\)

Genes themselves are segments of DNA, which, as Thomas Hunt Morgan first observed, are located on the chromosomes. Each gene has a unique location on the chromosome, and each consists of a unique sequence of nucleotides (A, T, C, and G). About 5% of our genes code for (that is, provide the instructions for the building of) proteins: the DNA sequence is transcribed by the cell into RNA, which in turn is translated into specific proteins. The proteins so produced fulfil a variety of functions, including structural (qua building components of tissues), chemical (qua enzymes), messengers (e.g. hormones) and gene expression regulation. The remaining 95% of our genes – sometimes referred to as non-coding DNA or ncDNA – are involved in a variety of tasks, some of which include the control of physical structure of chromosomes themselves, regulation of genes, and many other functions which are not presently understood (NHMRC 2010; CRS 2007; Government of Ontario 2002).

Technically speaking, a genetic test is simply an examination of a gene (that is, the DNA sequence in the chromosome) with the aim of identifying a mutation that is associated with a particular disease (Government of Ontario 2002). However, not all definitions are

\(^{83}\) Although the nucleus contains the majority of DNA in the cell, some DNA is also contained in the mitochondria of each cell. Mitochondrial DNA codes for protein necessary for mitochondrial functioning (NHMRC 2010).
circumscribed only to the examination of DNA; some definitions also include within the scope of genetic testing the analysis of other elements. One of the most often cited, broad definitions stipulates that genetic testing is “the analysis of human DNA, RNA, chromosomes, proteins, and certain metabolites in order to detect heritable disease-related genotypes, mutations, phenotypes, or karyotypes for clinical purposes” (Holtzman and Watson, 1997).84

Variations in DNA sequence from one person to another are fairly small – they are estimated to range from 0.1 to 0.2% (NHMRC 2003) – and only some of those variations have implications for health. Variations are usually categorised as one of three types. The first type are the harmless variations – i.e. polymorphisms. These include differences between individuals with regard to height, hair colour, eye colour and so forth; they generally have no effect on the individual’s health. The second type of DNA sequence variations includes those variations whose effects on health are not certain – the susceptibility-creating or disease-modifying variations. Finally, the third type of DNA sequence variations – those most salient to medicine – are the disease-causing mutations (NHMRC 2003).

‘Genetic diseases’ have traditionally been divided into monogenic and multifactorial diseases. In monogenic (also referred to as Mendelian) diseases, a variation in DNA sequence will usually cause a disease. Monogenic or Mendelian diseases are fairly rare, although over 2,300 of them have been identified – Huntington’s chorea, cystic fibrosis and haemophilia fall into this category. Multifactorial diseases, on the other hand, result from a combination of causes, which include both genetic mutations and a variety of environmental factors (e.g. smoking, pollution, etc.). Cancers are often counted in this category, as are diabetes, dementia and obesity. Identifying the genes involved in these diseases is very challenging, and genome-wide association studies which involve thousands of subjects are required. Currently, it accepted that the majority of diseases fall into this category (NHMRC 2003; RCPA 2008; NHMRC 2010).

84 The breadth of the definition adopted has implications for policy development. For the purposes of health policy development, it may be preferable to define genetic tests narrowly, as this will have implications for financial coverage of the tests (by the publicly funded system), downstream costs etc. On the other hand, a broad construal of genetic tests may be preferable – for example, when anti-genetic discrimination policies are being developed (CRS 2007).
Thus, although traditionally, the term ‘genetic disease’ has referred to largely the monogenic or Mendelian diseases, it is now estimated that most diseases can be categorised as ‘genetic’ insofar as common chronic conditions (cancers, cardiovascular disease, diabetes, etc.) involve a combination of both genetic and environmental factors (CRS 2007). Consequently, for example, the Royal College of Pathologists of Australasia estimates that 71% of paediatric admissions to Australian hospitals have a genetic component and 60% of Australian adults will develop a disorder with a genetic component by the age of 60 (RCPA 2008). In Canada, the Government of Ontario similarly estimates that approximately 60% of adults will experience a disease with a genetic component over their lifetime (Government of Ontario, 2002). The estimated public health burden of genetic diseases is therefore quite substantial.

### 6.3.2. The Technology…

Genetic tests play a crucial role in the identification of mutations or genetic variants. ‘Genetic tests’ however, do not constitute a monolithic group; a variety of genetic tests exist, which may be categorised in a number of ways – e.g. according to the methodology adopted, the gene that is tested, or the purpose for which a genetic test is performed.

The most commonly encountered taxonomy of genetic tests is based on the last approach (NHMRC 2010). Based on their purpose, genetic tests may be divided into the following types:

1. **Diagnostic tests**: performed to diagnose a person already displaying symptoms of a particular disorder (e.g. testing for Fragile-X in a boy with mental retardation)
2. **Predictive or presymptomatic testing**: performed on an asymptomatic person in order to determine the likelihood that this person may develop a particular disease in the future (e.g. testing a healthy person for Huntington’s disease, BRCA1 or BRCA2 mutations). The risk of the disease materialising is contingent on the gene involved, the mutation involved and environmental factors
3. **Genetic carrier testing**: performed to determine whether a person has a genetic abnormality (which does not impact that person’s health) that may be passed on to children (e.g. Tay-Sachs, Cystic fibrosis)
4. **Screening testing**: performed on persons not usually known to face increased risk of a particular genetic disease; can be conducted on individuals or groups of varying size (e.g. PKU screening of newborns)
5. **Pre-implantation and prenatal testing**: performed, respectively, prior to implantation or in-utero, particularly where parents are at risk (e.g. they are carriers of Cystic fibrosis or Tay-Sachs disease)
6. **Pharmacogenetic testing**: tests for a person’s response to a medication (e.g. tests a genetic variant which affects the way a person metabolises drugs)
7. **Research testing**: analyse genetic information in medical or scientific context, to determine the influence of genes on health, or interaction of genes and environment in disease.

8. **Identification testing or forensic testing**: performed in criminal investigations, (e.g. in missing persons cases and for the purpose of victim or suspect identification).

9. **Parentage or kinship testing**: performed to establish the biological relationship between individuals (e.g. for custody purposes or in immigration proceedings). (ALRC 2003; RCPA 2008; RCPA 2006; NHMRC 2010)

It is worth emphasising here, that the same test may be classified as a different type of test depending on its timing. For example, Huntington’s Chorea testing on an asymptomatic patient would be categorised as a predictive test or possibly a carrier test; the same test on a patient displaying symptoms of the disease would be categorised as a diagnostic test.

### 6.3.3. … and the process of genetic testing.

Prior to undergoing any type of genetic testing, however, the patient will often first have a genetic counselling session, in order to ensure that the individual is acquainted with the ramifications of the possible results and to obtain informed consent. Should the testing proceed, the test is normally ordered by a physician (e.g., a general practitioner, medical geneticist, or other specialist) or a nurse practitioner (Mayo Clinic 2011; NIH 2012).

In order to procure the DNA required for testing, a blood sample is normally obtained from the patient. However, a buccal swab, skin, hair, or other tissue may also be used; in prenatal testing, amniotic fluid or a sample from the placenta is used for this purpose (Mayo Clinic, 2011; Centre for Genetics Education, 2007). The sample is sent to a laboratory, where the testing process involves the isolation of the DNA from sample, amplification (using Polymerase Chain Reaction or PCR) of the segment of interest, sequencing and analysis (NIH 2012, Centre for Genetics Education 2007).

The test results are sent back to the ordering physician (or nurse). The following types of results may be obtained:

- a) **Positive result**: indicates that the mutation that was tested for was detected
- b) **Negative result**: indicates that the mutation was not detected
- c) **Inconclusive result**: a mutation of unclear significance was detected
Different results and different genetic test types will have different implications for the patient. For example, a positive result from a diagnostic test will require a discussion of the disease management plan while a positive result on a carrier test will necessitate an assessment of the risk of disease to the offspring. A negative result does not guarantee that the patient lacks the mutation (as the test may be insufficiently sensitive). An inconclusive result indicates that it is not clear whether the mutation detected is disease-causing or not – and additional follow-up testing may be required (NIH 2012; Mayo Clinic 2011).

6.4. Conclusion

The public health burden of genetic disease is quite substantial. Moreover, genetic testing can potentially impact all life-stages, ranging from pre-conception (via carrier screening of the parents and pre-implantation testing of the embryos) through pregnancy (prenatal testing of the foetus), at birth (newborn screening), to childhood and adulthood (diagnostic, predictive testing, carrier testing, pharmacogenetic testing, identification or forensic testing and parentage or kinship testing).

Genetic tests, in providing us with a means of identifying genetic mutations that underlie genetic diseases, therefore have the potential for playing a crucial clinical role at all life stages. Although ‘genetic testing’ has so far turned out to be more hype than reality – predictive testing for the most part is probabilistic rather than certain, many genetic diseases lack treatment, diagnostic testing does not yet exist for many mutations and so on – some progress continues to be made in their clinical applications. This is evinced, for example, by integration of several genetic tests into healthcare systems in countries such as Canada, Australia and the United Kingdom.

What is not clear, however, is the evaluative process that genetic tests undergo prior to their integration into the health systems – and whether that evaluative process is adequate. The next section of the dissertation will therefore focus on answering these questions. In particular, it will focus on establishing whether genetic tests are ‘health technologies’ as the field of health technology assessment understands this concept, and, contingent on the answer to this question, how should genetic tests be assessed.
CHAPTER 7: Testing the Proposed Concept of Evidence Part II: Genetic Testing.

Summary: Due to the absence of a clear conceptualisation of ‘evidence’ in the field of HTA, chapter 4 proposed that, in HTA, ‘evidence’ ought to be construed as: available data that is relevant to the issue being addressed, question asked, etc. Chapter 5 tested this proposed conceptualisation of evidence against HTA agencies’ evidentiary behaviours vis-à-vis health technologies generally speaking. The present chapter tests the proposed conceptualisation against the HTA agencies’ evidentiary behaviours with respect to genetic tests in particular.

To this end, I will begin by, first, considering whether genetic are a sui generis health technology. Subsequently, Canada’s, Australia’s and UK’s approaches to the evaluation of genetic tests will be described and compared. In the final section, I will show that in context of genetic testing, the agencies adopt the same conceptualisation of ‘evidence’ as they do in context of health technologies more generally – that is, as of data from stipulated domains. Considering a hypothetical case study of a TTR-FAP genetic test evaluation, I will argue that the conceptualisation of evidence proposed in chapter 4 would handle the evidentiary issues around that assessment more adequately than would the agencies’ conceptualisation of evidence.
7.1. Are genetic tests a sui generis health technology?

Before assessing the HTA agencies’ evidentiary practices vis-à-vis genetic tests, a crucial issue first needs to be resolved. This is whether genetic tests ought to be considered ‘standard’ health technologies\textsuperscript{85} or, whether some property (or a set of properties) relegates them to a separate class of health technologies (call them sui generis health technologies). Resolving this issue is important, as a sui generis status could potentially justify a different evidentiary approach to genetic tests than to ‘regular’ health technologies. Conversely, establishing that genetic tests are a ‘regular’ health technology would justify an expectation of a similar evaluative behaviour on the agencies’ parts vis-à-vis genetic tests as vis-à-vis other health technologies.

I will therefore begin by considering three arguments in favour of the view that genetic tests are sui generis health technologies: organisational, evaluative, and arguments from genetic exceptionalism. I will argue that none of these arguments succeed; moreover, I will argue that the definitions of ‘health technologies’ adopted by the agencies themselves comfortably capture most types of genetic tests. Thus, the agencies’ evidentiary approaches vis-à-vis genetic tests can reasonably be expected to be similar to those vis-à-vis other health technologies.

7.1.1. In favour of sui generis status of genetic tests

The arguments for the sui generis status of genetic tests can be put forward on three grounds: organisational, evaluative and arguments from genetic exceptionalism. Consider each in turn.

7.1.1.1. Organisational considerations

The first argument in favour of sui generis status of genetic tests is that some jurisdictions separate out the evaluation of genetic tests from the evaluation of other health technologies by assigning this task to a separate agency. This is the case in the UK, for example.\textsuperscript{86} Whilst the

\begin{footnotesize}
\textsuperscript{85} ‘Standard’ is intended here in the descriptive sense of the word (rather than a normative one), conveying a meaning of ‘ordinary’ or ‘usual.’

\textsuperscript{86} As will be discussed in more detail shortly.
\end{footnotesize}
HTAs for other health technologies are carried out (at the national level) by NIHR, the evaluation of genetic tests is carried out by the United Kingdom Genetic Testing Network (hereafter, UK GTN). One could argue here that the underlying reason for the separation of the evaluation of genetic tests from the evaluation of ‘standard’ health technologies is the recognition that genetic tests are importantly different from other health technologies, and thus require a different evaluative process and different evaluata.

Two responses can be made in response to this argument, however. First, of course, it is true that some jurisdictions do separate out the assessment of genetic tests from the assessment of ‘standard’ health technologies. However, it is equally true that other jurisdictions do not separate out the evaluation of genetic testing from the evaluation of ‘standard’ health technologies. Canada and Australia can serve as examples here – at the national level, both jurisdictions rely on the same agencies (respectively, CADTH and MSAC) for the evaluation of genetic testing as they do for the evaluation of other health technologies.

Second, even health technologies whose status as ‘standard’ health technologies is not in question are sometimes separated out for assessments by HTA agencies – this is frequently the case with pharmaceuticals. For example, as noted in chapter 2, Australia separates its assessment of pharmaceuticals from the assessment of other health technologies – whilst MSAC evaluates devices and procedures, pharmaceuticals fall under PBAC’s remit. Thus, an evaluation of a health technology by a separate HTA agency does not decisively establish that health technology’s status as sui generis.

7.1.1.2. Evaluative framework (ACCE)

Nevertheless, one could argue on different grounds for the sui generis status of genetic tests. That is, one could argue that genetic tests are sui generis not because their evaluation is sometimes conducted by a separate agency – but rather, on the grounds that a unique evaluative framework has been designed for assessment of genetic tests prior to their integration into a health system. This is the ACCE framework, which takes its name from the first letter of each main evaluative category: Analytic validity, Clinical validity, Clinical utility, and Ethical, legal and social issues. The framework is a result of a collaboration, in 2000,
between the National Office of Public Health Genomics (an office within the United States’ Centre for Disease Control) and the Foundation for Blood Research (CDC 2010a).

The framework defines **analytic validity** of a genetic test as its ability to measure the genotype of interest accurately and reliably. The analytic validity category can be further sub-divided into four constitutive elements, which include: analytic sensitivity, analytic specificity, laboratory quality control and assay robustness (CDC 2010b).

The next component of the ACCE evaluative framework – **clinical validity** – is defined as the test’s ability to detect (or predict) the disorder (or phenotype) that is associated with the genotype of interest. Just like in the case of analytic validity, clinical validity can be further differentiated. Clinical validity’s constitutive elements include: clinical sensitivity, clinical specificity, prevalence of the disorder, positive and negative predictive values, penetrance, and gene or environment modifiers (CDC 2010b).

The ACCE framework defines **clinical utility** as the risks and benefits which are associated with the introduction of the genetic test into clinical practice. The concepts of risks and benefits are not themselves formally defined by the framework’s authors, although ACCE includes under this section the questions about the test’s impact on patient care, financial costs associated with the testing, economic benefits ensuing from testing, and so forth (CDC 2010b).

The picture that emerges with regard to the **ELSI** (that is, Ethical, Legal and Social Issues) component of the ACCE framework is similar to the picture that emerges with respect to its clinical utility component – that is, rather than formally defining ELSI, the framework offers a series of questions which are meant to convey this category’s scope. The questions centre

---

87 The definitional gap has been filled in somewhat in the literature, although no consensus exists. Definitions range from fairly narrow to quite broad, with the narrow ones focusing solely on clinical considerations (e.g. whether the test will lead to improved health outcomes), and the broader ones also encompassing other considerations (e.g. costs associated with the test). For an example of a narrower approach, see Burke et al (2002); broader approach is evident in Grosse and Khoury (2006).
around issues such as consent, privacy, discrimination, personal and family social issues, safeguards in place, and so on (CDC 2010b). The fact that such an extensive evaluative framework was developed specifically for genetic tests could be taken as suggestive of the view that genetic tests are sui generis health technologies. Nothing like these categories was encountered in chapter 5 which discussed the assessments of other, ‘regular’ health technologies. What dominated there, were the categories of: safety, effectiveness and cost-effectiveness. Thus, one could argue that the very existence of such an extensive, health technology-specific evaluative framework tells against the fact that genetic tests are ‘regular’ health technologies.

However, as was established in chapter 5, the vast majority of CADTH’s, MSAC’s and NIHR’s evaluata were M&S procedures and pharmaceuticals. Yet, it is precisely vis-à-vis diagnostic tests that categories such as clinical validity and clinical utility become salient. Hence, the fact that we have not yet encountered these evaluative categories is simply a reflection of the agencies’ evaluative foci. Furthermore, for this argument to go through, it would need to be shown that the ACCE categories do not fit within the framework of the standardly assessed HTA evaluata (such as safety, effectiveness, etc.). That they do, in fact, fit within the standard framework, will be shown shortly.

7.1.1.3. Genetic exceptionalism issues

Of course the arguments that genetic tests are sui generis health technologies only go so far. The question is how these tests are purportedly different – and the arguments for genetic exceptionalism shed some light on this issue.

Health information obtained as a result of genetic testing is sometimes argued to be in its own, unique medical category. This view – known as genetic exceptionalism – was first put forth by Thomas Murray (Murray, 1997). While subtle – and not so subtle – differences in conceptualisations of genetic exceptionalism are evident in the literature, the main idea behind

---

There is a fair degree of divergence in the literature with regard to which issues are regarded as salient under the ELSI category. Scholars and government bodies include varying issues here. GE3LS programme – Genome Canada’s equivalent of Human Genome Project’s ELSI programme – for example, broadens the relevant considerations to also include economic and environmental issues (Genome Canada 2009).
genetic exceptionalism is that genetic health information is “intrinsically” unlike other personal, including medical, data” (O’Neil, 2004, 101; emphasis in original). Insofar as genetic tests are the means of providing this unique information, they may be argued to thereby fall into a category of health technologies that is sui generis.

Arguments in favour of genetic exceptionalism have been advanced on a wide variety of grounds in the literature. However, the most frequently encountered arguments are succinctly summarised in a report by Privacy New South Wales (hereafter Privacy NSW), an office of the New South Wales Privacy Commissioner in Australia. These include:

1. uniqueness of the genetic health information to the individual
2. genetic health information’s implications both for an individual and that individual’s genetic kin
3. the predictive value of genetic health information, and
4. the potentially devastating effects of the disclosure of this information on the individual and/or their family (Privacy NSW, 2002).

However, none of these arguments satisfactorily establish the uniqueness of genetic information (and – what follows from it – the uniqueness of genetic tests as health technologies). In response to argument (1), which emphasises the traceability of health information to a particular individual, one may point out that other types of data share this property. For example, as Gostin and Hodge (1999) note, features such as fingerprints, facial features, voice spectrograms, irises, and birthmarks have the same capability. If genetic tests are to be declared sui generis health technologies on the grounds that they provide information that is traceable to a particular individual, then so, too, should be the eye scanners, fingerprint scanners, voice recognition software, and so forth.

Broadening the notion of a ‘patient’ from an individual to a larger group (which is the focus of arguments falling into category 2) is, likewise, scarcely unique to genetic tests. Diagnosis with an infectious disease will also have implications both for the patient and those around her. Whether it is an air transmissible disease (e.g. measles or tuberculosis, which will have implications for those in the patient’s physical vicinity) or a sexually transmitted disease (e.g. HIV or syphilis, which will have implications for that patient’s sexual partners and their partners, in turn), diseases other than genetic broaden the notion of a ‘patient’ to a larger group. The only difference between the genetic and non-genetic cases is that the larger group
consists of those related by blood. However, the salient issue here is the broadening of the notion of the patient to a larger group of people – which obtains in both the genetic and non-genetic cases. To argue that genetic tests are a sui generis health technology on the grounds that they broaden our notion of a patient from one individual to a group, then, is a non-starter.

**Argument 3** for genetic exceptionalism holds that genetic tests are unique insofar as they offer predictive health information. Recall, however, that as discussed in chapter 6, a number of different types of genetic tests exist, including: diagnostic, predictive, screening, prenatal, paternity, and so forth. At most, then, this argument would establish sui generis status only of predictive genetic tests. However, few predictive genetic tests offer information that indicates a **high probability** of the disease materialising. Falling into this category are predictive tests for Huntington’s Chorea and multiple endocrine neoplasia type 2, for example. The vast majority of predictive genetic tests, however, are not as predictive. At most, then, this argument would apply only to one subset of genetic tests (the predictive tests) and an even smaller subset of tests within this category (predictive tests for diseases with a very high penetrance).

Finally, **arguments falling into category 4** advance the claim that genetic tests merit special consideration because they have the potential to offer devastating information to the patient and/or their family. However, this property, again, is not unique to medical information acquired via genetic tests; medical information acquired through a variety of other diagnostic means has the same potential. One might consider here, for example, the results of biopsies indicating familial amyloid polyneuropathy (typically terminal within 10 years) or stage IV melanoma (5 year survival rates of 15-20%). To argue that genetic tests are a sui generis health technology on the grounds that they have the potential to deliver devastating information is, again, an untenable position to hold.

**7.1.2. Against sui generis status of genetic tests**

In short, none of the genetic exceptionalism arguments considered above successfully differentiate genetic tests from other health technologies. Nor, as we have seen, do the organisational and evaluative framework (ACCE) considerations establish genetic tests as a sui generis health technology. However, showing that arguments for the sui generis position fail, does not yet establish genetic tests to be ‘regular’ health technologies. What would succeed in
establishing this is a successful demonstration that genetic tests clearly fall within the scope of the term ‘health technologies’ as the term is understood in HTA. Let us therefore consider how each of the three agencies under consideration here understand the term ‘health technology,’ and whether genetic tests can be captured within that term’s scope.

Consider, first, CADTH’s definition of ‘health technologies.’ CADTH holds that a health technology is:

any method or intervention that is used to promote health; prevent, diagnose, or treat disease; or improve rehabilitation and long-term care. Technologies include drugs, devices, diagnostic agents, equipment, and medical and surgical procedures. The definition also includes organisational and service systems that provide health care, such as tele-health. (Canadian Agency for Drugs and Technologies in Health, 2011)

CADTH’s definition certainly seems to accommodate genetic tests insofar as they are used to diagnose and prevent diseases. The prevention of disease aspect would arise in cases of the predictive, presymptomatic, or screening testing, where treatment exists. The diagnosis stipulation would be met by diagnostic testing, carrier testing, pre-implantation/prenatal testing and pharmacogenetic testing (regardless of the existence of treatment). Nevertheless, not all of the types of genetic testing discussed in chapter 6 are captured by this definition – research testing, identification/forensic testing, and parentage/kinship testing are excluded, insofar as they fail to promote health; prevent, diagnose or treat disease; or improve rehabilitation and long-term care.

CADTH’s Australian counterpart – that is, MSAC – does not itself offer a definition of ‘health technologies.’ However, the Australian federal government recently conducted a large-scale review of its three national level HTA programmes. That review stipulated the following definition:

Health technology includes medicines; diagnostics, devices, equipment and supplies; medical and surgical procedures; support systems; and organisational and managerial systems used in prevention, screening, diagnosis, treatment and rehabilitation (Government of Australia, 2009a).

Australia’s construal of health technologies as “diagnostics, devices, equipment […] used in prevention, screening, diagnosis, treatment” seems to accommodate many kinds of genetic
tests. ‘Diagnostics’ captures diagnostic genetic tests, as well as genetic carrier screening, screening testing, pre-implantation/prenatal testing, and pharmacogenetic testing. As in the case of CADTH, the prevention aspect of the definition would capture predictive/presymptomatic testing (in cases where treatment exists); alternately, the predictive/presymptomatic testing may be captured by under the ‘devices used in treatment’ stipulation (insofar as diagnosis is a part of the treatment pathway). The following types of genetic tests, however, are excluded from the definition: research testing, identification/forensic testing, and parentage/kinship testing – this is because these types of genetic tests are not used in prevention, screening, diagnosis, treatment or rehabilitation.

Finally, NIHR’s understanding of health technology “covers a range of methods used to promote health, prevent and treat disease and improve rehabilitation and long term care including:

- Drugs: such as antidepressants, contraceptives, antibiotics
- Devices: such as pacemakers, dialysis machines, hearing aids
- Procedures: such as surgical techniques, acupuncture, counselling
- Settings of care: such as general practice, hospitals, care homes
- Screening: for cancer, sexually transmitted diseases, stroke”89

At the most general level, ‘genetic tests’ may be captured here under the term ‘devices,’ although the term itself is not further defined, so it is difficult to assert this with certainty. However, even if this approach were untenable, an alternate means of capturing genetic tests is available under this definition. Diagnostic genetic tests can be captured under ‘a method used to treat disease’ (insofar as diagnosis is part of the treatment pathway). On the same grounds, one would also include here: genetic carrier testing, screening testing, pre-implantation/prenatal testing and pharmacogenetic testing. Predictive/presymptomatic genetic testing would be captured under prevention (if treatment exists) or improvement of long term care (if treatment does not exist). As in the case of CADTH and MSAC, the following types of genetic tests fall outside of the scope of NIHR’s conceptualisation of health technologies: research testing, identification/forensic testing, and parentage/kinship testing. This is because these types of tests are not related to health, prevention and treatment of disease, or improvement of rehabilitation or long-term care, and their status as devices is – as noted above – unclear under NIHR’s definition.

89 http://www.hta.ac.uk/about/index.shtml
7.1.3. In summary

To sum up, three sets of arguments in favour of considering genetic tests to be sui generis health technologies were considered and discharged here: the arguments from organisational arrangements, the arguments from evaluative differences (ACCE), and the arguments from genetic exceptionalism. Instead, it is apparent that the three agencies’ understanding of ‘health technologies’ readily accommodates many types of genetic testing. CADTH, MSAC and NIHR capture the following types of genetic tests within the scope of their definitions of ‘health technologies’: diagnostic, predictive/presymptomatic, screening, carrier, pre-implantation/prenatal, and pharmacogenetic testing. All three agencies uniformly exclude research testing, identification/forensic testing, and parentage/kinship testing from their construal of ‘health technologies.’

As genetic tests are a ‘standard’ health technology, we can therefore compare the agencies’ evidentiary behaviour vis-à-vis genetic tests to each other (this chapter) and to their behaviour vis-à-vis other health technologies (in Chapter 8). Let us therefore focus on the former, beginning with Canada’s approach to genetic testing.

7.2. Canada and genetic testing

7.2.1. Current status of genetic testing in Canada

Canada’s population numbered 34.7 million in January 2012. Approximately 500 million medical tests are conducted annually in Canada and it is estimated that between 50,000 and 100,000 of those are DNA tests (Rousseau 2009).90 If those numbers are assumed to have held at relatively constant levels over the past 3 years, this translates into one genetic test per approximately 350 to 700 persons.

There is nothing unique about the evaluation of genetic tests from the Canadian national level HTA perspective. That is, just like other health technologies, genetic tests are evaluated by the same agency (CADTH) and undergo the same evaluative process as other health technologies.

---

90 It is worth emphasising here that the fact that genetic tests are carried out within the framework of the Canadian healthcare system is not due to the efforts at federal level; as noted in chapter 2, the decentralisation of the Canadian healthcare system means that decisions about what particular health technology is funded are made by the provinces.
7.2.2. Method

Because, as discussed previously, CADTH does not make explicit its understanding of the concept of evidence, a similar approach to the one adopted in chapter 5 was adopted here. That is, a scan of the agency’s website was carried out, in order to identify 10 full health technology assessments of genetic tests, and draw out inferences about ‘evidence’ from their content. However, only three (3) full assessments of genetic tests were identified (listed in Table 9, below).

In order to ascertain the understanding of ‘evidence’ implicit in the reports, the ‘Objectives’ section of each report was analysed for statements indicating the report’s construal of what it means by ‘evidence.’ (Due to the absence of the ‘Objectives’ section in the 1999 report, the ‘Purpose and Scope’ section was used, instead). The understanding of ‘evidence’ suggested by each report’s Objectives (or Purpose and Scope) section, was verified for coherence with the body of the report. The search yielded the following reports:

<table>
<thead>
<tr>
<th>Report Title</th>
<th>Test type</th>
<th>Date</th>
<th>Focus of assessment</th>
</tr>
</thead>
</table>
| 1 BRCA1 and BRCA2 Predictive Genetic Testing for Breast and Ovarian Cancers: A Systematic Review of Clinical Evidence | Predictive      | 2006  | - Analytical validity  
- Clinical validity  
- Contribution of testing to genetic counselling  
- Contribution of testing to clinical management  
- Ethical and psychosocial issues inherent in BRCA1/2 testing |
| 2 Molecular Diagnosis for Hereditary Cancer Predisposing Syndromes: Genetic Testing and Clinical Impact | Diagnostic and screening | 2003  | - Analytical validity  
- Clinical validity  
- Availability of genetic test  
- Cost of the genetic test  
- Impact of genetic test on the clinical management of patient |
| 3 Predictive Genetic Testing for Breast and Prostate Cancer                   | Predictive      | 1999  | - Clinical relevance (including predictive ability, genetic counselling)  
- Ethical implications (informed consent, privacy and confidentiality)  
- Psycho-social implications (interest and attitudes, psychological distress)  
- Policy implications (policy statements and guidelines, genetic discrimination, genetic literacy and legislative awareness, genetic services laboratory, cost-utility data and normative evaluations, payment for genetic testing services) |
7.2.3. How CADTH understands ‘evidence’ in context of genetic test evaluation

(a) Number of genetic tests evaluated

As is evident from the above table, very few HTAs for genetic testing have been carried out in Canada – only three full assessments (in 2006, 2003 and 1999) have been completed. It is also worth noting here that all 3 of these reports were carried out while the Canadian federal HTA agency was still CCOHTA – in other words, it seems that CADTH (qua CADTH) has not thus far prioritised the assessments of genetic testing. However, it must also be acknowledged here that more recently, CADTH has issued three brief reports on genetic tests. In 2011, an environmental scan was carried out for a newborn screening programme; in 2006, a report was produced on emerging technologies (focusing on CYP450 genotyping for determining drug metaboliser status); and, also in 2006, a technology overview focusing on BRCA1 and BRCA2 was conducted. As none of these were full health technology assessments, however, they were discarded. Thus, CADTH’s (and CCOHTA’s) combined output of full health technology assessments of genetic tests amounts to a total of 3 assessments over a span of 13 years (from 1999 to present).

(b) Types of genetic tests evaluated

Of the three assessments that met the inclusion criteria here, two focused on predictive tests – the 1999 report assessed testing for breast cancer and prostate cancer, and the 2006 report, again, assessed breast cancer, as well as ovarian cancer. The 2003 assessment – which evaluated hereditary cancer predisposing syndromes – evaluated both screening and diagnostic tests.

Although the time-span between the first and last HTA of a genetic test is 7 years, the sample here is very small (3 assessments) which mandates caution in postulating trends. However, it is noteworthy that, for such a small sample, the range of genetic test types that is assessed is actually moderately broad, encompassing diagnostic, predictive and screening tests. (No carrier, prenatal or pharmacogenetic tests were assessed).
(c) Evaluata assessed (in ACCE terms)

In order to facilitate subsequent comparisons between the three jurisdictions’ evidentiary approaches to genetic testing, a common nomenclature is required. To this end, each jurisdiction’s evaluata (with respect to genetic testing) will be considered, here, in terms of the previously discussed ACCE framework.\(^{91}\)

The first of ACCE’s four categories, that is, analytic validity is a measure of the genetic test’s ability to accurately detect the genotype of interest. This category was quite clearly assessed in the more recent two reports (2006 and 2003); the oldest report (1999) failed to assess data on analytic validity.

Clinical validity is the genetic test’s ability to detect (in case of, for example, a diagnostic test) or predict (in case of a predictive test) the disorder (or phenotype) of interest. This category – just like the previous one – is explicitly assessed by the reports from 2006 and 2003. The assessment from 1999 – focusing on predictive testing for breast and prostate cancers – also evaluated this, insofar as it considered the predictive value of the test. (The 1999 report never explicitly uses the term ‘clinical validity’ – possibly because it predates the ACCE framework).

Recall that clinical utility concept was not clearly defined by the ACCE framework. It can capture a variety of issues, although typically focuses on the test’s impact on clinical management of the patient and the financial costs associated with the testing. All three reports assessed by CADTH evaluated the patient management dimension, here. The cost dimension was also assessed by two of the three reports – the 2003 and the 1999 reports. The most recent, 2006 report, failed to assess this.

ACCE’s fourth category – that is, ELSI – encompasses the ethical, legal and social issues around genetic testing. Amongst those, included are issues of: consent, privacy, family and

---

\(^{91}\) Two sets of comparisons will be necessary and carried out throughout this chapter because UK’s ‘regular’ national level HTA agency – that is, NIHR – does not evaluate genetic tests. UK GTN, which does carry out UK’s evaluations of the genetic tests, does so in terms of ACCE categories. Thus, comparisons between CADTH, MSAC and UK GTN will be carried out in terms of ACCE nomenclature (analytic validity, clinical validity, etc.). Comparisons between CADTH, MSAC and NIHR will be carried out using standard HTA inputs (safety, effectiveness, etc.).
social issues, discrimination, and so forth. 2006 report assessed this dimension, insofar as it evaluated both the ethical and psychosocial issues around BRCA 1 and 2 testing. Amongst 2003 report’s evaluata, the consideration of the availability of genetic tests may be counted here; the report construed ‘availability of the genetic tests’ as their availability in the clinical practice (as opposed to only in research setting). Just like the 2006 report, the 1999 report devoted a fair amount of space to ELSI issues, assessing issues around ethics and psychosocial implications of genetic testing. Some of what is listed by the 1999 report under the ‘policy implications’ is captured here as well, insofar as that report considered issues of genetic discrimination and genetic literacy.

To sum up, all three reports – with one exception – assessed evidence in all four of the evaluative categories stipulated by the ACCE evaluative framework. Analytic validity was assessed by 2 of the 3 reports. Clinical validity was assessed by all 3 reports. Clinical utility considerations, overall, were captured by all 3 reports (with the patient management dimension being assessed by all reports, and the financial issues dimension being assessed by 2 of the 3 reports). Finally, the ELSI category was assessed by all 3 reports, with the earliest (1999) report capturing the broadest range of issues under this category, and the 2003 report capturing the narrowest range. This is summarised in Table 10, below.

<table>
<thead>
<tr>
<th>ACCE categories</th>
<th>No. of reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic validity</td>
<td>2 of 3</td>
</tr>
<tr>
<td>Clinical validity</td>
<td>3 of 3</td>
</tr>
<tr>
<td>Clinical utility – patient management</td>
<td>3 of 3</td>
</tr>
<tr>
<td>Clinical utility – financial issues</td>
<td>2 of 3</td>
</tr>
<tr>
<td>ELSI</td>
<td>3 of 3</td>
</tr>
</tbody>
</table>

As Table 10 suggests, in context of genetic test evaluation, CADTH conceptualises ‘evidence’ as data on: analytic validity (some of the time), clinical validity (all of the time), clinical utility/patient management issues (all of the time), clinical utility/financial dimension (some of the time) and ELSI (all of the time).

92 The psycho-social included: the impact of counselling on test uptake (and the contribution of testing to genetic counselling, that is, the rates of post-test counselling), perceived risk and anxiety; knowledge about the association between cancer and genetics; psychological impact of test results (distress, depression); communication of test result to family members. Ethical issues centred around: informed consent, and privacy and confidentiality concerns.
(d) Evaluata assessed (in standard HTA terms)

As established above, the evaluata for the genetic tests assessed by the Canadian federal HTA agency generally map well onto the 4 evaluative categories stipulated by the ACCE framework. How, on the other hand, do these evaluata cohere with the standard HTA categories? To establish this, consider how CADTH’s evaluative practices with respect to genetic tests match up against its stated evaluata (viz. effectiveness, cost-effectiveness and broader impact of health technologies).

Clinical effectiveness has been defined as the extent to which a health technology achieves what it is intended to achieve under standard circumstances, as opposed to under ideal conditions (INAHTA 2006; Cochrane Collaboration, 2005/2010 offers a very similar definition). By that definition, clinical validity would map onto this category because clinical validity assesses the test’s performance in the clinical setting. The test’s impact on patient management would also be included here, insofar as this is something that ‘the health technology is intended to achieve.’ As all 3 of the reports assessed clinical validity and patient management issues, they can all be said to have assessed clinical effectiveness.

Cost-effectiveness can be understood as an economic analysis which describes the costs per health gain (e.g. cost per stroke prevented) (Cochrane Collaboration, 2005/2010). Only 2 of 3 reports have evaluated this category. 2003 report, because it considered the cost of the genetic test; 1999 report, as it assessed the cost-utility data and issues around payment for genetic test services. 2006 report did not consider these issues. In short, only 2 of the 3 reports assessed the cost-effectiveness domain.

The ‘broader impact of health technologies’ category poses some challenges here. Insofar as it is potentially conceptually very broad (both with respect to the type of impacts and level of impacts – as discussed in chapter 5), it may map onto the ELSI category of the ACCE framework. If this is the case, then 2006 and 1999 reports captured this dimension insofar as they evaluated both ethical and social issues around genetic testing. 2003 report would also be captured, insofar as it considered the availability of genetic tests – availability of a health technology in a publicly funded health system can reasonably be interpreted as an ethical issue.
In short, all 3 reports assess CADTH’s ‘broader impact’ category. In standard HTA terms, then, CADTH’s conceptualisation of ‘evidence’ thus includes data on clinical effectiveness and broader impact of health technology (all of the time), as well as cost-effectiveness (some of the time). This is summarised in Table 11, below.

Table 11: CADTH: evaluata considered in HTAs of genetic tests (standard HTA nomenclature)

<table>
<thead>
<tr>
<th>HTA categories</th>
<th>No. of reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effectiveness (qua clinical validity and clinical utility/patient management)</td>
<td>3 of 3</td>
</tr>
<tr>
<td>Cost-effectiveness</td>
<td>2 of 3</td>
</tr>
<tr>
<td>Broader impact of health technologies</td>
<td>3 of 3</td>
</tr>
</tbody>
</table>

7.2.4. In summary

In summing up CADTH’s conceptualisation of evidence in context of genetic tests, one must bear in mind two caveats. First, the sample here is extremely small – genetic testing simply has not been prioritised by CADTH. Second, the completed assessments were all carried out whilst CADTH was still CCOHTA. That acknowledged, CADTH’s understanding of ‘evidence’ here, has been fleshed out both in the general HTA terms and in ACCE terms. In HTA terms, CADTH seems to understand evidence in context of genetic testing as data on: clinical effectiveness and broader impact of the health technology (all of the time) and cost-effectiveness (some of the time). In ACCE terms, CADTH appears to construe evidence as data on: clinical validity, ELSI, and clinical utility (patient management dimension) all of the time. Some of the time, data on analytic validity and the financial dimension of clinical utility are also included in CADTH’s conceptualisation of ‘evidence.’

7.3. Australia and genetic testing

7.3.1. Current status of genetic testing in Australia

As of January 2012, Australia’s population numbered just over 22.8 million. The most recent national data on prevalence of genetic testing in Australia, however, is from 2007 (when the population was slightly lower, at 21.2 million). During 2007, a total of nearly 161,000 genetic tests were carried out in Australia, predominantly consisting of screening (nearly 64,500) and
diagnostic (over 45,000) tests (RCPA 2008). This translates into approximately one genetic
test performed per 130 persons.

Neither a separate assessment process nor a unique agency has been established for the HTA
of genetic tests in Australia. In other words, genetic tests – just like other health technologies
that are neither pharmaceuticals nor prostheses – undergo an evaluation by MSAC, and are
subject to its usual evaluative process.

7.3.2. Method

As MSAC fails to make explicit its construal of the concept of evidence, the approach adopted
in chapter 5 was also utilised here. That is, in order to establish how MSAC construes
‘evidence’ in context of its evaluations of genetic testing, a scan of the contents of MSAC’s
website was undertaken, in order to identify 10 most recent reports that assessed genetic tests.
Ten (10) full reports (assessing 9 genetic tests) were identified and are listed below, in Table
12, below.

The ‘Objectives’ section of each located report was identified, and its contents were analysed
for statements suggesting the report’s construal of ‘evidence.’ The construal of evidence
suggested by each report’s Objectives section was verified for coherence with the body of the
report.
<table>
<thead>
<tr>
<th>Report Title</th>
<th>Test type</th>
<th>Date</th>
<th>Focus of assessment</th>
</tr>
</thead>
</table>
| 1 Hepatitis C viral load testing (Application 1021)                          | Diagnostic (for access to therapy)     | 2000  | - safety (adverse effects)  
- effectiveness (predictive value of the test)  
- cost-effectiveness (cost per test)                                                                                                                                                                                                                                                                 |
| 2 Genetic test for Fragile X syndrome (Application 1035)                     | Diagnostic and relative carrier test   | 2002  | - safety (adverse effects, psychological burden)  
- effectiveness (diagnostic accuracy, incl. sensitivity and specificity; usefulness of the test in improving outcomes),  
- cost-effectiveness (cost per case detected, annual cost to the system)                                                                                                                                                                                                                   |
| 3 Antenatal screening for heritable thrombophilia (MSAC reference 9b)        | Prenatal screening                     | 2003  | - risks associated with heritable thrombophilia (mortality, thrombotic disease, etc.)  
- effectiveness (diagnostic accuracy of the tests, incl. sensitivity and specificity; effectiveness of the treatment),  
- cost-effectiveness (cost per case, annual cost to the system)                                                                                                                                                                                                                           |
| 4 Genotypic resistance testing of antiretrovirals in HIV (Application 1067)  | Diagnostic (for access to therapy)     | 2004  | - safety (adverse events),  
- effectiveness (diagnostic accuracy of the test incl. specificity and sensitivity; usefulness of the test in improving patient outcomes),  
- cost-effectiveness (incremental cost per QALY, cost to the system)                                                                                                                                                                                                                  |
| 5 Hepatitis B DNA testing (Application 1096)                                | Diagnostic (for access to therapy)     | 2007  | - safety  
- effectiveness (diagnostic performance incl. sensitivity, specificity; the impact of diagnosis on changes in patient management; improvement in health outcomes)  
- cost-effectiveness (annual cost to the system assessed; cost-effectiveness)                                                                                                                                                                                                                       |
| 6 Molecular testing for myeloproliferative disease (Application 1125)       | Diagnostic                              | 2009  | - safety (adverse events),  
- effectiveness (diagnostic accuracy incl. sensitivity, specificity; evidence of change in patient health outcomes, incl. QoL, mortality, reduction in life-threatening events, psychosocial outcomes),  
- cost-effectiveness (unit cost; cost implications to the system)                                                                                                                                                                                                                          |
| 7 Genotypic resistance testing of antiretrovirals in HIV (Application 1127)  | Diagnostic (for access to therapy)     | 2009  | - safety (adverse events)  
- effectiveness (change in clinical outcomes, change in clinical management, diagnostic accuracy incl. sensitivity and specificity),  
- cost-effectiveness (cost per case, cost to the system annually)                                                                                                                                                                                                                               |
| 8 K-RAS mutation testing for cetuximab (Reference 44)                      | Diagnostic (for access to therapy)     | 2010  | - safety (adverse events),  
- effectiveness (analytic performance; impact on patient management),  
- cost-effectiveness (unit cost; annual cost to the system)                                                                                                                                                                                                                               |
| 9 Epidermal Growth Factor Receptor Testing and Access to PBS listed Gefitinib (Reference 41) | Diagnostic (for access to therapy)     | 2010  | - safety (adverse events);  
- clinical effectiveness (test performance incl. specificity and sensitivity; impact on patient management)  
- cost-effectiveness (ICER; annual cost to the system)                                                                                                                                                                                                                                           |
| 10 Genetic testing for hereditary mutations in the VHL gene that cause von Hippel-Lindau syndrome (app 1153) | Diagnostic and relative carrier test | 2011  | - safety,  
- effectiveness (change in patient health outcomes; diagnostic accuracy, incl. sensitivity, specificity; changes in patient management),  
- cost-effectiveness (cost/case, budget impacts)                                                                                                                                                                                                                                              |
7.3.3. How MSAC understands ‘evidence’ in context of genetic test evaluation

(a) Number of genetic tests evaluated

As Table 12 shows, MSAC conducted its first genetic test evaluation in 2000 – that is, whilst the Human Genome Project was still ongoing. Between 2000 and 2012, 10 genetic test assessments have been conducted and 9 genetic tests were assessed. (One genetic test was assessed twice: first as application 1067, which was rejected for funding, and subsequently as application 1127 which was recommended for public funding). In short, the table above contains the entirety of HTAs for genetic tests that have thus far been completed by MSAC.

These numbers, however, do not convey the full picture, as at present (March 2012), over a dozen further genetic tests are in various stages of assessment by MSAC. These include:

1. ALK gene testing for patients with non-small cell lung cancer for eligibility for crizotinib (application 1250)
2. Diagnostic test for TTR-FAP for access to tafamidis meglumine (application 1222)
3. V600 status testing in patients with locally advanced or metastatic melanoma for access to appropriate therapies (application 1207),
4. HER-2 status testing in advanced or metastatic breast cancer (stage IIIA-IV) to access lapatinib (application 1175)
5. CCR5 tropic HIV-1 testing, for access to maraviroc (application 1174)
6. Testing for epidermal growth factor receptor (EGFR) status in patients with locally advanced (stage IIIB) or metastatic (stage IV) non-small cell lung cancer (NSCLC) for access to erlotinib (application 1173)
7. BRAF mutation testing in patients with metastatic melanoma for access to vemurafenib (application 1172)
8. HER2 status testing for patients with gastric cancer for access to trastuzumab (application 1163)
9. HER2 status testing in breast cancer patients (application 38).
10. First line testing for mutations of epidermal growth factor receptor (EGFR) in patients with metastatic non-small cell lung cancer (NSCLC) (application 1161),
11. Hepatitis C testing for access to Telaprevir (application 1160)
12. Genetic testing for mutation in RET gene (application 1152)
13. Six genetic tests for the diagnosis of heritable cardiac conduction disorders (application 1151)

Thus, although MSAC may justifiably be charged with having failed to prioritise genetic tests in the past, the fact that the number of currently ongoing assessments (13) exceeds the number of assessments completed in the 12 years prior (10 evaluations), suggests that MSAC’s emphasis on the evaluation of genetic tests is very much increasing.
(b) Types of genetic tests evaluated

The ten (10) completed genetic test evaluations heavily focused on assessing diagnostic tests for eligibility for and access to a particular therapy – that is, pharmacogenetic tests (7 assessments). However, other test types were also assessed: 1 prenatal screening test, 1 (strictly) diagnostic test, and 1 diagnostic and carrier test. The assessments that are currently pending also focus heavily on pharmacogenomics testing. Of the 13 assessments, 9 are pharmacogenetic, 3 are strictly diagnostic (that is, they fail to stipulate that the test is a condition of access to particular therapies) and 1 genetic test type under evaluation is unclear (genetic testing for mutation in RET gene).

While MSAC’s heavy emphasis on pharmacogenetic testing may be surprising, one must keep in mind here the centralised nature of the Australian health system. The decisions about what is – and is not – publicly funded (either via the listing on the MBS in case of devices, procedures, etc., or on the PBS in case of pharmaceuticals) are made at the federal level. The emphasis on evaluating genetic testing for access to therapies is underpinned by cost-effectiveness considerations; with the use of pharmacogenetic testing, the – often quite expensive – pharmaceutical therapies are restricted only to those patients who are most likely to benefit from them.

(c) Evaluata assessed (in ACCE terms)

Let us now turn our attention to how the inputs assessed by MSAC cohere with the four main evaluative categories of the ACCE framework. Only the 10 reports listed in Table 12, above, will be considered here, as full reports are not yet available for the genetic tests currently undergoing the evaluation process.

Consider, first, the ACCE’s analytic validity category, which is the test’s ability to measure the genotype of interest. None of MSAC’s reports assessed this dimension.

Clinical validity is the test’s ability to detect the disorder of interest. All 10 reports assessed this. Report 1 clearly assesses it, insofar as the predictive value of the test is one of the dimensions of clinical validity, according to the ACCE framework. Reports 2 through 10
assess the diagnostic accuracy of their respective tests (including sensitivity and specificity), which is also constitutive of clinical validity.

**Clinical utility** encompasses the test’s impact on patient management and costs associated with the testing itself. The ‘patient management’ dimension is assessed frequently in the reports, and is typically framed as ‘patient outcomes’ (sometimes more specifically as ‘improvement in health outcomes’). 8 reports clearly evaluate this (reports 2, 4-10). Arguably, report 3 may also be included here, insofar as it assesses the ‘effectiveness of treatment’ (which is an aspect of patient management associated with the genetic test). Only report 1 failed to assess the patient management aspect. Turning to the cost dimension of clinical utility, all 10 reports assessed this, although the specific type of assessments varied. In some instances, cost per genetic test was assessed, while in others, it was the cost per case detected, or the cost to the system overall. In sum, clinical utility dimension as a whole was evaluated by all 10 reports; all 10 assessed the cost dimension and 8 (possibly 9) assessed the patient management dimension.

The **ELSI** category, on the other hand, was assessed infrequently by MSAC. Only report 6 evaluated this, insofar as it considered issues such as quality of life and psychosocial outcomes. The remaining 9 reports failed to consider the elements of the ELSI category altogether, suggesting that the category generally does not engender much consideration by MSAC.93

| Table 13: MSAC: evaluta considered in HTAs of genetic tests (ACCE nomenclature) |
|---------------------------------|----------------------------------|
| ACCE categories                 | No. of reports                   |
| Analytic validity               | 0 of 10                          |
| Clinical validity               | 10 of 10                         |
| Clinical utility – patient management | 8 (or 9) of 10                  |
| Clinical utility – financial issues | 10 of 10                        |
| ELSI                            | 1 of 10                          |

In summary, in ACCE vernacular, MSAC’s conceptualisation of evidence encompasses the test’s clinical validity and clinical utility. As no reports assessed analytic validity, MSAC’s conceptualisation of evidence cannot be said to encompass that element. On the other hand, as

93 A slight shift seems to be taking place in this regard. Several of the currently ongoing evidence assessments (in particular, the ones carried out by the AHTA group at the University of Adelaide) do incorporate a discussion of the ethical issues around genetic testing. This may, however, raise a potential problem vis-à-vis these issues’ status as ‘evidence’ given that ethical issues are not indicated in MSAC’s terms of reference as an evaluative category for HTAs.
ELSI was assessed in 1 report, MSAC’s conceptualisation may be said to encompass the ELSI element albeit only rarely.

(d) Evaluata assessed (in standard HTA terms)

In order to establish MSAC’s conceptualisation of evidence in standard HTA terms, consider how well MSAC’s ACCE evaluata map onto MSAC’s stated evaluata (viz. safety, clinical effectiveness, cost-effectiveness, and budget impact, as of 2010).

Although not listed among ACCE’s categories, safety was consistently assessed in all 10 reports. (Although it is worth noting here that each report’s ‘safety’ consideration was typically very brief; it was limited only to noting that the safety issues around genetic testing are those associated with the procurement of the blood sample, i.e. they are normally minimal).

As established previously, ‘effectiveness’ encompasses clinical validity and clinical utility. Under the clinical validity rubric, it captures the diagnostic performance of the test (typically comprising specificity and sensitivity). Under the clinical utility, on the other hand, it encompasses the impact of the test on the patient management. All 10 of MSAC’s reports assessed clinical validity; 8 (or 9) of 10 reports assessed the patient management issues.

Cost-effectiveness was also considered in all 10 reports, taking on the form of the assessment of: cost per test, cost per case detected, cost per QALY, and incremental cost-effectiveness ratio.

Finally, recall that ‘budgetary impact’ category was introduced into MSAC’s Terms of Reference in 2010. Therefore, one would expect to see this dimension assessed in 3 of the most recent reports listed in Table 12. It is therefore somewhat surprising that the category was considered in 9 of the 10 reports (that is, all but the first report, from 2000). The category was framed as: annual cost to the system, cost implications to the system, or budgetary impact.
Table 14: MSAC: evaluata considered in HTAs of genetic tests (standard HTA nomenclature)

<table>
<thead>
<tr>
<th>HTA categories</th>
<th>No. of reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safety</td>
<td>10 of 10</td>
</tr>
<tr>
<td>Effectiveness (qua clinical validity and clinical utility/patient management)</td>
<td>10 of 10</td>
</tr>
<tr>
<td>Cost-effectiveness</td>
<td>10 of 10</td>
</tr>
<tr>
<td>Impact on budget</td>
<td>9 of 3\textsuperscript{94}</td>
</tr>
</tbody>
</table>

As summarised in Table 14, then, in standard HTA terms MSAC’s conceptualisation of evidence is as of data on: safety, effectiveness, cost-effectiveness and budgetary impact of genetic tests.

7.3.4. In summary

To summarise, then, in ACCE terminology, MSAC construes evidence as data on: clinical validity and the financial aspect of clinical utility; most of the time the patient management aspect of clinical utility is also included. The ELSI dimension is included very rarely, and the analytic validity category not at all. In HTA vernacular, on the other hand, MSAC seems to understand evidence to be data on: safety, clinical effectiveness, cost-effectiveness and budget impact.

7.4. UK and genetic testing

7.4.1. Current status of genetic testing in the UK

The data on the genetic testing activity undertaken by the laboratories in the United Kingdom is collected by the Clinical Molecular Genetics Society (CMGS). According to CMGS, approximately 210,000 samples were received for DNA-based testing by laboratories in the UK during 2009/2010 (PHG Foundation, 2011). As UK population was 61.1 million in 2010,\textsuperscript{95} this translates into approximately 1 test per 300 persons.

As noted previously, NHS’s goals – set out at its inception in 1948 – were: to meet everyone’s needs, to be based on clinical need rather than ability to pay, and to be free of charge at the point of delivery (NHS 2009). Those principles remain at the heart of the NHS, although a

\textsuperscript{94} Given that ‘impact of budget’ was a category introduced into MSAC’s terms of reference only as of 2010, only 3 genetic test HTAs (those from 2010 onwards) should have assessed this. Yet, 9 of 10 (that is, nearly all HTAs) assessed this category.

\textsuperscript{95} As of March 2012, it is 62.2 million.
review in 2000 resulted in an addition of several further objectives. One of these newly added objectives acknowledged the importance of genetic technologies, stating that:

> Developments in science, such as genetics, offer important possibilities for disease prevention and treatment in the future. The NHS will take advantage of the opportunities offered by scientific developments and will ensure that new technologies are used in the interests of society as a whole and available to all on the basis of need. (NHS 2009)

This objective drove the creation, in 2002, of the UK Genetic Test Network (hereafter UK GTN), a sub-group of the National Genetics Commissioning Advisory Group within the Department of Health. Among UK GTN’s key functions are: the evaluation of new genetic tests for provision by the NHS, the promotion of equity of access to genetic testing services, the approval and audit of genetic testing laboratories, and maintenance of the NHS Directory of Molecular Genetic Testing (UK GTN 2008; Kroese et al 2007). NIHR – that is, UK’s ‘regular’ national level HTA body – is not involved in the health technological assessments of genetic tests.96

For an assessment of a genetic test by UK GTN to take place, a laboratory has to complete and submit a Gene Dossier to the UK GTN.97 The Dossier is evaluated by the Gene Dossier Working Group, which puts forward its recommendation to the Steering Group. The Steering Group, in turn, may either endorse or reject the Working Group’s recommendation. Steering Group’s own recommendation is subsequently submitted to Genetics Commissioning Advisory Group. Genetics Commissioning Advisory Group ratifies UK GTN recommendation and makes arrangements for funding; the accepted gene dossiers are subsequently placed on the NHS Directory of Molecular Genetic Testing (UK GTN 2003; UK GTN 2009)

The tool used by the UK GTN for the evaluation of genetic tests prior to their adoption into the NHS framework is known as the ‘Gene Dossier.’ It is based on the aforementioned ACCE model, and focuses on the assessment of analytic validity, clinical validity, clinical utility and ELSI issues (Kroese et al, 2007; Sanderson et al 2005).

96 It is not entirely clear why this decision was made. This is not discussed in the literature, and UK GTN itself has not been forthcoming with this information.
97 More on the Dossier, below.
7.4.2. Method

Although the assessment of genetic tests is carried out by a separate agency in the UK, some of the conceptual problems identified previously around NIHR recur. More specifically, just like NIHR, UK GTN is not clear on its understanding of the concept of ‘evidence.’ The Gene Dossier itself uses the term ‘evidence’ precisely once, asking if evidence exists with regard to the test’s affecting the prognosis and management of a disease whilst failing to explain the term (evidence). The agency also fails to state (either implicitly or explicitly) that its evaluative process is evidence-based in the First Report (published in April 2008) (UK GTN 2008). Only in the Second Report (published November 2010) is the term ‘evidence’ used to (implicitly) suggest that the tests included on the Directory of Genetic testing are based on the evidence of the tests’ usefulness and validity (UK GTN 2010).

Nevertheless, it has been made clear elsewhere that UK GTN does regard the evaluata contained in the Gene Dossier as ‘evidence.’ This view has been made clear by Kroese, who points out that “the development of the gene dossier has provided the UK GTN with a framework to evaluate genetic tests in an evidence-based manner” (Kroese, 2007, 921; emphasis added). Insofar as Kroese is UK GTN’s public health advisor, we may take that claim to be authoritative. Moreover, in a 2010 submission to the National Institutes of Health, UK GTN notes that “the evidence base for making recommendations [to list tests on the Directory] is contained in the Gene Dossier”98 (emphasis added).

In short, although its concept of evidence is not formally defined by UK GTN, the clues to its understanding lie in the contents of the Gene Dossier. Thus, for the purposes here, 10 Gene Dossiers were, first, located by searching UK GTN’s website in January and February 2012, and second, their contents were analysed. Effort was made to identify genetic tests (on the UK GTN) that were also assessed by either MSAC or CADTH. However, in the majority of cases,

---

98 UK GTN (January 7, 2010). NIH, USA: Request for Information on the National Institutes for Health Plan to develop the genetic testing registry. Submission by UK GTN, Jacqui Westwood.
the same tests were not assessed, and – where the same tests were assessed – Gene Dossiers were not available.99

As no evaluations in common (between UK GTN and either CADTH or MSAC) were identified, a sample of randomly chosen genetic tests whose Gene Dossiers were available were selected for inclusion here. The included Gene Dossiers are listed in Table 15, below. Dates of evaluations are not provided in the table, as they are not indicated in the Gene Dossiers themselves. However, insofar as the Gene Dossiers for tests listed on the Directory prior to 2007 are not posted online, it can be assumed that all of these evaluations are no older than 2007.

99 Gene Dossiers are available online only for evaluations carried out since 2007. Pre-2007 evaluations were carried out using the Gene Dossier form, but the Gene Dossiers themselves are not available on the UK GTN website.
<table>
<thead>
<tr>
<th>Report Title</th>
<th>Test type</th>
<th>Focus of assessment</th>
<th>Date</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Colorectal cancer (HNPCC)</td>
<td>Diagnostic; Presymp.; Carrier</td>
<td>Epidemiology - prev. of condition in general population - gene frequency - penetrance - target population - prev. of condition in target population</td>
<td>Not listed</td>
<td>Test characteristic - analytical sensitivity and specificity - clinical sensitivity &amp; specif. (in target pop.) - clinical validity (positive and negative predictive value) in target population</td>
</tr>
<tr>
<td>2 Hemachromatosis, type 4</td>
<td>Diagnostic, Presymp.</td>
<td>As above, in HNPCC</td>
<td>Not listed</td>
<td>Clinical utility - clinical utility of test in target population - how will test add to patient mg/t alter clinical outcome - test’s impact on the NHS (e.g. remove need for alternate management) - availability of alt. diagnostic means? - ELSI issues around this particular test</td>
</tr>
<tr>
<td>3 Amyotrophic lateral sclerosis 6 with or without frontotemporal dementia</td>
<td>Diagnostic; Presymp.; Carrier; Prenatal</td>
<td>As above, in HNPCC; Under ‘clinical utility’ adds: - what are the consequences of not testing - utility of the test in the NHS</td>
<td>Not listed</td>
<td></td>
</tr>
<tr>
<td>4 Angioedema, hereditary, type i</td>
<td>Diagnostic; Progn&amp;Mgt Presymp.; Carrier</td>
<td>As above, in HNPCC; Under ‘clinical utility’ adds: - what are the consequences of not testing - utility of the test in the NHS</td>
<td>Not listed</td>
<td></td>
</tr>
<tr>
<td>5 Ellis-van creveld syndrome</td>
<td>Diagnostic; Progn&amp;Mgt Presymp.; Carrier; Prenatal</td>
<td>As above, in HNPCC; Under ‘clinical utility’ adds: - what are the consequences of not testing - utility of the test in the NHS</td>
<td>Not listed</td>
<td></td>
</tr>
<tr>
<td>6 Monosaccharide Malabsorption</td>
<td>Diagnostic; Progn&amp;Mgt Presymp.; Carrier; Prenatal</td>
<td>As above, in HNPCC; Under ‘clinical utility’ adds: - what are the consequences of not testing - utility of the test in the NHS</td>
<td>Not listed</td>
<td></td>
</tr>
<tr>
<td>7 Pancreatistis (hereditary)</td>
<td>Diagnostic; Progn&amp;Mgt Presymp.; Carrier</td>
<td>As above, in HNPCC; Under ‘clinical utility’ adds: - what are the consequences of not testing - utility of the test in the NHS</td>
<td>Not listed</td>
<td></td>
</tr>
<tr>
<td>8 Polycystic kidney disease</td>
<td>Diagnostic; Progn&amp;Mgt Presymp.; Carrier; Prenatal</td>
<td>As above, in HNPCC</td>
<td>Not listed</td>
<td></td>
</tr>
<tr>
<td>9 TAR syndrome</td>
<td>Diagnostic; Progn&amp;Mgt Carrier; Prenatal</td>
<td>As above, in HNPCC; Under ‘clinical utility’ adds: - what are the consequences of not testing - utility of the test in the NHS</td>
<td>Not listed</td>
<td></td>
</tr>
<tr>
<td>10 Macular degeneration</td>
<td>Diagnostic; Progn&amp;Mgt Presymp.; Carrier; Prenatal</td>
<td>As above, in HNPCC; Under ‘clinical utility’ adds: - what are the consequences of not testing - utility of the test in the NHS</td>
<td>Not listed</td>
<td></td>
</tr>
</tbody>
</table>
7.3.3. How UK GTN understands ‘evidence’ in context of genetic test evaluation

(a) Number of genetic tests evaluated

Each year, the UK GTN publishes a directory of all of the genetic tests that have been favourably evaluated by the UK GTN and ratified by the Genetics Commissioning Advisory Group. According to the most recent numbers available, from its establishment in 2002 to 2010, UK GTN has listed on the NHS Directory of Genetic Testing 688 tests for 503 diseases (UK GTN 2010). This translates into approximately 86 test evaluations each year (and would suggest that approximately 860 tests will be on the directory by end-2012).

(b) Types of genetic tests evaluated

The Gene Dossier form differentiates the following test types: diagnostic, prognosis and management,100 presymptomatic, carrier testing, and prenatal testing. Each Gene Dossier submission must indicate which test type (or types) the submission concerns.

It is noteworthy that each of the Gene Dossiers listed in Table 15 indicated multiple test types. The number of test types indicated on the form ranged from 2-5 types, with the average being approximately 4. Diagnostic testing dominated; this test type was indicated by all 10 Dossiers, with carrier screening being indicated by 8, and the remaining categories (prognosis and management, presymptomatic, and prenatal) all being indicated by 7 Gene Dossiers.

(c) Evaluata assessed (in ACCE terms)

The information in Table 15, above, shows that the evaluata assessed by the UK GTN in its evaluation of genetic tests is very consistent. This is, of course, due to the fact that a standardised form – the Gene Dossier – is utilised.101

The domains assessed by the Gene Dossier are: epidemiology, test characteristics and clinical utility. Under the epidemiology heading, issues pertaining both to the test and the disease are

100 Which focuses on whether a particular mutation will affect prognosis and management of the disease, i.e. seems to map onto what other jurisdictions regard as pharmacogenetic tests.
101 The only – small – deviation from this consistency that must be flagged, here, is that some Gene Dossiers assess two additional questions under the clinical utility category. This is due to the slight changes in the Gene Dossier form over the years.
addressed. These include: the test’s target population; condition’s prevalence (both in the general population and in the target population); and its penetrance.

Under the ‘test characteristic’ category, the test’s performance is assessed. More specifically, its analytical sensitivity and specificity are considered, as is the clinical sensitivity and specificity, and the positive and negative predictive values in the target population.

Finally, the clinical utility category is also considered as part of the Gene Dossier. Here, data is evaluated on: the clinical utility of the test in the target population, how the test affects patient management and alters the patient’s clinical outcomes, how the test impacts the NHS (e.g. whether it removes the need for an alternative means of patient management), whether there are alternative diagnostic means available, and finally, it is asked whether any ELSI issues arise with respect to the test. As noted previously, some version of the Gene Dossier also included the following two additional considerations: what are the consequences of not genetic testing (if any), and what is the utility of the test in the NHS specifically.

Insofar as the Gene Dossier is an adopted and adapted version of the ACCE framework one would expect – and indeed, does see – a great degree of coherence between Gene Dossier’s evaluata and the ACCE categories, although some adjustments are also evident.

ACCE’s analytic validity and clinical validity categories are captured under UK GTN’s ‘test characteristics’ section. All 10 of the evaluations provide data on these characteristics.

ACCE’s ‘clinical utility’ category and UK GTN’s ‘clinical utility’ section cohere reasonably well, with both encompassing the impact of the test on patient care (with UK GTN also including here the test’s impact on the NHS generally). The financial costs around genetic tests are not considered in the Gene Dossier, whilst the patient management issue is included. Thus, all 10 Gene Dossiers provided data pertaining to the patient management dimension (but all also failed to provide data on the financial aspect).

UK GTN categorises ELSI issues under ‘clinical utility’ (rather than splitting them into a separate category, as ACCE does). The ELSI section was generally only very cursorily
completed in the Gene Dossiers analysed. In response to the question whether any ELSI issues arise with respect to the test, 6 of the 10 Gene Dossiers responded ‘none’ or ‘unknown.’ The remaining four Gene Dossiers addressed the question very briefly. Answers to this question included the following: the test’s ethical issues centre around predictive testing for disease with no current treatment (ALSI test); test will contribute to good clinical management (Ellis-van-Creveld; TAR syndrome); and identification of potential consent issues (polycystic kidney disease).

The Gene Dossier seems prima facie to also contain an evaluative category that extends beyond the ACCE framework – namely, epidemiology. However, insofar as that category captures data on: test’s target population; condition’s prevalence (both in the general population and in the target population); and its penetrance, it may be considered to be an assessment of ACCE’s clinical validity (which encompasses prevalence of the disorder and its penetrance). All 10 Gene Dossiers listed above provided data on this. In short, this suggests that UK GTN’s conceptualisation of ‘evidence’ in context of genetic tests understands evidence to be data on: analytic validity, clinical validity, clinical utility (patient management dimension only), and ELSI (superficially considered).

Table 16: UK GTN: evaluata considered in HTAs of genetic tests (ACCE nomenclature)

<table>
<thead>
<tr>
<th>ACCE categories</th>
<th>No. of reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic validity</td>
<td>10 of 10</td>
</tr>
<tr>
<td>Clinical validity</td>
<td>10 of 10</td>
</tr>
<tr>
<td>Clinical utility – patient management</td>
<td>10 of 10</td>
</tr>
<tr>
<td>Clinical utility – financial issues</td>
<td>0 of 10</td>
</tr>
<tr>
<td>ELSI</td>
<td>10 of 10 (v. superficially)</td>
</tr>
</tbody>
</table>

(d) Evaluata assessed (in standard HTA terms)

Insofar as UK GTN does not evaluate any technologies other than genetic tests, it does not conduct evaluations using standard HTA categories. (Discussion of how NIHR’s conceptualisation of evidence – in standard HTA terms – would handle a genetic test assessment is provided below).

7.3.4. In summary

In the United Kingdom, the evaluation of genetic testing is carried out by an agency other than the ‘regular’ national level HTA agency. This introduces a potential complication into the
potential comparisons between UK’s approach to the evaluation of genetic testing and Canada’s and Australia’s approach thereto. Nevertheless, it is quite clear that, in ACCE terms, UK GTN construes ‘evidence’ as data on analytic validity, clinical validity, clinical utility (patient management dimension only; financial aspect is not considered) and ELSI (superficially).

7.5. Evidentiary issues around genetic tests: Canada, Australia and UK

7.5.1. Comparisons of the 3 jurisdictions’ practices

(a) Prevalence of genetic testing

Of the three jurisdictions, UK has the largest population (at over 62 million), followed by Canada (over 34 million) and Australia (over 22 million). Yet, the rates of genetic testing per capita tell a different story. More Australians have undergone genetic testing (approx. 1 in 130 persons) than either Britons or Canadians (1 per 300 and 1 per 350-700, respectively). The relatively high (in comparison) level of genetic testing by Australians may, as noted, be linked to the fact that in Australia, access to particular (publicly funded) pharmaceutical therapies is contingent on having a particular genetic mutation. Nevertheless, the numbers here are somewhat surprising, as UK has the largest number of genetic tests evaluated and available under its publicly funded scheme, yet its testing rate is more comparable to Canada’s (whose national agency’s track record in evaluation of genetic tests is practically non-existent).

(b) Evaluative process

In two of the three jurisdictions under examination here, that is Canada and Australia, the same body carries out the HTA assessments for both genetic testing and other health technologies (CADTH and MSAC, respectively). Likewise, no unique evaluation process has been put in place for the evaluation of genetic test by either agency. United Kingdom is the outlier here, in that it has created both a separate agency and a separate process for evaluation of genetic tests: UK GTN rather than NIHR carries out the evaluation of genetic tests, and the process used for the evaluation relies on the Gene Dossier – an adopted and adapted version of the ACCE framework.
(c) Number of genetic tests assessed

Substantial discrepancies exist between the three jurisdictions vis-à-vis the number of genetic tests evaluations that have been completed. Since 1999, Canada has completed only 3 full genetic test assessments. In comparison, over approximately the same period (that is, between 2000 and 2012), MSAC has completed 10 genetic test evaluations (of 9 genetic tests, as one genetic test was assessed twice), and a further 13 tests are currently in the process of undergoing assessments.

One might expect that part of the reason why CADTH and MSAC have completed a relatively low number of genetic test assessments (compared to the UK) is that these agencies’ resources are allocated to the assessments of a number of health technologies – genetic tests included. But even taking this into consideration, the discrepancy between CADTH and MSAC’s evaluative output vis-à-vis genetic tests is substantial: over three times as many evaluations have been completed by MSAC. Thus, it seems that at least part of the explanation must centre around prioritising – Australia can (and does) prioritise genetic test assessments because results of genetic tests are linked to eligibility for access to publicly-funded pharmaceuticals (and other therapies). The disconnect between CADTH and funding decisions (which are made at the provincial level), on the other hand, might explain its relatively low prioritisation of the evaluations of genetic tests.

The NHS Directory of Molecular Testing currently lists over 680 genetic tests. This is a result of UK’s having made genetic testing an NHS priority (as reflected both in the updated NHS objectives and a creation of an agency dedicated to genetic testing – the UK GTN). The extremely high (in comparison to both CADTH and MSAC) number of completed genetic test evaluations can be concluded to be the result of those two factors.

(d) Types of genetic tests evaluated

Recall, chapter 6 differentiates the following types of genetic tests: Diagnostic, Predictive or presymptomatic, Carrier tests, Screening tests, Prenatal tests, and Pharmacogenetic tests.
Of those, CADTH assessed predictive tests, screening tests and diagnostic tests. No carrier, prenatal or pharmacogenetic tests were assessed. Unlike CADTH, MSAC’s evaluative endeavours focused on pharmacogenetic tests (8 of 10 assessments), although prenatal, diagnostic and carrier tests were also assessed. A random sample of 10 Gene Dossiers evaluated by UK GTN showed that each Gene Dossier indicates multiple test types (typically 4 types), with diagnostic tests predominating, although carrier screening as well as prognosis/management (i.e. pharmacogenetic), presymptomatic and prenatal testing have also been indicated in 7 or more Gene Dossiers.

It is difficult to discuss trends with regard to the tests evaluated by CADTH as the sample is so small. It is interesting to note that Australia prioritises pharmacogenetic tests, whilst simultaneously de-emphasising strictly diagnostic tests. In the UK, on the other hand, strictly diagnostic tests predominate, although pharmacogenetic tests were also assessed in 7 of 10 reports considered here. (It needs to be noted here, however, that a pharmacogenetic test is, at its base, a diagnostic test – it is simply a diagnostic test on whose results the access to a particular therapy is contingent, in Australia’s case.)

**(e) Evaluata assessed (in ACCE terms)**

The greatest variability with respect to the ACCE categories assessed by all 3 jurisdictions arises in context of analytic validity. Analytic validity was assessed by 2 of the 3 CADTH reports and by all 10 of UK GTN’s Gene Dossiers (under the ‘test characteristics’ rubric). MSAC did not assess this category at all; none of the reports considered it.

Clinical validity, on the other hand, was uniformly evaluated. All of CADTH’s reports, all of MSAC reports and all of the Gene Dossiers (under the ‘test characteristics’ category) evaluated this.

Clinical utility was, overall, also assessed by all of the reports from all of the jurisdictions. The picture is more complex, however, if the category is further differentiated into its clinical management and costs sub-components. The patient dimension was assessed by: 3 CADTH
reports, 8 (possibly 9) MSAC reports and all 10 Gene Dossiers. The financial component was assessed by: 2 CADTH reports, 10 MSAC reports and no (0) Gene Dossiers.

Finally, great variability was evident with respect to the ELSI category. ELSI issues were assessed by all of CADTH reports, 1 of 10 MSAC reports and, nominally, all 10 Gene Dossiers (although in practice, this section of the Gene Dossiers was completed very poorly). This is summarised in Table 17, below.

(f) Evaluata assessed (in standard HTA terms)

How do the agencies’ evaluata (in context of genetic tests) map onto the standardly assessed HTA categories, however?

Safety is one of the most oft-cited HTA evaluata. This was not considered by any of the CADTH reports, nor any of the Gene Dossiers. MSAC did, consistently, evaluate this, however – all 10 of its reports assessed safety of the genetic test under consideration (even if briefly). UK GTN, on the other hand, did not evaluate this dimension at all.

Clinical effectiveness was evaluated by all 3 CADTH reports, 10 MSAC reports and all 10 Gene Dossiers (insofar as cost-effectiveness encompasses clinical validity and the patient management component of clinical utility).

Cost-effectiveness was assessed by 2 of the 3 CADTH reports, all 10 of MSAC’s reports and none of the Gene Dossiers (insofar as cost-effectiveness maps onto the financial issues’ aspect of clinical utility).

If CADTH’s ‘broader impact of health technologies’ category is interpreted as an ELSI category, then all 3 of its reports considered it. ELSI issues were also considered by 1 of MSAC’s 10 reports, and (briefly) by all 10 of the Gene Dossiers.

MSAC’s reports also assessed (one might say, over-assessed) its ‘budget impact’ category. Budgetary impact was introduced into MSAC’s terms of reference in 2010, and thus, should
only have been considered in 3 of the most recent reports. However, this category was, in fact, evaluated in 9 of 10 reports.

This is summarised in Table 17, below.

7.5.2. Conceptualising ‘evidence’ in context of genetic test assessment.

(a) How do the 3 agencies conceptualise ‘evidence’?

Recall the distinction drawn, in chapter 5, between the conceptualisation of ‘evidence’ in theory and in practice. The understanding of evidence in theory centred on what the agencies’ statements conveyed about their understanding of the concept of ‘evidence.’ Understanding of ‘evidence’ in practice, conversely, was meant to convey what the agencies’ evaluative practices suggested about their understanding of that term. As the agencies’ theoretical conceptualisation was discussed in more detail in chapter 5 (section 5.4.2.1), it need not be reiterated here. It is therefore the understanding of ‘evidence’ in practice that will be the focus, here. The agencies’ evaluative practices vis-à-vis genetic tests, which ground the agencies’ understanding of evidence in practice is summarised in Table 17, and described in more detail, below.

Table 17: The agencies’ understanding of ‘evidence’ in HTAs of genetic tests.

<table>
<thead>
<tr>
<th>Evaluata: ACCE framework terminology</th>
<th>CADTH</th>
<th>MSAC</th>
<th>UK GTN</th>
<th>NIHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic validity</td>
<td>2 of 3 reports</td>
<td>0 of 10 reports</td>
<td>10 of 10 reports</td>
<td>--</td>
</tr>
<tr>
<td>Clinical validity</td>
<td>3 of 3 reports</td>
<td>10 of 10 reports</td>
<td>10 of 10 reports</td>
<td>--</td>
</tr>
<tr>
<td>Clin. util: px mgmt</td>
<td>3 of 3 reports</td>
<td>8-9 of 10 reports</td>
<td>10 of 10 reports</td>
<td>--</td>
</tr>
<tr>
<td>Clin. util: $ issues</td>
<td>2 of 3 reports</td>
<td>10 of 10 reports</td>
<td>0 of 10 reports</td>
<td>--</td>
</tr>
<tr>
<td>ELSI</td>
<td>3 of 3 reports</td>
<td>1 of 10 reports</td>
<td>10 of 10 reports (brief)</td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluata: general HTA terminology</th>
<th>CADTH</th>
<th>MSAC</th>
<th>UK GTN</th>
<th>NIHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>--</td>
<td>Safety: 10 of 10</td>
<td>--</td>
<td>Safety: 0 of 10</td>
<td></td>
</tr>
<tr>
<td>Clin eff: 3 of 3</td>
<td>Clin. eff: 10 of 10</td>
<td>--</td>
<td>Clin. eff: 10 of 10</td>
<td></td>
</tr>
<tr>
<td>Cost eff: 2 of 3</td>
<td>Cost-eff: 10 of 10</td>
<td>--</td>
<td>Cost-eff 10 of 10</td>
<td></td>
</tr>
<tr>
<td>Broader imp: 3 of 3</td>
<td>Budget imp: 9 of 3</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

102 Although UK GTN is a separate agency from NIHR, it shares NIHR’s problem in failing to specify its theoretical understanding of evidence. Nevertheless, its statements suggest that its understanding of evidence is as of data contained in the Gene Dossier. Analysis of 10 Gene Dossiers suggests that UK GTN’s theoretical understanding of evidence is therefore: of data on epidemiology, test characteristics and clinical utility of genetic tests. That is, UK GTN appears to adopt the same theoretical conceptualisation of evidence as do CADTH, MSAC and NIHR – as of data from stipulated domains.

103 The comparison will be carried out both in terms of the standardly evaluated HTA categories (safety, effectiveness, etc.) and in terms of the ACCE categories, in order to permit comparisons between all three jurisdictions. The comparison in general HTA terms will be made between CADTH, MSAC and NIHR; the comparison in terms of ACCE nomenclature will be made between CADTH, MSAC and UK GTN.

104 As established in chapter 5.
Consider, first, the agencies’ practical understanding of ‘evidence’ in terms of the ACCE framework. In those terms, CADTH’s practical understanding of ‘evidence’ is as of data on: analytic validity (some of the time), clinical validity (patient management dimension all of the time; financial dimension some of the time), clinical utility (all of the time) and ELSI (all of the time).

MSAC’s practical understanding of ‘evidence’ here is as of data on: clinical validity and clinical utility (patient management dimension most of the time; financial dimension all of the time); very rarely it also encompasses ELSI issues and never the data on analytic validity.

UK GTN’s practical construal of evidence in ACCE terms encompasses data on: analytic validity, clinical validity, clinical utility (patient management aspect only; it omits the financial dimension) and ELSI issues (included under the clinical utility category; typically covered very superficially).

In terms of the general HTA terminology, CADTH’s practical understanding of ‘evidence’ includes clinical effectiveness and the broader impact of the genetic tests all of the time (3 of 3 reports), and the tests’ cost-effectiveness some of the time (2 of 3 reports). MSAC’s practical understanding encompasses, here: safety, effectiveness and cost-effectiveness (10 of 10 reports), and budget impact (in 9 reports, although given the timing of the introduction of the expanded Terms of Reference, it should have only been assessed in 3 reports). Finally, recall, that in NIHR’s case, its practical conceptualisation of evidence encompassed data on effectiveness and cost-effectiveness of a health technology under assessment.

(b) Challenges for the ‘stipulated domains’ approach

Recall that, in chapter 5, it was argued that the agencies’ understanding of evidence – which understood evidence as data from particular domains (i.e. the ‘stipulated domains’ approach) – was vulnerable to the ‘cart before the horse’ objection. Briefly, this was the objection that stipulating particular domains as sources of relevant data before a particular HTA question, issue, or evaluation, is even considered, amounts to putting the cart before the horse. Insofar each HTA will have different evidentiary needs, the ‘cart before the horse’ approach is problematic.
One might point out, here, however, that the stipulated domains approach was problematic in chapter 5 insofar as chapter 5 dealt with ‘health technologies’ more generally. However, this problem should dissolve here, insofar as we are focusing on a single health technology type (genetic testing), and assessments of a single technology type would have similar (if not identical) evidentiary needs.

Yet, the ‘cart before the horse’ problem remains a problem even here. This is because even within a single health technology type, different evidentiary needs are apparent. CADTH’s assessment of two predictive genetic tests (in 2006 and 1999, respectively), both evaluated: clinical validity, clinical utility (patient management issues) and ELSI issues. Differences arose, however, with regard to analytic validity (2006 report evaluated this, 1999 did not) and the financial dimension of clinical utility (2006 did not consider it; 1999 did). In MSAC’s case, reports 2 and 6 assess diagnostic tests. Yet, the differences here centre around ELSI issues – report 6 evaluates them (even if briefly) whilst report 2 does not. The examples cited here are intra-agency differences; if inter-agency evaluative differences were to be considered (e.g. UK GTN versus CADTH), differences would multiply yet more.

In short, even within a single category of health technologies (viz. genetic tests) – and even within a single sub-category (e.g. predictive tests) of that single category – differences remain with regard to what is deemed as salient evaluata. That is, different evaluations will have different evidentiary needs – and the ‘cart before the horse’ approach remains problematic.

7.5.3. Advantages of adopting the proposed conceptualisation of evidence: the FAP-TTR genetic test case study

That the stipulated domains approach to conceptualisation of evidence is generally a problem, however, does not yet establish that the conceptualisation of ‘evidence’ proposed in chapter 5 handles potential problems any better. In order to demonstrate that this is the case, let us consider the following, hypothetical genetic test assessment.105

105 Strictly speaking, the assessment is only hypothetical with respect to Canada’s and UK’s potential handling thereof; this genetic test is currently in process of assessment by MSAC. Some of this section’s contents is adopted from the Decision Analytic Protocol (DAP) document produced as part of that assessment, and available on the MSAC website at http://www.msac.gov.au/internet/msac/publishing.nsf/Content/1222
(a) Background

Transthyretin familial amyloid polyneuropathy (hereafter, TTR-FAP) is an extremely rare disease. The European Medicines Agency estimates the overall population of TTR-FAP patients worldwide at 5,000-10,000 (EMA 2011). Penetrance rate of the disease is less than 100%, although it varies by mutation, geographic region and/or ethnic group. In Sweden, the penetrance rate by age 50 is 11%; at the same age, penetrance is 60% among Portuguese patients and 18% among French patients (Plante-Bordeneuve and Norgren, 2011; Pagon et al 2012).

TTR-FAP is transmitted in an autosomal dominant manner (i.e. first degree relatives have a 50% chance of inheriting the same mutation). The disease is caused by mutations in the transthyretin (TTR) gene – over a hundred disease-causing mutations have been identified in the TTR gene so far, with Val30Met and Ile84Se being the most common mutations (NHSC 2010). The primary symptoms of TTR-FAP are the progressive loss of nerve functions (including sensory, autonomic and motor), and the disease is typically fatal within 10 years of onset. There is presently no cure; treatment includes liver transplant (as liver is the primary site of TTR-FAP synthesis), pacemaker insertion (for cardiac dysfunctions), treatment of pain with pharmaceuticals, surgical decompression for carpal tunnel syndrome, therapeutic measures for orthostatic hypotention and correction of hydration (Plante-Bordeneuve and Said, 2011; Bittencourt et al 2002).

A genetic test for TTR-FAP currently exists. Testing is typically performed using Polymerase Chain Reaction (PCR) to amplify the DNA sample (contained in the blood collected from the patient), and a full gene sequence analysis. Diagnostic testing is typically carried out on symptomatic patients and asymptomatic first degree relatives of patients with confirmed TTR-FAP.

(b) The agencies’ conceptualisations of evidence vis-à-vis TTR-FAP assessment?

To lay the groundwork for the claim that the agencies’ evaluation of the TTR-FAP genetic test will be incomplete under their ‘stipulated domains’ conceptualisations of evidence, two comparisons will be made here. The first comparison will utilise the HTA agencies’ general
evaluative categories (e.g. safety, effectiveness, etc.), so that a comparison can be made between CADTH, MSAC and NIHR. Second, a comparison will be drawn out utilising the ACCE framework’s evidentiary categories, so that CADTH, MSAC and UK GTN can be compared.

(i) Comparison in terms of general HTA categories

The following evaluative categories are typically included HTA reports: safety, effectiveness, cost-effectiveness, etc. The three agencies’ evaluative patterns (using those categories) are listed in Table 18, below.

<table>
<thead>
<tr>
<th>Evaluative HTA categories</th>
<th>CADTH</th>
<th>MSAC</th>
<th>NIHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safety</td>
<td>n/a</td>
<td>10 of 10</td>
<td>0 of 10</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>3 of 3</td>
<td>10 of 10</td>
<td>10 of 10</td>
</tr>
<tr>
<td>Cost-effectiveness</td>
<td>2 of 3</td>
<td>10 of 10</td>
<td>10 of 10</td>
</tr>
<tr>
<td>Broader impact</td>
<td>3 of 3</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Budget impact</td>
<td>n/a</td>
<td>9 of 3</td>
<td>n/a</td>
</tr>
</tbody>
</table>

In context of genetic testing, CADTH understood ‘evidence’ as data on effectiveness (all of the time), cost-effectiveness (some of the time) and the broader impact of health technology (construed as ELSI issues; all of the time). This suggests that in evaluating the TTR-FAP genetic test, CADTH would evaluate the test’s effectiveness and broader impact (qua ELSI) but it is not certain whether it would evaluate its cost-effectiveness.

Recall that, in practice, in context of genetic testing, MSAC understood ‘evidence’ as data on safety (10 reports), effectiveness (10 reports on clinical validity, 8-9 of 10 reports on patient management), cost-effectiveness (costs dimension of clinical utility; 10 of 10 reports) and the budget impact of the test (9 reports, although applicable only to 3). This suggests that an evaluation of TTR-FAP test would consider data on the following: safety (around issues with blood sample procurement), cost-effectiveness (the cost dimension of clinical utility), and the budget impact of the test. With respect to the effectiveness dimension, the clinical validity would be considered, and patient management issues would most likely (albeit not certainly) also be considered.

193
Although NIHR does not assess genetic tests, were it to assess one, its practical understanding of evidence (as established in chapter 5) suggests that data on effectiveness and cost-effectiveness would be assessed. NIHR’s evaluation would therefore likely focus on the test’s clinical validity, as well as the patient management dimension of clinical utility (as those are typically considered under clinical effectiveness). Cost-effectiveness evaluation would likely consider the cost issues around the test itself (i.e. the cost dimension of clinical utility). Safety would likely not be assessed at all, as NIHR was shown to never assess this domain in practice.

Thus, insofar as CADTH’s practical understanding of evidence includes cost-effectiveness only some of the time, this dimension could well be omitted in its evaluation. Neither safety nor budget impact constitute any part of its construal of evidence in context of genetic tests, and therefore neither of those categories would be assessed. MSAC, on the other hand, would consider the safety, effectiveness, cost-effectiveness and budget impact of the test. Nevertheless, it might well fail to consider patient management issues (under effectiveness) insofar as its practical construal of evidence includes this dimension some (but not all) of the time. Finally, NIHR’s evaluation would likely assess effectiveness and cost-effectiveness; safety and budget impact do not constitute a part of its practical conceptualisation of evidence, and thus would likely not be assessed here. Each agency’s conceptualisation of evidence would therefore fail to consider at least one dimension of an evaluation.

(ii) Comparison in terms of ACCE categories

The following are the main categories of evaluata according to ACCE: Analytic validity; Clinical validity; clinical utility (patient management issues and cost issues), and ELSI. The three agencies’ evaluative patterns (in ACCE terms) are listed in Table 19, below.

<table>
<thead>
<tr>
<th>Evaluata: ACCE categories</th>
<th>CADTH</th>
<th>MSAC</th>
<th>UK GTN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic validity</td>
<td>2 of 3 reports</td>
<td>0 of 10 reports</td>
<td>10 of 10 reports</td>
</tr>
<tr>
<td>Clinical validity</td>
<td>3 of 3 reports</td>
<td>10 of 10 reports</td>
<td>10 of 10 reports</td>
</tr>
<tr>
<td>Clin. util: Px mgt</td>
<td>3 of 3 reports</td>
<td>8-9 of 10 reports</td>
<td>10 of 10 reports</td>
</tr>
<tr>
<td>Clin. util: $ issues</td>
<td>2 of 3 reports</td>
<td>10 of 10 reports</td>
<td>0 of 10 reports</td>
</tr>
<tr>
<td>ELSI</td>
<td>3 of 3 reports</td>
<td>1 of 10 reports</td>
<td>10 of 10 reports (brief)</td>
</tr>
</tbody>
</table>
In light of the conceptualisations of evidence suggested by the contents of Table 19, above, CADTH’s evaluation of TTR-FAP would likely assess its clinical validity, clinical utility (patient management dimension) and ELSI. It is likely (although not certain), that CADTH would also evaluate this test’s analytic validity and clinical utility (cost dimension). MSAC, on the other hand, would evaluate TTR-FAP’s: clinical validity and the cost dimension of clinical utility. It would very likely (although not certainly) also evaluate the patient management dimension of clinical utility. It is unlikely that it would evaluate either the ELSI issues around the test, or its analytic validity. UK GTN’s practical conceptualisation of evidence would suggest that an evaluation of the TTR-FAP test would assess all four of the ACCE categories, except the financial aspects of the clinical utility criterion. ELSI would be evaluated, although in all likelihood, very superficially.

Just like in the case of their conceptualisations in terms of general HTA categories, the agencies’ practical conceptualisations of evidence (fleshed out in ACCE terms) are would all also omit some dimension of ACCE. CADTH (likely) and MSAC (certainly) would fail to assess analytic validity. MSAC would (possibly) omit the patient management dimension of clinical utility. CADTH would (also possibly) omit the cost dimension of clinical utility; UK GTN would certainly miss this. Finally, MSAC would almost certainly fail to evaluate ELSI issues; UK GTN would include ELSI but quite likely, its consideration thereof would be superficial.

(c) Evaluata that should be assessed in an adequate assessment of TTR-FAP

In short, it appears that the practical conceptualisations of evidence adopted by the agencies – whether fleshed out in standard HTA terms or in terms of the categories of the ACCE framework – would all miss something. Of course showing that the agencies’ conceptualisations would miss something does not yet establish that these conceptualisations would miss something crucial. Let us consider, then, what an adequate assessment of TTR should include – first, in standard HTA terms and subsequently, in ACCE terms.

(i) In terms of general HTA categories

A proper assessment of TTR-FAP, in general HTA terms, ought to assess the test’s: effectiveness, cost-effectiveness, budget impact, and ethical issues. Its effectiveness must be
assessed because whether a technology performs as it should is a consideration that cannot be
omitted. Cost-effectiveness is equally crucial, insofar as all three jurisdictions under evaluation
here have publicly funded health systems, which are subject to resource limitations. For the
same reason, the budget impact of the adoption of the test must be evaluated – cost-
effectiveness will not appropriately capture this dimension, as it focuses on the costs vis-à-vis
a particular patient, whilst the ‘budget impact’ category takes a system-level view. Finally,
ethical issues around this test should also be considered. Although it has been previously
argued that arguments for genetic exceptionalism fail, that section did not – nor was it intended
to – show that genetic tests are free of ethical considerations (nor are other health technologies,
for that matter). Ethical issues are important to assess here, because although treatment for
TTR-FAP exists, a cure, presently, does not; moreover, the disease’s autosomal dominant
heritability raises reproductive considerations, as well as the potential obligations to warn first-
degree relatives. Ethics should, therefore, be assessed here. On the other hand, safety need not
be assessed because this particular test requires only a small sample of blood. Thus, typically,
no safety issues will arise here (barring very unusual cases, like haemophiliac patients, for
example).

(ii) In terms of ACCE categories

In ACCE vernacular, a proper assessment of TTR-FAP would require a consideration of the
following. Clinical validity – insofar as it considers the test’s actual diagnostic performance –
must be assessed. Clinical utility category, recall, typically encompasses both the
considerations around patient management and financial issues around the test itself. The
former would require assessment here, as there is nothing to be gained from a diagnostic test
that fails to impact patient management in any way – it would be a waste of resources to do so.
That there is some gain to be had from the testing (in terms of the clinical management of the
patient) would need to be established, in other words. The cost issues dimension of clinical
utility would have to be construed as both a cost-effectiveness assessment and a system-level
budget assessment, for reasons outlined above. Finally, the ELSI category – in particular, the
ethics dimension – would require assessment – because, as noted before, this particular test is
for a condition which is terminal and for which only treatment (but not cure) presently exists.
Insofar as clinical validity is predicated on analytic validity, the latter does not require an
evaluation here.
(d) What the agencies’ conceptualisations would miss

(i) In terms of general HTA categories

As established above, an adequate assessment of TTR-FAP, in general HTA terms, ought to assess the test’s effectiveness, cost-effectiveness, budget impact, and ethical issues. Safety would not require an evaluation. All three agencies’ conceptualisations of evidence, however, would fail to capture at least some of these evaluata.

CADTH’s conceptualisation of evidence in HTA terms would capture the test’s effectiveness and ethical issues. On the other hand, it would fail to capture the budget impact of the test (as it does not assess this category); and it may or may not capture the test’s cost-effectiveness.

MSAC’s conceptualisation, on the other hand, would capture the test’s effectiveness, cost-effectiveness and budget impact. It would, nevertheless, fail to assess the ethical issues around the test. Moreover, its current conceptualisation of evidence would unnecessarily dictate assessing the test’s safety issues.

Finally, NIHR’s conceptualisation – were that agency to assess genetic tests – would not capture much of what was established as necessary for an adequate assessment here. NIHR would capture the data around the test’s effectiveness and cost-effectiveness. It would (rightly) fail to capture the data on safety, as that is unnecessary here. NIHR’s conceptualisation, however, would (wrongly) fail to capture the budget impact, and it would (wrongly) fail to consider the ethical issues around the test only superficially.

(ii) In terms of ACCE categories

As noted above, an adequate assessment of TTR-FAP in terms of ACCE’s categories, would require an assessment of the test’s: clinical validity, clinical utility (both patient management and cost issues; the latter encompassing both budget impact and cost-effectiveness dimensions). ELSI issues would also require an assessment; analytic validity, on the other hand, would not.
CADTH’s conceptualisation of evidence in ACCE terms would therefore face here the problem of possibly (albeit not certainly) capturing the clinical utility’s cost dimension only qua cost-effectiveness. It would fail to capture the budget impact issues, and it would also likely (but unnecessarily) capture the analytic validity.

MSAC’s conceptualisation would be problematic vis-à-vis patient management issues (its conceptualisation captures them most of the time, but not all of the time) and ELSI issues (its conceptualisation almost always fails to capture those).

Finally, UK GTN’s conceptualisation of evidence would be vulnerable to the following three problems. First, it would fail to capture the economic dimensions of clinical utility (both cost-effectiveness and budget impact). Second, its ELSI consideration would in all likelihood be inappropriately superficial. Third, it would unnecessarily evaluate analytic validity.

(e) How the proposed conceptualisation of evidence would fare vis-à-vis a TTR-FAP genetic test assessment?

How, on the other hand, would the proposed conceptualisation of evidence fare in adequately assessing the TTR-FAP genetic test? Consider this issue, first, in terms of the general HTA categories and second, in ACCE terms.

In terms of the general HTA categories, it was established, above, that an adequate evaluation of the TTR-FAP test ought to assess the test’s: effectiveness, cost-effectiveness, budget impact, and ethical issues. Recall that the proposed conceptualisation of evidence stipulated that ‘evidence’ is available data that is relevant to the issues being addressed, question asked, decision made, etc. Insofar as it is available data that is assessed with regard to an issue being addressed (the health technological assessment of the TTR-FAP test), those two conditions appear to be met. What of the relevance stipulation? Insofar as it was shown that those are the relevant evaluata here, it appears that this criterion, too, is met by the proposed conceptualisation.

In terms of the ACCE categories, an adequate evaluation of the TTR-FAP test ought to assess the test’s clinical validity, clinical utility (both patient management and cost issues; with the
latter encompassing both budget impact and cost-effectiveness dimensions). ELSI issues would also require an assessment, whilst analytic validity would not. How does the proposed conceptualisation of evidence fare here? Again, insofar as it is the available data that is assessed with regard to an issue being addressed (the health technological assessment of the TTR-FAP test), those two stipulations of the proposed conceptualisation are met. Additionally, insofar as these evaluata were established to be the relevant ones, here, it appears that this stipulation is met by the proposed conceptualisation, as well.

In sum, whether cashed out in general HTA terms or in terms of ACCE nomenclature, the proposed conceptualisation captures all – and only – the evaluata required for an adequate assessment of the TTR-FAP genetic test. An important objection needs to be considered and discharged here, however. This is that it appears that the proposed conceptualisation seems to rely, here, on the ‘stipulated domains’ approach, which was previously argued to be problematic. However, I do not think this is as great a challenge as it prima facie appears. This is because it is not being proposed here that the domains indicated above will be salient for every genetic test assessment (as the agencies’ stipulated domains conceptualisation would require). It is only proposed here that these domains are salient for the purposes of this assessment – i.e., that they meet the relevance criterion of the proposed conceptualisation in this particular evaluation.

7.6 Conclusion

As argued in chapter 4, the field of HTA currently lacks a clear conceptualisation of ‘evidence’. This is problematic insofar as HTA agencies claim to be evidence-based yet it is not clear what, precisely, this either means or entails. A context-based analysis, in chapter 5, suggested that the agencies’ conceptualisation of ‘evidence’ vis-à-vis health technologies considered generally, understands ‘evidence’ to be data from stipulated domains (which vary by jurisdiction). In this chapter, I have shown that the ‘stipulated domains’ conceptualisation of evidence also underlies the agencies’ evidentiary behaviour vis-à-vis a single health technology type, viz. genetic testing.

To demonstrate that this conceptualisation is problematic, I considered a hypothetical evaluation of the TTR-FAP genetic test. Each of the agencies’ conceptualisations of evidence
would be inadequate for the purposes of this assessment, since each would miss an important
dimension(s) and/or evaluate superfluous data. The conceptualisation of evidence proposed in
chapter 4, on the other hand, was shown to be able to capture all – and only – the evaluata
required for an adequate assessment of the TTR-FAP genetic test, thus clearly establishing
both its tenability and utility in practice.
CHAPTER 8: Conclusions

This dissertation began with an introduction to the field of Health Technology Assessment. Chapter 1 described both the nomenclature and the historical factors that gave rise to the field, as well as discussing the field’s general aims, methodologies and evaluative practices. It became quite apparent that two views are commonly held: that the role of HTA is to provide input into health policy and that the field regards itself as evidence-based. It also became apparent that a variety of differences arise – e.g., with respect to which technologies are assessed (i.e., how broadly the scope of ‘health technologies’ is construed), the methodologies adopted for that purpose, whether health technology assessments influence funding decisions, and so forth.

In order to flesh out the cross-jurisdictional similarities and differences in approaches to HTA in more detail, chapter 2 focused on three jurisdictions in particular: Canada, Australia and the UK. Because HTA activities do not exist in a vacuum, but rather, they are subject to the structural and policy constraints of the health system within which they operate, all three health systems were described and, subsequently, the general picture of HTA activities at both national and sub-national level was sketched for each jurisdiction.

The jurisdictions were compared in chapter 3. Due to the dearth of comparative, cross-jurisdictional work in the HTA literature, no heuristically useful tool for cross-jurisdictional analysis of HTA approaches currently exists. Consequently, a tool was devised here for this purpose. It consists of 8 categories: 2 external categories (focusing on the health system context within which the HTA system is embedded) and 6 internal categories (focusing on the minutiae of the HTA itself).

The 3 national level HTA programmes described in Chapter 2 – viz., Canada’s CADTH, Australia’s MSAC and UK’s NIHR – were subsequently analysed and compared using that tool. The analysis revealed that although both the HTA literature and HTA bodies consistently claim that HTA is an evidence-based endeavour, the agencies’ assessments utilise a variety of probative inputs. Thus, for example, CADTH claims to evaluate in its HTAs the effectiveness, cost-effectiveness and broader impact of health technologies. MSAC, in turn, claims to assess
the health technologies’ safety, effectiveness, cost-effectiveness and overall budgetary impact. NIHR, finally, claims to assess safety, effectiveness and cost-effectiveness of health technologies. This, however, raises the following question: if all three agencies conduct assessments that are evidence-based – as they all claim to – then how to account for the differences in evaluative inputs considered? Why is it, for example, that MSAC and NIHR appear to include safety in its understanding of ‘evidence’ whilst CADTH does not? Why does CADTH include the technologies’ broader impact whilst NIHR fails to do so?

Differences such as these raise a more fundamental question: how, exactly, is the concept of ‘evidence’ understood in the field of HTA? This issue was taken up in chapter 4. As a thorough scan of both academic and non-academic literature established, however, little to no conceptual work exists on this issue in HTA. It was argued that this is a gap in urgent need of filling and, towards that end, literature was analysed in two cognate fields: evidence-based medicine and evidence-based health policy. Those fields were chosen since they are both evidence-based, and both centre around issues in health – similarly to HTA. Unlike in HTA, however, conceptual work on the concept of evidence exists in those two fields.

The state of play was found quite similar in EBM and EBHP: that is, both offer definitions of evidence that can be characterised as either narrow or broad. In light of the problems that emerge around the narrow conceptualisation, it was argued in chapter 4 that a broad conceptualisation of ‘evidence’ would be more appropriate for HTA purposes. Specifically, it was argued that ‘evidence’ ought to be construed as: available data that is relevant to the issue being addressed, question asked, decision made, etc. This conceptualisation enjoys numerous advantages, including: recognition that each health technology assessment has different evidentiary needs; allowing evidence-based assessments of health technologies to be conducted at various stages in the health technologies’ development cycle; capturing assessments of ‘health technologies’ understood both narrowly and broadly; taking into consideration that the same data is not ‘evidence’ in all cases; and finally, the conception’s invulnerability to the problems around the ‘stipulated domains’ approach to conceptualising evidence.
The worth of a definition is measured in its tenability and usefulness. Chapter 5 therefore tested the definition’s tenability and usefulness against the actual evaluative practices in the field of HTA. Because no explicit definition of evidence was offered by any of the agencies of interest here, a contextual analysis established the three HTA agencies’ current theoretical and practical understanding of evidence. The theoretical understanding of evidence was established from the agencies’ statements about the probative inputs they consider. The agencies understood ‘evidence’ as data from stipulated domains (the ‘stipulated domains’ approach), as follows:

**CADTH:** evidence qua data on effectiveness, cost-effectiveness, the broader impact of health technology

**MSAC:** evidence qua data on safety, effectiveness, cost-effectiveness and budget impact (‘budget impact’ as of 2010 only)

**NIHR:** evidence qua data on safety, effectiveness, and cost-effectiveness of health technologies.

The ‘stipulated domains’ approach, however, was argued to be problematic, in that it effectively amounts to putting the cart before the horse. Different assessments (or questions, issues, or decisions) will require different evidentiary inputs. Thus, the ‘stipulated domains’ approach runs the risk of including either too much or not enough of what is ‘evidence’ for a particular assessment.

That this is a pressing problem became very clear when the agencies’ practical understanding of evidence was considered. Practical understanding of evidence was established from the agencies’ actual evaluative behaviour – a representative sample of each agency’s 10 sequential HTAs (spanning 3-4 years in each case) was taken, and the evaluata considered in each were identified. As in the case of the theoretical understanding, the agencies’ practical understanding of evidence was as of data from stipulated domains. These included:

**CADTH:** evidence qua data on effectiveness, cost-effectiveness, broader impact of health technologies (approximately one-half of the time) and safety (approximately one-third of the time)

**MSAC:** evidence qua data on safety, effectiveness, cost-effectiveness, and budget impact of health technologies as of 2010 (all assessed 100% of the time)
NIHR: evidence qua data on effectiveness and cost-effectiveness (both consistently assessed 100% of the time); safety was omitted from all of the assessments

It is immediately apparent that the ‘stated domains’ generally don’t quite match in theory and in practice. For example, it is clear that CADTH sometimes regards data on safety as evidence (even though ‘safety’ is not among the stipulated evidentiary domains); NIHR never considers data on safety (although safety is indicated as one of its stipulated evaluative domains); and so on. It was therefore argued that the proposed definition would be preferable here, as it would both bypass the mismatch problems and would allow the agencies’ evaluative practices to meet the evidentiary-based assessment standard no matter from what domains the data assessed originated – as long as the data that was assessed was relevant to the issue at hand.

Chapter 5 addressed the agencies’ evaluative behaviour vis-à-vis health technologies generally speaking – that is, it spanned the health technology assessments of multiple health technology types (including procedures, pharmaceuticals, diagnostics, etc.). Thus, one could argue that some of the problems around the ‘stated domains’ approach to understanding evidence might disappear if a single health technology type were considered. In order to address this worry, assessments of only genetic tests were selected for further analysis.

Chapter 6 therefore established the historical and conceptual background for chapter 7’s discussion of genetic testing as an HTA issue. Thus, chapter 6 discussed some of the key events in the history of modern genetics that led to the inception of the Human Genome Project, and which, in turn, led to the emergence of genetic testing as a health technology.

Subsequently, in chapter 7, genetic tests were considered as an HTA issue. Although a variety of approaches to the evaluation of genetic tests was adopted by the jurisdictions of interest here, all three jurisdictions shared the same conceptualisation of ‘evidence’ as in context of health technologies more generally – that is, as of data from specific domains. For comparison purposes, each agency’s understanding of ‘evidence’ was cashed out both in terms of standard HTA categories and in terms of ACCE categories. The following understandings were evident:
Understanding of ‘evidence’ in terms of ACCE nomenclature:

**CADTH:** evidence qua data on analytic validity (most of the time), clinical validity, clinical utility/patient management, clinical utility/financial issues (most of the time), ELSI

**MSAC:** evidence qua data on clinical validity; clinical utility/patient management (most of the time); clinical utility/financial issues; ELSI (rarely)

**UK GTN:** evidence qua data on analytic validity, clinical validity, clinical utility/patient management; ELSI (very superficially); clinical utility/financial issues category was never assessed

Understanding of ‘evidence’ in terms of HTA categories:

**CADTH:** evidence qua data on effectiveness; cost-effectiveness (most of the time); broader impact of technologies

**MSAC:** evidence qua data on safety; effectiveness; cost-effectiveness; budget impact (inappropriately considered in 6 cases)

Thus, although one might have suspected that limiting the agencies’ assessments to a single health technology type might dissipate some of the problems around the agencies’ understanding of evidence as data from stipulated domains, this was found not to be the case. Mismatches and inconsistencies between stipulated domains were evident both intra-agency (i.e., assessment to assessment) and inter-agency (CADTH vs. MSAC, for example). Beyond that, it was argued that each agency’s ‘stipulated domains’ approach would miss at least one crucial probative input in a hypothetical assessment of a TTR-FAP genetic test. Adopting the proposed definition of evidence, on the other hand, would allow the agencies to capture all – and only – the inputs required for an adequate, evidence-based assessment of TTR-FAP genetic test. In short, whether considered against health technologies more generally, or against a particular health technology type (here, genetic testing) more specifically, the proposed definition of evidence was established to be both tenable and useful.

Beyond the immediate conclusions about the concept of evidence in field of HTA, two further, Canada-specific, conclusions can be drawn from the analysis offered in this dissertation – first, with regards to genetic testing specifically and second, with regards to the Canadian approach to health technologies more generally.
With respect to genetic tests specifically, it is quite evident that genetic tests are not currently prioritised by Canada’s national level HTA body. As a result, the potential benefits of this technology are not being fully realised. This may, in part, be due to the absence of a federal framework for regulation and evaluation of genetic tests. In its absence, funding decisions regarding genetic tests are currently made on a local and ad-hoc basis in Canada (Adair et al, 2009). Whilst regulatory framework was not a subject of this dissertation, precedents for successful evaluative frameworks for genetic tests do exist – either the Australian MSAC process or the United Kingdom’s UK GTN process could be a good candidate for adoption and adaption to Canadian purposes.

More generally speaking, as became evident in chapter 2, the HTA process for non-pharmaceutical health technologies in Canada is presently quite fragmented. In light of this, Canada’s Federal/Provincial/Territorial Policy Forum have recently prioritised the consideration of the feasibility of a centralised non-drug health technologies review. It is evident that successful models of centralised non-drug health technologies exist in other jurisdictions, such as Australia or the UK. Whilst both these jurisdictions have also adopted a centralised approach to health system funding – unlike Canada – the extant Canadian Common Drug Review process shows that a centralised health technology assessment process can be implemented within a decentralised health system. What would be required for an establishment of a successful centralised non-drug health technologies review in Canada, would thus be an amalgamation of its Common Drug Review process with elements of the Australian and UK approaches to HTA.
REFERENCES


213


219


