

2012

Empirically Evaluating the Impact of Adjudicative Tribunals in the Health Sector: Context, Challenges and Opportunities

Lorne Sossin

Osgoode Hall Law School of York University, lsossin@osgoode.yorku.ca

Steven J. Hoffman

Follow this and additional works at: http://digitalcommons.osgoode.yorku.ca/scholarly_works



Part of the [Health Law and Policy Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Sossin, Lorne, and Steven J. Hoffman. "Empirically Evaluating the Impact of Adjudicative Tribunals in the Health Sector: Context, Challenges and Opportunities." *Health Economics, Policy and Law* 7 (2012): 147-174.

This Article is brought to you for free and open access by the Faculty Scholarship at Osgoode Digital Commons. It has been accepted for inclusion in Articles & Book Chapters by an authorized administrator of Osgoode Digital Commons.

Empirically evaluating the impact of adjudicative tribunals in the health sector: context, challenges and opportunities

STEVEN J. HOFFMAN*

Assistant Professor, Department of Clinical Epidemiology & Biostatistics, and Adjunct Faculty, McMaster Health Forum, McMaster University, Hamilton, Ontario, Canada
Research Fellow, Global Health Diplomacy Program, Munk School of Global Affairs, University of Toronto, Toronto, Ontario, Canada

LORNE SOSSIN

Dean and Professor, Osgoode Hall Law School, York University, Toronto, Ontario, Canada

Abstract: Adjudicative tribunals are an integral part of health system governance, yet their real-world impact remains largely unknown. Most assessments focus on internal accountability and use anecdotal methodologies; few, studies if any, empirically evaluate their external impact and use these data to test effectiveness, track performance, inform service improvements and ultimately strengthen health systems. Given that such assessments would yield important benefits and have been conducted successfully in similar settings (e.g. specialist courts), their absence is likely attributable to complexity in the health system, methodological difficulties and the legal environment within which tribunals operate. We suggest practical steps for potential evaluators to conduct empirical impact evaluations along with an evaluation matrix template featuring possible target outcomes and corresponding surrogate endpoints, performance indicators and empirical methodologies. Several system-level strategies for supporting such assessments have also been suggested for academics, health system institutions, health planners and research funders. Action is necessary to ensure that policymakers do not continue operating without evidence but can rather pursue data-driven strategies that are more likely to achieve their health system goals in a cost-effective way.

Introduction

Adjudicative tribunals are an integral component of health system governance. Created by legislation, these quasi-judicial bodies are granted regulatory, oversight and dispute resolution powers to promote fairness in health decision-making, build confidence in health system management, strengthen governance structures,

*Correspondence to: Steven J. Hoffman, BHSc, MA, JD, McMaster University, 1280 Main Street West, MML-417, Hamilton, Ontario, Canada L8S 4L6. Email: hoffmans@mcmaster.ca

and improve health services and patient safety.¹ Their actual impact on the health system, however, remains largely unknown. Most evaluations of their work have focused on internal measures of accountability and independence rather than external indicators of societal influence. When their effectiveness is examined, assessors tend to utilize anecdotes from various experts and stakeholders rather than the rigorous empirical data that are almost certainly better suited for the purpose. As efforts to reform health systems continue internationally, it will be increasingly important to truly understand the benefits, costs and implications of adjudicative tribunals for providers and consumers of health-care services as well as the institutional structures on which they rely. The dynamic, independent and powerful oversight mechanism of administrative bodies, and their dispute resolution potential, may only be realized with further information on the ways in which they interact with the rest of the complex health system and the impact they have within it. A strong and accountable health system may depend upon it.

In this context, empirical evaluations are an opportunity to inform health policymaking through the collection of objective data regarding the impact of adjudicative tribunals on the health system. Such research includes both quantitative and qualitative investigations on the effects of enacted or proposed interventions – including laws, regulations and policies – on economic, social or health outcomes. Empirical evaluations may be distinguished from other types of research by their reliance on data and use of the scientific method of inquiry (Mello and Zeiler, 2008). Empirical study designs range from experimental (e.g. randomized controlled trials) to quasi-experimental (e.g. interrupted time-series studies) to observational (e.g. cohort, case-control and cross-sectional studies), with data often gathered from surveys, interviews, focus groups, statistical inventories, performance data or documentary analyses.

Empirical research, however, is not new to the health or legal spheres. For health, experimentation, observation and the scientific method have at least notionally been at its core for over a hundred years, with modern ‘evidence-based medicine’ (Evidence-Based Medicine Working Group, 1992) even going as far as prioritizing the ‘conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients’ over all other possible inputs (Sackett *et al.*, 1996). In the legal arena, empirical research has also started to expand both in general (George, 2006) and in health-related

1 Adjudicative tribunals may be defined in a number of ways. This category could include (1) any administrative body engaged in adjudication, including regulatory bodies whose principle function is policymaking but who also engage in adjudication; (2) both administrative and judicial bodies, which engage in adjudication; or (3) only those bodies whose primary or only function is adjudication. We adopt the latter interpretation, but rely on studies and empirical approaches drawn from the regulatory and judicial environment as well, with necessary adaptation to the sphere of administrative bodies whose primary statutory function is adjudication. Ron Ellis, for example, has identified 27 of such tribunals in the Province of Ontario in Canada that engage in ‘rights adjudication’ (Ellis, 2009: 77).

studies specifically (Mello and Zeiler, 2008). However, when compared with the health sphere, it is clear that empirical methodologies in studies of legal institutions have been relatively underutilized. Existing literature explains this observation as a consequence of the complexity of legal interventions, the dearth of large-scale accessible data sets upon which to rely and the heterogeneity of legal interventions, which prevent natural experiments of cross-jurisdictional comparisons (Mello and Zeiler, 2008).

In addition to these general challenges faced by all empirical legal researchers, any attempt to evaluate the impact of a health-related adjudicative tribunal faces additional hurdles. Not only has such an assessment never before been comprehensively undertaken but the most suitable research methodology to do so also remains elusive. Much of empirical health research, for example, relates to patient outcomes and the costs associated with achieving these outcomes. In the setting of adjudicative tribunals, these metrics may not apply. A proceeding before a health tribunal may take place after the outcome for the patient already has occurred, and for this reason the tribunals in fact may impose additional costs on the health system without directly yielding improved health outcomes. Although these additional costs may well lead to better practices and procedures on the part of other actors in the health system (e.g. regulatory colleges, insurance plans, hospitals), this type of benefit is indirect, may only become apparent over time and is inherently difficult to measure.

Distinctions in statutory mandate and the absence of clear statutory language setting out the purposes of adjudicative tribunals often leave no final target outcomes against which services can be evaluated. Further, as creatures of statute that serve quasi-judicial functions, adjudicative tribunals sit at the intersection of the legal and health spheres. These tribunals operate within these two paradigms – a dichotomy of process and outcomes – whose goals may sometimes diverge. Indeed, these administrative bodies are expected to preserve the legal focus on process, fairness and individual-level dispute resolution while at the same time working to improve health-related outcomes by enhancing the overall effectiveness of the health system.² The tension between a process and a substance-based mandate presents distinct challenges for empirical evaluation, especially as evaluations of either dimension would be difficult. The complex co-dependence and interconnectedness of these tribunals with the health system's constituent elements ensure that simple appraisal techniques cannot be effectively utilized. To the extent that adjudicative tribunals have an impact on the health system – which has yet to be empirically proven – this impact is likely linked to a host of other variables. The fact that evaluation is not easy, however, does not detract from its importance.

2 Although this process–outcomes dichotomy between the legal and health spheres is certainly evident when comparing their respective research literature, it is important to recognize that both types of work are conducted within both realms. Mello and Zeiler (2008), for example, highlight several socio-legal studies that gathered outcome-related data, and health researchers frequently address questions of ethics and resource allocation that are more procedural in nature.

The absence of comprehensive empirical evaluations on the impact of health-related adjudicative tribunals, and the potentially significant benefits of doing so, certainly provides sufficient justification for further exploring this possibility. Empirical research, for example, is one of the most accurate ways to systematically assess the population's needs, capture stakeholders' perceptions, test the effectiveness of new initiatives and verify improvements over time. It can help to identify areas of strength and weakness, point to opportunities for growth or improvement, and facilitate a continual process of enhancements so as to better serve the tribunals' constituents and strengthen the health system. Empirical evaluations may also allow for greater accountability for the investment of public resources in health-related adjudicative tribunals. They can help to assess value for money and to determine whether and to what extent such tribunals have any impact at all, which may be particularly important in light of the recent global economic crisis and the austerity measures that have followed.

Despite these benefits and recognized importance, the evaluation and accountability of adjudicative tribunals is also one of the least scrutinized areas of administrative and health law. The topic necessarily engages the issue of administrative independence, the statutory environment within which all adjudicative tribunals operate, the policy priorities of government that funds tribunals, the complexity of the health system and the role of the court in supervising health-related adjudicative tribunals through the mechanisms of judicial review. Evaluating impact in the health sector is also necessarily a contextual exercise. As Peter Cane observed in the administrative law context, "the impact of judicial review needs to be studied in a contextualised way by reference to judicial review's objectives and functions. Also, it should not be assumed that, when we discuss the impact of judicial review, we are all talking about impact of the same thing or, at least, of a single institution with a single set of objectives and functions" (Cane, 2004; see also Sunkin, 2004; Pearson, 2007). A similar approach is necessary for health-related adjudicative tribunals but has never been systematically followed.

This paper aims to explore the context, challenges and opportunities for empirically evaluating the impact of adjudicative tribunals in the health sector. First, we will discuss the purpose, function and importance of these bodies within the health system, including their statutory mandates and policy goals. Second, we will examine the various ways in which their performance could potentially be assessed and will justify why there is a need to develop empirical approaches for the assessment of adjudicative decision-making. This part of the study will be illuminated with examples from a focused scoping review of past evaluations that we identified through searches in legal, medical and general databases (i.e. Westlaw, Quicklaw, MEDLINE, Cochrane Library, Campbell Library, Google Scholar), an existing review of empirical evaluations relating to administrative tribunals (Partington *et al.*, 2007) and other studies that were already known to us and the various experts and colleagues we consulted. Special effort was made to find empirical studies, with non-empirical works

collected and reviewed as found. Third, we will identify the extensive barriers to empirically evaluating the societal impact of adjudicative tribunals, which we situate into three distinct categories: (1) complexity of the health system, (2) methodological difficulties and (3) realities of the legal profession and the environment in which it currently operates. Finally, based on this analysis, we will advance what we believe to be the most constructive path forward for the empirical assessment of adjudicative decision-making. We hope that this work will encourage and inform future empirical evaluations of adjudicative tribunals in the health sector that will help to assess their worth, improve their performance, enhance health decision-making, advance patient safety and facilitate the achievement of population health goals. Such evaluations would also help to confirm or dispel current beliefs and opinions on adjudicative tribunals, which in most instances, including our own, are based predominantly on personal experience rather than more objective and systematic assessments.

To enhance the utility of this paper for health policymakers in any context, each of these sections will be brought to life by focusing on their application to two prominent health tribunals in Canada, namely, the Ontario Health Professions Appeals and Review Board (HPARB) and the Health Services Appeals and Review Board (HSARB). Both HPARB and HSARB have statutory mandates to review important health decisions that intimately affect the lives of their constituents. Using these two bodies as case studies for exploring the context, challenges and opportunities for evaluating adjudicative tribunals may enrich our understanding of administrative tribunals throughout other sectors as well. More broadly, we believe that this analysis may resonate with the experiences of health-related adjudicative tribunals in other common law jurisdictions including the United States (e.g. U.S. Department of Health & Human Services Appeals Board, Medicare Appeals Council, California Department of Managed Health Care's Independent Medical Review Program), United Kingdom (e.g. Health Professions Council, Family Health Services Appeal Authority, Mental Health Review Tribunal), Australia (e.g. Health Professions Licensing Authority, Occupational Safety & Health Tribunal), New Zealand (e.g. Health Practitioners Disciplinary Tribunal, Pharmacy Board of Appeal) and South Africa (e.g. Health Professions Council, Mental Health Review Board). The various shared features among all adjudicative tribunals – including their creation by legislation, use of delegated powers, dispute resolution services, legal status and oversight function – ensures cross-border utility of this analysis and the relevance of case studies from one jurisdiction to another.

The purpose, function and importance of adjudicative tribunals in the health sector

Adjudicative tribunals are administrative bodies that are created by statutes and that exercise delegated decision-making powers of the executive branch for the purposes of achieving certain policy goals. They serve as an oversight mechanism

for lower-level decision-makers and apply legal and normative principles to resolve disputes between conflicting parties. They are independent – operating at arm’s length from the government – and serve quasi-judicial functions otherwise fulfilled by the formal judicial system. This independence, however, also has limits; their members are appointed by the executive branch of government (in the case of HPARB and HSARB, the power of appointment is effectively in the hands of the Minister of Health), which also sets their staffing allowances and budgets. Their decisions, although often final, must be authorized by their enabling statute and are subject to judicial review by the courts. Governments pursue policies in relation to these bodies for a host of reasons. The government may, for example, wish to remove the need for court intervention, facilitate opportunities for settlement, enhance access to efficient dispute resolution mechanisms or promote fairness.

In the health sector, adjudicative tribunals may be involved with resolving disputes regarding medical malpractice claims, insurance coverage for health-care services, determination of mental capacity, licensing decisions for health-care facilities and patient safety procedures. They serve as an accountability mechanism that ensures health decision-makers follow appropriate processes and act according to their respective statutory mandates. Adjudicative tribunals aim to boost public confidence in the credibility of decision-making within the health system, facilitate better and more consistent decisions, and reduce the risk of errors that in this context can have deadly consequences. Finally, they can help to promote independence, fairness and justice within health care, militate against self-interest and corruption, and provide opportunities to address wrongs through redress.

The HPARB, for example, is an integral part of Ontario’s self-regulating health professional system. It helps to ensure that the health professions are regulated in the public interest, that appropriate standards of practice are created and maintained, that patients have access to the health professional of their choice and that they are treated with respect and sensitivity by health professionals. HPARB was established as a response to two related phenomena in the late 1960s and early 1970s: first, the recommendation arising out of the Report by the Honourable James Chalmers McRuer’s Royal Commission Inquiry into Civil Rights, which emphasized the need for public interest oversight over self-regulating professional bodies (McRuer, 1968); and second, the Committee on the Healing Arts tabled by the government on 28 April 1970, which also emphasized the primacy of public interest regulation of health professionals (Ontario Committee on the Healing Arts, 1970). Under the Province of Ontario’s *Regulated Health Professions Act* (RHPA), people may appeal the decision of a self-regulated health professional college to not pursue a disciplinary proceeding before HPARB.³ If the appropriate statutory processes

³ *Regulated Health Professions Act, 1991*, S.O. 1991, c. 18, s. 3. Also see Steinecke, R. Looseleaf. *A Complete Guide to the Regulated Health Professions Act*. Toronto: Canada Law Book. The RHPA is one of several statutes administered by HPARB.

were not followed by the relevant college, then the board is empowered to send the matter back to the college for reconsideration. HPARB also hears appeals from adverse decisions by the colleges in relation to registration requests.

The HSARB, similarly, is a part-time board providing oversight for the decisions of various actors within the health system. Its broad jurisdiction arises from 14 different statutes and includes reviewing decisions concerning payment for health-care services under the Ontario Health Insurance Plan (OHIP), eligibility for housing in long-term care facilities, licensing of nursing homes and other independent health facilities and the decisions of public health officials (Pitfield and Flood, 2005).

Both HPARB and HSARB have a full-time chair,⁴ and a roster of part-time members, some of whom have legal training (and in the case of HSARB, medical training) and some who do not. Each board is supported by a shared secretariat, which forms part of the Ontario Ministry of Health. Both Boards have been held to be expert bodies by reviewing courts, which warrant deference. Their substantive decisions may only be overturned if found to be ‘unreasonable.’⁵

As indicated above, a key aspect of evaluating tribunals created by statute is to assess whether a tribunal is fulfilling its statutory objective(s). This may be especially challenging, for example, if the specific goals of the relevant tribunal are diffuse and ambiguous in their enabling legislation. Ontario’s RHPA, for example, does not detail the purposes of the boards, so this must be inferred from the powers and authority they have been provided. For example, HPARB has the power to review decisions of regulated health colleges not to refer complaints to discipline on grounds of the reasonableness of the college’s decision and the adequacy of the college’s investigation.⁶ HPARB has broader jurisdiction to review decisions by colleges to deny registration to applicants.⁷ Thus, although HPARB’s role is generally to ensure public interest accountability over decision-making by regulated health colleges, HPARB’s role in reviewing complaints suggests a different purpose, and a more deferential standard of review, than its role in reviewing denials of registration. Evaluation needs to be responsive to these differences of statutory mandate.

Approaches to evaluating health-related adjudicative tribunals

Assessing the work of these adjudicative tribunals and others in the health sector is an inherently complex enterprise. However, evaluations can be thought of and categorized according to their orientation and methodology (see Table 1).

4 Since 2008, the same individual has served as chair of both boards.

5 See, with respect to HPARB, *Botros v. Beadle* (2007), 71 Admin L.R. (4th) 225, and with respect to HSARB, *Flora v. Ontario (Health Insurance Plan, General Manager)* 2008 ONCA 538.

6 See s. 29(2) of the *Health Professions Procedure Code*, Schedule 2 to the RHPA.

7 See s. 22(1) of the *Health Professions Procedure Code*, Schedule 2 to the RHPA.

Table 1. Categorizing questions about health-related adjudicative tribunals according to their orientation and the possible methodologies that can be used to answer them

	Methodology	
	Expert reviews	Empirical evaluations
Orientation		
Internal procedural analyses	<ul style="list-style-type: none"> ● Are tribunal adjudications fair? ● Are procedures transparent? ● Is there sufficient independence in the body's decision-making? ● What management structures should be in place? ● Are there sufficient accountability mechanisms in place? ● What barriers exist to prevent potential users from accessing the services of a tribunal? ● What bottlenecks exist in the provision of services? ● What steps can be taken to increase the consistency of decision-making? 	<ul style="list-style-type: none"> ● How many disputes are resolved? ● How many employees are needed for optimal efficiency of operations? ● What costs are involved? ● What factors influence independence and impartiality? ● How satisfied are users with the body's operations? ● What are users' perceptions of the quality of tribunals' services and decisions? ● What is the impact of having legal representation for the parties? ● How long is the average hearing?
External impact evaluations	<ul style="list-style-type: none"> ● How do tribunals interact with one another? ● How do they fit into the larger health system? ● To what extent do they affect the work of the judicial system? ● Do they provide sufficient opportunities for individuals to pursue justice? ● Do they create a societal impression that justice is obtainable? ● What are some of the theoretical benefits of these bodies? ● How do they contribute to good governance and accountability in the health system? 	<ul style="list-style-type: none"> ● What is the perceived effectiveness and utility of tribunals among users, stakeholders and the general public? ● What impact do they have on decision-making and health outcomes? ● Under what circumstances do they positively influence the health system? ● How do these bodies enhance health care services? ● Is their public confidence with the services provided? ● How can the tribunal increase its impact on the community? ● Were past initiatives effective? ● What areas are most in need of improvement? ● What types of services are needed?

Examples of research questions that can be answered by empirical impact evaluations, the focus of this paper, are located in the bottom right-hand cell of this table.

In terms of orientation, evaluations of tribunals can be focused on how they function or what impact they have. The former would analyze the internal operations of a tribunal while the latter would assess the body's external effects on a specified population. Procedural analyses are important to promote coherent internal management structures, good governance, accountability, efficiency and efficacy. External impact evaluations, on the other hand, represent a way to assess

the real-world effectiveness of the adjudicative tribunal, its impact on others within the health-care system and the benefits (or consequences) that this impact yields. Such studies can determine whether or not these bodies support and/or enhance the functioning of various health system institutions and decision-makers and whether or not they ultimately influence service provision, access to justice in the health sector and health outcomes. External impact evaluations require expertise and independence – they are not traditionally conducted by auditors (e.g. McCarter, 2008), ombudsmen (e.g. Marin, 2008) or internal staff (e.g. Health Professions Regulatory Advisory Council, 2008).

Government evaluations of both health and non-health administrative bodies tend to focus on issues related to internal operations rather than external impact. The recent report of the Ontario Security Commission's Fairness Committee, for example, examined whether the agency's internal governance structure created a perception or reality of bias in its adjudicative responsibilities (Osborne *et al.*, 2004). The United Kingdom's National Audit Office similarly reviewed the procedures used by its Department of Work and Pensions to medically assess incapacity and disability (National Audit Office, 2003a) and to hear appeals of social security benefit decisions (National Audit Office, 2003b). Some reviews examine particular problems that had previously been identified (e.g. Blumenthal and Wessely, 1993; Blumenthal and Wessely, 1994; Society of Ontario Adjudicators and Regulators, 1995), while others focus on users' satisfaction with a tribunal's provision of services (e.g. Carscallen *et al.*, 2003; Confederation of British Industry, 2005; Employment Tribunals Service, 2005; Aston *et al.*, 2006). Several assessment efforts have even focused on the internal operations of multiple tribunals or a jurisdiction's entire tribunal system, including the report of Ontario's Agency Reform Commission (Guzzo *et al.*, 1998), the UK's Leggatt Review of Tribunals (Leggatt, 2001) and the report of the UK's former Council on Tribunals (Adler and Gulland, 2003). Academics also tend to focus on the internal operations of tribunals across various topics – whether they regulate securities (Moyer, 1997; Rousseau, 2008), medical malpractice claims (Siegal *et al.*, 2008), denials of health insurance coverage (Gresenz and Studdert, 2005), privacy (Jacobs, 2008), pensions (Sossin, 2007) or determinations of medical incapacity (Peay, 1989; Bradley *et al.*, 1995; Dolan *et al.*, 1999) – and often examine users' experience (Genn *et al.*, 2006a, 2006b). While not a single governmental evaluation could be found that focused on the external impact of adjudicative tribunals, at least one academic publication discusses the potential benefits that administrative 'health courts' (which resolve malpractice claims) can have on patient safety (Mello *et al.*, 2006).

In terms of methodology, assessments of tribunals can either be conducted through expert reviews or empirical evaluations. The first approach would take advantage of the personal experiences and perspectives of an investigator while the second approach harnesses the objectivity of the scientific method and the generalizability of data that were collected from many people. Expert reviews rely

upon the contextual and reflective expertise of the authors and are important for probing the etiology of complex challenges within the tribunal system, raising questions of possible concern or future inquiry, identifying structural problems and possible ways to overcome them, justifying political decisions (either from the past or those planned for the future) and suggesting recommendations for reform. This approach is also more likely to have fewer costs and a faster completion timeline. Empirical evaluations of tribunals, by contrast, seek scientific, evidence-based methods and can be used to, *inter alia*, quantitatively or qualitatively assess their impact on the health system, identify the factors that determine their successful operations, and track perceptions of them over time. An empirical evaluation may or may not lead to recommendations for reform. That said, these two methodological approaches cannot in reality be so strictly distinguished. Experts often utilize empirical methods and even the most scientifically rigorous and objective evaluations must be interpreted by individuals (often experts in the field) who may have stated or unstated agendas for reform.

Reviews of adjudicative tribunals have been conducted using both expert and empirical methodologies. Prominent observers, academics and practitioners, for example, have assessed various tribunals' organizational structures (Moyer, 1997; Osborne *et al.*, 2004), efficiency (Carscallen *et al.*, 2003), accessibility (Adler and Gulland, 2003), independence (Rousseau, 2008), performance standards (Society of Ontario Adjudicators and Regulators, 1995), accountability (Allsop and Jones, 2008) and overall effectiveness (Guzzo *et al.*, 1998; Leggatt, 2001; Thomas, 2005; Mello *et al.*, 2006; Sossin, 2007). Other reviews feature empirical elements such as (1) *surveys* that capture the perceived quality of services offered (Employment Tribunals Service, 2005), stakeholder attitudes toward the tribunal (Confederation of British Industry, 2005) and the functioning of a certain process (Corsi and Hurley, 1979; Toubman *et al.*, 1995; Latreille *et al.*, 2004, 2005); (2) *interviews* that probe users' experiences with the tribunal (Aston *et al.*, 2006; Genn *et al.*, 2006b), its perceived impartiality (Jacobs, 2008) and the effectiveness of a particular procedure (Bradley *et al.*, 1995; National Audit Office, 2003a; Siegal *et al.*, 2008; Urwin *et al.*, 2010); and (3) *performance data* and *documentary analyses* for examining key features of a tribunal's caseload (Hayward *et al.*, 2004) and arrangements for how, when and why it makes appeal decisions (Peay, 1989; National Audit Office, 2003b; Gresenz and Studdert, 2005).

The challenge in evaluating health-related adjudicative tribunals, therefore, seems to lie at the intersection of orientation and methodology. Assessments of adjudicative tribunals have focused on both process and impact, and have been conducted using both expert reviews and empirical methods; yet, not a single review could be found that empirically evaluated the external impact of an adjudicative tribunal, despite extensive searching (see Online Appendix A for 27 examples of adjudicative tribunal evaluations listed according to these categories). While this lack of research may indicate that such undertakings are not important, interesting or possible, the evidence suggests otherwise: the need

for external impact evaluations is evident (OECD Development Assistance Committee, 1998; Hertogh and Halliday, 2004; Center for Global Development's Evaluation Gap Working Group, 2006; World Bank, 2006) and such evaluations have been conducted with success in related settings, which also involve the nexus of the health and law sectors and beyond (Jones *et al.*, 2009).

The dearth of externally focused empirical evaluations is not only a missed opportunity but it may also pose a significant risk. On the one hand, the lack of an empirical rationale for the benefits of adjudicative tribunals may render them vulnerable to opposition or simply to general cost-cutting initiatives. On the other hand, ineffective tribunals may unjustifiably be receiving government funding at the expense of other programs, which cannot be supported as a result of limited resources. Without these data, adjudicative tribunals may also lack the baseline measures needed to track changes over time, evaluate the performance of decision-makers and staff, and engage in longer-term strategic planning.

Indeed, it is widely accepted that data-driven strategies are more likely to help decision-makers achieve their goals in a cost-effective way than policies pursued in the absence of evidence (World Health Organization, 2004; Maynard, 2006; Chalkidou *et al.*, 2009; Hoffman *et al.*, 2009). Information gathered by health-related adjudicative tribunals like HPARB and HSARB through empirical methods may be of particular interest to government officials as it can demonstrate performance benchmarks and ensure that public funds are being invested and spent effectively. If reform is called for, empirical data will be essential in identifying what needs to change. For academics, it is an under-scrutinized sphere of administrative law and health systems functioning that is both ripe for research and, potentially, reform.

Challenges for empirically evaluating adjudicative tribunals in the health sector

Yet despite the tremendous benefits, empirical impact evaluations of health-related adjudicative tribunals are not being conducted. This absence of assessment efforts is most likely attributable to the various challenges facing anyone who embarks on undertaking such a project. These obstacles can be divided into three categories: (1) complexity in the health system; (2) methodological complications; and (3) environmental realities in which the legal profession and tribunals operate (Table 2).

Challenges with complexity in the health system

Empirically evaluating the impact of any adjudicative tribunal is a naturally difficult enterprise as it requires the body's various effects to be isolated from the larger social context within which it operates. This is no doubt complicated for tribunals in every sector because their activities are usually only indirectly

Table 2. Summary of the complexity, methodological and environmental challenges associated with empirically evaluating the impact of adjudicative tribunals in the health sector

Complexity challenges	Methodological challenges	Environmental challenges
<ul style="list-style-type: none"> ● Health system is multi-layered, non-linear and highly sophisticated ● Countless independent actors and institutions that interact in unpredictable ways ● Nesting of sub-systems throughout the larger health system ● Inaccuracy of mechanistic “cause–effect” relationships ● Tribunals serve diverse functions for various players within different contexts 	<ul style="list-style-type: none"> ● Indirect relationship between adjudicative tribunals and their existential purpose ● Absence of legislatively defined desired outcomes ● Evolution within tribunals ● Difficulty in crafting suitable goals that are observable, measurable and testable ● Cannot randomize potential users into different “treatment” groups ● Few examples to emulate ● No obvious data set upon which to rely 	<ul style="list-style-type: none"> ● Focus of legal sector on process, administration and oversight rather than outcomes ● Concern for independence ● Limited statutory mandates ● Costs and resource limitations ● Dearth of empirical training and competence in the legal sector ● Diminished value of and prestige for empirical legal research within academia ● Culture of deference to authority and expertise

related to their existential goals. This challenge, however, may be further exacerbated in the health context due to its overwhelming complexity.

Indeed, health systems are increasingly being recognized as complex adaptive systems that are multi-layered, non-linear and highly sophisticated. They consist of countless sub-systems with immeasurable independent actors, established policies, zealously guarded interests, entrenched professional ‘silos’ and divergent cultures that can all influence each other and even alter their external environments. This web of elements, and the unpredictable interactions among them, ensures that conventional mechanistic or ‘cause–effect’ conceptualizations of the health system are inaccurate and oversimplifications of its complex dynamics (Zimmerman *et al.*, 1998).

While scientific knowledge has been greatly advanced by breaking big questions into smaller ones that can be observed, analyzed and understood through rational deduction, this process is severely limited when the studied phenomenon or intervention is located within a system whose constitutive parts are not independent, constant or predictable. The fact that the health system exhibits characteristics of distributed control, co-dependence and nesting of smaller systems within other larger systems further aggravates this challenge and makes it difficult to fully examine adjudicative tribunals without reference to other actors and institutions (such as its adjudicators, staff, government policymakers, regulatory colleges, relevant expert panels, the traditional court system and the public). Isolating and attributing impact are further problematized by the fact that health-related

adjudicative tribunals serve diverse functions according to various players within completely different contexts (Plsek and Greenhalgh, 2001).

Challenges with research methodology

Yet in addition to the daunting barriers of evaluating adjudicative tribunals caused by health system complexity, there are further methodological barriers associated with such an undertaking. The primary challenge, as highlighted above, is that simple research designs cannot be used to isolate adjudicative tribunals and to elegantly locate causal inferences between them and their goals. But above and beyond the various explanations illuminated by the complexity perspective is the fact that efforts of adjudicative tribunals are only indirectly related to their goals. Indeed, health services themselves only partially help meet their goal of improved health for people. Any legal, regulatory or oversight ‘intervention’ that serves to better structure these services would be even further removed from their ultimate goals. Empirical impact studies of such interventions must be expertly designed to account for this complexity.

However, even if simple methods did exist to observe the relationship between adjudicative tribunals and their goals, there is currently a lack of clear evaluative criteria against which particular adjudicative tribunals can be measured. This is because their goals are not easily articulated and have thus not been defined with adequate precision – if defined at all. Desired outcome measures are consequently absent, which ensures that suitable quantitative and/or qualitative research methodologies cannot be matched to them. This problem, however, cannot simply be overcome by brainstorming possible goals of adjudicative tribunals. Indeed, the existential purpose of these bodies may change and evolve over time with new legislators, government policymakers, adjudicators and tribunal staff who can each contribute toward a shift in the focus and priority of their operations over time.⁸ Various community stakeholders may also perceive the role of a particular adjudicative tribunal in their sector very differently depending upon their own mandate, ideological perspective and unique vantage point. While reference to a tribunal’s enabling statute may be informative in crafting an outcome measure, it is not always decisive. In the case of HPARB, legislative provisions suggest that this body was created to ensure effective regulation of the health professions in the public interest;⁹ yet, this goal is not easily quantifiable. Indeed, the ability to empirically evaluate a complex intervention like a health-related adjudicative tribunal depends upon having a desired outcome that is observable, measurable and testable against a null hypothesis.

⁸ For example, the evolution of Ontario’s HPARB over a period of 40 years was documented in the tribunal’s formal submission to the Health Professions Regulatory Advisory Council regarding inter-professional collaboration among health colleges and professionals (Health Professions Appeal and Review Board, 2008).

⁹ *Regulated Health Professions Act*, 1991, S.O. 1991, c. 18, s. 3.

Some intended outcomes may be impossible to measure or may depend on factors outside a tribunal's control, such as how others respond to their decisions and recommendations.

A desire to empirically 'prove' causal inferences between adjudicative tribunals and a particular outcome is also complicated by the impossibility of randomly allocating potential users of existing tribunals into groups that either receive or do not receive their services. Randomized controlled trials – the most rigorous of discrete empirical evaluations (GRADE Working Group, 2004) – assess the effect of an intervention on a test population in comparison with a theoretically identical population. This method, however, requires properly constituted (i.e. randomized) and adequately sized (i.e. large) treatment and control groups with both known and unknown confounding factors evenly distributed between them in order to isolate the impact of tribunal services and measure it against a benchmark. Non-randomized retrospective evaluations comparing users of tribunals to non-users (or the situation of the general public in jurisdictions with and without comparable tribunals) may not be an ideal solution to this challenge as this creates a situation where user-status and outcomes are measured at the same time. This makes causal determinations infinitely more challenging and requires evaluators to systematically identify and control for confounding factors in a separate process (Mello and Zeiler, 2008).

A penultimate methodological challenge for conducting external impact evaluations of health-related adjudicative tribunals is that there are few examples of past efforts upon which to emulate. As previously mentioned, many empirical studies have examined the internal processes of tribunals, but none could be found that focused on their societal impact. This is exacerbated by the dearth of obvious empirical data sets that can be analyzed and from which potential evaluators can draw (Mello and Zeiler, 2008). Although hospitals may be able to compare their patient population and its outcomes to those from neighboring hospitals, adjudicative tribunals may not be in a position to continually collect data about their past users or to compare this information with existing data sets from the same region or others.

Finally, the identity and background of the researchers evaluating the impact of an adjudicative tribunal must also be considered. While the goal of empirical study is to avoid bias and ideological assumptions, every researcher brings a particular matrix of perspective, orientation, experience and values to their work. Insiders, for example, may bring intuition and experiential judgment, while outsiders may bring independence, fresh eyes and objectivity.

Challenges with the legal environment

As institutions that function within both the health and legal systems, health-related adjudicative tribunals must also overcome the realities of the legal sector that may not be particularly nurturing for empirical impact evaluations. For example, legal actors are often focused more on achieving due process, transparency and good

governance than specific societal outcomes (like improved health status, which is the goal of direct clinical health care). Excellent process in the legal world is often thought to be the most likely way to achieve the best outcome and may indeed be constitutive of it in the absence of gold standards or clear benchmarks.

There is also greater concern for maintaining independence and avoiding any apprehension of bias within the legal sector than in other areas. Like impartiality, independence is a common law right of procedural fairness enjoyed by parties who come before administrative bodies in common law jurisdictions (including Canada, United States, United Kingdom, Australia and New Zealand). In Canada, independence for adjudicative tribunals is based on the categories of judicial independence identified by the Supreme Court of Canada in *Valente v. The Queen* (i.e. security of tenure, financial independence and administrative independence over adjudicative matters)¹⁰ and applied to administrative bodies in *Canadian Pacific Ltd. v. Matsqui Indian Band* – albeit in a more flexible and contextually sensitive manner.¹¹

Respecting this independence of adjudicative tribunals will naturally influence the process and content of any evaluation in multiple ways. For example, independence suggests that governments should refrain from evaluating tribunals' substantive decisions, lest reasonable observers reach the legally problematic conclusion that tribunals may adjust their decision-making to align with what the government of the day perceives as 'successful'. Similarly, it may also be difficult for a tribunal to establish evaluative criteria or outcome measures for itself as this might lead a reasonable observer to conclude that the tribunal may pursue these goals at the expense of fairness to the parties. This concern for independence even questions the extent to which tribunals' staff and members can be directly involved in any evaluation for fear of influencing or interfering with their services that must remain neutral at all times. In contrast to encouraging self-evaluation as is common within the health sphere, the legal environment may actually discourage adjudicative tribunals from assessing their own external impact, especially as such undertakings are not explicitly part of their statutory mandates.¹²

Finally, as recently highlighted by the Nuffield Inquiry on Empirical Legal Research, the legal academy also suffers from a dearth of empirical competence and capacity to conduct such studies (Genn *et al.*, 2006a). While the field of empirical health law scholarship has recently grown exponentially (Mello and Zeiler, 2008), it is generally accepted that the current capacity is inadequate and

10 *Valente v. The Queen* (1985), 24 D.L.R. (4th) 161 (S.C.C.).

11 *Canadian Pacific Ltd. v. Matsqui Indian Band* (1995), 122 D.L.R. (4th) 129 (S.C.C.). It should be noted that these standards of independence that are relevant in the adjudicative tribunal context are only a common law right, which may be displaced by statute, unlike judicial independence, which is a constitutional principle. See *Ocean Port Hotel v. British Columbia*, 2001 SCC 17.

12 On the other hand, a study that expresses respect for the adjudicative independence of tribunals will likely have greater credibility and attract broader 'buy in' than a study that is perceived as inconsistent with it.

that it may further diminish over time. Empirical legal methodologies are also not generally recognized to be as prestigious within the academic community as traditional doctrinal investigations (Genn *et al.*, 2006a). The pervasive culture of deference to experts and authority must further diminish the perceived value of objective empirical work and weaken any apparent need for more rigorous research that is higher on the hierarchy of evidence (GRADE Working Group, 2004). Again, the focus on elements of process (e.g. bias and independence) rather than impact (e.g. judicial decisions) as indicators of quality and performance must also deter legal scholars from conducting work in this area such that target outcomes are less likely to be assessed.

Reasons for optimism

However, despite the challenges faced by potential evaluators of adjudicative tribunals, there is reason for optimism: each of the various identified barriers can be overcome and have indeed been circumvented in similar evaluations. For example, as previously mentioned, many empirical evaluations have been conducted that focus on the internal operations of these bodies. A major literature review in 2007 highlighted much of the work that has been conducted and published in this area (Partington *et al.*, 2007). Yet in addition to these studies, empirical evaluations have also been undertaken to assess the external impact of similarly functioning specialty courts that operate within the judicial system. A systematic review of the research evidence has even been conducted on the societal impact of at least one type of these judicial organs (Wilson *et al.*, 2006).

Indeed, methodologically, there may be much to learn from external impact evaluations of specialist courts in the judicial sector (Plotnikoff and Woolfson, 2005). For example, ‘drug courts’ have been extensively evaluated in the United States and in other jurisdictions regarding their ability to increase treatment rates, lower criminal recidivism, and enhance cost-effectiveness of prosecution (United States Government Accountability Office, 2005; Wilson *et al.*, 2006; Green *et al.*, 2009). Domestic violence courts and community courts have similarly been assessed for compliance, cost-effectiveness, conviction rates and public perception, and mental health courts have been comprehensively examined for reducing criminal violence, enhancing community safety, conserving fiscal resources and improving clinical outcomes (see Online Appendix B for a list of 91 external impact evaluations of specialist courts). However, it must be recognized that the context within which these judicial bodies operate is very different from that of health-related administrative tribunals. Not only are they part of the judiciary rather than the executive branch of government, but their existential goals are usually related to diverting complex or special cases from traditional courtrooms rather than supporting the infrastructure of a completely different system (like that of the health system). Empirically tracking desired outcomes like cost-savings and reduced reoffending rates will naturally be easier

in this context when the intervention or service is more directly related to its goal. Yet, alternatively, it may actually be more difficult for these judicial organs to evaluate themselves due to their strict separation from the executive (which has the financial resources to fund such undertakings) and the likelihood of them to zealously guard their independence.

Researchers seeking to empirically evaluate health-related adjudicative tribunals can also benefit from earlier accounts of similar efforts. For example, Mello and Zeiler (2008) describe the diversity and comparative advantages of various empirical approaches that have been taken by scholars in the health law field to address issues as wide-ranging as medical malpractice reform and motor safety laws. Furthermore, on the use of randomized controlled trials, for which these two scholars are less optimistic, Pleasence (2008) provides an account of such an undertaking in the United Kingdom, which highlights the many technical, practical and ethical barriers that were faced, and suggests ways to overcome them in the future.

Opportunities for moving forward

Deliberate and concerted efforts, however, will be necessary to overcome the numerous barriers to empirical impact evaluations of health-related adjudicative tribunals among both individual evaluators and others that must support them. The analysis of challenges described above points to several strategies that can be pursued.

At the individual level, potential evaluators of adjudicative tribunals may need to assemble interdisciplinary teams to obtain the necessary methodological expertise, bring an aura of independence and credibility to their work, and save tribunal staff from the potentially uncomfortable situation of relinquishing their perceived independence by evaluating their own performance. Like the process for assessing the effectiveness of complex clinical interventions, evaluators of adjudicative tribunals may then be advised to conceptually map out the way in which their tribunal functions, its interactions and relationships with others in the health and legal systems, and its potential effects on each of them (Campbell *et al.*, 2007). This will aid in focusing the inquiry, identifying areas in which little is known, generating suitable research questions and determining the appropriate methodology.

Potential evaluators must also thoughtfully consider both the target audience of their research and the overall goal that their particular health-related adjudicative tribunal is expected to help achieve, and then identify the most important targeted outcomes that are relevant to the audience and important for the goal's fulfillment. When such outcomes cannot directly be measured, as may often be the case, evaluators must identify strong surrogate endpoints, which are measurements that reflect important outcomes even if they are of indirect or diminished practical importance. Performance indicators can then be developed followed by the corresponding methodologies for tracking changes to them (see Panel 1).

Panel 1. Steps for moving empirical impact evaluations forward at the individual level

The following seven steps are a suggested sequence of actions that researchers, and/or tribunal staff can take to empirically evaluate the impact of adjudicative tribunals in the health sector:

1. Assemble an interdisciplinary team of researchers that has the necessary methodological expertise, independent aura and credibility among target audiences.
 2. Conceptually map the way in which the adjudicative tribunal functions and its interactions with and effects on others within the health and legal systems.
 3. Identify the overall goal that the adjudicative tribunal is expected to help achieve and the most important targeted outcomes that are necessary for its fulfillment.
 4. Develop measurable surrogate endpoints and corresponding performance indicators that reflect important outcomes.
 5. Construct a realistic and ethical research methodology that is appropriate for tracking each surrogate endpoint.
 6. Conduct the empirical impact evaluation.
 7. Disseminate findings widely to decision-makers who can act upon them and other researchers whose work may be better informed by them.
-

In the case of Ontario's health-related adjudicative tribunals, both HPARB and HSARB can describe their overall goal as contributing to the health of Ontarians by enhancing decision-making within the health system. If government officials are the evaluation's intended audience, targeted outcomes could include (1) confidence in the health system, (2) equity, justice and fairness in health decision-making, (3) strengthened health system institutions and (4) better health services and patient safety via enhanced regulation and oversight. As these outcomes would be nearly impossible to measure directly, surrogate endpoints can be developed and could possibly include (1a) access to adjudicative mechanisms for dispute resolution, (1b) perceived legitimacy of adjudicative decisions, (2a) satisfaction with adjudicative services, (2b) perceived fairness and legitimacy of adjudicative services, (3a) interaction among health system institutions and decision-makers, (3b) existence of support mechanisms for primary health decision-makers, (3c) effective oversight of primary health decision-makers, (4a) better decisions by primary health decision-makers and (4b) respect for the tribunal's oversight function. Performance indicators and their corresponding empirical methodologies could then range from the public's awareness for the tribunal's existence to the perceived concern among primary health decision-makers that their decisions will be reversed. Suggestions for inclusion as part of a government-focused, empirical evaluation matrix of the external impact of health-related adjudicative tribunals such as Ontario's have been presented in Table 3 to illuminate this process in the context of a real-world example. Evaluations of these bodies for other target audiences such as regulatory colleges, insurance plans, hospitals, individual health professionals or patients would likely lead to different measurable outcomes.

Once a system of empirical observation is in place, potential evaluators can establish benchmarks according to which they can track and assess performance.

Table 3. Suggestions for an empirical impact evaluation matrix for Ontario’s health-related adjudicative tribunals

Targeted outcomes	Surrogate endpoints	Performance indicators	Empirical methodology
Confidence in the health system	<ul style="list-style-type: none"> ● Access to adjudicative mechanisms for dispute resolution (including technical, realistic and perceived access) 	<ul style="list-style-type: none"> ● Number of cases heard per year and percentage of decisions overturned ● Persons using accessibility options (e.g. translators, phone dial-in) ● Wait times for tribunal services 	<ul style="list-style-type: none"> ● Quantify and characterize services provided ● Audit technical and realistic accessibility by sub-group ● Solicit knowledge and perception of access among the public ● Access internal performance data
	<ul style="list-style-type: none"> ● Perceived legitimacy of adjudicative decisions 	<ul style="list-style-type: none"> ● Public knowledge of the tribunal’s existence ● Public and stakeholder confidence in the tribunal’s decision-making 	<ul style="list-style-type: none"> ● Solicit knowledge and perception among public, users and relevant experts
Equity, justice and fairness in health decision-making	<ul style="list-style-type: none"> ● Satisfaction with adjudicative services 	<ul style="list-style-type: none"> ● Satisfaction with the tribunal’s appeal/ review process ● Satisfaction with execution of the process 	<ul style="list-style-type: none"> ● Survey and/or interviews of users and stakeholders
	<ul style="list-style-type: none"> ● Perceived fairness and legitimacy of adjudicative services 	<ul style="list-style-type: none"> ● Public confidence in the fairness and legitimacy of tribunal operations 	<ul style="list-style-type: none"> ● Solicit perceptions of fairness and legitimacy among public, users and relevant experts
Strengthened health system institutions	<ul style="list-style-type: none"> ● Interaction among health system institutions and decision-makers 	<ul style="list-style-type: none"> ● Partnerships between the tribunal and others ● Joint ventures co-led by the tribunal and others ● Exchanges of information 	<ul style="list-style-type: none"> ● Audit of formal collaborations between tribunals and others ● Interviews with tribunal staff and primary decision-makers
	<ul style="list-style-type: none"> ● Existence of support mechanisms for health primary decision-makers 	<ul style="list-style-type: none"> ● Educational initiatives and opportunities ● Services targeting primary decision-makers 	<ul style="list-style-type: none"> ● Audit of initiatives and services targeting primary decision-makers ● Perceived utility of the tribunal among primary decision-makers

Table 3. (Continued)

Targeted outcomes	Surrogate endpoints	Performance indicators	Empirical methodology
	<ul style="list-style-type: none"> ● Effective oversight of primary health decision-makers 	<ul style="list-style-type: none"> ● Percentage of decisions overturned ● Direction provided for better decision-making (either explicitly or implicitly) 	<ul style="list-style-type: none"> ● Access internal performance data ● Audit appeal/review processes ● Solicit perceived impact of setting standards and expectations among primary decision-makers
Better health services and patient safety via enhanced regulation and oversight	<ul style="list-style-type: none"> ● Better decisions by primary health decision-makers ● Respect for the tribunal's oversight function 	<ul style="list-style-type: none"> ● Number of cases in which primary decisions appropriately exercised their statutory authority and did not exceed it ● Perceived concern or fear among primary health decision-makers that their decisions will be reversed 	<ul style="list-style-type: none"> ● Audit a sample of decisions ● Survey, interview or focus groups to uncover perception of tribunals among primary decision-makers and the consequences of reversed decisions

Note: The *intervention* to be examined in this evaluation matrix is the operation and services of Ontario's health-related adjudicative tribunals as part of the health system. The *population* under study is assumed to be Ontario residents in general and the users of the health-related adjudicative tribunals specifically, and the *system goal* is to promote better health among Ontarians through a strong health system that provides excellent health services. The *counterfactual* in this scenario would be the situation if Ontario's health system did not have any adjudicative tribunals. The *primary target audience* for this evaluation is assumed to be government officials.

Such benchmarks can be drawn from thoughtful consideration, aspirational goals of leaders, expert judgments on what is possible, data from similar tribunals in other jurisdictions (i.e. comparative analysis) or previous empirical observations from the same tribunal (i.e. interrupted time-series analysis). For experimental methods like randomized controlled trials that are rarer in socio-legal studies, the control group would serve as the comparative benchmark rather than any observational data that are external to the evaluation. Such comparisons are naturally better because they more accurately represent the counterfactual of what the situation would be like without the tribunal and can help lead to causal inferences.

However, overcoming the identified challenges and systematizing empirical impact evaluations of health-related adjudicative tribunals across time and jurisdictions require action from stakeholders throughout the health and legal systems. For example, scholars in the health law field must intensify their efforts to build capacity for utilizing empirical methodologies, enhance the status of such work within legal circles and overcome any real or perceived problems with maintaining independence. Health system institutions should also start to build policy-relevant databases that are rigorously compiled, comprehensive and publicly accessible. Finally, health planners and research funders must facilitate (or even catalyze) the continuous improvement of adjudicative tribunals by supporting undertakings to empirically evaluate their impact on society. Initial funding for small-scale evaluations and/or pilot projects would be particularly helpful, as would support for disseminating any lessons learned as widely as possible.

Syntheses of research evidence may be helpful in encouraging stakeholders to support empirical impact evaluations of health-related adjudicative tribunals, especially because they are likely to highlight the current dearth of knowledge in this area. A systematic review on the effect of adjudicative tribunals in the health sector, for example, would be a disciplined and rigorous approach to assess the current state of research evidence in this area and tracking developments in it over time (Lavis *et al.*, 2004; Petticrew and Roberts, 2008). This tool applies the scientific method to gathering, appraising and synthesizing what is known (and what is not known) on a particular topic such that publication and selection bias are limited (Lavis *et al.*, 2005; Rothstein *et al.*, 2005). Such a review has already been conducted, for example, to assess the impact of drug courts on criminal recidivism (Wilson *et al.*, 2006), and a protocol has been developed to evaluate the influence of these specialist courts on narcotics use in particular and criminal activity more broadly (Wilson *et al.*, 2007). User-friendly scoping reviews may also be extremely helpful for pushing this agenda forward in various contexts.

Conclusion

In summary, adjudicative tribunals serve an essential function within the health sector, yet their contributions and impact on the delivery of health services and

society in general are not usually evaluated empirically. The focus of past tribunal assessments on their internal operations to the near-total exclusion of their external impact limit these bodies' ability to inform continuous quality improvement efforts, enhance the public's confidence in them and maximize their societal impact. Three challenges, however, serve to complicate empirical impact evaluations of health-related adjudicative tribunals. First, the complexity of the health system and its countless independent actors prevents their mechanistic isolation, which is necessary to elegantly find causal inferences between them and their goals. Second, the indirect relationship between tribunal services and their existential purposes – exacerbated by uncertain objectives, difficulties with randomization and a lack of examples to follow – presents methodological barriers that cannot be easily overcome. Third, several realities of the legal profession and the environment in which tribunals currently operate further hinder evaluation efforts, including the dearth of empirical capacity, culture of deference to authority and focus on process and independence.

There are, however, two main reasons for optimism. Empirical evaluations of similar judicial bodies have been previously conducted and there is currently a rapid expansion of interest in empirical health law scholarship. This analysis of challenges to empirically evaluate the impact of adjudicative tribunals in the health sector highlights several potential ways to help move this agenda forward. Individual evaluators, for example, can assemble interdisciplinary teams, identify their tribunal's overall goal, develop surrogate endpoints and conduct a realistic evaluation that tracks each of them. Stakeholders within the health and legal systems, on the other hand, can support individual efforts by earmarking funds for such empirical impact evaluations, by building policy-relevant databases and by assisting with cross-jurisdictional learning and dissemination efforts. Syntheses of the research evidence on this topic, and systematic reviews in particular, may be helpful for highlighting the absence of knowledge in this area and for building support to capitalize on this otherwise missed opportunity.

Nevertheless, a preliminary question remains as to whether it is even the responsibility of adjudicative tribunals like Ontario's HPARB and HSARB to be empirically evaluating their own impact or to help others in doing so. Besides the limitations imposed upon them by their respective statutory mandates, these bodies and others may not necessarily be concerned about the impact of their decisions and could in fact be preoccupied with maintaining their independence. Further consideration must be given to these issues, and others, so that continuous quality improvement and self-evaluation can become part of tribunals' core mandates as has been widely promoted for all government undertakings (Mayne, 2001); otherwise, it may be impossible for these bodies to provide the best services possible to their users, stakeholders and larger constituencies. Any ambitions for self-improvement among health-related adjudicative tribunals, however, must obviously be balanced with the legislative, political and social realities within which they operate.

Acknowledgements

The ideas shared in this paper were first presented at the Workshop on the Impact of Health Tribunals that was held at the University of Toronto's Faculty of Law on 28 May 2009 and funded by the Canadian Institutes of Health Research. We are grateful to the attendees of this workshop for their insights and feedback, especially Louise Bélanger-Hardy, Ron Ellis, Irwin Fefergrad, Colleen Flood, Chris Gill, Joan Gilmour, Grant Huscroft, Laverne Jacobs, Linda Pearson, Gabrielle St. Hilaire, Mariana Valverde, Kevin Whitaker and Albert Yoon. Additional helpful suggestions were offered by Michelle Mello and Martin Partington. An earlier version of this paper was also presented at the Canadian Institute for the Administration of Justice's Annual Conference held in Ottawa from 3 September to 2 October 2009 and published in a monograph summarizing that event. At the time that this study was conducted, Steven J. Hoffman was employed by the World Health Organization and Lorne Sossin was (and remains) a Vice-Chair of the Health Professions Appeals and Review Board and the Health Services Appeal and Review Board. The opinions expressed in this paper are those of the individual authors writing in their academic capacities and do not necessarily represent the views of their affiliated institutions.

References

- Adler, M. and J. Gulland (2003), *Tribunal Users' Experiences, Perceptions and Expectations: A Literature Review*, London: Council on Tribunals, [http://www.council-on-tribunals.gov.uk/docs/other_adler\(2\).pdf](http://www.council-on-tribunals.gov.uk/docs/other_adler(2).pdf) [2 February 2009].
- Allsop, J. and K. Jones (2008), 'Withering the citizen managing the consumer: complaints in healthcare settings', *Social Policy and Society*, 7(2): 233–243.
- Aston, J., D. Hill and N. D. Tackey (2006), *The Experience of Claimants in Race Discrimination Employment Tribunal Cases. Employment Relations Research Series, No. 55*, London: UK Department of Trade and Industry, <http://www.dti.gov.uk/files/file27818.pdf> [12 March 2009].
- Blumenthal, S. and S. Wessely (1993), *The Pattern of Delays in Mental Health Review Tribunals*, London: Stationery Office.
- Blumenthal, S. and S. Wessely (1994), 'The pattern of delays in mental health review tribunals', *Psychiatric Bulletin*, 18(7): 398–400, <http://pb.rcpsych.org/cgi/reprint/18/7/398.pdf> [12 March 2009].
- Bradley, C., M. Marshall and D. Gath (1995), 'Why do so few patients appeal against detention under the mental health act?', *British Medical Journal*, 310: 364–367, <http://www.bmj.com/cgi/content/full/310/6976/364> [12 March 2009].
- Campbell, N. C., E. Murray, J. Darbyshire, J. Emery, A. Farmer, F. Griffiths, B. Guthrie, H. Lester, P. Wilson and A. L. Kinmonth (2007), 'Designing and evaluating complex interventions to improve health care', *British Medical Journal*, 334: 455–459.
- Cane, P. (2004), 'Understanding Judicial Review and its Impact', in M. Hertogh and S. Halliday (eds), *Judicial Review and Bureaucratic Impact: International and Interdisciplinary Perspectives*, Cambridge, UK: Cambridge University Press, 15–42.

- Carscallen, M. P., W. K. Gray and J. G. Pink (2003), *Regulatory Burden Task Force: Report to the Ontario Securities Commission*, Toronto: Ontario Securities Commission, http://www.osc.gov.on.ca/About/Governance/Accountability/ga_20031212_rbtfrpt.pdf [12 March 2009].
- Center for Global Development's Evaluation Gap Working Group (2006), *When Will We Ever Learn? Improving Lives Through Impact Evaluation*, Washington, DC: Center for Global Development, http://www.cgdev.org/files/7973_file_WillWeEverLearn.pdf [22 June 2009].
- Chalkidou, K., S. Tunis, R. Lopert, L. Rochaix, P. T. Sawicki, M. Nasser and B. Xerri (2009), 'Comparative effectiveness research and evidence-based health policy: experience from four countries', *Milbank Quarterly*, 87(2): 339–367, http://www.milbank.org/quarterly/milq_87_2-final-chalkidou.pdf [5 September 2009].
- Confederation of British Industry (2005), *A Matter of Confidence: Restoring Faith in Employment Tribunals*, London: Confederation of British Industry, [http://www.cbi.org.uk/ndbs/Press.nsf/0/33f9830ed75f765b8025708800523621/\\$FILE/Tribunals Brief—CBI.pdf](http://www.cbi.org.uk/ndbs/Press.nsf/0/33f9830ed75f765b8025708800523621/$FILE/Tribunals%20Brief-CBI.pdf) [12 March 2009].
- Corsi, J. R. and T. L. Hurley (1979), 'Pilot study report on the use of the telephone in administrative hearings', *Administrative Law Review*, 31: 485.
- Dolan, M., R. Gibb and P. Coorey (1999), 'Mental health review tribunals: a survey of special hospital patients' opinions', *Journal of Forensic Psychiatry*, 10: 264–275.
- Ellis, R. (2009), 'Executive Branch Justice: Canada's 'Official Courts'', PhD Dissertation. Toronto: York University.
- Employment Tribunals Service (2005), *Employment Tribunals Service User Survey 2005*, London: Employment Tribunals Service.
- Evidence-Based Medicine Working Group (1992), 'Evidence-based medicine: a new approach to teaching the practice of medicine', *Journal of the American Medical Association*, 268: 2420–2425.
- Genn, H., M. Partington and S. Wheeler (2006a), *Nuffield Inquiry on Empirical Legal Research: Law in the Real World: Improving our Understanding of How Law Works: Final Report and Recommendations*, London: Nuffield Foundation, http://www.ucl.ac.uk/laws/socio-legal/empirical/docs/inquiry_report.pdf [27 April 2009].
- Genn, H., B. Lever and L. Gray, National Centre for Social Research (2006b), *Tribunals for Diverse Users*, London: UK Department for Constitutional Affairs, http://www.dca.gov.uk/research/2006/01_2006.pdf [29 March 2009].
- George, T. E. (2006), 'An empirical study of empirical legal scholarship: the top law schools', *Indiana Law Journal*, 81(1): 141–161, http://www.law.indiana.edu/ilj/volumes/v81/no1/8_George.pdf [15 January 2009].
- GRADE Working Group (2004), 'Grading quality of evidence and strength of recommendations', *British Medical Journal*, 328: 1490, doi:10.1136/bmj.328.7454.1490. <http://www.bmj.com/cgi/content/full/328/7454/1490> [10 April 2009].
- Green, B. L., C. J. Furrer, S. D. Worcel, S. W. M. Burrus and M. W. Finigan (2009), 'Building the evidence base for family drug treatment courts: results from recent outcome studies', *Drug Court Review*, 6(2): 53–82, <http://www.ndci.org/sites/default/files/ncdc/DCR%2C%20Vol.%206%2C%20No.%202.pdf> [23 August 2009].
- Gresenz, C. R. and D. M. Studdert (2005), 'External review of coverage denials by managed care organizations in California', *Journal of Empirical Legal Studies*, 2(3): 449–468.
- Guzzo, G., J. Baird, B. Grimmett, G. Martiniuk and J. Flaherty (1998), *Everyday Justice: Report of the Agency Reform Commission on Ontario's Regulatory and Adjudicative*

- Agencies, Toronto: Government of Ontario, http://www.ccat-ctac.org/downloads/1998_Guzzo-report.pdf [accessed on 12 March 2009].
- Hayward, B., M. Peters, N. Rousseau and K. Seeds (2004), *Findings from the Survey of Employment Tribunal Applications 2003*. *Employment Relations Research Series*, No. 33, London: UK Department of Trade and Industry, <http://www.dti.gov.uk/files/file11455.pdf> [12 March 2009].
- Health Professions Appeal and Review Board (2008), *Recommendations to Health Professions Regulatory Advisory Council*, Toronto: Health Professions Appeal and Review Board, <http://www.hprac.org/en/projects/resources/hprac-1457May30HPARB.pdf> [28 March 2009].
- Health Professions Regulatory Advisory Council (2008), *HPRAC: Annual Report April 1, 2007–March 31, 2008*, Toronto: Health Professions Regulatory Advisory Council, <http://www.hprac.org/en/reports/resources/HPRACAnnualReport2007-2008.pdf> [22 June 2009].
- Hertogh, M. and S. Halliday (eds) (2004), *Judicial Review and Bureaucratic Impact: International and Interdisciplinary Perspectives*, Cambridge: Cambridge University Press.
- Hoffman, S. J., J. N. Lavis and S. Bennett (2009), 'The use of research evidence in two international organizations' recommendations about health systems', *Healthcare Policy*, 5(1): 66–86.
- Jacobs, L. (2008), *Reconciling Tribunal Independence and Expertise – Empirical Observations*. Presented at The Future of Administrative Justice Symposium, University of Toronto, 18 January 2008, http://www.law.utoronto.ca/documents/conferences/adminjustice08_Jacobs.pdf [29 November 2008].
- Jones, N., H. Jones, L. Steer and A. Datta (2009), *Working Paper 300: Improving Impact Evaluation Production and Use*, London: Overseas Development Institute, <http://www.odi.org.uk/resources/download/3177.pdf> [22 June 2009].
- Latreille, P. L., J. A. Latreille and K. G. Knight (2004), *Findings from the 1998 Survey of Representatives in Employment Tribunal Cases*. *Employment Relations Research Series*, No. 35, London: UK Department of Trade and Industry, <http://www.berr.gov.uk/files/file11453.pdf> [12 March 2009].
- Latreille, P. L., J. A. Latreille and K. G. Knight (2005), 'Making a difference? Legal representation in employment tribunal cases: evidence from a survey of representatives', *Industrial Law Journal*, 34(4): 308–330.
- Lavis, J. N., F. Becerra-Posada, A. Haines and E. Osei (2004), 'Use of research to inform public policymaking', *The Lancet*, 364(9445): 1615–1621.
- Lavis, J. N., H. T. O. Davies, A. Oxman, J. L. Densi, K. Golden-Biddle and E. Ferlie (2005), 'Towards systematic reviews that inform healthcare management and policymaking', *Journal of Health Services Research and Policy*, 10(Supplement 1): 35–48.
- Leggatt, A. (2001), *Tribunals for Users: One System, One Service*, London: UK Department for Constitutional Affairs, <http://www.tribunals-review.org.uk/index.htm> [12 March 2009].
- Marin, A. (2008), *Ombudsman Ontario: Annual Report 2007–2008*, Toronto: Office of the Ombudsman, http://www.ombudsman.on.ca/media/18971/ar08_eng.pdf [22 June 2009].
- Maynard, A. (2006), 'The "long and lonely road": translating evidence into policy', *Health Economics, Policy and Law*, 1(4): 415–421, doi:10.1017/S1744133106006049.
- Mayne, J. (2001), 'Addressing attribution through contribution analysis: using performance measures sensibly', *Canadian Journal of Program Evaluation*, 16(1): 1–24.

- McCarter, J. (2008), *Annual Report 2008: Office of the Auditor-General*, Toronto: Queen's Printer for Ontario, http://www.auditor.on.ca/en/reports_en/en08/ar_en08.pdf [22 June 2009].
- McRuer, J. C. (1968), *Ontario Royal Commission Inquiry Into Civil Rights: Report No. 1*, Toronto: Queen's Printer.
- Mello, M. M. and K. Zeiler (2008), 'Empirical health law scholarship: the state of the field', *Georgetown Law Journal*, 96: 649–702.
- Mello, M. M., D. M. Studdert, A. Kachalia and T. A. Brennan (2006), "'Health courts" and accountability for patient safety', *Milbank Quarterly*, 84: 459–492.
- Moyer, P. (1997), 'The regulation of corporate law by securities regulators: a comparison of Ontario and the United States', *University of Toronto Faculty of Law Review*, 55(1): 43–76.
- National Audit Office (2003a), *Progress in Improving the Medical Assessment of Incapacity and Disability Benefits: Report by the Comptroller and Auditor General*. HC 1141, Session 2002–2003, 17/10/2003, London: Stationery Office, http://www.nao.org.uk/publications/nao_reports/02-03/02031141.pdf [12 March 2009].
- National Audit Office (2003b), *Getting It Right, Putting It Right: Improving Decision-Making and Appeals in Social Security Benefits*. HC 1142, Session 2002–2003, 07/11/2003, London: Stationery Office, http://www.nao.org.uk/publications/nao_reports/02-03/02031142.pdf [12 March 2009].
- OECD Development Assistance Committee (1998), *Review of the DAC Principles for Evaluation of Development Assistance*, Paris: Organisation for Economic Co-operation and Development, <http://www.oecd.org/dataoecd/63/50/2065863.pdf> [22 June 2009].
- Ontario Committee on the Healing Arts (1970), *Report*, Toronto: Queen's Printer.
- Osborne, C. A., D. J. Mullan and B. Finlay (2004), *Report of the Fairness Committee to David A. Brown, Q.C., Chair of the Ontario Securities Commission*, Toronto: Ontario Securities Commission, http://www.osc.gov.on.ca/Regulation/FiveYearReview/fyr_20040818_fairness-committee.pdf [12 January 2009].
- Partington, M., E. Kirton-Darling and F. McClenaghan (2007), *Empirical Research on Tribunals: An Annotated Review of Research Published between 1992 and 2007*, London: Administrative Justice and Tribunals Council, <http://www.ajtc.gov.uk/docs/EmpiricalResearch.pdf> [9 March 2009].
- Pearson, L. (2007), *The Impact of External Administrative Law Review: Tribunals*. University of New South Wales Faculty of Law Research Series, Paper 53, Sydney: University of New South Wales, <http://law.bepress.com/unswwps/flrps/art53> [10 April 2011].
- Peay, J. (1989), *Tribunals on Trial: A Study of Decision Making Under the Mental Health Act 1983*, Oxford: Clarendon Press.
- Petticrew, M. and H. Roberts (2008), 'Systematic reviews – do they “work” in informing decision-making around health inequalities?', *Health Economics, Policy and Law*, 3(2): 197–211.
- Pitfield, C. and C. M. Flood (2005), 'Section 7 “safety valves”: appealing wait times within a one-tier system', in C. M. Flood, K. Roach and L. Sossin (eds), *Access to Care, Access to Justice: The Legal Debate Over Private Health Insurance in Canada*, Toronto: University of Toronto Press, 477–501.
- Pleasence, P. (2008), 'Trials and tribulations: conducting randomized experiments in a socio-legal setting', *Journal of Law and Society*, 35(1): 8–29.
- Plotnikoff, J. and R. Woolfson (2005), *Review of the Effectiveness of Specialist Courts in Other Jurisdictions*. DCA Report Series, No. 3/05, London: UK Department for Constitutional Affairs, http://www.dca.gov.uk/research/2005/3_2005.pdf [10 April 2009].

- Plsek, P. E. and T. Greenhalgh (2001), 'Complexity science: the challenge of complexity in health care', *British Medical Journal*, **323**: 625–628.
- Rothstein, H., A. Sutton and M. Borenstein (eds) (2005), *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, West Sussex, England: John Wiley & Sons.
- Rousseau, S. (2008), *The Québec Experience with an Independent Administrative Tribunal Specialized in Securities: A Study of the Bureau de décision et de révision en valeurs mobilières*, Ottawa: Expert Panel on Securities Regulation, <http://www.groupeexperts.ca/eng/reports/research-studies/quebec-independent-adjudicative-tribunal-rousseau.php> [12 March 2009].
- Sackett, D. L., W. M. Rosenberg, J. A. Gray, R. B. Haynes and W. S. Richardson (1996), 'Evidence based medicine: what it is and what it isn't', *British Medical Journal*, **312**(7023): 71–72, <http://www.bmj.com/cgi/content/full/312/7023/71> [8 January 2009].
- Siegal, G., M. M. Mello and D. M. Studdert (2008), 'Adjudicating severe birth injury claims in Florida and Virginia: the experience of a landmark experiment in personal injury compensation', *American Journal of Law and Medicine*, **34**: 489–533, http://www.hsph.harvard.edu/faculty/michelle-mello/files/FL-VA_PDF.pdf [2 February 2009].
- Society of Ontario Adjudicators and Regulators (1995), *Towards Maintaining and Improving the Quality of Adjudication: SOAR Recommendations for Performance Management in Ontario's Administrative Justice Tribunals*, Toronto: Society of Ontario Adjudicators and Regulators, http://www.soar.on.ca/soar-perf_man.htm [12 March 2009].
- Sossin, L. (2007), *The Effectiveness and Efficiency of Pension Regulation in Ontario and in Comparative Perspective*, Toronto: Expert Commission on Pensions, <http://www.pensionreview.on.ca/english/summaries/8Sossin.html> [accessed on 29 November 2008].
- Sunkin, M. (2004), 'Conceptual issues in researching the impact of judicial review on government bureaucracies', in M. Hertogh and S. Halliday (eds), *Judicial Review and Bureaucratic Impact: International and Interdisciplinary Perspectives*, Cambridge, UK: Cambridge University Press, 43–75.
- Thomas, R. (2005), 'Evaluating tribunal adjudication: administrative justice and asylum appeals', *Legal Studies*, **25**(3): 462–498.
- Toubman, A. A., T. McArdle and L. Rogers-Tomer (1995), 'Due process implications of telephone hearings: the case for an individualized approach to scheduling telephone hearings', *University of Michigan Journal of Law Reform*, **29**: 407–474.
- United States Government Accountability Office (2005), *Adult Drug Courts: Evidence Indicates Recidivism Reductions and Mixed Results for Other Outcomes*, Washington, DC: United States Government Accountability Office, <http://www.gao.gov/new.items/d05219.pdf> [23 August 2009].
- Urwin, P., V. Karuk, P. Latreille, E. Michielsens, L. Page, B. Siara, S. Speckesser, A. Boon and P. A. Chevalier (2010), *Evaluating the Use of Judicial Mediation in Employment Tribunals. Ministry of Justice Research Series 7/10*, London: Ministry of Justice, <http://www.justice.gov.uk/publications/docs/evaluating-judicial-mediation-march10.pdf> [10 April 2011].
- Wilson, D. B., O. Mitchell and D. L. Mackenzie (2006), 'A systematic review of drug court effects on recidivism', *Journal of Experimental Criminology*, **2**: 459–487, doi:10.1007/s11292-006-9019-4.
- Wilson, D. B., O. Mitchell and D. L. MacKenzie (2007), *Effects of Drug Courts on Criminal Offending and Drug Use: A Campbell Collaboration Systematic Review Protocol*,

- Oslo, Norway: Campbell Library of Systematic Reviews, <http://campbellcollaboration.org/lib/download/367> [10 April 2011].
- World Bank (2006), *Conducting Quality Impact Evaluations under Budget, Time and Data Constraints*, Washington, DC: World Bank, [http://lnweb90.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/757A5CC0BAE22558852571770059D89C/\\$file/conduct_qual_impact.pdf](http://lnweb90.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/757A5CC0BAE22558852571770059D89C/$file/conduct_qual_impact.pdf) [22 June 2009].
- World Health Organization (2004), *World Report on Knowledge for Better Health*, Geneva: World Health Organization, http://www.who.int/rpc/meetings/world_report_on_knowledge_for_better_health.pdf [15 April 2009].
- Zimmerman, B., P. Plsek and C. Lindberg (1998), *Edgeware: Insights from Complexity Science for Health Care Leaders*, Irving, TX: VHA Inc.

Online Appendices

Appendix A: Twenty-Seven Examples of Evaluations of Adjudicative Tribunals.
Appendix B: Ninety-One Examples of External Impact Evaluations of Specialist Courts.

Available at <http://www.journals.cambridge.org/hep>.