Editorial

Dear readers,

This quarter's edition of *Health Law in Canada* explores what seem like two very divergent topics: naturopathy and artificial intelligence ("AI").

In our first article, the author notes that over the past 15 years, across parts of Canada, the health profession of naturopathy was regulated, either via new recognition, or via integration into existing legislation. The effect of this recognition is reviewed from the perspective of the naturopaths themselves; their demographics, their relationships with their regulators and professional advocacy associations, and their perceptions of how the medical profession, the media, and most Canadians interact with and view the profession. The authors argue that policy development and legislative evolution will be informed by this information, which has not before been collected.

In our second article the authors offer recommendations for reforming tort law to ensure that Canadian health care providers can meaningfully integrate into their practices, AI clinical decision support tools that are intended to transform and improve patient outcomes.

Is there any connection between naturopathy and machine learning? Will machine-generated models based on data derived from human subjects replace human judgment? Will these AI tools yield better patient outcomes than human decision-making? The near future will tell to what extent humans might be replaced but the bigger question, and the more important question, is how humans can stay relevant when AI-based tools may be much more efficient at generating better patient outcomes.

The COVID-19 pandemic has demonstrated that virtual connections are helpful for maintaining social stability when physical or in-person connections are impossible. However, we have also learned, collectively, that telephone, video and other communication methods are not an adequate substitute for human, in-person interactions. Perhaps the most important connection one can find

between the regulation of naturopathy and AI tools that support clinical decision-making is a human one, rooted in improving patient care.

Thank you for your continued interest in *Health Law in Canada*, where we explore the nexus between

health policy, and law. Please continue to submit thought-provoking articles and suggestions for policy and legislative reform.

Editorially yours,
Simmie Palter
Deputy Editor-in-Chief

A Survey of Regulated Naturopathic Doctors in Canada: Demographics, Governance, and Public Understanding

Dave Snow

Abstract

Naturopathic medicine is now regulated or semi-regulated in six of Canada's seven most populous provinces, yet there has been minimal research on the beliefs and attitudes of naturopathic doctors ("NDs"). This multidisciplinary paper begins with a systematic review of the laws governing naturopathic medicine in Canada's six regulated provinces. It then examines the results from an original dataset based on a 2019 survey of Canadian NDs in the six provinces with some level of regulation. NDs were asked questions about demographics, governance and representation, and the public understanding of naturopathic medicine. Demographically, most respondents were young, female, relatively new to practice, and science-educated prior to entering their naturopathic medical program. In terms of governance, most respondents believe recent regulatory changes have been positive, especially for their patients, though Ontario respondents were the most critical. Likewise, most respondents expressed positive attitudes about their national and provincial promotional organizations, and satisfaction was strongly associated with membership. However, respondents did not believe naturopathic medicine is understood by the Canadian public, medical doctors, and especially the media. For all the integration of naturopathic medicine into provincial legislation governing health professions over the past 15 years, Canadian NDs still perceive that their profession is poorly understood. As naturopathic medicine has become more professionalized across Canada and globally, future health policy and legal researchers should focus on how naturopathic medicine is viewed by patients and other medical professionals within the Canadian health care system.

Introduction

The World Health Organization describes traditional and complementary medicine ("TCM") as "an important and often underestimated health resource".[1] TCM typically refers to health care products, practices, and practitioners that "are not fully integrated into the dominant health care system" and are not considered part of conventional medicine within a given country.[2] While there is debate regarding the extent to which TCM should be integrated into public health care systems,[3] there is lit-

tle doubt that its use is growing worldwide. Among TCM professions, naturopathic medicine (also known as naturopathy) has undergone considerable professionalization in the 21st century, especially in Canada. Naturopathic medicine is now regulated or semi-regulated in six of Canada's 10 provinces, with regulatory changes over the last 15 years further integrating the profession into provincial policy structures governing health professions. While some have criticized these regulatory changes for potentially legitimizing unsafe practices,[4] there has been limited empirical research on the beliefs and

attitudes of Canadian naturopathic doctors ("NDs") since those policy changes. Existing scholarly surveys of Canadian NDs have either predated regulatory changes or focused on aspects other than regulation.[5]

This paper's objective is to understand how the regulation of naturopathic medicine is perceived by Canadian NDs themselves. After conducting a systematic review of all laws and policies in Canada's regulated provinces, it examines the results from a survey of regulated Canadian NDs conducted in 2019. It seeks to answer three questions regarding demographics, naturopathic organizations and regulation, and the public understanding of naturopathic medicine: What are the demographic and educational characteristics of Canadian NDs? How do they perceive the way they are governed and represented? Finally, how well do they believe their profession is understood?

Demographically, the 426 ND respondents were primarily young, female, and science-educated prior to entering their naturopathic medical program. In terms of governance and representation, most respondents believe regulation has been positive, especially for naturopathic patients. Respondents who are not regulated want to be regulated, and those who are currently regulated under their provincial health professions framework support that framework. Respondents were also highly satisfied with their promotional associations, both at the provincial and national level. However, respondents displayed a strong belief that naturopathic medicine is not well understood by the Canadian public, medical doctors, and especially the Canadian media.

This study offers several empirical and theoretical contributions for better understanding the role of naturopathic medicine in Canadian and international health care systems. Empirically, it is the first scholarly study to systematically review and analyse the laws and policies concerning naturopathic medicine in the Canada's six regulated provinces,

to determine how many regulated NDs are practicing in Canada, to isolate naturopathic attitudes toward regulation after that regulation occurred, and to explore attitudes from Canadian NDs from outside the province of Ontario. Theoretically, the survey data contributes to understanding the relationship between policy design, the implementation of health policies regulating TCM, and social perceptions of health care professions. Although the survey data show respondents were generally satisfied with regulation, NDs' perception that they are not well understood by medical doctors and the media provides further evidence that naturopathic medicine ought to be understood as a "repressed structural interest" in the Canadian health care system, existing outside the public health care system in perception and in practice.[6] Future scholarship in public policy, bioethics, and the health sciences should explore how dominant actors in the health care system, namely medical doctors, perceive naturopathic medicine and interact with naturopathic doctors.

This paper unfolds as follows. First, I define the scope and regulation of naturopathic medicine in Canada and distinguish between the various forms of legal and associational governance in the six provinces with some level of regulation. After a brief explanation of the survey methods, I discuss the survey results, in particular responses regarding demographics, promotional associations, regulation, and the public understanding of naturopathic medicine. I then discuss the paper's theoretical and empirical contributions, before concluding with an exploration of the future directions for scholarship pertaining to naturopathic medicine in Canada and abroad.

Defining the Scope and Regulation of Naturopathic Medicine in Canada

The Canadian Association of Naturopathic Doctors defines naturopathic medicine as "a distinct primary health care system that blends modern scientific

knowledge with traditional and natural forms of medicine".[7] According to Bradley, *et al*.,[8] naturopathic medicine shares a foundation with traditional western medicine in terms of biomedical physiology and diagnostics. However, it de-emphasizes prescription drugs and surgical interventions and emphasizes preventative techniques, health promotion, physical activity, herbal medicine, and homeopathy—the latter of which is especially controversial, including among many NDs.[9] Naturopathic medicine is defined by a set of six guiding principles: first do no harm; the healing power of nature; identify and treat the causes; doctor as teacher; treat the whole person; and prevention.[10] Although the terms "naturopathy" and "naturopathic medicine" are typically used interchangeably in Canada, this paper uses the latter term for clarity, as it is most commonly used by the naturopathic organizations being discussed.

In Canada and the United States, naturopathic medical programs must be accredited by the Association of Accredited Naturopathic Medical Colleges. All Canadian and American naturopathic medical programs must fulfil requirements set by the Council on Naturopathic Medical Education, an accrediting body. There are five accredited naturopathic medical programs in the United States and two in Canada, although in 2020, the two Canadian programs (the British Columbia-based Boucher Institute of Naturopathic Medicine and the Toronto-based Canadian College of Naturopathic Medicine) announced a merger that would maintain both campuses.[11] Prior to entry into the Canadian programs, naturopathic students must have completed a three- or four-year undergraduate university degree, including prerequisite courses in biology, chemistry, and psychology.[12]

The Canadian Association of Naturopathic Doctors ("CAND") is the national organization that represents Canadian NDs. However, because health care is primarily set by provincial governments, naturopathic scope of practice is determined by a combination of provincial statutes, provincial regulations, and self-regulatory bylaws created by provincial naturopathic organizations. Among the six provinces with some regulation for naturopathic medicine, provincial policies vary: the three more populous provinces of British Columbia, Ontario, and Alberta have separate promotional associations and self-regulatory colleges (the "college model"), with NDs regulated under the same provincial legislation that governs other health professions; the less populous provinces of Manitoba and Saskatchewan each have a single, self-regulatory body, though both passed laws that will move the profession to the college model; and Nova Scotia, a small province, only has a promotional association but grants title protection to members who are licensed by one of the five regulated provinces. All six provinces grant title protection for terms such as "naturopath" and "naturopathic doctor" to regulated/licensed members only. There are naturopaths who practice in the other four Canadian provinces, but naturopathic medicine is effectively unregulated in those provinces and there is no title protection.

Each of the six provinces with some level of regulation has made recent changes to their naturopathic regulatory structures, with a trend towards the college model (it should be noted that a regulatory "college" is not an educational institution, but instead is the name for the provincial self-regulatory organization that governs an individual health profession). In 2007, only British Columbia had a regulatory college; once the Saskatchewan and Manitoba legislation comes into force, there will be five provinces with a regulatory college. However, scholars have yet to explore how regulation has affected naturopathic practice, nor how NDs perceive their regulatory structure. When Ontario NDs were surveyed prior to that province's move to the college model, most NDs were supportive of forthcoming regulations, though some were worried about increased financial burdens, a diminishing of scope of practice, and a shift towards a more con-

ventional biomedical model of care.[9] Post-regulation surveys of Canadian NDs have focused on integration with medical doctors,[13] naturopathic research,[14] paediatric practice[15] and cancer care.[16] However, little is known about NDs' views toward regulation, policy implementation, naturopathic organizations, and the public understanding of naturopathic medicine.

Methods

Prior to this study, the precise number of regulated NDs in Canada was not known; although CAND claims over 2,400 members, this number includes naturopaths practicing in unregulated provinces without title protection and does not include non-CAND members (membership is optional in many provinces). To determine how many regulated NDs were currently practicing in the six regulated or semi-regulated provinces, a database of active NDs in British Columbia, Alberta, Saskatchewan, Manitoba, Ontario, and Nova Scotia was created in Summer 2019. NDs' names were first retrieved from the official online directories of CAND, provincial regulators, and provincial associations, and this list was supplemented using contact information from publicly available websites. After removing those who had retired, passed away, or had duplicate entries, the number of active NDs in the six provinces was determined to be 2,287 as of August 31, 2019.

Survey questions were drafted and sent to every naturopathic association and provincial regulator for feedback. Organizations were then re-notified one week in advance of the survey, a link for which was distributed to NDs via email using Qualtrics XM on October 29, 2019. With one week remaining, NDs were emailed a reminder, and the survey closed on November 26, 2019. After excluding those who could not be contacted electronically, in total 2,248 NDs were emailed the survey. Because the survey was voluntary and only available to those for whom an individual or clinic email address was available, the possibility of volunteer bias exists. The survey received ethics approval from the University of Guelph Research Ethics Board (REB #18-08-022) and was conducted in accordance with the Canadian Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans. Informed consent was obtained from all human participants and responses were anonymous.

The survey was divided into three sections. The first section asked demographic and educational questions; the second section focused on naturopathic organizations and regulation; and the third asked respondents about the public understanding of naturopathic medicine. Depending on their selected primary province of practice, respondents were asked between 31-35 questions, which included open-ended text boxes, multiple choice, and a five-point Likert scale from "strongly agree" to "strongly disagree". Apart from selecting their primary province, no questions were mandatory, and respondents were given the option of selecting "NA / cannot answer" to all non-text box questions; those responses are excluded from the tables below. Once the survey was complete, data was analyzed quantitatively using SPSS Statistics software. Qualitative content from the text responses was analyzed and coded using Braun and Clarke's thematic analysis, a method for identifying and reporting patterns within a dataset.[17] The analysis below focuses primarily on quantitative results from the multiple choice and Likert questions, though some qualitative text responses are discussed briefly.

Results

In total, 426 NDs completed the survey, a 19.0% response rate. Most respondents (83%) practiced in Ontario and British Columbia, the two most populous regulated provinces. Table 1 shows the sample is broadly representative of the population of NDs in the six provinces, although respondents from Ontario and Manitoba are slightly underrepresented while those from British Columbia, Alberta, Saskatchewan, and Nova Scotia are slightly overrepresented.

Table 1 Active NDs vs. Respondents							
	BC	**AB**	**ON**	**SK**	**MB**	**NS**	**Total**
Active NDs (% of total)	601 (26.3%)	199 (8.7%)	1354 (59.2%)	39 (1.7%)	33 (1.4%)	61 (2.7%)	2287 (100%)
Respondents (% of total)	131 (30.8%)	39 (9.2%)	223 (52.3%)	9 (2.1%)	5 (1.2%)	19 (4.5%)	426 (100%)
Net % difference	+4.5%	+0.5%	-6.9%	+0.4%	-0.2%	+1.8%	

Respondents were asked about their education before, during, and after their naturopathic medical program. Most respondents (95.3%, n=406) had completed at least an undergraduate degree prior to their naturopathic education. Nearly two-thirds of respondents (65.3%) had a Bachelor of Science prior to entering, with a Bachelor of Arts (15.3%) and Kinesiology (10.3%) the next-most common degree types. Five respondents (1.2%) had obtained a medical doctorate (MD) outside of Canada before entering their naturopathic medical program, and 15 respondents (3.5%) had completed or were in the process of completing a master's degree after their naturopathic education. Most respondents (90.1%) completed their naturopathic education at a Canadian naturopathic medical program, 71.8% from the Canadian College of Naturopathic Medicine and 18.3% from the Boucher Institute of Naturopathic Medicine. The remaining 9.9% of respondents completed their education at a naturopathic program in the United States.

In terms of gender, 78.5% of respondents identified as female and 21.5% as male; no respondents chose another gender identity. Respondents were young: 58.5% were aged 40 and under, while 87.7% were aged 50 and under. These results were consistent with demographic data collected by naturopathic organizations and other scholars: one 2011-2012 survey of Ontario NDs found 79% of respondents were women, while the 2018 annual report of the College of Naturopaths of Ontario ("CONO") showed 59% of members were aged 40 or under, and 87% were aged 50 or under.[18] This suggests that the sample is broadly representative of the overall ND population despite the possibility for volunteer bias.

Most respondents were also relatively new to practice: 31.4% of respondents had been practicing for fewer than five years, and 57.5% had been practicing for fewer than 10 years. Only 11% of respondents had been practicing for 20 or more years. Regional variation for age, gender, and number of years practicing was minimal, with British Columbian respondents slightly older. When cross-tabulating demographic characteristics with the questions described below, there was either no relationship or a very weak relationship. In short, respondents' age, gender, and years of practice did not have a substantive effect on their views about naturopathic organizations, regulation, and the public understanding of naturopathic medicine.

Promotional Associations

Respondents were asked about membership in provincial and national promotional associations. Every respondent in Nova Scotia (where there is a provincial promotional association, but no regulator) and Saskatchewan and Manitoba (where there is a provincial regulator but no promotional association) held membership in their lone provincial organization, which is a provincial requirement to practice. Every respondent from these three provinces (33 total) were also CAND members.

Of the three more populous provinces of Ontario, British Columbia, and Alberta where associational membership is optional, there was some variation. Associational membership was highest in British Co-

lumbia (93.9% in the British Columbia Naturopathic Association and 94.6% in CAND), with Ontario's slightly lower (75.3% in the Ontario Association of Naturopathic Doctors and 76.7% in CAND). The most interesting outlier was Alberta. Every Albertan respondent (39/39) held CAND membership even though it is not mandatory, yet only 43.6% (n=17) held membership in the Alberta Association of Naturopathic Doctors ("AAND"), the provincial association that was only formed in 2018. The recency of the creation of the new provincial organization likely explains the low membership rate. When it came to non-membership, by far the most common reason was cost. Of those who provided text responses, 70.3% (45/64) of respondents cited cost as a reason for not joining CAND, and 74.7% (65/87) cited cost as a reason for not joining their provincial association.

Table 2 describes respondents' attitudes towards their national and provincial promotional associations. After being given a quote from the provincial association's website, respondents from the four provinces with a promotional association (British Columbia, Alberta, Ontario, and Nova Scotia) were asked whether they agreed that their provincial association "is doing a good job promoting the naturopathic profession". Respondents were generally favourable: 65.2% somewhat or strongly agreed that their provincial association was doing a good job with promotion, compared with 25.1% who somewhat or strongly disagreed (see Table 2). Respondents from British Columbia (83.2%) and Nova Scotia (84.2%) were the most likely to agree, with respondents from Ontario (57.7%) less likely, but still more likely to agree than disagree. Albertan respondents had an equal number of respondents agree (38.5%, n=15) as those who neither agreed nor disagreed (38.5%, compared with 6.7% in the other three provinces), which likely reflects uncertainty over (and low membership in) Alberta's new provincial association. There was a strong relationship (Cramer's V = .391, p < 0.001) between membership and a belief that the provincial association was doing a good job promoting the profession: 72.8% of members agreed, compared with only 27.1% of non-members. Even in Alberta, where the AAND was new, 58.8% of members agreed the organization was doing a good job promoting the profession, compared with 27.3% of non-members.

Table 2				
Associations: Promotion and Understanding				
	CAND: Good Job Promoting	Provincial Association: Good Job Promoting	CAND Improved Public Understanding	Provincial Association Improved Understanding
Strongly Agree	22.1%	29.9%	17.1%	17.5%
Somewhat Agree	46.2%	35.3%	49.1%	45.9%
Neither Agree nor Disagree	10.1%	9.7%	16.0%	15.8%
Somewhat Disagree	14.8%	16.3%	12.2%	14.8%
Strongly Disagree	6.8%	8.8%	5.6%	6.1%
	n=426	n=411	n=426	n=412

Similar to the above question, respondents were given a quote from the CAND website and then asked whether the national association was doing "a good job promoting the naturopathic profession". Over two-thirds (68.3%) of respondents agreed while 21.6% disagreed. Agreement with this statement ranged from 100% in Manitoba to 57.8% in Ontario, demonstrating an overall high level of support for CAND's promotional activities by province. There was also a strong relationship between membership and agreement that CAND was doing a good job: 72.8% of CAND members somewhat or strongly agreed, versus only 28.8% of non-members (V = .345, p < 0.001).

Respondents were also asked whether their provincial and national associations had "improved public understanding of naturopathic medicine". Again, responses were broadly positive, with some regional variation. For provincial associations, 63.3% of respondents agreed that their association has improved public understanding, and 20.9% disagreed. British Columbian (84%) and Nova Scotian (78.9%) respondents were more likely to agree than those from Ontario (53.8%) and Alberta (41.0%). All four provinces had considerably more respondents agree than disagree, including Ontario (57.7% agree versus 33.3% disagree) and Alberta (38.5% agree versus 23.1% disagree). For CAND, 66.2% of respondents either somewhat or

strongly agreed that the national organization had improved public understanding, while 17.8% somewhat or strongly disagreed.

Regulation and Public Understanding of Naturopathic Medicine

Table 3 describes respondents' responses to questions about regulation generally and regulatory organizations specifically. Respondents were given a brief statement outlining recent or proposed regulatory changes specific to their province, and then asked two questions: whether those regulatory changes had been "positive or negative for *naturopathic patients in [province]*," and whether those regulatory changes had been "positive or negative for *your own naturopathic practice*" (emphasis in survey). The regulatory changes differed from province-to-province: respondents from Alberta and Ontario were asked about new regulatory colleges that had been operating since 2012 and 2015, respectively; respondents in Saskatchewan and Manitoba were asked about the forthcoming creation of regulatory colleges that had been legislated but were not yet operational; respondents in Nova Scotia were asked about a future "robust regulatory framework" as advocated by the Nova Scotia Association of Naturopathic Doctors; and respondents in British Columbia, where a regulatory college has existed for decades, were asked about the 2008 *Naturopathic Physicians Regulation* that, among other things, expanded prescribing authority.

Table 3						
Regulatory Changes						
	Regulatory changes "positive or negative for <u>naturopathic patients</u>"?					
	Existing Changes			**Prospective Changes**		
	BC	**AB**	**ON**	**SK**	**MB**	**NS**	**All Provinces**
Positive	95.4%	71.8%	56.8%	66.7%	50.0%	100.0%	72.5%
Neutral	3.8%	10.3%	29.6%	33.3%	25.0%	0.0%	18.4%

	BC	AB	ON	SK	MB	NS	All Provinces
Negative	0.8%	17.9%	13.6%	0.0%	25.0%	0.0%	9.2%
	n=131	n=39	n=213	n=9	n=4	n=18	n=414
Regulatory changes "positive or negative for <u>your own naturopathic practice</u>"?							
	Existing Changes			Prospective Changes			
	BC	AB	ON	SK	MB	NS	All Provinces
Positive	85.8%	44.4%	33.3%	44.4%	25.0%	73.7%	53.1%
Neutral	11.8%	25.0%	35.3%	55.6%	50.0%	21.1%	26.8%
Negative	2.4%	30.6%	31.4%	0.0%	25.0%	5.3%	20.1%
	n=127	n=36	n=204	n=9	n=4	n=19	n=399

Because the regulatory changes varied by province, caution should be taken when comparing provincial responses. Nevertheless, Table 3 does demonstrate some trends. First, a majority of NDs felt that regulatory changes were (or would be) positive for their naturopathic patients (72.5% positive, 9.2% negative) and their own naturopathic practice (53.1% positive, 20.1% negative). This is true whether the changes had already occurred or were prospective. Respondents were most positive in British Columbia, where the regulatory changes described were the smallest (adjusting scope of practice rather than creating a new framework). In their open-ended text responses regarding the regulatory changes' effect on patients and practice, respondents were asked to give the "most important reason" for their answer. With respect to patients, protection of the public / patient safety was mentioned the most (by 21.6% of respondents who gave an answer, n=88), with other common positive answers including better oversight, quality control, and professional legitimacy.

Interestingly, many respondents believe that the regulatory changes were positive for naturopathic patients, but negative for their own naturopathic practice. This is especially notable in Alberta (71.8% positive for patients, 44.4% positive for own practice) and Ontario (56.8% positive for patients, 33.3% positive for own practice), the two provinces who became regulated under a college model most recently. Ontario respondents were the most negative about the effect on their own practice, with a roughly three-way split between positive, neutral, and negative. The most common reason for a negative impact on practice in the text responses was decreased scope of practice, which was given by 20.9% (n=76) of respondents who answered, including 33.1% (n=61) of Ontario respondents who answered. Several respondents from Ontario specifically mentioned losing the ability to use specific therapies pertaining to biopuncture, prolotherapy, mesotherapy, and the injection of platelet-rich plasma.

Table 4					
Regulatory Organizations					
Regulator "doing a good job protecting the public interest"?					
	BC	**AB**	**ON**	**SK**	**MB**	**Total**
Strongly Agree	67.2%	48.7%	42.6%	77.8%	60%	**52.1%**
Somewhat Agree	21.4%	38.5%	39.0%	22.2%	40%	**32.9%**
Neither Agree nor Disagree	5.3%	0%	12.1%	0%	0%	**8.4%**
Somewhat Disagree	4.6%	2.6%	4.9%	0%	0%	**4.4%**
Strongly Disagree	1.5%	10.3%	1.3%	0%	0%	**2.2%**
	n=131	n=39	n=223	n=9	n=5	**n=407**
Regulator "improved public understanding of naturopathic medicine"?					
	BC	**AB**	**ON**	**SK**	**MB**	**Total**
Strongly Agree	13.7%	2.6%	4%	0%	0%	**6.9%**
Somewhat Agree	31.3%	35.9%	17.5%	55.6%	20%	**24.6%**
Neither Agree nor Disagree	24.4%	25.6%	29.1%	11.1%	20%	**26.8%**
Somewhat Disagree	22.9%	20.5%	28.3%	22.2%	40%	**25.8%**
Strongly Disagree	7.9%	15.4%	21.1%	11.1%	20%	**16%**
	n=131	n=39	n=223	n=9	n=5	**n=407**

Respondents from the five fully regulated provinces were also asked about their regulatory organizations. They were first given a quote directly from the regulator's online material describing its role in protecting patients and/or the public, and then asked whether they agreed if their regulator was "doing a good job protecting the public interest" and whether the regulator had "improved public understanding of naturopathic medicine." Table 4 shows that respondents consistently felt their regulator improved the public interest: 85% agreed that their regulator was protecting the public interest, compared with only 6.6% who disagreed. This was consistent across the provinces, with respondents from British Columbia the most positive. However, when it came to whether their regulator had improved public understanding of naturopathic medicine, respondents were far less positive. Only 31.7% somewhat or strongly agreed that

their regulator had improved public understanding of naturopathic medicine, compared with 41.8% somewhat or strongly disagreeing. Ontario respondents were the most likely to disagree. It should be noted that improving the public understanding of naturopathic medicine is more traditionally associated with promotional associations rather than regulators. Nevertheless, the negative responses regarding regulators' improvement of the public understanding of naturopathic medicine do stand in contrast to respondents' responses regarding their promotional associations.

After being prompted with a brief description of the provincial policy that regulates the investigation and discipline of NDs, respondents from the five fully regulated provinces were then asked whether they were satisfied with their province's process for investigation and discipline. Respondents could choose from three options: satisfied; dissatisfied because the process was *too strict*; and dissatisfied because the process was *not strict enough*. Most respondents (83.6%) who provided a substantive answer were satisfied with the process; 13.1% were dissatisfied because the process was too strict, while only 3.3% of respondents felt that the process was not strict enough. Notably, a high number of respondents (n=70) chose "NA / cannot answer" to this question.

	"The Canadian news media understands naturopathic medicine"	**"Canadian media portrayals of naturopathic medicine are fair and balanced"**	**"Medical Doctors (MDs) understand naturopathic medicine"**	**"Most Canadians understand naturopathic medicine"**
Strongly Agree	0%	0%	0%	0.2%
Somewhat Agree	2.6%	1.6%	6.3%	11.7%
Neither Agree nor Disagree	4.2%	3.3%	9.6%	20.4%
Somewhat Disagree	29.6%	22.3%	43.9%	51.4%
Strongly Disagree	63.6%	72.8%	40.1%	16.2%
	n=426	n=426	n=426	n=426

Table 5

Public Understanding of Naturopathic Medicine

Finally, respondents were asked whether they agreed or disagreed with whether the following groups "understand naturopathic medicine": the Canadian news media; Medical Doctors ("MDs"); and most Canadians. They were also asked whether they agreed with the statement "Canadian media portrayals of naturopathic medicine are fair and balanced". Table 5 shows that respondents overwhelmingly disagreed that these various groups/institutions understood naturopathic medicine. Over 93% of respondents disagreed that Canadian news media understands naturopathic medicine, while 95.1% disagreed that media portrayals of naturopathic medicine are fair and balanced. Likewise, 84% of respondents disagreed that medical doctors understand naturopathic medicine,

and 67.6% disagreed with the statement that most Canadians understand naturopathic medicine. These results are discussed below.

Discussion

This study's policy review, database, and survey results provide several empirical and theoretical contributions to understanding the current regulation of naturopathic medicine in Canada. The review of provincial policies governing naturopathic medicine shows that six provinces are either fully or semi-regulated, and that there has been a movement to the college model over the last two decades. The database of naturopathic doctors shows that, as of August 2019, there were approximately 2,287 regulated NDs practicing in Canada. The survey results demonstrate several demographic trends; notably, most Canadian NDs are young and female, and nearly two-thirds of respondents were educated with a Bachelor of Science before they entered their naturopathic medical program.

The survey results also provide the first glimpse into how NDs view naturopathic organizations and regulation more generally. Respondents were quite positive about their promotional associations: approximately two-thirds agreed that their national organization was doing a good job promoting naturopathic medicine (68.3% agree, 21.6% disagree) and had improved public understanding (66.2% agree, 17.8% disagree) of the profession. Numbers were similar for provincial associations, with British Columbian and Nova Scotian respondents the most positive. Notably, Ontario respondents were the most likely to disagree that all three organizations—CAND, the Ontario Association of Naturopathic Doctors, and the provincial regulator, CONO—had improved public understanding, suggesting more generalized organizational dissatisfaction among NDs practicing in Canada's most populous province.

Overall, respondents were positive about their regulatory framework: NDs currently regulated under a

college model (British Columbia, Ontario, and Alberta) support that model; NDs whose province is moving to a college model (Saskatchewan and Manitoba) support that move; and NDs whose province is not fully regulated (Nova Scotia) support a "robust regulatory framework" for their province. Majorities agreed that regulatory changes had been or would be positive for naturopathic patients (72.5% positive, 9.2% negative) and for their own naturopathic practice (53.1% positive, 20.1% negative); that their regulator protects the public interest (85% agree, 6.6% disagree); and that their provincial rules for investigation and discipline were satisfactory (83.6% satisfied, 16.4% dissatisfied).

One especially notable finding concerns the difference between whether respondents agreed that provincial regulatory changes were positive or negative for naturopathic patients (72.5% agree) compared with their own naturopathic practice (53.1% agree). Some NDs clearly believe that regulatory changes that benefit the public do not benefit them professionally. In their qualitative responses, many NDs agreed that the regulations had protected the public, increased standards of care, and increased the legitimacy of the profession. By contrast, NDs who disagreed that regulations would be good for their own practice most frequently mentioned a reduced scope of practice (20.9% of respondents who answered), increased costs (9.9%), and increased restrictions (4.1%). These responses reflect the bioethical trade-offs that health professionals face when it comes to regulation, as financial and professional autonomy for individual practitioners does not always align with the public interest. Future qualitative studies should explore specifically which aspects of regulation NDs (and other health professionals) believe are in patients' best interest, but not necessarily in their own professional self-interest.

The last set of findings regard the public understanding of naturopathic medicine. Even though respondents agreed that their promotional organiza-

tions had improved understanding of naturopathic medicine, 95.1% did not believe that media portrayals of naturopathic medicine are fair and balanced. My previous research has found that naturopathic medicine was "subject to far more negative than positive social constructions in Canadian newspapers" between 2013-2017, particularly in the *Globe and Mail*, one of Canada's two national newspapers.[19] More recent articles about naturopathic medicine in Canada's other national newspaper, the *National Post*, were publicly criticized by naturopathic organizations and municipal politicians for inaccurate reporting.[20] In their open-ended text responses explaining why they felt the media did not understand naturopathic medicine, 22 NDs (5.9% of those who offered an answer) specifically mentioned the *National Post*, and 30 (8.1%) mentioned the Canadian Broadcasting Corporation, Canada's public broadcaster. There is clearly a sense among Canadian NDs that national media outlets do not portray their profession accurately.

Likewise, a vast majority of respondents disagreed (84%) that medical doctors (MDs) understand naturopathic medicine, a finding that complements Meyer's study of the integration of NDs and MDs in Ontario. Meyer surveyed NDs and their patients, and found NDs viewed integration with MDs as beneficial, particularly with respect to effectiveness in diagnosis and patient convenience. However, nearly two-thirds of Meyer's ND respondents (64.7%) said they has received some form of hostility from MDs, and 94.1% claimed MDs do not understand what NDs do and/or the rigorousness of their training. Meyer also surveyed patients of NDs, who believed that the lack of integration was in part due to MDs' "often negative views towards naturopathic approaches and/or the very different philosophies of MDs and NDs".[21] However, our results did show minor evidence of collaboration between NDs and MDs in Canada: in the open-ended text responses, 23 respondents (6.5% of those who provided an answer) indicated that some—though not most—MDs understand naturopathic medicine, with the following answer representative of that view: "Some MDs seem to understand naturopathic medicine, in large part, but many seem to have only minimal understanding of the profession". Future research ought to explore where, when, and to what extent this collaboration between NDs and MDs is happening in Canada.

Overall, the survey results suggest that even as NDs view regulation positively, they do not believe this has translated into accurate portrayals of their profession in the media or among MDs. For all the policy changes over the past 15 years, and despite the fact that they believe their own promotional organizations have improved public understanding of naturopathic medicine, Canadian NDs thus still exhibit the characteristics of a "repressed structural interest" rather than a "dominant structural interest" in the Canadian health care system, perceived as lacking in medical legitimacy by dominant medical actors and by the media.[22]

Conclusion

In addition to providing a systematic review of the laws and policies regulating naturopathic medicine in the Canadian provinces, this multidisciplinary study sought to survey regulated NDs in Canada to better understand three things: the demographics and education of NDs; their views on representation and governance; and their views on how well the public understands naturopathic medicine. Demographically, respondents were primarily young, female, relatively new to practice, and most likely to have entered their naturopathic medical program with a Bachelor of Science degree. Respondents were generally positive about the role played by their promotional associations and regulatory bodies. However, the vast majority of respondents did not believe the media, medical doctors, and the Canadian public understand naturopathic medicine.

These results demonstrate the need for scholars of health policy to further explore the regulation, edu-

cation, and practice of naturopathic medicine in Canada and internationally. One limitation of this study was that it did not explore naturopathic *practice*—what naturopathic doctors do, what they are taught, and which diagnostic tools and treatments they use. Future research on naturopathic medicine, whether in the form of surveys or in-depth interviews, can and should explore these questions, including whether and how naturopathic practice has substantively changed due to new regulations in Canada and elsewhere. This study also demonstrates the need to isolate naturopaths as a distinct profession in order to foster better understanding about how they are trained and governed. Much recent scholarship has examined naturopathic medicine as part of a broader study on TCM, including those with different scopes of practice such as homeopaths, chiropractors, and midwives.[23] While naturopathic medicine is certainly part of TCM, the recent professionalization and growth of the profession highlights the need for naturopathic medicine to be studied as a unique object of inquiry.

This study also highlights the importance of hearing directly from health professionals. While past scholarship using Canadian NDs' websites has provided valuable information about the way they advertise their practice,[24] it is crucial for future scholarship to speak with regulators, associations, and NDs themselves to measure naturopathic practice and beliefs about how the profession operates and how regulations have affected that operation. Given the similar curriculum for schools in Canada and the United States accredited by the Association of Accredited Naturopathic Medical Colleges, comparative survey research on the regulation and attitudes of regulated NDs in the United States would provide an opportunity to explore whether similar policy trends exist in each country. Recent comparative research has begun in this vein, with Dunn, *et al.* finding considerable regulatory heterogeneity across the world, but also finding that jurisdictions with regulatory frameworks have higher standards.[25]

Another avenue for research is the extent to which the profession copes with internal divisions about what naturopathic medicine should be. NDs have long been subject to "in-fighting between self-identified naturopaths of different persuasions… which has for a long time weakened their public identity and their political impact".[26] Such a split—between those concerned regulations would move the profession away from its naturalistic roots, and those who want the profession to become more evidence-based and integrated with biomedicine and pharmacology—has been highlighted by previous studies of Canadians NDs.[27] Recent interview research has similarly suggested that the younger generation of North American NDs are more "science-oriented" and likely to view evidence-based medicine as an essential part of their practice.[28]

While the survey data presented here did not find evidence of a generational divide regarding views on regulation, it confirmed that the profession in Canada is quite young and that respondents were most likely to enter their naturopathic medical program with a Bachelor of Science degree. Moreover, the desire for a movement away from more controversial aspects of naturopathic medicine's past and present has manifested in recent scholarship. A group of scholars including several NDs have publicly argued for naturopathic medicine to adopt a seventh principle—*scientia critica*, the ability to critically analyse accumulated knowledge—to guide the training and practice of naturopathic medicine in North America.[29] Those scholars critiqued the "legacy of vaccine hesitancy [that] may remain in some quarters of the naturopathic profession", while other NDs have recommended that naturopathic medical programs should de-emphasize homeopathy due to a lack of scientific evidence for its utility beyond placebo.[30] Especially insofar as provincial standards of practice increasingly prevent NDs from offering vaccine alternatives—indeed, NDs can administer vaccines in

British Columbia—future scholarship should explore the interaction between generational differences, regulations, and changes in naturopathic policy and practice. Moreover, the tremendous medical breakthrough that emerged with the development of COVID-19 vaccines in 2020 provides an opportunity for scholars to see how naturopathic organizations and NDs themselves have responded to emerging technological vaccine developments in a post-COVID-19 world.

Across the country and across the world, naturopathic medicine is becoming increasingly professionalized and regulated. As the conventional health care system faces growing issues surrounding funding and physician shortages, and as an increasing number of patients visit NDs, the role of legal regimes in permitting or proscribing naturopathic medicine in health care delivery will only continue to grow. Those who research at the intersection of medicine, law, and public policy should continue to examine naturopathic medicine to understand more about its role in health policy, management, and delivery.

[**Dave Snow** is an associate professor of political science at the University of Guelph who specializes in public law and health care governance. Dave can be contacted at <snow@uoguelph.ca>.]

1. World Health Organization, *WHO Global Report on Traditional and Complementary Medicine 2019*, <https://www.who.int/publications/i/item/978924151536> (Accessed October 16, 2021), 5.

2. World Health Organization, *Global Report*, 8, *ibid.*

3. Timothy Caulfield, *et al.*, "Health misinformation and the power of narrative messaging in the Public Sphere" (2019) 2:2 Canadian Journal of Bioethics 52-60; Blake Murdoch, Stuart Carr and Timothy Caulfield, "Selling falsehoods? A cross-sectional study of Canadian naturopathy, homeopathy, chiropractic and acupuncture clinic website claims relating to allergy and asthma" (2016) 6:12 BMJ Open: e014028.

4. Blake Murdoch, Timothy Caulfield and Robyn Hyde-Lay, "Commentary: An Examination of the Public Justifications for the Expansion of Canadian Naturopaths' Scope of Practice" (2011) 19 Health Law Journal 215-236; Peter McKnight, "Message to naturopaths: magic isn't medicine", *The Globe and Mail* (9 May 2016), <https://www.theglobeandmail.com/opinion/message-to-naturopaths-magic-isnt-medicine/article29929971/> (Accessed August 20, 2021); Timothy Caulfield, "Don't legitimize the witch doctors", *National Post* (22 January 2013) <https://nationalpost.com/opinion/timothy-caulfield-dont-legitimize-the-witch-doctors> (Accessed August 20, 2021); Alheli Picazo, "Alberta shares the blame", *National Post* (24 March 2016) A11-12.

5. Healther Boon, "Canadian naturopathic practitioners: holistic and scientific world views" (1998) 46:9 Social Science & Medicine 1213-1225. doi: 10.1016/s0277-9536(97)10050-8 [hereinafter, "Boon, Canadian naturopathic practitioners"]; Nadine Ijaz, *et al.*, "Supportive but 'worried': Perceptions of naturopaths, homeopaths and Chinese medicine practitioners through a regulatory transition in Ontario, Canada" (2015) 15 BMC Complementary and Alternative Medicine 312 [hereinafter, "Ijaz, Supportive but worried"]; Stephen P. Meyer, "Naturopaths in Ontario, Canada: Geographic patterns in intermediately-sized metropolitan Areas and integration implications" (2017) 14:1 Journal of Complementary and Integrative Medicine [hereinafter, "Meyer, Naturopaths in Ontario"] 1-13; Marja J. Verhoef, Heather Boon and Donatus R. Mutasingwa, "The scope of naturopathic medicine in Canada: An emerging profession" (2006) 63 Social Science & Medicine 409-417.

6. Jeremy Ng, "The regulation of complementary and alternative medicine professions in Ontario, Canada" (2020) 9 Integrative Medicine Research 12-16 [hereinafter, "Ng, The regulation of complementary and alternative medicine"]; Dave Snow, "The social construction of naturopathic medicine in Canadian newspapers" (2019) Policy Studies 1-21, doi: 10.1080/01442872.2019.1704234 [hereinafter "Snow, The social construction of naturopathic medicine"].

7. Canadian Association of Naturopathic Doctors, "Common questions", <https://www.cand.ca/common-questions-about-naturopathic-medicine/> (Accessed August 20, 2021).

8. Ryan Bradley, *et al.*, "Naturopathy as a model of prevention-oriented, patient-centered primary care: a disruptive innovation in health care" (2019) 55:9 Medicina 603-616.

9. David H. Nelson, *et al.*, "The bell tolls for homeopathy: time for change in the training and practice of North American naturopathic physicians" (2019) 24 Journal of

Evidence-Based Integrative Medicine 1-11 [hereinafter, "Nelson, The bell tolls for homeopathy"]; Jonathan E. Prousky, "Repositioning individualized homeopathy as a psychotherapeutic technique with resolvable ethical dilemmas" (2018) 23:4 Journal of Evidence-Based Integrative Medicine 1-4.

10. Association of Accredited Naturopathic Medical Colleges, "The Six Principles of Naturopathic Medicine", <https://aanmc.org/6-principles/> (Accessed August 20, 2021).

11. Boucher Institute of Naturopathic Medicine, "Merger Establishes a Global Leader in Naturopathic Medical Education and Research in Canada", November 17, 2020, <https://binm.org/blog/merger-establishes-a-global-leader-in-naturopathic-medical-education-and-research-in-canada/> (Accessed October 16, 2021).

12. Monique Aucoin, *et al.*, "Naturopathy Special Interest Group Research Capacity and Needs Assessment Survey" (2019) 25:2 Journal of Alternative and Complementary Medicine 189-195 [hereinafter "Aucoin, *et al.*, Naturopathy Special Interest Group"].

13. *Supra*, note 5, Meyer, "Naturopaths in Ontario".

14. *Supra*, note 12, Aucoin, *et al.*, "Naturopathy Special Interest Group".

15. Antony Porcino, *et al.*, "Pediatric training and practice of Canadian chiropractic and naturopathic doctors: a 2004–2014 comparative study" (2017) 17:512 BMC Complementary and Alternative Medicine 1-8.

16. Dugald Seely, *et al.*, "Naturopathic oncology care for thoracic cancers: a practice survey" (2018) 17:3 Integrative Cancer Therapies 793-805.

17. Virginia Braun and Victoria Clarke, "Using thematic analysis in psychology" (2006) 3 Qualitative Research in Psychology 77-101.

18. Ijaz, *et al.*, "Supportive but Worried", *supra*, note 5; College of Naturopaths of Ontario [CONO], *Protecting the Public. Supporting Safe Practice: Annual Report 2017-2018* (2018): <https://cono.wpengine.com/wp-content/uploads/2020/11/2017-2018-CONO-Annual-Report.pdf> (Accessed August 20, 2021).

19. Snow, "The social construction of naturopathic medicine", 12, *supra*, note 6.

20. Sharon Kirkey, "Naturopathy is poised to 'disrupt' health-care status quo, proponents of controversial practice say," *National Post* (27 September 2019), <https://nationalpost.com/news/naturopathy-is-poised-to-disrupt-health-care-status-quo-proponents-of-controversial-practice-say> (Accessed August 20, 2021); Sharon Kirkey, "Criticism for Ontario city's decision to grant $45K toward naturopathic clinic", *National Post* (6 November 2019), <https://nationalpost.com/news/confused-ontario-city-spends-45k-toward-naturopathic-clinic-because-it-has-a-shortage-of-doctors/wcm/06ea4596-5d4b-4464-b202-d6cbb1df799a/> (Accessed August 20, 2021); Mark Fontes and Shawn O'Reilly, "Naturopathic medicine plays credible complementary role in improving Canadians' health," *Canadian Association of Naturopathic Doctors*, September 30, 2019, <https://www.cand.ca/wp-content/uploads/2019/10/CANDNPSept302019-4.pdf> (Accessed August 20, 2021); Mark Fontes, "Statement by chair of the Canadian Association of Naturopathic Doctors Dr. Mark Fontes, ND," *Canadian Association of Naturopathic Doctors*, November 6, 2019, <https://www.cand.ca/wp-content/uploads/2019/11/CAND-statementNov6.pdf> (Accessed August 20, 2021); Nick Seebruch, "Councillor points to missing facts in National Post naturopath story", *Cornwall Seaway News* (6 November 2019), <https://www.cornwallseawaynews.com/2019/11/06/councillor-points-to-missing-facts-in-national-post-naturopath-story/> (Accessed August 20, 2021).

21. *Supra*, note 5, Meyer, "Naturopaths in Ontario", 9.

22. *Supra*, note 6, Ng, "The regulation of complementary and alternative medicine"; Snow, "The social construction of naturopathic medicine", *supra*, note 6.

23. Ijaz, *et al.*, "Supportive but worried", *supra*, note 5; Ng, "The regulation of complementary and alternative medicine", *supra*, note 6; Daniel Hollenberg, *et al.*, "Addressing provider shortage in underserviced areas: The role of traditional, complementary and alternative (TCAM) providers in Canadian rural healthcare" (2013) 5:1 European Journal of Integrative Medicine 15-26.

24. Timothy Caulfield, Alessandro R. Marcon, and Blake Murdoch, "Injecting Doubt: Responding to the Naturopathic Anti-Vaccination Rhetoric" (2017) 4:2 Journal of Law and the Biosciences 229-249; Caulfield, Timothy and Christen Rachul, "Supported by Science? What Canadian Naturopaths Advertise to the Public" (2011) 7:14 Allergy, Asthma and Clinical Immunology 1-4.

25. J.M. Dunn, *et al.*, "Characteristics of global naturopathic education, regulation, and practice frameworks: results from an international survey" (2021) 21:67 BMC Complementary Medicine and Therapies 1-19.

26. Mike Saks, "Book Review of *Nature's Path: A History of Naturopathic Healing in America* by Susan E. Cayleff, (2017) 30:3 Social History of Medicine 711.

27. *Supra*, note 5, Boon, "Canadian naturopathic practitioners"; Ijaz, "Supportive but worried", *supra*, note 5.

28. Joshua Z. Goldenberg, *et al.*, "Shifting attitudes towards research and evidence-based medicine within the naturopathic medical community: the power of people, money and acceptance" (2017) 4 Advances in Integrative Medicine 49-55.

29. Alan C. Logan, *et al.*, "North American Naturopathic Medicine in the 21st Century: Time for a Seventh Guiding Principle—*Scientia Critica*" (2018) 14:5 Explore 367-372.

30. Logan, *et al.*, "Time for a Seventh Guiding Principle",370, *ibid.*; Nelson, "The bell tolls for homeopathy", *supra*, note 9.

Trust, Tort Law and The Integration of Black Box Artificial Intelligence into Clinical Care

Ian Stedman & Michael Brudno

Abstract

Optimism abounds that artificial intelligence ("AI") will help us transform our health care systems for the better. There are of course many different methodological approaches to AI and many different possible applications within health care. This paper focuses specifically on whether and how opaque (*i.e.*, black box) artificial intelligence decision support ("DS") tools can be integrated into clinical care pathways. We begin by explaining what these tools are, what they can do, why they are important and how they work. We next explain that clinical decision support ("CDS") tools are not stand-alone or autonomous decision-makers and there are accordingly several barriers to their adoption, including three that specifically emerge from the law of torts. We propose two solutions to help us overcome these legal barriers in order to facilitate the use of black box CDS tools, including tools that learn in real time from new data without the need for ongoing human intervention. The success of our proposed solutions is contingent on the implementation of an innovative multi-disciplinary governance framework that considers the risks associated with each use of black box AI-CDS and requires a robust approach to oversight and accountability. As things stand, it is unlikely that many AI-based black box DS tools will make it into broad clinical use unless there is little or no associated risk to patients and/or the physicians. Our governance framework moves us closer to achieving the goal of clinical integration by, among other things, ensuring that black box AI-CDS tools are designed in a way that allows administrators to immediately identify and address problems as soon as they arise.

I. Introduction

Optimism abounds that artificial intelligence ("AI") will help us transform our health care systems for the better. There are of course many different methodological approaches to AI and many different possible applications within health care. This paper focuses specifically on whether and how opaque (*i.e.*, black box) AI decision support ("DS") tools can be integrated into clinical care pathways.[1] We briefly explain what these tools can do, why they are important and how they work. We then identify several potential barriers to the adoption of black box clinical decision support ("CDS") tools, focusing primarily on barriers that specifically emerge from the law of torts. We propose two solutions to help overcome these barriers and facilitate the use of black box CDS tools, including especially those tools that continuously learn from new data without the need for ongoing human intervention. The success of our proposed solutions is contingent

on the implementation of an innovative multi-disciplinary governance framework that considers the risks associated with each tool and requires a robust approach to oversight and accountability. Our governance framework leverages technical computer science expertise and requires black box AI-CDS tools to be designed in a manner that allows them to continually learn but prevents them from providing recommendations if their accuracy wavers as a result of their real time learning. Our framework also allows administrators to immediately identify, address and potentially learn from the challenges that new data can sometimes give rise to.

II. Types of Clinical Decision Support Tools

Different types of decision support are needed for different areas of clinical health care practice. How we design and integrate tools of this nature will ac-

cordingly vary depending not only on their intended use(s), but also on the risk levels that we assign to them. We will return to a discussion of risk profiles in Part VI, but for now it is important to identify some of the use cases that are commanding broad interest and acceptance. The following is a non-exhaustive list of types of CDS tools that are being developed and/or already adopted within clinical health care settings (*i.e.*, at the point of care):

1. Tools that can assist with diagnostics (*e.g.*, a CDS tool might track an in-patient's vitals in order to predict a catastrophic event, or a tool might suggest a differential diagnosis to a physician after they enter their patient's symptoms into the electronic health record ("EHR"));[2]

2. Tools that can suggest treatments, therapies, specialist referrals, *etc*. These recommendations could be responsive to confirmed diagnoses or could be predictive in nature; and,

3. Tools that can suggest a care plan, including a schedule for return visits and/or specific tests.

Clinical practice requires physicians (and other clinicians, such as registered nurses) to use their best professional judgment to determine what each patient might benefit from within the specific context of their individual health care needs. Clinicians have access to many resources when making these judgments, including research materials, patient data, consultations with colleagues, tests they can order, *etc*. They also have access to data-driven DS tools that leverage the simplicity of things like decision trees and other more direct, easy to understand analytic techniques. As different ways to track and record health-related data become more ubiquitous, there will be countless situations in which a clinician could benefit from having computational assistance in determining the most effective course of action when caring for a patient.

The drive to integrate machine learning into health care is reflective of the general expectation that high-powered analytic tools will allow physicians

to consider a broader range of data and to benefit from a more detailed data analysis when making clinical decisions. An integrated AI-CDS tool could monitor each individual patient (*e.g.*, the prescriptions they have taken or care plan they have followed and any reported outcomes) in order to learn something about that individual and (potentially, depending on the data collected) about broader populations. An AI-CDS tool could make recommendations to clinicians about how to consider varying their follow-up and treatment plan(s) or what courses of action to consider for a new patient who presents with similar symptoms to ones the CDS has already learned from.[3] These tools could also be particularly useful in emergency care[4] or general/family practice settings, where care providers serve patients with wide ranging needs and can have limited time in which to make decisions.

III. AI, Black Boxes and Trust

There are various methodological approaches to AI in the health care context. In general, however, AI tools are designed, trained and validated using retrospective data in order to analyze new data in real time and provide prospective analysis to support decision-making (*e.g.*, by physicians, patients and administrators). These tools allow their users to make decisions that are informed by the increasingly large amount of data that is available in modern society. Tools that analyze clinical data can reveal patterns, make predictions, enable better disease surveillance, facilitate early disease detection, create workflow efficiencies, and propose solutions to other complex problems.[5] These analyses can then be used to improve institutional decision-making, help to reduce health care costs[6] and generally help improve outcomes for patients and their families.

There are a growing number of examples of AI-based CDS tools that analyze EHR data in real time.[7] Machine learning ("ML") is one example of an AI methodology where algorithms learn from retrospective data in order to construct mathemati-

cal models that can help users better understand real time, prospective clinical data. Multi-disciplinary teams around the world are working to integrate these tools directly into care pathways and in doing so are also being asked to address some important challenges. For example, while the analysis conducted by some simpler ML tools might be easy for humans to understand, others are incredibly complex. ML methodologies like deep neural networks conduct analysis that is opaque and inaccessible to the end-user. The complexity of the model means that it is capable of considering a higher volume of data and drawing more inferences in its analysis. The trade-off is of course that the greater complexity leads to greater opacity and greater opacity means that the step-by-step analysis that underpins the tool's output/recommendation is impossible for any human user to understand. This opacity can serve as a significant barrier that prevents users from trusting and adopting these tools into their workflow. The design of CDS tools like these and the data that their "black box" algorithms learn from is accordingly very important.

This underlying data is important because it has been well established that how we collect, analyze and use data can serve to create and/or perpetuate biases. For example, bias can exist in data because it is incomplete (*e.g.*, it does not provide a full picture of what is clinically relevant regarding a particular matter) or because it is not representative (*e.g.*, data for a particular demographic group was never collected, but the AI tool is nevertheless being applied for analysis of that group).[8] Given that a black box AI does not explain its analysis, its users will have no way to discern whether it is relying on biased data. Because of this, most DS tools that have been designed for clinical integration are not opaque and can only learn from (*i.e.*, train on) new data that has first been reviewed (*i.e.*, validated) by a human.[9] The focus of this paper is on algorithms that analyze such vast amounts of data that the analytic process undertaken, and conclusions arrived at

are impossible for a person to understand. Until we see significant advances in clinically-relevant approaches to explainability,[10] users of black box AI will have to choose whether or not to incorporate its analysis into their assessments even though they do not fully understand it.[11] Adoption of opaque CDS recommendations could lead to clinical decision-making that is based on incomplete information and that requires a user to trust the black box CDS tool rather than being able to fully or adequately understand it and explain it.
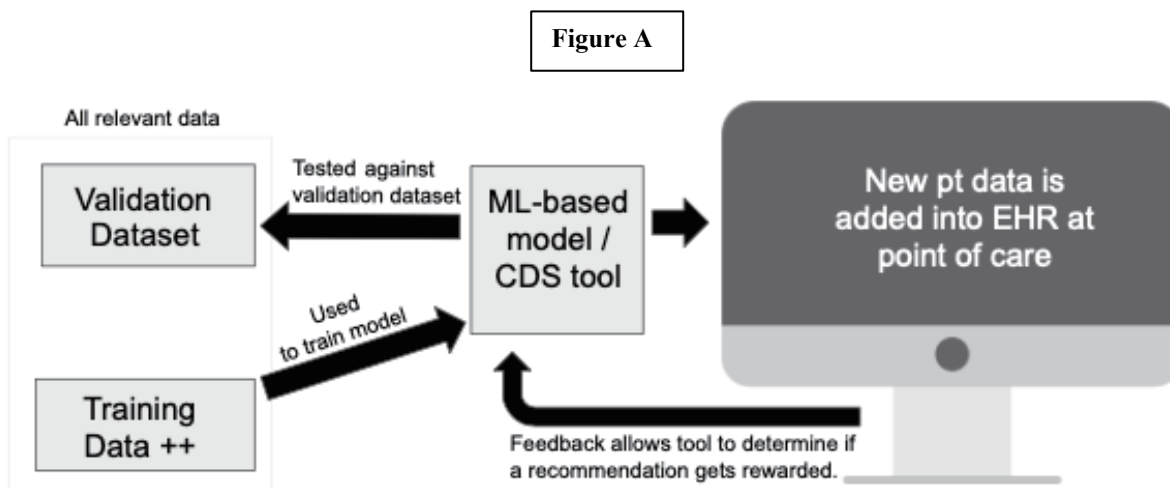
Despite the above, it is not unheard of to trust someone or something else to undertake the analysis that supports our health care decisions. Physicians have knowledge and skills that most patients do not, and patients generally respond to this imbalance by forming beliefs and expectations that cause them to place greater trust in the professional judgment of their physician.[12] Most physicians do not have a full understanding of many of the tools that they rely on to inform their care decisions. For example, most do not understand the physics that underlies Magnetic Resonance Imaging ("MRI") and take for granted that MRI machines have been properly constructed and maintained in order to produce valid results that they can rely upon. Health care systems would grind to a standstill if physicians could not trust that the tools of their trade are being continuously validated by someone else who has the requisite expertise. The layers of trust and abstraction that have always been integrated into these complex systems cannot be ignored when we talk about black box ML-based CDS tools. The important difference between relying on an MRI analysis and in integrating a black box ML-based CDS tool into clinical care, however, is that an expert actually exists who can provide a step-by-step explanation of the math and science behind how the MRI machine works. That is not typically the case with black box AI. An expert can explain how a black box AI was constructed, but they simply cannot walk another person step-by-

step through its analysis and explain what data was relied upon and why. A black box AI does not provide its end-user with that information.

IV. Applying Reinforcement Learning

Black box ML tools, including those that can continuously learn in real time from the feedback they receive in response to the recommendations they provide[13] represent the most complicated, challenging and powerful approach to AI-CDS. One way this can be done is through a process called reinforcement learning. Reinforcement learning is a methodology where the algo-

rithm seeks to maximize whatever notion of a reward it is programmed to strive for. The algorithm will provide analysis (*i.e.*, a recommendation) to its user and then learn whether that analysis was good or bad based on the feedback it receives (*i.e.*, the health outcome information that is subsequently inputted into the EHR). Because the algorithm is programmed with a policy to reward those actions that feedback tells it will lead to better health outcomes, it will learn to recommend actions that maximize its ability to meet its policy objectives. This type of system is explained in 'Figure A'.[14]

Figure A

All relevant data

Validation Dataset

Tested against validation dataset

ML-based model / CDS tool

Used to train model

Training Data ++

New pt data is added into EHR at point of care

Feedback allows tool to determine if a recommendation gets rewarded.

A model like this could be integrated into the clinical setting at the point of care, likely via the EHR. Each new patient encounter within a hospital system would also give rise to new data. Using an AI methodology called reinforced continual learning, that new data inputted into the EHR could be automatically added back into the training data (provided, of course, that it is labelled properly)[15] and used to immediately update/retrain the algorithm.[16] The newly updated algorithm would then test itself against the validation set in order to ensure that its suggestions/predictions remain accurate (*i.e.*, that

its predictions continue to conform with the model's policy objectives). It is this "capability to continuously evolve that underlies much of the potential benefit of AI/ML".[17] A system using reinforced continual learning in this way would be almost identical to one using reinforcement learning, except that any new patient data entered into the EHR could also be added to the training data. The continually updated training data would then be used to continually retrain the underlying model. This type of system would work as described in Figure B.

Figure B
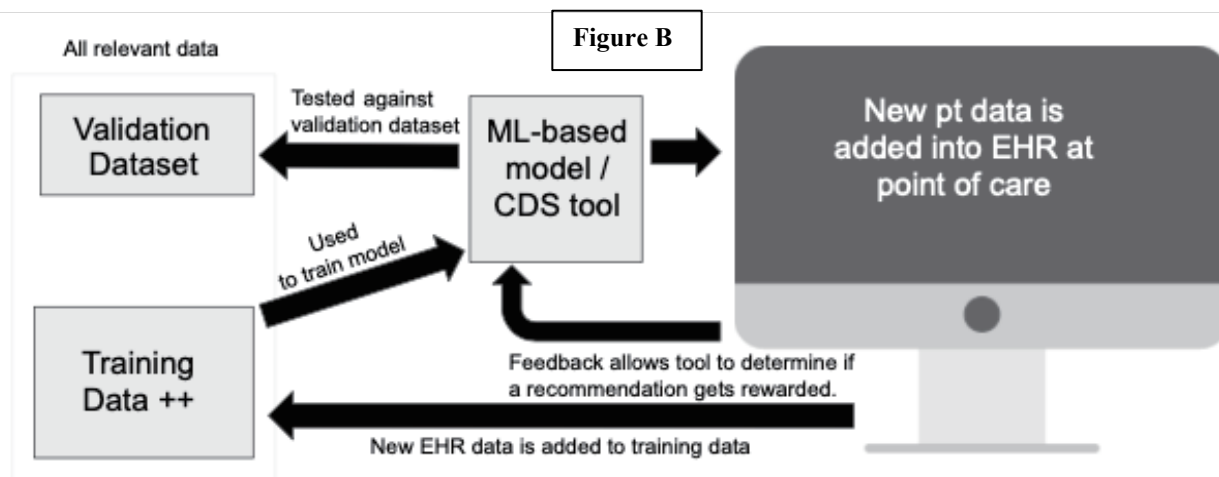
If properly constructed, these tools should not need new data to be cleaned and validated by humans and should be capable of remaining active while new data is automatically added to the training set.[18] Reinforced continual learning[19] can play an important role in our health care systems by allowing us to learn from new data in order to recognize mistakes and/or gaps in our initial assumptions and give us the information needed to take action to correct them. For example, a simple AI-CDS model might be built that learns from retrospective data about clinical decisions. That model will be trained on data relating to decisions that were made before any high-powered analytic decision support was available to the clinicians whose decisions are being learned from. By allowing the model to begin to learn from new clinical data as well, we open up the possibility that it will recommend new courses of action that could modify clinical decision-making over time in a manner that improves health outcomes. In other words, we allow the algorithm to identify and learn from our mistakes. Not only can this help us to improve the quality of care offered to patients, but these decision support tools also promise to have a significant impact in the public health context.

The use of reinforced continual learning should allow better detection, tracking and prediction during public health risks and/or outbreaks. An AI-CDS tool that assists emergency room physicians with making di-

agnoses may not have been built with COVID-19 in mind, for example, but a person with symptoms that correspond to COVID-19 is more likely to be accurately diagnosed if the algorithm has already learned in real time that the disease's prevalence has increased in that location during that time period. Imagine also that the AI-CDS could learn from data being collected at different sites, whether locally or abroad. The potential that reinforced continual learning AI-CDS tools hold to help us improve our response to public health crises is apparent.

V. Standard of Care, Duty to Warn and the Learned Intermediary Rule

One of the barriers to integrating black box AI within the clinical setting that has been identified in the literature is that of determining what the standard of care is and ought to be for physicians who wish to use these tools, and how liability might be apportioned in the event of a negative outcome for a patient.[20] These are challenging issues, in large part because of all the potential locations (*i.e.*, components, actors, processes) in the AI development and integration pipeline where things could go wrong and liability could potentially be apportioned.[21] Concerns about the standard of care arise in the context of negligence actions being brought against physicians. A plaintiff (*e.g.*, a patient) who brings an action must establish four elements to be

successful in proving that a defendant physician was negligent: (1) that the defendant owed a duty of care, (2) that the defendant breached that duty, (3) that the plaintiff sustained a loss, and (4) that the defendant's breach caused that loss. The overall claim must be proven on a balance of probabilities, which simply means that it is determined to be more likely than not that the physician was negligent. This is important because the standard of proof to be met in civil claims is not as high as the standard needed to establish criminal culpability (*i.e.*, guilt beyond a reasonable doubt). After the plaintiff has presented their case and had an opportunity to argue that the four elements of negligence can be established, the defendant then has an opportunity to raise defenses.

The first step of establishing that a duty of care exists between the parties is generally quite straightforward in the context of a physician and their patient. Although beyond the scope of this paper, establishing this relationship and/or a related duty of care could potentially become more difficult in the future as more processes and procedures become automated, leading to physicians being removed from some patient encounters. The second element of a negligence claim can be significantly more complicated. Assuming a duty of care has been proven, the court must then establish what a reasonable physician would have done in the circumstances in question. This stage of a negligence claim is often very adversarial, with both the plaintiff and defendant potentially calling expert witnesses who will provide competing testimony about what the standard of care is and how a reasonable physician ought to have acted in that specific context. To establish that a duty of care was breached, the plaintiff must prove that the defendant did not meet the standard of care, *i.e.*, that they did not provide care with the skill level of a competent physician of the same specialty and with access to the same resources.[22] It should come as no surprise

then that the legal standard of care for physicians is subjective and can evolve.

In the context of black box AI-CDS tools, what we have is a new resource that very few physicians will have access to in the early stages of its adoption. Some physicians at resource-rich health care institutions may have access to these tools, but many others will not until they become more cost-effective and/or are more widely available. It is easy to imagine then that a breach of the standard of care might be difficult to establish when the defendant is the only physician who can be held up as an example of what a reasonable physician might do in similar circumstances and with similar resources.

There remains no caselaw in Canada or the United States that tells us what standard of care is expected of physicians when using AI-CDS tools.[23] Recognizing this gap, the Canadian Medical Protective Association ("CMPA"), which insures all physicians and defends them against negligence claims, has released informal guidance. This is important because it could very well be the case that a court's decision about whether an insurance policy applies to indemnify a physician in the context of a negligence/malpractice claim is what will lay the groundwork for a legally recognized standard of care in this area:

> Physicians using AI need to be mindful of their legal and medical professional obligations, and discuss with the patient the appropriateness of using AI technology and privacy risks...

> While endorsement of an AI technology from a reputable professional or regulatory organization may be a factor to consider in evaluating whether you have complied with your professional and legal obligations, you should still review and seek advice on its suitability in clinical practice, including consideration of the following, among other things: What are the terms of use? Has the AI technology been subject to rigorous evaluation of its accuracy, consistency, and reliability? Does it use appropriate privacy and confidentiality safeguards and policies (e.g. patient consent, encryption, password protection)?[24]

Physicians regularly use their professional judgment to determine what technologies, tests and methods to employ in their care decisions. While AI is a new technology, the general concerns about risk and physician liability are not new. What is new here is simply the nature of the knowledge that physicians must have in order to exercise informed judgment about whether to use these tools. The CMPA is suggesting that physicians should learn about accuracy, consistency, reliability, privacy, *etc*., for any AI technology that they wish to use in practice. They need to be informed enough to be able to discuss privacy risks with their patients and, more generally, "the appropriateness of using AI technology".[25]

In considering the appropriateness of the technology, physicians will also have to consider things like whether uncertainty, misalignment and over-reliance could give rise to risk. First, even if a CDS tool's predictions are compelling, there may be some uncertainty about how much weight is given to different data classifiers to render a specific prediction. Not knowing how data is weighed will make it harder for the physician to understand and explain the tool's prediction, if necessary.[26] It will also make it harder to decide whether that prediction should be relied upon. To that end, physicians must also consider whether a CDS tool is capable of offering some sort of explanation when its prediction does not seem to align with the clinical data. A doctor would not be exercising good clinical judgment if they simply adopted a black box prediction that did not clearly align with the clinical data nor their own clinical judgment. Finally, there is a real risk that physicians could over-trust and develop an over-reliance on these tools instead of exercising their own independent clinical judgment.[27] There is no legal or regulatory approach to AI-CDS that allows a physician to fully delegate their decision-making responsibilities to a machine. Physicians must be attentive to this risk and have a

strategy in place to help sidestep this potential intellectual trap.

To be able to meet the standard of care expected of them, a physician will need to know whether they are legally responsible for ensuring the proper functioning of the tools they use in their practice. Physicians must accordingly consider the legal obligations they have by virtue of their role as 'learned intermediaries'. The learned intermediary rule can only be understood in the context of the 'duty to warn' that applies to manufacturers as part of the law of tort. If a manufacturer of a product knows or ought to know there are dangers inherent in the use of their product, they have a duty to warn consumers of those dangers. The nature and extent of the warning required will depend on what is reasonable in the circumstances (*i.e.*, the more dangerous the product, the more explicit the warning). A manufacturer of a product that is inherently dangerous could be held liable if its warning is not sufficiently explicit.[28] This duty does not simply apply at a single point in time (*e.g.*, at the point of sale) but is instead an ongoing duty and is the reason why, for example, we regularly see product recalls when new risks become known. The policy purpose underlying the duty to warn is that it helps to correct a knowledge imbalance that exists between manufacturers and consumers by notifying consumers about dangers and "allowing them to make informed decisions concerning the safe use of the product".[29]

In some cases, it may not be necessary for the manufacturer to warn the end consumer. It may instead be possible for them to satisfy this duty by warning a learned intermediary of the risks inherent in the use of their product. This exception applies either when a product is highly technical in nature and it is only intended to be used under the supervision of an expert or "where the nature of the product is such that the consumer will not realistically receive a direct warning from the manufacturer before using it".[30] In circumstances where a consumer is placing primary

reliance on the judgment of an expert, and where it is established that the manufacturer discharged its duty to warn that expert, that expert, or learned intermediary, may be found solely liable for failing to provide adequate and timely warning to the consumer (*e.g.*, the health care patient).[31]

This tort law concept of the learned intermediary arises in the context of CDS tools because manufacturers will always want to minimize their potential liability by disclosing possible risks and predictive limitations to their customers, who are in most cases the physician users of their products. Assuming that a manufacturer discloses known risks and limitations accurately, the legal burden is then transferred to the physician who must warn their patients of all known material risks associated with the medical device or product they are relying on.[32] A physician could be found to have breached the standard of care solely by virtue of not having discharged their duty under the learned intermediary rule by failing to provide their patient with a timely warning about risks.[33] This analysis seems to support the CMPA's guidance to physicians with respect to needing to be knowledgeable about the technology they are using.

We will return now to the final two elements of a negligence claim: establishing (3) that the plaintiff sustained a loss, and (4) that the defendant's breach of their established duty of care (*i.e.*, not meeting the expected standard of care) is what caused that loss. Although it can sometimes be difficult to establish a loss or injury that is recognized by the law and for which damages can be assessed,[34] we will assume the plaintiff's success for the purposes of our analysis. Where things might become complicated again with clinical AI is in the task of apportioning damages when more than one defendant has been deemed negligent. As we noted above, there are many potential locations in the AI development and integration pipeline where things could go wrong and where liability could then be ascribed. This means that a negligence action might involve

more than two parties. Deciding how to apportion liability and assign responsibility for any related damages will be a deeply contextual fact-finding exercise that is again beyond the scope of this paper. We will instead turn our attention now to the need to establish causation between the breached duty of care and the compensable injury being claimed.

It is well recognized that plaintiffs "face an uphill battle in satisfying the causation test in medical negligence claims".[35] A claim cannot succeed unless the plaintiff can prove that they would not have sustained their injury *but for* the negligent actions of the defendant.[36] Imagine then that we are using a black box ML-based CDS tool to help diagnose a patient who might otherwise be impossible to diagnose by a reasonable physician in similar circumstances *without* similar resources (because, as you will recall, this is a novel technology). If the physician uses a black box ML-based CDS tool and the analysis/prediction is wrong, that does not necessarily mean that the wrong prediction caused the injury. That person would have still been sick and sustained an injury even without the AI-aided diagnosis.[37] Where it becomes complicated is when the AI's prediction is wrong, but the physician trusts it anyway (even when another reasonable physician might have interrogated the AI's analysis before acting) and then acts on it in a way that results in injury. Prescribing a therapeutic, for example, that is not appropriate and that clearly worsens the patient's original symptoms leading to further health complications. A physician will typically argue that a plaintiff did not establish their negligence claim on the balance of probabilities because the physician either acted reasonably (*i.e.*, met the standard of care) or even if they did not there were no resulting damages that can be causally linked to the actions that have been determined unreasonable.

Although there are still some missing pieces to the puzzle, it seems evident that black box clinical AI could challenge the traditional legal paradigm used

to establish physician negligence. Courts will need to wrestle with important questions about the appropriate standard of care and about how to determine whether and when inaccurate AI being used by an over-trusting physician can in fact cause compensable damage to a patient who was already injured. Does a physician who has access to a black box tool of this nature have a legal obligation to disclose to patients that the tool is available for use and then to require those patients to consent to its use?[38] If so, can obtaining patient consent vitiate physician liability when the standard of care is otherwise not met? Would it be reasonable for a physician to trust an AI tool if it has been validated as having greater predictive accuracy than a competent physician of the same specialty in that same setting but who does not have that tool? Could a physician ever have a positive legal obligation to use a tool that has been proven more accurate than a competent physician of the same specialty or will they always be allowed to exercise their discretion on a case-by-case basis? To date, the answer to all these questions is quite simply that physicians must exercise their own professional judgment and discretion in order to determine for themselves, in their precise circumstances, what they think the best approach is to using AI-CDS tools in their practice.[39]

The clinical integration of black box ML-based CDS tools, particularly those that employ reinforcement or reinforced continual learning, adds a new layer of complication to the risk-mitigation relationship between manufacturer and physician. If we want physicians to be confident in their ability meet the standard of care expected of them when using these tools, they must either be able to understand the recommendations being offered in a clinically relevant way (*i.e.*, those recommendations must be explainable to them) or, if they can't understand the reason for a recommendation, they must be confident that the tools are validated on an ongoing basis and that they will only offer recommendations that conform with accepted ethical, legal and medical standards of practice.

VI. Overcoming the Explainability Barrier

It is clear at this point that we do not know what it would take for a physician relying on analysis from a black box CDS tool to meet the standard of care expected of them. No caselaw explains whether a physician can justifiably rely on AI without also understanding its analysis or what advice and support should be made available to users of black box AI-CDS systems. What we have is a reasonably unique scenario where there is very little guidance about how to use a tool that many believe has a tremendous potential to transform clinical medicine.

What we do know is that even though it can be difficult for a plaintiff to prove compensable negligence, physicians still have an obligation to provide care to a certain standard. Generally speaking, and as long as they are safe and effective, the instrumentation and tools they use to provide care to that standard are not as important as the fact that they meet the standard. AI is a tool that can help its users get at new information, but it is not always the only way to get at that information. The real challenge here arises when the black box AI-CDS suggests a course of action that the physician might not have otherwise considered. Following that CDS' advice could lead to a situation where harm is caused to a patient. That being said, it is also the case that things can go wrong without a doctor being negligent. A poor outcome may lead to a negligence action but is not itself necessarily an indication of negligence.[40]

As A. Michael Froomkin, *et al*. have argued, a physician who uses an AI-CDS tool is in a better position vis-à-vis potential negligence claims if they actually understand how that AI works. We would accordingly also do well to define the standard of care for AI-assisted decision-making in clinical care as requiring the physician to be "meaningfully

involved in reviewing the diagnostic decision".[41] This leaves us in the same impossible place when it comes to black box AI however as no physician can meaningfully review the decision. At best a physician can ask if they would have arrived at the same conclusion as the AI, but they cannot compare their actual analysis with that of the AI.

Given that the standard of care at present seems to require physicians to have a deep understanding of the individual AI-CDS tools they use, we are in a position where we are asking already over-burdened physicians to do and know more, over and above that which they already must know to practice medicine at a high level. When it comes to using black box AI, we are in fact asking them to do the impossible if they want to meet the standard of care that currently seems to be expected of them.

The best way to minimize the burden on physicians will be to ensure that there is proper regulatory oversight of black box AI tools intended for clinical practice. Much like it does with pharmaceuticals and new health technologies (*e.g.*, by the FDA in the U.S. or by Health Canada in Canada), this oversight would be focused on testing and verifying the safety and efficacy of these tools before they are approved and made available for clinical use. This regulatory oversight is important because it signals to users (*i.e.*, physicians and other health care providers) that the tool has been tested by experts and they have verified it does what it says it will do, in a manner that is acceptable under standards/rules that are transparent. Given that standards of care in medicine often evolve, sometimes rapidly,[42] to reflect developments in science and/or how medicine is practiced, including how technologies are used,[43] regulatory oversight can also move the needle on what is expected of AI users. We allow physicians to prescribe new medicines for their approved indications without asking them to also read through and critically analyze the background clinical trial data. The courts generally accept that health technol-

ogies that meet safety and efficacy standards are acceptable for physicians to use. As proper regulatory testing and oversight come into place for black box AI then,[44] physicians should begin to feel more comfortable relying on them, even if the tool's analysis does not conform with the physician's own.

If, by chance, black box AI cannot be tested and approved in the same way as non-black box AI is expected to be, a further alternative to physicians becoming AI experts is for lawmakers to pass legislation that shields physicians from liability.[45] Legislation of this nature could, for example, specify that physicians will not be held liable for harm caused by reliance on a black box AI-CDS tool if they can demonstrate that they followed generally agreed-upon standards of practice. The existence and acceptance of these standards would of course have to be proven to a court of law if and when a lawsuit was brought against a physician. Legislation of this nature would serve to limit the chilling effect negligence law might otherwise have on the adoption of these technologies and would incentivize the creation of widely agreed upon standards before a tool is used clinically. How to establish those standards and who to include in that process is beyond the scope of this paper, but incredibly important.

A further challenge with legislation of this nature is of course that guidelines often become outdated and the cost to maintain them is very high. When new oncology drugs are approved by Health Canada, for example, a provisional algorithm has been developed by the Canadian Agency for Drugs and Technologies in Health that specifies when and how that pharmaceutical product should be used in various different care scenarios and pathways.[46] These algorithms are very specific and are updated every time something new comes into the market that impacts those scenarios or pathways. Because the work is done centrally by a regulator, it is much more manageable than if it were to be done in a de-

centralized manner by a community of practitioners who are using a particular black box AI tool. It can be incredibly resource-intensive and burdensome to even attempt to stay abreast of all other developments in the broader marketplace that could impact the appropriateness of how and when a specific technology is used.

Consider also that many AI tools are being developed and deployed in a site-specific (or health system-specific) manner. This is the case in the United States because the data that is relevant to individual health care sites is generally very different and so too are the types of tools that are most useful. It will accordingly take longer to develop broad standards of practice rather than site-specific ones for each individual tool. In Canada, by contrast, much of the work on AI is being done in the larger hospitals in the major cities that have access to the most donor money and the best data infrastructure. Innovation of this nature requires a hospital to assume some risk and hospitals that are publicly funded generally do not receive the amount of discretionary funds needed to innovate in this way. It is the Canadian hospitals that are closely affiliated with post-secondary institutions (with computer science and/or health science programs) and that have strong foundations to fundraise the capital needed to take risks of this nature that will be the first to advance AI initiatives.

If a regulatory testing and oversight scheme does not emerge soon and if lawmakers do not pass legislation to provide indemnity to physicians, we could possibly see insurers looking to reduce costs by requiring the use of AI products. Both patients' private health insurers and physicians' professional negligence insurers are certainly monitoring developments in AI closely and could require the use of particular AI tools in certain situations. If, for example, a physician does not use a diagnostic aid that their insurer thinks is safe and effective and if that decision results in a misdiagnosis leading to injury, the insurer may refuse coverage when the

patient brings a negligence action. It requires very little effort to imagine that insurance coverage exclusions could be based on which diagnostic tools were used in a particular case.

It could also be the case that hospitals step in to indemnify physicians in order to encourage them to use specific black box AI-CDS tools. Hospitals might want to do this if they are the ones who are building the AI tools to help improve care or even to save money at their site(s). Given that hospitals can be held vicariously liable for their physicians, depending on how the relationship is structured (*e.g.*, employment relationship vs. the physician being an independent contractor),[47] they should be reluctant to force physicians to use AI when a physician is not comfortable doing so. If physicians are resistant however, hospitals can assure them they will be indemnified if something goes wrong. It seems to us that in the absence of regulatory oversight or legislative innovation regarding liability for negligence, the likely short-term approach to facilitating broader use of black box AI in clinical health care that will be adopted by some hospitals, though not all, is to indemnify those who use AI tools that have been approved by the institution.

Each of the above scenarios allows the standard of care for using AI to relax because work is being done to ensure that the AI tools are tested, capable of being used in a manner that is safe and effective and that physicians are protected from liability when using them. It will likely continue to be the case that hospitals with the money to take risks will be where the most progress is made on clinical integration of black box AI. The regulation of black box AI is also a bit of a chicken and egg situation because regulators will not know what to look for when it comes to safety and efficacy (*e.g.*, whether AI ought to be considered as a site-specific tool) until they can observe a black box AI tool that has actually been integrated into clinical practice. Insurers will likewise not know what to require of their insured until they see how a tool actually per-

forms and legislators will not be able to rely on practice guidelines until a tool is actually being used in practice.

We should expect to see regulatory oversight continue to lag behind actual clinical implementation of AI-CDS in both the United States and in Canada. A black box AI-CDS tool is more likely to be integrated into clinical care if a hospital decides to develop and test that tool on its own and then to take responsibility (*i.e.*, indemnify users) for its use. Legal standards of care related to AI in health care are likely to evolve slowly for these reasons. In the interim, a technology-based solution may be the best way to mitigate the concerns physicians will likely have about using black box AI. Given that the alternative is to wait for regulators to catch-up to the hospitals that are innovating in-house, developers should be keen to explore innovative technical solutions. If developers ensure that tools are designed with an emphasis on explainability and/or bounded variability, as appropriate, physicians will be able to feel confident that they can accurately understand and explain the inherent limitations, risks and benefits of the tools they are using to their patients. A technical solution of this nature may be a good way to inspire physicians to move forward with the adoption of innovative AI technologies before robust regulatory testing and oversight is in place and appropriate standards of practice and care are firmly established.

VII.   Technical Solutions

Until one of the above solutions is in place, it will be important for physicians to receive robust advance training about how individual CDS tools are designed and how they work.[48] This training is a due diligence step that would help physicians to understand what a tool is looking for, how it looks for that data and how it analyzes relevant data once it is uncovered (*i.e.*, the policies built into the algorithm). Having this knowledge could help to instill confidence in physicians that they will be able to

meet the high standard of care the courts expect of them while leveraging new black box CDS tools in their practice. The requisite depth and breadth of this training could lessen over time, as individual tools gain wider use and their track records become more well-known.

In the interim, however, the black box nature of these tools will make things complicated. A black box CDS could suggest a course of action that does not conform with a physician's own assessment, yet the physician could still choose to proceed by adopting that suggestion. When this scenario leads to a patient being harmed, it is unlikely that a court will agree that the physician met the standard of care by simply trusting the CDS tool's analysis without also being able to independently verify that analysis (*e.g.*, such as by having a mechanism by which that CDS tool could offer a meaningful explanation). As long as there are potential risks associated with a black box CDS tool's use, it is highly unlikely that any degree of predictive accuracy will be high enough for a doctor to feel comfortable blindly trusting and acting on its analysis. As Madam Justice Cory A. Gilmore of the Ontario Superior Court of Justice explained in *Campbell v. Roberts*:

> Clinical judgment is not guesswork based on incomplete information. Exercises of clinical judgment are decisions made by medical practitioners once they have considered all of the relevant information it is reasonably possible to obtain under the circumstances. This includes the results of tests and consultations.[49]

If some of that information contradicts or simply does not accord with the physician's own judgment or prior knowledge, that physician will have to justify relying on it as part of their exercise of good clinical judgment.

All this uncertainty will work against manufacturers' interest in having their products used by physicians, but it could also give rise to new opportunities for creativity and context-specific design thinking. Manufacturers should want to design and build their black box CDS products so that

physicians can clearly understand how they will be able to meet the standard of care while using them. Being able to have this confidence in a black box CDS tool will make clinical adoption more likely, even if its recommendations do not accord with that physician's own judgment or they do not understand its underlying analysis.

There are at least two ways that black box CDS tools can be designed to help give physicians the confidence and information they need. The first way is that their user interfaces can provide a clinically meaningful explanation. What it means to be clinically meaningful is not so obvious however and will certainly be application specific. Researchers are actively looking for ways to make complicated black box analyses intelligible so that physicians can rely on them while still exercising good clinical judgment. The second approach applies only to black box CDS tools that are allowed to continually learn from and adapt to new data. These tools must be allowed to waver in their accuracy as they learn and retrain but must also be constrained from offering recommendations that fall outside a given range of accuracy. This constrained variability approach will only be acceptable for CDS tools in certain application areas. Specifically, application areas where there are already established medical, ethical and legal standards that can be integrated into the algorithm in order to establish outer bounds or limits. We will further explore the idea of risk classification for CDS tools in Part VI, and how risk classification can help us determine which design approach is optimal.

a. Explainability

Early advances in the science of artificial intelligence primarily placed emphasis on building models that could make accurate predictions from increasingly complex data. As the science has reached new heights and as the public has become fascinated by its real-world potential, use cases have emerged for which stakeholders have come to agree that AI must do more than simply conduct analysis and spit out recommendations. For decision-makers to want to rely on AI when making decisions for which they can be held accountable (*i.e.*, liable), an AI's analysis must somehow be understandable, and its recommendations must be defensible.[50]

In the context of using black box AI in health care then, it is important to be clear about the difference between two terms of art: interpretability and explainability. Black box algorithms conduct analysis that can involve complex, compounding, non-linear combinations of a broad assortment of variables. This analysis is so complicated that it would be impossible for a human to work through and trace (this is what we refer to when we use the language of 'interpretability'),[51] much less for a human to understand (this is 'explainability').[52] We must simply accept that interpretability is impossible when we use black box CDS tools. Given the way that standard of care analysis operates however, we cannot ignore the need for there to be some sort of meaningful explanation. Again, physicians must be able to demonstrate that they exercised reasonable clinical judgment and it is not obvious what this would look like if they were to rely on an analysis that they could not explain. We will of course not be able to fully explain an analysis that we already know we cannot interpret, but something else useful must still be possible if our goal is for physicians to want to use these tools. A physician will not be able to unpack a deep learning algorithm's analysis in order to fully explain it to a patient, but they must be able to understand and communicate the basis for the algorithm's recommendation if there is any risk that adopting that recommendation could potentially harm the patient.[53] What a useful explanation might look like will of course depend on each individual clinical setting and application that we are considering.

There are already some excellent examples of black box CDS tools built by multi-disciplinary teams that were attentive to the needs of clinical practice

and the importance of tackling the challenge of explainability through a well thought out user interface.[54] The difference between a good algorithm and a good clinical decision support tool is, to put it crudely, thorough design thinking that seriously considers actual clinical workflows and needs, human factors engineering, user interface and the user experience (*i.e.*, anything that is essential for successful clinical integration).[55] To assess whether a CDS tool is able to explain itself meaningfully, we must first acknowledge that there is a difference between understanding the whole of an analysis from start to finish and understanding and explaining the parts of an analysis that, once present, signal that a particular conclusion is likely (*i.e.*, that tells the user what essential information justifies the machine's recommendation, regardless of how the machine got there). If the user has been trained on how the CDS tool works and what the algorithm is looking for, the explanation that tool offers via its user interface might only need to clarify what the exact data was (*i.e.*, the signal) that ultimately triggered the recommendation. Again, what exactly that looks like will depend on the area of clinical practice in question and the use being made of the recommendation.

Two rather straight-forward examples of ways that a user interface can offer a visual explanation (as opposed to audible, mathematical, or tactile explanations, for example) could be by:

(1) displaying physiological signal data (*e.g.*, from an ECG or heartbeat monitor) and highlighting with a coloured visual overlay what part of the signal triggered the tool to make its prediction/recommendation; and,

(2) using coloured visual overlay on an image of a person (*e.g.*, their face or skin) or some other medically relevant image (*e.g.*, a radiograph or MRI) to indicate what the signal (*i.e.*, part of the image) is that caused the tool to make its prediction/recommendation.

An example of how this latter approach to explanation could work is found on the website of a DS tool called Face2Gene.[56] This company's software uses machine vision to analyze a picture of a person's face to detect known genetic disorders. It then uses colour coding to highlight the areas of the face that led it to its conclusion. The colour of the overlay corresponds to how heavily weighted (*i.e.*, how statistically significant) that area of the face is to the algorithm in its analysis. What is important for both these examples is that the information being highlighted to explain the tool's analysis is the type of clinically relevant information that a physician would be looking for if they did not have access to these tools. The tools themselves might have the advantage of picking up finer nuances in the case of Face2Gene or, in the first example, being able to efficiently analyze a greater amount of signal data than a physician could.

Unfortunately, visual explanations like these do not generally disclose or draw attention to risks and limitations inherent in the CDS tool and/or its predictions (*e.g.*, problems with the data the model relies upon and/or to what extent inaccuracy in the prediction could cause harm to a patient, *etc.*). These concerns might generally be included in a physician's training on how to use a tool, but continually drawing attention to prediction-specific risks with every recommendation offered would make these tools burdensome and would severely inhibit clinical integration. Because of the tort law duty to warn and the learned intermediary rule however, manufacturers of CDS tools will want to disclose all known risks and limitations in order to mitigate risk. One way we are starting to see this accomplished is through the use of something similar in kind to a material safety data sheet (AI-MSDS).[57] A tool-specific AI-MSDS might appear upon login to a user interface or might be physically located at a workstation where the tool is accessed and would include important information. This information, both simple and more advanced,

should be sufficient to allow the user to explain the tool's known use cases (*e.g.*, indications and contraindications), risks and limitations, whether certain important data is missing from the tool's analysis (*i.e.*, whether the analysis has to be supplemented with further clinical data that is unable to be digitized); whether training data is site-specific and/or fundamentally different than what would be normally encountered in a clinical setting; and whether there is over or underfitting of certain populations,[58] *etc*. How much information needs to be disclosed will of course depend on the nature of the risk inherent in the tool's use. Given the standard of care analysis presented above, coupled with the risks presented by the learned intermediary rule, physicians will want to know as much information as possible about risks and will want to relay that information to patients as part of the process of obtaining informed consent for care.

Challenges

There are also some important challenges with explainability that must be addressed if we hope to move forward with broad integration of these tools. First, research suggests that users of CDS tools that are more transparent will make more obvious mistakes because they will place more trust in the tool's analysis.[59] These transparent tools create what we might call a 'fake trust' that can only be guarded against through diligence and training. Physicians (and or the health care systems they work in, depending on who is sourcing the tool) must not take for granted that a good explanation is the same as an accurate analysis.

A second, yet complementary challenge, is that research has demonstrated there are ways to manipulate explanations to make them more trustworthy, but not necessarily more accurate or true to what the model is actually doing.[60] Clinicians must continue to be mindful of what information they would want to have available to them if they did not have access to a particular CDS tool. It will be difficult

to argue that a physician has met the standard of care expected of them if they are blindly/mindlessly over-relying on black box analysis and not trying to understand everything they can about that tool before integrating it into their practice.

As an alternative to the AI-MSDS, which could become so dense as to dissuade users from using particular CDS tools, and as a way to mitigate these two challenges with explainability that we have illustrated, there might also be blanket liability waivers between manufacturers and purchasers. These waivers could even be entered into between physicians and the hospitals in which they work, particularly since those hospitals might be held vicariously liable for the negligence of health care providers who are considered agents of that institution.[61] Physicians might refuse to use black box CDS tools unless the hospital supports them through training and indemnifies them for any resulting harm caused to patients. Given also that the duty to warn is ongoing, manufacturers that design black box CDS tools that continually learn will not be able to, nor want to, take their product off-line in order to verify and validate all new training data before using it to retrain the algorithm, assessing new risks and limitations, and disclosing those to the user. If we want black box CDS tools that continuously learn in real time to be integrated into clinical practice, manufacturers will need assurance that they will not be liable for failing to meet their ongoing duty to warn.

A blanket indemnity might also allow manufacturers to offer higher risk products, such as those that seek to understand what can be learned from the natural history of a patient with an undiagnosed rare disease.[62] Most rare diseases are poorly understood and accordingly have no standards available to help guide physicians or to help train algorithms. Explainability standards would have to be very high for physicians to want to rely on these types of tools in clinical practice relating to rare diseases. Some sort of blanket indemnity between the manufacturer and

the hospital as purchaser might incentivize more creativity and risk-taking by manufacturers when it comes to applications of this nature.

b. Constrained Variability

Addressing the need for explainability and the on-going duty to warn will prove to be slightly more challenging when we use AI models that can continuously learn. If continual learning CDS tools are built correctly and tested adequately, new data should not generally cause much variability in their accuracy or recommendations. That being said, we should want these kinds of tools to be able to explore and learn from new data so they can challenge our assumptions and alert us to something we might otherwise have missed. As W. Nicholson Price II explains, "black-box medical algorithms should not be artificially limited to only those applications that confirm what providers already know…Black-box medical algorithms provide tremendous possibilities for using big health data in ways that are not merely incremental but transformative".[63] Allowing for variability, depending on the classifiers you are inputting, *e.g.*, age, race, demographic information, *etc.*, may alert you to something you did not otherwise know, including that there is bias in your data collection, missing data, or inaccuracies in the algorithm's predictions for certain populations.[64] Two very simple examples of how this might arise is if the CDS tool suggests changing a drug's dosage in a way that does not conform with existing medical knowledge or if it suggests taking an unrelated drug based on a secondary effect that had not yet been noted in the literature but that is evident to the algorithm from the raw data.[65]

To minimize risk to the user, a notification that there is discrepancy between a real time prediction and an *ex ante* expectation of that tool's predictive outputs could be provided to the clinician (who might then have to consider whether to deviate from the CDS tool's recommendation due, for example, to socioeconomic and/or cultural considera-

tions relevant to the individual patient) or a hospital administrator who could prioritize further analysis. This type of notification system will be further explained below but is in line with the total product lifecycle ("TPLC") regulatory approach that the U.S. Food & Drug Administration has proposed for Artificial Intelligence/Machine Learning ("AI/ML") based software as a medical device ("AI/ML SaMD"). Specifically, the FDA's TPLC proposal requires ongoing monitoring and evaluation of an AI/ML SaMD product's performance.[66] If new data causes the algorithm to learn that the predictions upon which regulatory approval of the CDS tool were based are no longer accurate, then it will not be acceptable to continue to use that tool.

An effective way to design these types of CDS tools would be to ensure that they are only permitted to suggest courses of action that they predict, if adopted, will lead to outcomes that fall within a pre-defined range of accuracy (*i.e.*, a range of predictive accuracy that corresponds with that which is known to be acceptable within a particular clinical setting). This is the range for which they would have no doubt been approved by the FDA or Health Canada in which to operate. The permitted range of accuracy would be based on accepted standards of practice (*e.g.*, practice guidelines set by professional organizations,[67] ethics rules,[68] legal standards of care) and would be integrated into the underlying model. Because clinical practice guidelines are merely instructive of legal standards of care,[69] model design would also have to integrate any other relevant legal and ethical standards. It is accordingly important to ensure that developers work in multidisciplinary teams to create and properly integrate all relevant social, legal and technical standards and guidelines into their models whenever possible. If a model aims to accomplish a task for which there are no established standards or guidelines that can be used to limit its permitted range of accuracy, then the related CDS tool must not be permitted to learn from new data in the manner that

we are contemplating in this paper. Care must also be taken to ensure that CDS tools are not capable of circumventing legal and ethical standards and that developers do not inadvertently bypass undertaking rigorous legal and ethical analysis during the development phase and delegate those matters to the algorithms themselves.

If a model acquires new data that causes its predictive accuracy to fall outside the range of values for which it has been programmed and approved,[70] then oversight will become imperative. Recently added training data may have to be removed from the CDS tool (or the tool as a whole may have to be disabled) and the tool may have to either revert to working on the original training data or using the training data that was most recently validated. Poor accuracy could also indicate that new data added to the training set is evidence that the original model was not accurate or that there is a problem with the data (*e.g.*, bias or incompleteness). CDS tools must be designed so that they can allow and also recognize these fluctuations in accuracy and notify a designated system administrator who can then facilitate the analysis of the updated training data.[71] A notification would also be provided at the point of clinical care to alert the user to the fact that administrators are auditing the tool in order to determine what new data led to a change in the tool's predictive accuracy. It would then be important to undertake a broader consultation with the multidisciplinary development team before considering whether and how to re-deploy the CDS tool. Transparent and open communication with all impacted stakeholders would be incredibly important throughout this process.[72]

VIII. Risk Classification, Oversight and Accountability

AI-based CDS tools must be designed with input from multi-disciplinary teams, including legal experts, patients and families, clinicians, ethicists, computer scientists, *etc*.[73] Each of these contribu-

tors must bring with them an understanding of how to assess the level of risk tolerance that should be integrated into a particular tool. For example, a lawyer will understand the legal standard of care, an ethicist will understand clinical and research ethics guidelines and a clinician will understand practice guidelines. A CDS tool's risk level will be higher if its inaccuracy can potentially cause harm to a patient than if its inaccuracy would simply lead to something that carries no known associated risk of harm,[74] *e.g.*, an unnecessary, but harmless or low risk test is conducted.

The governance goals we strive for must focus on enabling the use of continual learning CDS tools without sacrificing quality of care. If new information causes unacceptable variations in a CDS tool's accuracy, that tool's pre-determined risk classification will indicate what the appropriate systems-level response needs to be (*e.g.*, alerting administrators to possible concerns about model accuracy or even immediately deactivating the tool). We do not propose here to provide a comprehensive list of what different types of AI-CDS tools should be associated with which risk levels in which settings. Such an analysis would need to be undertaken on a case-by-case basis and with an enterprise's risk management needs in mind.

Generally speaking, if a low or no-risk CDS tool acquires new data that causes its predictive accuracy to stray, we may not want to disengage the tool unless we believe the risk of relying on its recommendations has increased.[75] Instead, we may wish to keep the model active and conduct something like A/B testing. A/B testing is a way to compare two versions of something (*i.e.*, A and B), that is traditionally used in marketing and website design but could also prove to be an effective tool in this setting. Assume that both A and B are identical except for one variation that might affect a user's behaviour. Version A might be the version of the tool currently in use (*i.e.*, the control or status quo), while version B would be the modified version

(*i.e.*, the black box recommendation that differs from the status quo).[76] In the event that both A and B are low or no risk scenarios, the system could seamlessly recommend either course of action to its user (*e.g.*, discharge a patient today with a referral or monitor for one more day and discharge tomorrow with no referral). The tool could then track both A and B and learn from all reported outcomes. To be clear, no new data about either A or B would be added into the CDS tool's ongoing real time analysis until the results of those choices (*i.e.*, the patient's associated health outcomes) are known and inputted. This user update step is important because it allows the CDS tool to learn whether its recommended variation from the status quo did in fact prove effective for that patient. It could also be the case that neither A nor B have an impact on patient health, but that A saves health care dollars. A/B testing could therefore be a very effective way to help uncover spending inefficiencies without exposing patients to unnecessary risk.

Not every novel recommendation that a CDS tool makes will be low risk. As noted above, a CDS tool must alert administrators and/or users to significant changes in its accuracy resulting from the integration of new training data. This notification should engage an oversight mechanism that causes a team member to analyze the updated training data in order to determine what data was in the training set when it was accurate and what was added to cause it to become inaccurate. Working with clinicians and researchers to understand what the new data is telling us, it could then be possible to leverage these new insights in order to improve the model or even conceptualize new ideas for research that could be conducted, including novel clinical trials. The results of that research might then contribute to the body of knowledge used to train the CDS tool's underlying algorithm.

Depending upon their expected variability and the agreed upon risk levels that have been integrated into their design, it may also be important to have

mandatory technical reviews form part of the administrative oversight of certain CDS tools. This is especially true for those tools that are acquiring a large volume of new data or that are being deployed in areas of clinical practice about which relatively little is known (*e.g.*, for many rare diseases). This is again in line with the U.S. FDA's proposed TPLC approach for AI/ML SaMD. If periodic review is deemed necessary, then either the original design team, an institutional governance team, independent consultants or some combination of the above could be consulted during those reviews. A systems governance framework might also require that CDS tools at all risk levels be regularly audited so that any concerns about bias, data completeness, clinical utility, *etc.*, could be proactively addressed. As their track record for safety and effectiveness grows, the need for this oversight might lessen. Until then, mandatory periodic review would help to build trust in AI-CDS tools.

IV. Conclusion

Broad integration of AI-DS tools into clinical care is inevitable. With the huge investments being made into AI and computational medicine throughout the world, the future of health care may actually be closer than the casual observer can appreciate. It is accordingly important for us to begin to consider the implications of the fact that some very useful and accurate artificial intelligence methodologies conduct analysis that is opaque to humans. We should still want to use these systems despite their opacity, but we must work diligently to understand what the ethical, legal and social implications of this might be so that we can work towards efficient and effective system design and governance without sacrificing patient trust or care. We must be able to use AI-based CDS tools that are flexible enough to allow us to learn and benefit from new data we are collecting. To be able to do this, the manner in which these tools are integrated into care must be wholly above reproach.

In this paper, we have argued that black box AI-CDS tools have an important role to play in the future of health care. This is especially the case for those tools that can continually learn in real time. Manufacturers will be eager to get these products to market, but physicians may be hesitant to use them. This hesitation will come in part from the fact that physicians act as learned intermediaries and assume responsibility for risks disclosed to them by the manufacturers of these tools. This principle of tort law could have a chilling effect on the uptake of black box CDS tools until physicians feel confident that they will not be taking on unnecessary risk when using them in their practice.

We have argued that there are two important ways manufacturers can meet this challenge. The first is to provide users with clinically meaningful explanations of a black box CDS tool's reasons for its recommendation(s), and the second is to ensure that those black box tools that can continuously learn are programmed so that variability in the accuracy of their predictions is constrained by all applicable standards and guidelines. The risk classification levels assigned to each tool must dictate how we handle oversight and accountability, and immediate action must be taken by administrators if, for example, new data causes a tool with associated risk to offer a recommendation that falls outside of the range of options that it has been programmed to deem acceptable. Notifying administrators who oversee a particular tool will allow those administrators to work with the design team to identify any issues with the tool's design, the data it is learning from, and/or to uncover ideas for innovative research.

Finally, it is important to remember that the CDS tools contemplated in this paper are to be characterized and treated as support tools in our current medical and regulatory environments. They are not stand-alone or autonomous decision-makers. Although this regulatory environment may change in the future, we should consider these tools as merely enhancing and not supplanting a physician's professional judgment. How we manage the learned intermediary rule could accordingly play an important role in the uptake of AI into clinical care. As things stand, it is unlikely that many AI-based black box DS tools will make it into clinical use unless there is either no or very little risk to the patient and/or the clinician. That being said, it is also possible that diligent and well-trained physicians will find ways to meet the (as yet unclear) standard of care expected of them or will arrange to be indemnified if meeting that standard is not in fact possible.

[**Ian Stedman**, LLB, MA, PhD, is an Assistant Professor, Canadian Public Law and Governance in the School of Public Policy and Administration at York University. His research focuses on law, public policy and ethics related to technologies that are driving greater personalization in health care. He currently sits as a member of the CIHR Institute of Genetics' Institute Advisory Board and as a legal member on the Research Ethics Board at The Hospital for Sick Children. His work in the area of law and technology earned him the IP Osgoode David Vaver Medal for Excellence in IP Law in 2020. Ian can be contacted at <istedman@yorku.ca>.]

[**Michael Brudno**, PhD, is a Computer Scientist specializing in computational methods for the analysis of medical data, working as a Professor at the University of Toronto, the Chief Data Scientist for the University Health Network (UHN), and as a faculty member at the Vector Institute for Artificial Intelligence. His research investigates the capture of structured phenotypic data from clinical encounters, using both refined user interfaces, and mining of unstructured data (mostly text mining, based on machine learning methodology), and the analysis of omics data (genome, transcriptome, epigenome) in the context of the structured patient phenotypes, mostly for rare diseases. His goal is to streamline clinical workflows and enable faster and better treatments by creating seamless automated analyses

of patient omics data based on automatically captured information from clinical encounters. Michael can be contacted at <michael.brudno@uhn.ca>.]

1. The concept of "black-box medicine" was introduced into legal scholarship in 2015 by W. Nicholson Price II in "Black-Box Medicine" (2015) 28:2 Harvard J. of L. & Tech. 420.

2. See Mark P. Sendak, *et al.*, "A Path for Translation of Machine Learning Products into Healthcare Delivery" (2020) European Medical Journal – Innovations, DOI/10.33590/emjinnov/19-00172 (where the authors note that "[t]o date, no machine learning product ingesting EHR data has successfully diffused and scaled across healthcare").

3. W. Nicholson Price II, "Regulating Black-Box Medicine" 116 Mich. L. Rev. 421 at 425.

4. Eric Topol, *Deep Medicine* (New York: Basic Books, 2019) at 44.

5. Eric J. Topol, "High-performance medicine: the convergence of human and artificial intelligence" (2019) 25 Nature Medicine 44.

6. See *e.g.*, Frank Pasquale, "Grand Bargains for Big Data: The Emerging Law of Health Information" (2013) 72:3 Maryland L. Rev. 682.

7. See *e.g.*, Suchi Saria, Atul Butte & Aziz Sheikh, "Better medicine through machine learning: What's real, and what's artificial?" (2018) 15:12 PLoS Medicine.

8. Milena A. Gianfrancesco, *et al.*, "Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data" (2018) 178:11 Journal of American Medical Association - Internal Medicine (November) 1.

9. This is called supervised machine learning. See Sendak, *et al.*, *supra*, note 2 (where the authors detail 21 different models and explain at p. 3 that "[m]ost models integrated into clinical workflows as clinical decision support are supervised machine learning models").

10. Pelaccia Thierry Pelaccia, Germain Forestier & Cédric Wemmert, PhD, "Deconstructing the diagnostic reasoning of human versus artificial intelligence" (2019) Canadian Medical Association Journal (2 December), doi: 10.1503/cmaj.190506.

11. Christopher C. Kelly, *et al.*, "Key challenges for delivering clinical impact with artificial intelligence" (2019) 17 BMC Medicine 195, doi:10.1186/s12916-019-1426-2.

12. Emily LaRosa & David Danks, "Impacts on Trust of Healthcare AI", Proceedings of the 2018 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society, online: Artificial Intelligence, Ethics, and Society <www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_132.pdf>.

13. We explain why these types of CDS tools are important below.

14. Omer Gottesman, *et al.*, "Guidelines for reinforcement learning in healthcare" (January 2019) Nature Medicine 25.

15. Cecilia S. Lee & Aaron Y. Lee, "Clinical applications of continual learning machine learning" (June 2020) 2 Lancet Digi Health.

16. The technical benefit of reinforced continual learning is that it can be used to help neural networks overcome a well-known problem they have of forgetting previously learned information. The term of art used to describe this problem is "catastrophic forgetting" or "catastrophic interference".

17. Boris Babic, *et al.*, "Algorithms on regulatory lockdown in medicine" (6 December 2019) 366:6470 *Science Magazine* 1202.

18. This is the difference between methodologies that are called 'supervised' and 'unsupervised' learning.

19. Continual learning has been described as "the ability to learn consecutive tasks without forgetting how to perform previously trained tasks". See Ju Xu & Zhanxing Zhu, "Reinforced Continual Learning" (preprint), online: arXiv <https://arxiv.org/pdf/1805.12369.pdf>.

20. W.N. Price, S. Gerke & I.G. Cohen, "Potential Liability for Physicians Using Artificial Intelligence" (2019) Journal of the American Medical Association (4 October), doi: <https://doi.org/10.1001/jama.2019.15064>.

21. Yaniv Benhamou & Justine Ferland, "Artificial Intelligence and Damages: Assessing Liability and Calculating Damages" in G. D'Agostino, A. Gaon & C. Piovesan, eds., *Leading Legal Disruption: Artificial Intelligence and a Toolkit for Lawyers and the Law* (Canada: Thomson Reuters, 2021) 165.

22. A. Michael Froomkin, Ian Kerr & Joëlle Pineau, "When AIs outperform doctors: confronting the challenges of a tort-induced over-reliance on machine learning" (2019) 61 Ariz. Law Rev. 33 [hereinafter, "*Froomkin, et al.*"].

23. See *e.g.*, *supra*, note 20, at E1.

24. The Canadian Medical Protective Association, "Can I get an [artificial] second opinion?: Benefits and risks of AI technologies in medicine" (2019), online: Canadian Medical Protective Association <https://www.cmpa-acpm.ca/static-assets/pdf/about/annual-meeting/19_annual_meeting_ai_paper-e.pdf> (This CMPA article is adapted from "A Primer on Law, Risk and AI in Health Care," published in *Healthcare and Life Sciences Law Committee Update* (Vol. 3 no. 1, Sept. 2018)).

25. *Ibid*.

26. See Andrew D. Selbst & Solon Barocas, "The Intuitive Appeal of Explainable Machines" (2018) 87 Ford L. Rev. 1085 (for a discussion of how important it might be for a

physician to have specialized knowledge to allow them to understand an algorithm's source code).

27. See *e.g.*, *Froomkin, et al.*, *supra*, note 22; Maia Jacobs, *et al.*, "How machine-learning recommendations influence clinician treatment selections: the example of the antidepressant selection" (2021) 11:108 Translational Psychiatry 1.

28. *Lambert v. Lastoplex Chemicals Co.*, [1972] S.C.R. 569.

29. *Hollis v. Dow Corning*, [1995] 4 S.C.R. 634 at 635-636.

30. *Ibid.*

31. *Buchan v. Ortho Pharmaceuticals (Canada) Ltd.*, [1986] O.J. No. 2331 (C.A.) (in which the plaintiff argued that Ortho Pharma had a duty to warn her directly about known risks, and not simply to warn her physician. Her argument was that taking birth control pills is normally a decision that a woman makes on her own (*i.e.*, to avoid getting pregnant) rather than simply being a prescription that she is given by her doctor in response to health concerns).

32. *Reyes v. Wyeth Laboratories*, 498 F.2d 1264 at 1276 (5th Cir. 1974), followed by the Supreme Court of Canada decision in *Hollis v. Dow Corning Corp.*, [1995] S.C.J. No. 104, 129 D.L.R. (4th) 609 at 622.

33. Michael A. Jones, *Medical Negligence*, 5th edition (London, UK: Thomson Reuters, 2018).

34. Lorain Hardcastle, *Introduction to Health Law in Canada* (Toronto: Emond Montgomery Publications Ltd, 2019) at 21.

35. *Ibid.*, at 27.

36. See *e.g.*, *Barnett v. Chelsea and Kensington Hospital Management Committee*, [1969] 1 Q.B. 428.

37. See *e.g.*, *Cottrelle v. Gerrard*, [2003] O.J. No. 4194, 67 O.R. (3d) 737 (C.A.) (where experts could not definitively say that a mistake in diagnosis in treatment by the physician defendant would have in fact made a difference in the plaintiff's health outcomes).

38. I. Glenn Cohen, "Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?" (forthcoming 2020) 108 Georgetown L.J.

39. W. Nicholson Price II, *et al.*, "Potential Liability for Physicians Using Artificial Intelligence" (2019) 322:18 Journal of the American Medical Association 1765.

40. Michael D. Greenberg, "Medical Malpractice and New Devices: Defining an Elusive Standard of Care" (2009) 19:2 Health Matrix: The Journal of Law-Medicine 423 at 429 [hereinafter "*Greenberg*"]; see also *Froomkin, et al.*, *supra*, note 22 at 61.

41. *Froomkin, et al.*, *ibid.*, at 99.

42. *Ibid.*, at 56.

43. See *e.g.*, *Greenberg*, *supra*, note 40 at 424 ("Negligence and malpractice doctrine generally make it clear that standards of care are evolutionary rather than static, and that providers have an obligation to stay abreast of new techniques and developments").

44. This is something that both the FDA and Health Canada were actively working on in 2020.

45. See W. Nicholson Price II, "Artificial Intelligence in Health Care: Applications and Legal Implications" (2017) 14 Sci. Tech Lawyer 10 at 13 (where the author suggest that it is worth considering whether the learned intermediary rule may need to "bow to the recognition that doctors cannot fully understand all the technologies they use or the choices it helps them make when they are not provided the needed and/or necessary information").

46. "CADTH Provisional Funding Algorithms" (13 February 2021), online: Canadian Agency for Drugs and Technologies in Health <https://www.cadth.ca/cadth-provisional-funding-algorithms>.

47. *Greenberg*, *supra*, note 40 at 437-438.

48. See *e.g.*, *Greenberg*, *ibid.*, at 435 (where the author argues that "[t]he first, obvious criterion for using a new device appropriately involves knowing something about its safety and effectiveness".)

49. *Campbell v. Roberts*, [2014] O.J. No. 4934, 2014 ONSC 5922 at para. 100.

50. This discussion comes up often in the context of public sector uses of AI, where the principles of administrative law require that a decision-maker who is exercising discretionary power must also be prepared to give reasons. See *e.g.*, Marion Oswald, "Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power" (2018) 376:2128 Philosophical Transactions of the Royal Society A, online: <https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2017.0359>.

51. See *e.g.*, Finale Doshi-Velez & Been Kim, "Towards a Rigorous Science of Interpretable Machine Learning" (2017), online: arXiv <https://arxiv.org/pdf/1702.08608.pdf> (where the authors define interpretability as "the ability to explain or to present in understandable terms to a human").

52. See *supra*, note 22 at 48 (for a discussion of explainability in clinical AI).

53. Jianxing He, *et al.*, "The practical implementation of artificial intelligence technologies in medicine" (January 2019) 25 Nature Medicine 30 at 36 (where the authors argue that "...in certain instances, enforcing transparency and interpretability can potentially result in decreased accuracy or predictive performance of a model. And certain scenarios may not require interpretability, such as if the model has no significant impact or if the problem is well-studied".)

54. See *e.g.*, Mark P. Sendak, *et al.*, "Real-World Integration of a Sepsis Deep Learning Technology Into Routine

Clinical Care: Implementation Study" (2020) 8:7 JMIR Medical Informatics e15182.

55. See *e.g.*, Edward H. Shortliffe & Martin J. Sepúlveda, "Clinical Decision Support in the Era of Artificial Intelligence" (2018) Journal of the American Medical Association (5 November) (for a discussion of what capabilities and characteristics the authors believe must be incorporated into a CDs tool if it is to be accepted and integrated directly into clinical care); Sendak, *et al.*, *supra*, note 2 (where the authors explain that "[m]ost implementations of clinical decision support do not have the intended effect because of the difficulty with clinical integration"); Sujay Nagaraj, *et al.*, "From Clinic to Computer and Back Again: Practical Considerations When Designing and Implementing Machine Learning Solutions for Pediatrics" (2020) 6:4 Current Treatment Options in Pediatrics 336 (where the authors chart a path forward for the clinical integration of AI in pediatric health care).

56. Face2Gene (accessed 13 February 2021), online: Face2Gene <https://www.face2gene.com>.

57. An example of this type of risk information sheet has been associated with Duke Health Systems' ML CDS tool called Sepsis Watch, which monitors patients for signs of sepsis.

58. Overfitting means that the analysis is based on data that corresponds too closely with, or exactly to, some classifier(s), *e.g.*, a particular demographic population. A model like this would almost invariably have less predictive accuracy when applied to other matters/classifiers/populations.

59. Forough Poursabzi-Sangdeh, *et al.*, :Manipulating and Measuring Model Interpretability" (8 November 2019), online: arXiv <https://arxiv.org/abs/1802.07810>.

60. Himabindu Lakkaraju & Osbert Bastani, "How do I fool you?: Manipulating User Trust via Misleading Black Box Explanations" (15 November 2019), online: arXiv <https://arxiv.org/pdf/1911.06473.pdf>.

61. W. Nicholson Price II, "Medical Malpractice and Black-Box Medicine" in I Glenn Cohen, *et al.*, eds., *Big Data, Health Law, and Bioethics* (UK: Cambridge University Press, 2018) at 12.

62. See *e.g.*, W. Nicholson Price II & Arti K. Rai, "Clearing Opacity Through Machine Learning" (2021) 106 Iowa L. Rev. 775 at 787; Emma Pierson, *et al.*, "An algorithmic approach to reducing unexplained pain disparities in underserved populations" (2021) 27 Nature 136.

63. W. Nicholson Price II, "Big data and black-box medical algorithms" (12 December 2018) 10 Science - Translational Medicine 3.

64. See *e.g.*, Boris Babic, *et al.*, "Algorithms on regulatory lockdown in medicine" (6 December 2019) 366:6470 Science Magazine 1202. (For a discussion of some of the risks that can be associated with adaptive algorithms).

65. *Supra*, note 61, at 8.

66. U.S. Food and Drug Administration, "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML) Based Software as a Medical Device (SaMD): Discussion Paper and Request for Feedback (April 2, 2019), online: U.S. Food and Drug Administration <https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/SoftwareasaMedicalDevice/UCM635052.pdf>.

67. Patricia R. Recupero, "Clinical Practice Guidelines as Learned Treatises: Understanding Their Use as Evidence in the Courtroom" (2008) 36:3 Journal of the American Academy of Psychiatry and the Law 290.

68. Including, for example, those established by institutional review boards / research ethics boards.

69. See *e.g.*, Brian K. Cooke, *et al.*, "The Elusive Standard of Care" (2017) 45:3 Journal of the American Academy of Psychiatry and the Law 358; Angela Campbell & Kathleen Cranley Glass, "The Legal Status of Clinical and Ethics Policies, Codes, and Guidelines in Medical Practice and Research" (2011) 46 McGill L.J. 473 (for discussions about caselaw in both Canada and the United States that support the conclusion that although clinical practice guidelines may be accepted by the court as the legal standard of care, it is not automatically the case simply because a Guideline has some organization's seal of approval).

70. The FDA and Health Canada are responsible for deciding whether and how to regulate medical software, including that which adapts in real time to new data. Depending on the use case and the risk levels inherent therein, black box AI may need to be approved before it can be deployed in health care settings. Software that is then allowed to change may have to be withdrawn from use and reconsidered by the relevant regulator. See Thomas J. Hwang, Aaron S. Kesselheim & Kerstin N. Vokinger, "Lifecycle Regulation of Artificial Intelligence and Machine Learning-Based Software Devices in Medicine" (22 November 2019) Journal of the American Medical Association E1 (for a good discussion of the FDA's role in pre-market approval and post-market monitoring/regulation).

71. There may be privacy implications depending on who is responsible for this task, but any such concerns are beyond the scope of this paper. See Roger Allan Ford & W. Nicholson Price II, "Privacy and Accountability in Black-Box Medicine (2016) 23:1 Mich. Telecom & Tech. L. Rev. at 3 (for a discussion of the privacy implications of black box medicine).

72. See *supra*, note 70 (where the authors argue with respect to modifications made to learning algorithms that "The FDA should also require a high standard of transparency…manufacturers should be required to document and publicly post a summary in an understandable format");

See also Sendak *et al.*, *supra*, note 2 (where the authors describe 21 different CDS tools that have been deployed by health care institutions across the United States and explain that many "are pursuing 'stealth science' to protect trade secrets and avoiding regulatory and academic scrutiny…many of the figures marketed by product developers have no peer-reviewed evidence".

73. See Jenna Wiens, *et al.*, "Do no harm: a roadmap for responsible machine learning for health care" (2019) 25 Nature Medicine 1337, doi:10.1038/s41591-019-0548-6; Jianxing He, *et al.*, *supra*, note 54 at 36.

74. *Supra*, note 66 at 5 (Figure 1).

75. Tools to help us understand things like undiagnosed rare diseases may in fact be very low risk, depending on the intervention they are recommending. There might otherwise be no established standard approach to care for these patients and an unconstrained ML-driven CDS tool could open up new ideas for research, treatment, symptom management, *etc*.

76. Michael Barber, "Data science you need to know! A/B testing" (2018) Towards Data Science (28 January), online: Medium <https://towardsdatascience.com/data-science-you-need-to-know-a-b-testing-f2f12aff619a>.

Note: The articles included in *Health Law in Canada* reflect the views of the individual authors. *Health Law in Canada* is not intended to provide legal or other professional advice, and readers should not act on information contained in this publication without seeking specific advice on the particular matters with which they are concerned.