Article details: 2019-0151	
Title	Ethical concerns around AI and healthcare from the perspectives of meningioma patients, caregivers, and healthcare providers: a qualitative study
Authors	Melissa D McCradden PhD, Ami Baba MRes, Ashirbani Saha PhD, Sidra Ahmad, Kanwar Boparai, Pantea Fadaiefard, Michael D. Cusimano MD PhD
Reviewer 1	Dr. Donald J. Willison
Institution	Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ont.
General comments (author response in bold)	Thank you for the opportunity to review this paper. The authors seek to extend the extant literature around secondary use of personal health information to address use of this information for AI-related healthcare applications. MAJOR COMMENTS: 1. METHODS: It was a bit of a challenge to piece together exactly how the authors conducted the study. There were a number of gaps in describing their methods that should be clarified. Given the CMAJ audience, it may be that not all of these need to be included in the body of the manuscript. However, they are important issues when reviewing the paper: Sampling Frame: 1.1. The authors indicate that participants were “consecutively recruited”. It is not clear what this means. Did the authors recruit from the chosen clinic on a particular day and approach each patient that day? Was it one day only or over a number of days? If patients were selected from one clinic on one day, then this may introduce additional selection biases. **This study was part of a larger study aimed at developing a meningioma-specific quality of life questionnaire. For this larger study, we recruited patients diagnosed with a meningioma, their informal caregivers, and healthcare providers involved in the care of meningioma patients. Recruitment took place in the senior author’s (MDC) neurosurgery clinic during the period of August 2018 – February 2019. Patients were consecutively recruited, based on the inclusion/exclusion criteria of the larger study. In other words, recruitment took place on a number of days, whenever the senior author held an outpatient clinic during the specified study period. We clarified this within the manuscript, under Methods – Design and setting (page 3) by adding the study period, along with the clarification that recruitment took place within MDC’s clinic.** 1.2. Health care providers were sampled through “snowball sampling” but there is no information as to how this was done. For example, were these HCPs from the same clinic? Were they all in the same specialty? This may help to account for the homogeneity of responses in this group. **As this study was part of a larger study as stated earlier, we were confined to recruit HCP who were normally part of the circle of care of meningioma patients. The views and responses of this group may reflect those commonly held by individuals who work within neurosurgery; however, we aimed to recruit a diverse range of healthcare professionals. The healthcare providers that were participated in the study came from a variety of backgrounds, and included neurosurgical residents, medical administrative assistants, nurse practitioners, physiotherapists. This is a bias for sure, but we have limited reason to think the responses are substantively different from other HCPs as most relied heavily on clinicians’ primary duties of care**

to patients in their responses. At the time there were no AI tools, no AI research going on that would have given them a unique insight into the issue.

Scenarios and questions asked of participants:
1.3. It is not entirely clear what questions were asked of the participants. While Supplementary file #2 provides the vignette guide, this appears to not include all the questions which were asked, as the body of the manuscript discusses issues that do not appear to be discussed in the vignettes. In particular, I found no discussion of de-identification in the scenarios. Yet, there appeared to be substantial discussion of this topic. Supplementary file #2 should include the interview guide.
Supplementary file number 2 has been provided and consists of the vignette guide. The interview guide acted as a starting point for initiating discussion with participants and the prompts served to ensure that all participants spoke to the same concepts. We have clarified in the manuscript that this was our procedure. To the reviewers comment, the prompt under scenario 1 following the question about whether Jaimie should ask all Atlantans for permission to see their information is “what if there was no identifying information included in the database?” From this question participants spoke to issues of de-identification, and in fact most explicitly called it this without us having to introduce the term. Those who brought it up themselves were asked what they thought was meant by the term, whereas anyone who asked us was given the definition which is now denoted next to the prompt. This is likely because most of these participants also consented to another study by the same PI which involved significant data linkage (including biobanking, EHR data, etc) and so de-identification was a strong element within the consent process for that study. We found generally that without prompting most participants repeatedly brought up ‘de-identification’. We acknowledge that this level of knowledge may introduce a bias and have noted this in the limitations section. Further, in Scenario 3 we asked participants about the acceptability of selling ‘identifiable’ data.

1.4. I was surprised to see no explanation in scenario 1 of the conditions surrounding the secondary use of the data – for example, that it would involve use of de-identified records without the need for anyone to go through the actual records. In fact, the end of the paragraph implies to the naïve reader that the researcher would actually go through individual records as it reads “Jamie is thinking about whether or not to ask… for permission to see their information.” People’s understanding of how the data are being accessed and used has an important bearing on their responses. Given that the interviewers provided no additional information if asked, there is considerable room for misunderstanding the use conditions. This is a particular concern as it sets the scene for all discussions around secondary uses.
We thank the reviewer for pointing out this potential limitation as we agree it could change the way people interpret the scenario. However, we might point out that we intentionally left it vague because in some cases with AI researchers are looking at actual records with identifiable data. As this is a preliminary study on general attitudes toward AI, and because AI has barely yet encroached into clinical care, we thought it a bit premature at this point to try and narrow down exactly what sort of uses participants would and would not be accepting of.

Interviews:
1.5. Participants were asked to “divine the perspectives of a hypothetical public as a means of assessing their own opinions.” This is very confusing. Were they asked to provide their own opinion or to imagine what the average person may think?
Participants were asked to consider their own opinions, while also thinking about how they think the characters as presented within the scenarios would react to each situation. For the sake of clarity, we have revised the wording of this specific sentence on page 4, under “Development of AI vignettes.” The sentence now reads: “Participants were told these scenarios were examples of realistic but hypothetical AI-enabled research. Participants were asked to share their own opinions, while also considering how characters in the vignette would react; in addition, we asked about their current knowledge of AI.”

1.6. Were the scenarios read to the participants or did they have the opportunity to read the scenarios themselves? If the former, were the questions revealed in staged fashion?
The scenarios were read to the participants. Participants were not given the opportunity to read the scenarios themselves, though they were told by the research personnel conducting the interview that they could ask for portions to be re-read to them should they need to hear it again. Each scenario was presented to participants in a staged fashion.

Analysis:
1.7. In the Analysis paragraph, the authors indicate that they conducted a “directed content analysis”. Does this mean that they looked only for things that fit into their pre-defined concepts? Taking this deductive approach runs the risk that the investigators would ignore messages emerging that did not fit their pre-conceptions. Please clarify.
We agree with the reviewer that we took a deliberate deductive approach in this case. This is because the literature on the issue of health data research is extremely well-developed and the main themes are consistently reproduced even recently (e.g., Paprica et al., 2019). We thus sought to reproduce the work with the explicit lens of AI research to assess whether AI might change how participants thought of the same concepts that have been explored previously. The only new theme was ‘deference to computer outputs’ as this was the direct goal of Scenario 2. Exploring a more fulsome sense of AI-specific considerations of these concepts could be the topic of future work in this area.

2. RESULTS
2.1. There is no indication of the response rate on the part of patients or caregivers. There is an indication in Supplemental file #1 that one patient chose not to participate. Was that the only person who declined to participate?
All patients and caregivers who were approached to participate consented to doing so, with the exception of the one patient who declined to participate.

2.2. In paragraph 1, lines 3 and 4, the authors state that patients and caregivers held more diverse views than did the providers. However, they do not state whether or not they achieved theoretical saturation among patients and caregivers. This information may be found at the very end of Supplementary file 1, but should

be included in the main text.
As suggested, we have taken the section from Supplementary file 1 where saturation is discussed, to be included within the Results section on page 4.

2.3. Page 5, lines 18-19 indicates that “Most participants cited ‘de-identification’ as a satisfactory condition for non-consented use of health data for research.” It is not clear how the term “de-identification” was introduced into the discussion. Given that a quantitative response (76%) was provided as to opinion on this topic, the authors should indicate how this was raised in the interview. Were participants advised that deidentification is not synonymous with anonymization – i.e. that it may still possible to re-identify individuals?
As described above, the quantitative value for this question was in response to the first question prompt of scenario 1 regarding ‘identifiable information’. We chose this term because we did not think it useful in this context to get into granularity about de-identification vs anonymization etc., and particularly these terms are less relevant to AI research. We are grateful for the reviewer’s comment as it has allowed us to clarify as per the revised manuscript under the section “conditions of the use of data…”

3. LIMITATIONS
3.1. Page 8, lines 23-26 indicate that the study population had heavy involvement in healthcare. Beyond heavy involvement, these were patients and caregivers with “high stakes” health conditions. So, this does constrain even more the generalizability of the findings.
Thank you for pointing this out. We agree that this is a unique population of participants and have acknowledged that the patient and caregiver participants are those who are managing with a serious health condition, which may potentially affect their viewpoints on the involvement of AI in healthcare. The impact that this has on the generalizability has been acknowledged within the limitations section on page 8.

3.2. There is no consideration of the generalizability of the findings from the health care providers. But we also know very little about these participants beyond what we learn from Table 1. And from here, we learn they were all female and all but one under the age of 50. Given the homogeneity of responses from the HCPs, there may be some limitations due to selection bias.
As this study was part of a larger study as stated earlier, we recruited HCP that were normally part of the circle of care of meningioma patients. The healthcare providers that participated in the study came from a variety of backgrounds, and included neurosurgical residents, medical administrative assistants, nurse practitioners, and physiotherapists. Due to the small number of healthcare providers (n = 5), we elected not to disclose how many participants fell within each category to protect the identity of the participants. The views and responses of this group may reflect those commonly held by individuals who work within neurosurgery; however, we aimed to recruit a diverse range of healthcare professionals. However, as we did recruit healthcare professionals through snowball sampling, we acknowledged the limitations associated with this recruitment strategy within the Limitations section on page 8.

3.3. Given the issues raised under “Methods” above, it may be necessary to discuss limitations about our inference on participants’ understanding of the

	scenario, particularly around how the records are being used. **We thank the reviewer for this note and have revised the limitations section on page 8.** 4. CONCLUSIONS 4.1. Page 8, line 40, the authors state “Overwhelmingly, there is broad support for health data use in research…” This generalization is overly-broad. The statement implies unconditional support. At the very least, the authors identified at least 2 restrictions: (1) under conditions of de-identification of the data; for research purposes “for the purpose of improving medical care”. **Thank you for pointing this out. We have revised this section to reflect the restrictions that were identified within this study. The conclusions section on page 8 now reads: “While there is broad support for health data use in research not just in Canada but elsewhere as well, this study has identified certain caveats to the use of data, such as the desire for de-identification of data, in addition to the overarching objective of improving health and medical care. This is consistent with other studies, which cite privacy as the main concern from the public.”** 4.2. The last sentence reads: “As such, AI’s endorsement by society is overdue for public education initiative to earn trust.” This statement suggests that any limitations to people’s trust in AI is due to a knowledge-deficit. That is a very sweeping generalization which is not supported by either the empirical findings or anything raised in the Interpretation section of the manuscript. **Thank you for pointing this out. To align better with the Interpretation section within the manuscript, we have removed that last sentence in the Conclusions on page 8. Instead, we have stated that the study is consistent with the findings present in other studies, but that there are certain caveats concerned with the use of health data.** MINOR COMMENTS: 1. Abstract, page 2, line 25. “could” should probably be replaced with “should” **Thank you for this suggestion. We have changed the “could” to “should” in line 25 on page 2.** 2. The document needs some copy-editing to address some missing words and syntactical errors. **Thank you for pointing this out. We have read through the manuscript to correct syntactical errors.**
Reviewer 2	Dr. DC Engel
Institution	Department of Neurosurgery, Bern University Hospital, Bern, Switzerland
General comments (author response in bold)	This manuscript deals with and shows different aspects of medical research. AI is probably often not thought of as a topic to deal with in day to day life for many researchers, patients and health care providers. The topic is current and innovative and certainly deserves to be written about. 1. It took me a while to get into this theme whilst reading. Maybe it is because English is not my first language, or because the topic of your manuscript isn't my daily bread and butter, but the manuscript might benefit of a more easier and language to pick up and entice the reader to read on in the introduction. **Thank you for this feedback. We have gone through the manuscript and**

	revised certain sections to ensure ease of readability. 2. Some of the referred manuscripts were written in a different country, with huge differences in health care systems. For example, Banal et al. is a US paper. We all know that that free market and for-profit companies play a large role in US health care, and as far as my knowledge goes, Canada has a much more social system. This leads to patients in ie. a US county hospital, often uninsured, will only receive chemotherapy if they enroll in a research project. This aspect of health care background might very well explain some of the differences found in the literature. I am missing this "ethical parameter" in the manuscript. **We agree entirely that the scope of this field varies by the healthcare system. We note the reviewer's concern in the second paragraph of 'Interpretation' (page 7) in reference to Bansal's work and note the ethical issue of intersecting vulnerability and obtaining informed consent. We feel the main ethical themes are consistent across the literature, yet they may present differently.**
Reviewer 3	Dr. Sally Bean
Institution	Ethics Centre, Sunnybrook Health Sciences Centre, Toronto, Ont.
General comments (author response in bold)	This is important research to help inform the AI dialogue. The paper is well-written and easy to read. The authors attempted the difficult task of succinctly presenting qualitative date. 1. In terms of opportunities for improvement: given that this is qualitative research, it's a shame to not have powerful quotes embedded, rather just reference to table 3. [Ed note: The presentation of quotes in boxes is consistent with CMAJ Open's style.] **Quotes from participants are embedded within the main manuscript text in the Results section (pages 4-6). Particularly impactful quotes were included, and Table 3 further highlights the study findings. Within the main text we were limited by the word count but have tried to highlight more quotes in Table 3.** 2. In table 3 you include illustrative responses but these are all single participant quotes so it affords no opportunity to contrast them among the participant groups. I appreciate how difficult this is to balance with word limitations, but the article's powerfulness is hindered, in my opinion, by this limitation. [Ed note: Quotes in boxes do not count towards the word limit] **Within the manuscript, particularly in the Results section (page 4-6), we have included quotes from participants within the text to illustrate their responses, particularly to highlight and support the findings from the study. With each quote, we have also labelled the participant type (e.g. patient, caregiver, healthcare provider), along with the study ID number. We have included quotes from all participant types, which we believe afford the opportunity for comparing/contrasting the types of responses received from different participant groups.** 3. While there is nothing that can be tweaked now, I found the vignette choices a bit odd and would like to have more details around vignette formulation. Was there any piloting of the vignettes? Perhaps add a qualifier indicating that the authors created extreme forms of the vignettes to elicit reactions versus more subtle, nuanced case studies, i.e. a clinician that is supplemented or enhanced with an

	algorithm versus a true replacement with vignette 2. **Before the vignettes were presented to the participants, it was first shown to a diverse group of people, which included students (undergraduates/graduates), post-doctoral fellows, data scientists, and healthcare practitioners. Feedback was obtained from this group of individuals in regard to specific use of language, neutrality (to prevent swaying participant responses towards one direction), accuracy, and thoughts on construct validity of the questions. Feedback was then incorporated to revise the vignettes, leading to another round of revisions before the vignettes were finalized and presented to the participants. This process is detailed within Supplementary file 1.** 4. Also, there is very little information about the demographic information of the participants. Were they a diverse age group? It simply indicates participants recruited from a private clinic so it is difficult to gauge the participant population diversity. The generational views on privacy and confidentiality can vary widely and would present a limitation. **Demographic information of the participants is included within Table 1. Recruited participants demonstrated a diverse age group, which we have outlined here: average age of patient participant was 60.5 (SD = 15.4; range = 52), caregiver average age was 60.8 (SD = 13.7; range = 37), and healthcare providers average age was 43.6 (SD = 11.5; range = 31). There was representation from each ten-year span from 20-90, which demonstrates a diverse sample of age groups.** **In addition to age, we also collected participant's ethnic backgrounds and educational backgrounds, which are listed in Table 1. Additionally, the researchers also collected current occupation from each participant; however, to protect the identity of each participant, we did not disclose it within the manuscript. A diverse range of backgrounds were represented, and none of the participants had extensive experience with AI and machine learning.**