THE UNIVERSITY OF CHICAGO | PEDIATRIC CANCER DATA COMMONS

PCDC Public Webinar - May 26, 2021
Q&A Session Transcript

Questions that were answered via the text Q&A function in Zoom are reproduced below. Questions that were answered live are linked directly to the point in the recording when they were answered. If you have any additional questions that are not addressed below, please reach out to us!

Q: Is St. Jude fully engaged in the PCDC process?
A: Answered live at 1:28:45

Q: How successful has machine learning been when applied to the data you've collected?
A: Answered live at 1:30:40

Q: Towards global harmonisation, are the clinical data dictionaries readily available across the range of cancer types mentioned at the start?
A: Answered live at 1:31:51

Q: You mentioned downloadable data dictionaries. Can you please provide a link?
A: On our website at commons.uchicago.edu, under the "PCDC" drop-down are disease group pages. Each disease group with a completed data dictionary has it linked from their page.

Q: I'm most interested in the "common" dictionary. Is that one available?
A: Email Sam - slv@uchicago.edu.

Q: When do you plan to incorporate electronic health records?
A: Answered live at 1:33:29

Q: Are the software tools you are building open source, under what licensing model?
A: Answered live at 1:34:42

Q: The GEARBOx tool sounds incredible! Will it include trials in specific regions e.g. UK or just America?
A: Answered live at 1:35:34

Q: What is the source of the protocol information in GEARBOx and how often is it updated?
A: Right now, mainly cancer.gov and other public sources. Ultimately, we will want to get protocol information from sponsors (with the proper governance, etc. in place).

Q: I couldn't agree with you more on how critical it is to create one database and integrating the information available throughout the various platforms that exist. How do you see that coming together? What are the biggest obstacles? What kind of time frame to make that happen? Seems like a daunting task but so important, it would be huge not just for now but for the future.
A: Answered live at 1:37:17

Q: Is there a standardised technology (equivalent to Gen3) developed for the national patient registry?
A: Answered live at 1:40:02

Q: How can we increase understanding of the importance of data collection to help accelerate research, especially within rare disease patient communities?
A: Answered live at 1:41:10

Q: Is PCDC/CCDH in touch with the OHDSI/OMOP and/or mCode/CODEX initiatives, which are also working on cancer data standards?
A: Answered live at 1:43:07

Q: Are you concerned or worried about security and patient privacy? How do you protect data from misuse?
A: Answered live at 1:44:47

Q: Is data access granted institutional based, or individual licences can be issued as well?
A: Answered live at 1:48:08

Q: With many separate data access "committees," how do you foresee making sure timely approvals are returned for data use?
A: Answered live at 1:49:56

Q: Doesn't stage generally mean "stage at diagnosis" so stage 2 is not mutually exclusive with metastasis?
A: This is a good point. We should have been more clear that we're talking about a particular time point. So, at diagnosis, if a patient has been diagnosed as Stage 1 disease, they should not have metastatic sites listed (at diagnosis). Happy to discuss further.

Q: For the clinical trial selection, why not just link to clinicaltrials.gov (which I assume is a database) so that the list is more conclusive and up to date?
A: We plan to link to clinicaltrials.gov for trials in GEARBOx. Eligibility criteria are still stored in a text blob that requires parsing out using some form of natural language processing. From what we've observed and heard from others, there is often a delay in updating clinicaltrials.gov listings and so when trials get amended we would likely need to have a separate process in place to keep our records up-to-date.

Q: Brian, are the delays in CT.gov on the cancer centers updating their records or on NLM's side of approving the amendments (which honestly I'm not sure they even do after the original submission is posted)?
A: That's a great question and I'm not certain precisely where the delays come in as the whole workflow is a little opaque to me. I would assume the delays could be introduced both on the trials administration side as well as NLM, but I don't have visibility into that unfortunately. Would love to learn more

Q: A comment: Although important for data sharing and standardization, the NCIt codes do not lend themselves to statistical analysis. The plan in the INRG is to map the standardized data NCIt items back to an ordinal, binary, or categorical coding system for analyses in SAS and R.
A: It would be good to continue discussions around this topic. Ultimately, statistical software like R should be capable of performing analyses where categorical variables are mapped to any given coding system, whether that coding system is bound to an ontology that is interoperable, like NCIt, or something more bespoke. It may be that there are customs among statisticians in this space as to how values get coded (e.g. using 0,1,2, etc.) and we can look for ways to address this.