Cyber Seminar Transcript
Date: 06/16/2015

Series: VIReC Clinical Informatics
Session: Mixed Method Finings and Implications for Future Informatics Research: PTSD Project

Presenter: Ruth Reeves
 *This is an unedited transcript of this session. As such, it may contain omissions or errors due to sound quality or misinterpretation. For clarification or verification of any points in the transcript, please refer to the audio version posted at www.hsrd.research.va.gov/cyberseminars/catalog-archive.cfm.*

Unidentified Female: So everyone welcome to VIREC Clinical Informatics Cyber Seminar. Thank you to Sider for providing technical promotional support for this series. Today’s speakers are Drs. Ruth Reeves and \_\_\_\_\_ [00:00:14]. Dr. Reeves is a research scientist at the VA Tennessee Valley Healthcare System. She has Doctorate in linguistics and has pursued a career in applied computational linguistics. Since entering the field of medical informatics, she has spoken to research on event semantics within the temporal frame of medical narrative. Deriving differences in language models represented in clinical narrative reports. Dr. Inez Tipcart (PH) our second speaker is a Sociologist and a Core Investigator with the HSRD Center to improve Veteran involvement and care at the VA Portland Healthcare System. Her interests include qualitative realist, and participatory research and evaluation methods, patient centered medical home implementation, social integration, displacement and health and access to/quality of care for marginalized populations. Questions will be monitored during the talk and will be presented to both speakers at the end of the session. I am pleased to welcome today’s first speaker Dr. Ruth Reeves.

Dr. Ruth Reeves: Hello. This is Dr. Ruth Reeves, and I am going to present the first part of the presentation today. Thanks everybody for coming. This will be a talk about the consulting for healthcare informatics and one of the projects focus is PTSD so thank you for showing up. So, the brief outline here is I am going to present the state of knowledge at the time that the PTSD project started and then we will also go through the overview of the aims of the project and the presentation of the qualitative analysis that had to do with provider interviews. Then there was an annotation project that descended from that work and then we will also go through some of the challenges and analyses that came from both of those parts of the mixed methods. So, right I would like to ask a question about your attendance so today you are attending primarily because, a, you have an interest in informatics, you have an interest in qualitative or mixed methods, you have an interest in PTSD or you are equally interested in two or more of the above. So I would like to acknowledge all of the persons that were involved in this project so from Portland – well I am not going to read through all of you. You can see them on your screen. There are quite a lot of people that did a lot of work.

Unidentified Female: We have the poll question up on the screen. We will give everyone just a few moments.

Dr. Ruth Reeves: I’m sorry.

Unidentified Female: I am sorry. I will go through that and go through the results and then we will bring it right back to your screen so you can move onto your next slide. It looks like we have actually slowed down here so I am going to close it out and the results we are seeing are 20% have an interest in informatics, 6% have an interest in qualitative or mixed methods, 6% have an interest in PTSD and 69% are equally encrusted into one more. Thank you everyone for participating.

Dr. Ruth Reeves: Okay. So, once again, I am going to just skip over all of the names involved in the acknowledgements but you can see them there and that will be available after the presentation as well. So, briefly, the state of knowledge at the time that the project was started, this is a sort of abbreviated very selected context to sort of say where we were. So, in 2006, the institute of medicine in collaboration with the VA and the Department of Defense came out with a study, a really expensive review of the PTSD clinical literature with randomized trials, pharmaceutical therapy studies and psychotherapy studies. It was a very thorough view, and basically, what they said was, that based on all of their reviews that the treatment modalities for PTSD did not reach acceptable level of uncertainty. So, the committee reached the strong consensus that a high-quality research was essential for every treatment modality and then it was already in the making at the time of that study. The VA and Department of Defense also were getting together a guideline for the management of PTSD and came to their consensus in 2010. So both of those things were in the background as this study was starting up and then two very, well actually three very important studies that I want to highlight that were sort of part of what was known and reacted to. So in the PTSD treatment investigation there was study just had to do with the practices and the attitudes of people that were doing compensation and pension for PTSD-related disability. This study showed that there was a really high percentage of variability and some of the practices actually conflicted with best practice guidelines. Then, this other study with Shiner as the first author this basically came out of Boston that actually looked at the evidence based on psychotherapy for PTSD where they compared administrative data for those practices against the clinical notes of those same patients to make a comparison and treating the clinical notes as gold standard. The admin data is unreliable at least for making that decision and mandated use for many, many things so it is not that surprising. Then, one final study that I want to highlight that influence where the direction of the chair study was. There was a terminology investigation of PTSD and basically what they wanted to know was are PTSD terms found in SNOMED CT Medical Terminology and they used the whole battery of terms collected from a lot of different expert sources that you can see there on the screen. So, if you just took the actual terms that occurred in SNOMED, there was only very modest coverage of those terms. If you did as the study tried to post coordinate the terms, so in other words, things like if you took a term like avoid and you took a term like recollection you could get an important PTSD concept which is avoid memory of but of course that is not in SNOMED itself. So, if you get all that then you could get a high percentage of coverage so the conclusion was analogical work and description logic which is just sort of how do coordinate, what other kinds of rules. Those will be required in order to actually have coverage from an electronic source. So, that was where the state of knowledge at the time on the inception of the project handed over the Dr. Tipcart now.

Dr. Tipcart: Thank you very much, and I am apologize. I forgot that I am going to be the one that has to show my screen now so I having a little difficulty getting to my slides. Give me just a second. Here we go. So, thank you. I apologize for the delay. Let me get this into slideshow. Is that displaying correctly now Heidi?

Heidi: Yes it is.

Dr. Tipcart: Thank you very much. So Ruth, I am Inez Tipcart. Thanks again for everyone for listening in. Ruth has just given an overview of where the project was and sort of the rational. This is not advancing. Here we go. So, and she already gave you an outline of the entire presentation. I am going to just talk now about the piece of this research that I was most involved in together with Dr. Susan Zickmund who was part of the team out of Philadelphia. We were the qualitative investigator leads on this project. What the project wanted to do. We have done the poll. We have acknowledged everyone. So, the aim of this particular project. I just put them here. One was to identify the vocabulary used by clinicians to describe the clinical course of PTSD among veterans with a PTSD diagnosis and enrolled in VA care, right. Obviously, we ideally like to know about that for veterans who are not receiving care within the VA but the data we can work with was for veterans who were within the VA. Then, having done this, we wanted to improve information extraction techniques for PTSD using computational linguistics in machine range needs and also through manual annotation of clinical text sets which Ruth is going to talk more about. Measure the performance of newly developed information extraction techniques for classifying clinical important concepts. So, that was the way that this project was laid out and that first piece identifying the vocabulary use by clinicians we wanted to do it in a couple of different ways part related to the manual annotation of looking at their notes but part of it was through using qualitative methods to talk to them directly about their notes and the rational behind their notes. This came out of the work of the original PI on the project, Dr. David Hickam who has since left the VA to go the Patient’s Center Outcome Research Institute who had a long interest in understanding sort of patient provider communication and the way patient and providers use tools. So, I did a very simple graph to just to kind of outline the way we thought about this project proceeding. There was a data collection generation and analysis piece where the clinical note annotation and ontology development were going on but simultaneously we were collecting qualitative data. We did this in three steps and the idea was that this would inform the ontology development and would interact with the clinical note annotation would be kind of a triangulation of findings and also potentially sort of open up and build on what was being found through the note annotation and that these things would have two products. One would be to improve natural language processing or NLP and other data extraction tools as we outlined in our aims but the other thing that we realized early on is that this was interesting opportunity to really do research more generally on how the electronic health record was being used and it could improve both care delivery and research. So, the qualitative data that we collected, we started off with an expert panel to help us and when had session with an expert panel. These were experts in PTSD to talk to them about the concepts that they thought would be very important. I will talk a little bit more about that in a moment. We then did focus groups with providers at each of the participating sites in this study. This was a multi-site study as the acknowledgements indicated so we talked to providers in Portland, in Pittsburgh, in Tampa, in Nashville, in West Haven, and in Salt Lake City. We held these focus groups. Those focus groups lead to some findings that Dr. Zickmund is working on publishing but they also then helped us, and I will touch on this again in a moment, they helped us to refine the sorts of things we wanted to ask about in the interviews. Then the final step, which is mainly what I am going to talk about here, was we then did cognitive interviews with individual providers at each of those sites. We ended up doing about 40 individual interviews. In those interviews, we asked providers to review notes for their patients with a diagnosis of PTSD to review an actual note or two or three if time permitted and talked to us both about the specific language they used in the note and their rational or their thinking behind the language that they chose. Then the decisions they made about what to include and what was important and so on. We then used a hybrid inducted, deducted coding scheme where we both had some directed topics that we knew that we were looking for and we also were open to finding unexpected themes. So, how did we use this information to help inform the ontology development? So, this is a fairly raw example. We wrote up reports that were much more structured but I think this might be interesting to show you. So, one of the things we did in our coding is in the directed piece of the coding is we knew we were looking for language on symptoms and so we went through and these are examples of the actual language that clinicians reported using in their notes to talk about specific symptoms. So here, you know, if I have a dash that indicates that it is actually the word sleep or sleeping so poor- means poor sleeping. These were the extracts from the notes that physicians chose. They were allowed to choose any note from a recent visit with a patient who had a diagnosis of PTSD as long as it was not an initial intake note. I should actually also point out that providers, when I say providers, we talked to a range of providers who see patients with PTSD and not necessarily to treat their PTSD. So, we also talked to primary care providers. We talked to social workers. We talked to psychiatrists. We talked to clinical psychologist. We were also interested in seeing if they work through different disciplinary approaches to writing the notes. So part of what is interesting here, you know, a lot of this language would overlap with what would be noted by the team doing the annotation of notes without the interview. But, some of it, they might have missed and this was why during these cognitive interviews was potentially useful because we asked providers to talk about what was describing symptoms. Sometimes they would report things. If you look towards the end, especially this whole cluster around anxiety things like no close relationships with family members, reports forming one good friendship. They described these things as being related to the expression of symptoms either decrement or improvement. So, those were from the directed coding some of the things that we passed along to the ontology development team to help make sure that everything that might be relevant to the language around symptoms for example were being captured. We also found some latent semantic or phonological findings that were quite informative. By that I mean, phonologically just generally that refers to an approach where you are trying to understand the experience of doing something and talking to providers about their experience of writing notes and the thinking behind it, there were some interesting incites that came out of that. So, one thing was that the concepts of interest are not always terms of interest. So, to give some specific examples, suicidality was something that our expert panel had been very interested in. They wanted to know how suicidality was documented in the note. They also wanted to know how resiliency was documented in the note. Well suicidality for various reasons is omnipresent references to whether a person is suicidal or not. Those are pretty consistently present in the notes and the language is fairly standardized so that maybe something that it will be possible to use in a relatively easy way for data extraction. Resiliency on the other hand was an important concept when we asked specifically about whether providers wrote anything about resiliency. They almost all endorsed that it was an important concept. It was relevant for thinking about treatment for patients but it was an absent word and the language that they used to describe it was that there was not a lot of language in the note to capture this idea of resiliency. The providers themselves in the focus groups, this is kind of interesting and this is one of the strengths of using the fairly exploratory qualitative methods. Providers brought up the importance of recovery and wanted to know if their colleagues used the word recovery. It turned out so we then asked in the interviews about this and the whole idea of whether recovery is a relevant concept to PTSD treatment sparks very different views from different providers. So there was not a lot of consensus on the realms of recovery and when they did use it, they used it in very diverse ways. Some providers were coming from a background in substance abuse treatment where recovery has specific connotations. Some of them used recovery within sort of the framework of the recovery model for mental health. So, if you think about why this matters for data extraction, a term that is used in very different ways by people may create some difficulties or making sense of any findings based on language related to that concept. Then the other thing we found that is sort of interesting is that really health providers use the note influences the content that is available for researchers or for any trying to extract data. We did a little graph there just in different ways but instead of explaining that I think I will just talk about some of the findings. These are all in a paper that is currently in revisions and I hope it will be out soon. But some of the important things to note were the providers really used these progress notes primarily to remember and access details for their patient care. They did not use them for conscious communication with other providers and so sometimes things that you think would be important were not in the note because did not feel that they actually contributed to patient care on making changes in how the patient was being cared for. So, that leads to the second point that providers infrequently recorded information that they judged would not directly contribute to improved care. This method, they would sometimes omit information that they thought might jeopardize patient’s access to or quality of care. An example of the kinds of things I am talking about included sexual trauma or nonmilitary trauma. Providers sometimes would leave this out because they were concerned about the effect of having this in the record, how this would potentially influence the patient’s relationship with other providers or how it would influence their access to service-connected care. Then, finally, a sort of overall conclusion from this and this relates to that second point that although this research was designed to contribute to the development of data extraction tools. Really, there was sort of a more broadly relevant finding, which is that understanding how providers think about their notes, and how they make choices in what they write can help other clinicians be aware of the limitations of the electronic health record as a tool for learning the histories of new patients. This is not to say that they are not useful for that but for people to be aware of some of the limitations that almost systematically present. Also, researchers who are relying on E-chart data for PTSD research should just be aware of likely areas of missing or ambiguous data. So, I am just briefly going to talk about some of the lessons learned and sort of reflecting on this process of trying to use this sort of early exploratory qualitative research to inform a larger informatics project and then I will turn it back over to Ruth to talk about the next steps that came in this project. But, the project designed for true mixed methods design can be really challenging especially in terms of timing. By that, true mixed methods design I mean where the different approach is to the data are really informing each other and this is difficult to do making sure that one part of the analysis is finished at the right time to truly inform and improve the other part. We did not do this but reflecting on it, I think we could have done better with more attention to timing so I would recommend that others keep that in mind. Learn each other’s terms and languages the best you can. If you working across different disciplines it certainly took me awhile to understand what the informatics were talking about in their ontology development. Having made that effort it actually made a lot of sense to me what ontology development means. But it does take some effort. One other thing to point out is just the importance of thinking ahead about whether you want to have a data repository. So this project actually did set up a data repository for some of the note annotation as well as the other data from this project but the interviews with providers did not go into that data repository because we had not thought to consent them about a data repository ahead of time. So if you are thinking about setting up the data repository, just make sure that you are informed consent process addresses this. That is the end of my part but I will be here for questions and to listen in and possibly comment on the rest of Ruth’s comments but there are my contact details will be in the handout and feel free to contact me about any of what was discussed today. So Ruth back to you.

Ruth: Great. Thank you. That was a great presentation. So let me advance to the next slide. Thanks for setting that all up and we certainly did take into account some of the things that we knew were likely to be missing and ambiguous that was part of what informed the clinical note annotation. Certainly all that work was called into what became our working process even being aware of some of the limitations since that time. A lot of us had thought about gathering notes from different types of providers to try to fill in some of the holes that Dr. Tipcart mentioned. So I am going to talk about the clinical note annotation. This effort was lead by Marion Circle at the Portland VAMC at the time. So, here are the characteristics of what we had at our disposal to work with. So the cohort definition you can see there. There were quite a few that were veterans that were with the hospital discharge of PTSD or CMP PTSD Service Connection Disability or PTSD was on the problem list or at least two outpatient visits coded with the ICD-9 for PTSD. So, that set of patients was then cross-referenced to the registry of OIF Veterans and that was really to limit to veterans were relatively recent within the course of the disorder. SORT did cut down on our data size and then these were all patients receiving care at any of 125 mental health clinics in the VA. So, then to get the corpus of records together, it was the electronic record of all those patients identified in the cohort and then what we were interesting in particular was the clinical notes and we selected across just two difficult years which resulted in a 1.6 million or so clinical notes. Then, we did a very first pilot annotation just for concept prevalent. I will show what the concepts were in a minute. To get a subset of notes where if there a relatively high density of the concept prevalence, then what we were looking for was the note type and so there was a whole set of logic that went with that to an example of the kinds of things that were selected for that note type. You can see there. So, then to set up the annotation, what we did was we made sure that we had quite a few annotators but for all documents they were double annotated. Then there was an independent clinical expert who adjudicated whatever disagreements there were, and then based on a few quality measures like count of note per patient and some other points of interest for us to get a clean data set, a little over 900 notes were selected for annotation. Then the eight annotations were just based on people’s availability and there were about 100 notes in each set. What the concept that we were after basically can be summarized there as either a symptom, a change in symptom, treatment, change in treatment, and then annotators were asked to relate for any given treatment that was found if there was a symptom in the same note. I should have said this annotation was all from the point of view of a note at a time. So, if they found a symptom and a treatment that was for that particular symptom then they would link them within the annotation task. They also look for treatment, noncompliance and negation of any of those terms. This is just a view of the tool that we used at the time that was annotator, which just allows you to set up the concepts of interest. This is a fake note by the way and annotate them so an annotator could see what is going on that the viewing pain. So, that was an annotator. That is what was used at the time. There is a newer development of an annotation tool that came out of Salt Lake City which is called E-Host, makes it a lot easier and it helps you do a few other quick-search terms and since that time there is actually another tool called \_\_\_\_\_ [00:30:59] which was developed by my colleague here at TBHS Gwen Goblin. It is something that allows a sort of learn as you go, because a lot of annotation is very repetitive and so what the sense of annotating the same thing over and over again. That is a very brief overview of those tools but I thought I should mention them since both of those had in part funding by here in different projects. So, here is a comparison. This is to just sort of to see what is the quality of the data that was collected. So remember I told you that each note was adjudicated by a clinical expert and this is the scores of each annotator as compared to the gold standard and it is expressed as a mean between the precision and the recall of the terms found. Anyway, you can see the scores improve once you get to batch two and that has to do with the fact that the annotators received a new training because we saw some problems in the first set. So, that was the data quality. I just want to give you some examples of the kind of terms so the terms that were actually found in the documents are the ones that you can see in italics there and the terms in bold are the terms that you can collect from looking at the DSM and they were in the category of either – well these are all symptom categories. So, just to give you a sort of idea of what kinds of things were collected and then remember I said that treatments and symptoms were linked and so these were some of the links between symptoms and treatments that were found in the annotation. So, that is a very quick overview of what was done not to mention Heidi but there is a lot to go through. Doing this manual annotation actually had a point to try to then use that to train an NLP system which was used to create the PTSD ontology because as we saw in the very first part of this talk, there just is not enough coverage in something like SNOMED to do the logical on your own. Basically, we are trying to create it by the annotation and then NLP tools to upgrade the ontology. So, I am going to talk about some work that came out of West Haven. The lead on this work was \_\_\_\_\_ [00:33:57] and she created something called MedCat which is a framework for high-level conceptualization of medical notes. This is a lot of work that she did. It used the PTSD annotations that I have just showed you about and in combination with YTEX, which is another tool that was funded partially through here. It is an NLP tool with a database front end, which makes it a lot easier to do work, and this was in combination with an NLP technique for clustering content that is found by topic detection. So, I am just going to really whiz through this because it was a quite complicated project that she did but you can see there what she needed to do was use obviously the data that notes themselves and then the annotations that were fed into her tool and then what comes out is abstracted information so why should you care about that? Well, just let me show you what you get out is taking the notes, going through her knowledge base, which does a sort of hierarchical treatment of the terms. That is the knowledge base with a blue tree and then those same notes are also fed through YTEX. YTEX spits out a bag of concepts and what MedCat spits out is a bag of categories. Together what you get is something that you can really use for building the ontology. So, those concepts that were extracted were organized into a knowledge base into the hierarchical structure. So, the very specific concepts are at the bottom and they feed all the way up to the more generalized ones. This was performed, this feeding into the various different hierarchical structures was performed by content experts and what was used was a sort of very simple part of hierarchical construction and you can see there the treatment categories like pharmacotherapy and psychological, psychosocial, psychoeducation case management. We also had another bag because some of them did not seem to fit into anything. So, this is an abbreviated tree of the category hierarchy that was created in this tool. So, this tool although it was developed independent of YTEX, it is being now made such that it is part of a feature engineering platform within YTEX. That is something that I am interested in using for my work and something that West Haven group has been doing extensive work with. So, this is just a sort of summary of what its capabilities are at the present time. So, the categories were as I said transformed into this hierarchy and what it does is it removes the redundant and irrelevant concepts and it reduces the dimensionality and that is what is actually needed for forming an ontology that is usable for doing any kind of description work on this piece concept that were found. So, going through this is a sort of standard thing. You take the data. You split it into well there is 2/3 taken for training and then 1/3 for the evaluation. This is just to tell you what the evaluation was. It was sort of at the level of the document itself so you just wanted to know were any treatments detected in a note and could you detect whether the note itself was a treatment note. So, there is detected and then missed where you failed to identify in at least one treatment type within a note. So, that is the sort of matrix there, so what is a good performance obviously is where you get detection, and you have not missed that is the B cell. So the results is out of the 911 notes, 445 were determined to be actual treatment notes by the clinical expert and so in the test data, 91% true positive raises did okay and so the error analysis there I have just made a sort of summary of the kind of things that at the time were not detected. These are being updated now so I mean it is a decent number but some of these things that are unexpected to have not worked like certain kinds of punctuation through it off but those are being updated now. Just to show breaking down performance per category that were listed, you can see there that the pharmacology and the psychological treatment types did the best and you can then basically the same, that is just another view of the same exact data. So, we are very encouraged by that performance and so I just wanted to talk about building outward from that. So, that work was done just on treatment types because it was really clear from the annotation from the beginning and also from the work that I talked about the beginning. It was really quite difficult to know what treatment type any patient was getting so that was the reason and that was the focus for Somas work. I am also interested in doing the same work for symptoms and so I have built on that in order to get to the next phase. So, this is almost the exact same set up, the hierarchical relation is. So, these are sort of hypernyms and hyponyms and the knowledge base is performed by experts and the team but what they were trying to use was the criteria outlined by the DSM as a category roll up. Oops, I think I went in the wrong direction. So there is a partial concept tree for working in the same way so the ones at the bottom are terms and then rolling up to, the concepts that we expect to find or that are named in the DSM and those are then used to create the categories. The same idea is what was used for the treatment that it removes the noise basically. The size of the terms was reduced from 1,037 to just four categories. So, the evaluation of that is underway. I have a lot more notes to go through before the whole evaluation is completed but I am confident that we are going to have pretty decent results because we have actually done a little bit more work on how YTEX is recognizing things. So, I wanted to say that the business of relating symptoms to the treatments that were found in the first step is something that is being added to another NLP tool which one that I am constructing that has to do with trying to get a chronology of symptoms and treatments in a \_\_\_\_\_ [00:43:25] pipeline. So, we are using the linking relationships between the symptoms and treatments that were done in the annotation as a way to train part of this sequence of events pipeline. This is far too much to discuss right now but that is the basic architecture of the tool I am using for the temporal reasoning portion of my build. One of the things that is being added is there is a classifier which actually relates events according to temporality and we are using the links that save that a particular treatment is to treat a particular symptom is being added to this temporal processing. It is actually part of both the classifier and the thing that is called B linker which I do not have enough time to talk about but that is where that is. There are further thing that is going on in my project now is we are taking the same notes. This is why Dr. Tipcart talked about having a research data repository because you need to build. You can see that there are main layers to this work so this is another fake note. Dr. Toohi does not exist and neither does this patient Huckenaeur. But this is an example of the kinds of extra annotations that will be done on the same notes. This scheme has already been created and the annotation is being constructed. The backend to the construction is being constructed now within my grant in order to do the temporal reasoning with this data and I just wanted to say one last thing. You will notice in my simplified pipeline that there was a need to collect temporal expressions and that module has already been modified and what you are looking at there is the red performances is the tool prior to any modifications that are just baseline and the blue boxes are the modified version. Those classes you see there are just time date duration set is just a word for various kinds of frequency data. So, it is quite a bit improved. There are some things that need to be worked on for duration mainly but I am encouraged by that. So, that is all I wanted to tell you about the analyses of the work that was done post annotation. Thank you. Lacy. I wanted to then turn this over to questions for the remainder of the time we have available.

Unidentified Female: Great! Thank you Ruth and Inez for your presentations. We have a few questions here from the audience and I will go through those now. So the first one, there has been some local discussion about how the content of notes has changed now that veterans have complete access to all their records. How do you think this impacts the utility of the medical record as a data source and what can we do about it?

Dr. Tipcart: I think that is a great question. One of the things I did not talk about today, was the focus group data and that was one of the things that came out of the focus groups was a real concern about potential – there were a lot of issues about people reading notes, actually other providers reading patient’s notes but also about the patients themselves and whether there were any concerns and how this was influencing how people write notes. I will just mention a colleague of mine in Portland, Dr. Steve Dobson. His team has really interesting project right now, looking specifically at how patients with mental health diagnoses respond to their notes and how we can sort of improve the whole open notes experience. I do think that the assumption of patient access will change how people write notes. Providers did talk about this somewhat but it is a difficult thing to predict what that means because some people really saw patient access as a therapeutic tool and so I guess the real answer is yes I think it is going to influence it. I am not sure that means the notes will be less useful which I think is the concern that people have that providers will be too careful about what they put in the notes. By the same token, they might actually write more thoughtful notes that include more information about what the patient considers to be important and for something like PTSD that may actually enrich the note. So, it is a very interesting question and I think somebody should be looking at that going forward.

Unidentified Female: Thank you. The next question I think this one might be for Ruth. Was there any measure of combat exposure accessed in the study?

Ruth: There were not actually. Since we sub-selected for patients that were already in treatment because those were the ones that were richest. They tended not to have as much of that data as would have been useful so that criteria which is exposure. It is the first DSM arm, was not included. Subsequent to that, actually, I have a colleague who is working on that but it is not included in this.

Unidentified Female: One more question. I think this is the last one we have right now. Can you discuss the reason for using annotation as research and its purpose?

Ruth: I am not exactly sure the intent of that question but there are two purposes that one usually uses annotation for. One is to train an NLP tool and evaluate it obviously. It is also and maybe this is the answer to the question. It is also sometimes used as a way of probing the data that you have. I mean you can ask one, hey why don’t you go through all these records and find for me, I do not know instances of some certain pharmacological treatment but if all you do is give them an excel sheet or something you may not record as much detail as you would like. I am not sure if that is getting at the intent of the question.

Unidentified Female: Okay. Those are all the questions we have right now. Thank you, Dr. Reeves and Dr. Tipcart for your presentation. Our next section in this series Clinical Informatics is scheduled for Tuesday, July 21, at 2:00 p.m. Eastern Time and it is entitled \_\_\_\_\_ [00:51:45] Natural Language Processing Tool for Extracting Ejection Fraction. It will be presented by Drs. Jennifer Garvin and Stephane Meister. Heidi will be pulling up the evaluation questionnaire when we close the session very soon and if possible if you can stay until the very end, it will just take a few minutes to fill it out that would be very helpful. Thank you so much for attending. Thank you for presenting and we hope to see you at the next session.