[music]

00:06 Sarah Crespi: Welcome to the Science Podcast for October 2nd, 2020. I'm Sarah Crespi. Each week, we feature the most interesting news and research published in Science and the sister journals. First up, we have investigative journalist, Charles Piller. He talks about his latest story on long-term failures at the FDA to regulate clinical trial safety and data integrity. We also hear from Machine Learning researcher, Klaus-Robert Muller, about an AI that can beat human experts in the sport of curling.

00:37 SC: Now, we have contributing correspondent, Charles Piller. He wrote an investigative feature on FDA's serious decline in protections for patients and for data integrity. Hi, Charlie.

00:48 Charles Piller: Hey, Sarah.

00:49 SC: I'm not sure everyone knows that FDA has oversight of clinical trials. What exactly is this agency's role with respect to human trials?

00:58 CP: The FDA is a vast agency with many responsibilities, and in this case, it has two elements of control over the quality of clinical trials. One is whether patients are properly protected during the trials, and also whether the data generated from these trials is reliable and is in-keeping with the important goal of the FDA of approving drugs and devices that actually work.

01:30 SC: But as you found in your investigation, the FDA enforcement has been light-handed, slow-moving, and secretive. Can you give us some examples that you found of shy clinical research practices for record-keeping that you came across in your deep dive into this?

01:46 CP: One example that comes to mind is a organization in Utah that is a contract research organization that does clinical trials on behalf of drug companies. This is an organization that had enormous problems spanning almost a decade. Problems with informed consent, problems associated with conduct of its trials that called into question the integrity of the data and also the safety of the volunteers. These are trials for pretty serious kinds of drugs and devices including opioids, antidepressants, antipsychotic drugs, drugs for Alzheimer's and Tourette's, and lots of other pretty serious ailments. What we found in looking at the FDA evaluations is that there were really serious problems that the agency itself warned the investigators might constitute fraud, scientific misconduct, significant human subjects protection violations. And perplexingly, despite four investigations, the agency never cited this organization for anything beyond just saying, "Look, you're doing a bad job here."

03:03 SC: Yeah.

03:03 CP: They never issued a warning letter, never required any changes in the procedures, they never insisted that the principal investigator change operations in a way that would safeguard

patients and data.

03:17 SC: I just wanna emphasize how important it is to protect people involved in clinical trials. These are medical experiments on human beings and there's a long history of unsafe practices.

03:30 CP: In the bad old days, there were a lot of abuses of human subjects. These days, there are pretty stringent requirements that there be informed consent of the patients who are participating in these trials, that there be a very rigorous review of the trial so that these patients are well protected. The FDA looks both at the investigators and the institutional review boards called IRBs, that are responsible for overseeing these experiments. And what they often learn is that not just are the investigators making mistakes or handling their experiments in a cavalier or unsafe way, but also, the investigation process, the review of that by the IRBs is often not very adequate. And the IRBs themselves, perhaps are run amateurishly or in a way that does not protect patients or the quality of the data that's generated by the experiments.

04:32 SC: This is an investigative piece you've spent almost a year gathering information. What kinds of evidence did you collect?

04:40 CP: Well, I went to basically every public source of data that the FDA publishes associated with their review of clinical trials. And so that includes summaries in their various databases online, it also includes warning letters, which is when serious mistakes are made, they might warn the investigator or the IRB that they've got a problem that needs to be fixed. I also looked at the inspection reports that were available through the Freedom of Information Act on hundreds of these cases. These are pretty difficult to obtain because the reports are only available under the FOIA, and as a result, it can take weeks, months, sometimes years to get access to them. So what I'm trying to convey here is that, this is a very difficult process to actually learn what happened in the FDA's evaluation of these organizations. Because oftentimes, they do not announce publicly what happened.

05:43 SC: What about the drug companies that hire these institutions to carry out this research?

05:48 CP: When there's really egregious problems going on, the worst case scenario is that a clinical scientist who is doing these experiments can be disqualified from future research. This is for someone who has broken the rules in such an egregious way that they've endanger patients or that they've intentionally submitted false data to the FDA or to the sponsor of the experiment. When that happens, there's no doubt that the companies who sponsored the specific experiments that were found to have been either fraudulent or so badly flawed, those times they're informed of it. The problem is that many times, even those experimenters, their long history of research is just sort of a black box. No one looks back to see if they made similar egregious errors or conducted experiments that were tainted by fraud.

06:41 SC: So if someone say, was flagged for five ongoing experiments, the FDA said, "This is all bad, you're done." If they had done 100 projects in the past couple of years, none of those were gonna be looked at.

06:53 CP: That's exactly right. And this is not just a theoretical question. I looked back at investigators who had been disqualified, and I contacted dozens of companies for whom those investigators did experiments in the past and found that none of those companies would say whether they ever looked back at the earlier experiments. We know the FDA didn't look back because that's not something they do. We know the investigators don't look back. In fact, what we have is a situation where there's unknown problems in the body of knowledge that contributes to drug approvals. There's unknown problems associated with the safety and health of patients who may have had trials conducted by investigators who're either inept or careless or perhaps engaged in fraud in certain cases, and we just don't know how they were harmed because no one is looking at the historical record.

07:51 SC: What are some of the trends that you saw in FDA's enforcement over time?

07:56 CP: I looked back to the beginning of the Obama administration, extending through the first three years of the Trump administration. What I found was that the FDA's budget for this sort of work, overall, it's risen over the last 11 years, and the amount of enforcement going on has declined sharply over that time. Since 2009, the FDA has issued 291 so-called OAI designations in its inspections. That refers to official action indicated, and that's important because it means that they found that clinical investigators or IRBs have broken the law or violated federal regulations in serious ways. And I found that about 6% of FDA inspections were classified OAI during the Obama administration. But during the first three years of the Trump administration, that fell to less than 1% of the total, a pretty dramatic decline.

09:00 CP: Now, what I found also was that the designation of warning letter, which is a kind of a step up from OAI, it shows that not only were there serious problems found, but the FDA escalated it for particularly important action, and these warning letters which demand the changes fell also dramatically between the Obama and Trump years. So under Obama, the first three years of Obama, there were 99 such letters, under the last three years of Obama, 36 warning letters, but only 12 during the first three years under President Trump, suggesting again that enforcement was dramatically down.

09:43 SC: There's no chance that everybody is just better at research now, that they're being more compliant?

09:50 CP: No, I don't think there's any chance of that, at least not, according to experts who track the FDA very closely and also others who are former FDA officials. What they told me was that this shows a really serious question about whether there's serious enforcement going on under the Trump administration, in particular, but I have to say they were also troubled by the decline in enforcement numbers that occurred during the latter years of the Obama administration.

10:19 SC: Not only do these numbers go down, fewer letters, fewer disbarments, but also after a letter is sent, is the FDA following up and making sure that adjustments are made? I think a really good example that you brought up is this trial of ketamine in emergency departments. Do you wanna talk a little bit about what the drug is and why it would be used there and what the problems were with the research going on?

10:43 CP: This was a very interesting case in Minnesota. It involved a series of experiments where an emergency department in a big hospital used ketamine to sedate patients and to calm them during emergency room procedures. This has sort of emerged as a big scandal in 2018 because local reporting and also concerns raised by the consumer advocacy group, Public Citizen, along with 60 clinicians and medical offices claimed that this hospital had used procedures for a clinical experiment involving ketamine that were grossly flawed and violated really standard informed consent requirements. Took advantage essentially of people who were in a compromised state.

11:32 CP: This resulted in a series of inspections, investigations by FDA. What was pretty interesting about this was that notwithstanding the investigations that validated the concerns raised about improper procedures for protecting patients, the agency never warned the hospital, never required them to make some changes. Even more perplexing is that in my investigation, I found that similar conclusions had been reached by FDA in a ketamine case years earlier at the same hospital involving some of the same people.

12:07 CP: So one of the most alarming elements of this was that there were 23 deaths in the very first ketamine trial that was looked at by FDA, and some of these were in patients who had been given the ketamine. Now the trial administrators said that these deaths were due to other causes, not due to the study drug, but they were not reported properly in that, our standard for a clinical trial, so it raised concerns among some experts that there was a kind of cavalier approach being used that was carried forward to the later experiments, and because the classification of adverse events such as deaths is so important to protecting patient safety and to understanding how best to run these trials in the future, that was a pretty big lapse that was really never addressed by the FDA in an adequate way.

13:01 SC: Do you see money as a big driver in these companies? And doctors get paid to perform these trials and then the drug companies get paid to sell these drugs. Running these clinical trials is a big business.

13:16 CP: You mean there's many clinical trial businesses and physicians who have a side business of conducting trials and because it's so big, the FDA can only touch a small portion of it. What's alarming about the results that we found is that even in this small fraction of the overall trials that are looked at year by year, there were so many abuses found and then not followed up on by the agency. So it suggests that the problem might be larger than we're able to characterize in our story. The other thing I wanted to mention about money is that the drug companies have so much money to give to physicians that they even give money to physicians who have been disgraced by FDA, who have been prohibited from doing additional clinical trials. These disqualified investigators sometimes earn hundreds of thousands of dollars from drug companies, even in a single year to give lectures, to teach other physicians and for other purposes that are separate from clinical trials, but still show that they can influence their fields and then they can benefit from the largest of drug companies. I think that demonstrates a kind of systemic corruption of the process of evaluating clinical experiments and preserving the body of knowledge effectively.

14:41 SC: What about patients? They are in these trials, they don't necessarily get any information if there was a problem with the drug trial they were in and if they were harmed, what kinda recourse

do they have?

14:54 CP: Yeah, it's the most tragic part of this entire story. I discovered about 1600 documents under the Freedom of Information Act and in very few, are there any of the names of trials. They're usually censored by the agency for commercial secrecy reasons and as a result, it's difficult or impossible for any patient who might have been in a tainted trial, who might have been subject to unsafe conditions to learn if indeed their trial was the one that was subject to FDA regulatory action. And so even if a patient suspects that their investigator has a problem or knows that their investigator had problems with the agency and request documents, they can't know for sure if it was their trial. As a result, patients may have been harmed, they may have been hospitalized, some may even have died, but they can't learn the outcome. Their loved ones can't know whether they were in any way subject to these flawed investigators and that's I think a basic fairness flaw in the regulatory system that seems to protect the commercial interests of drug companies at times more than it protects the safety of patients.

16:10 SC: Thank you so much, Charlie.

16:12 CP: Thank you.

16:13 SC: Charles Piller is an investigative journalist for Science and this story was supported by the Science Fund for Investigative Journalism. You can find a link to the feature and a related sidebar at sciencemag.org/podcast. Stay tuned for an interview with Klaus-Robert Muller about why it's so difficult to build an AI that can play the sport of curling.

[music]

16:38 SC: We've seen artificial intelligence take on chess, Go, poker. But did you know that people call curling, the ice based sport, a cross between chess and bowling? To play a sport like curling an artificial intelligence must do a lot more than it might when faced with cards or a board game. In a Science Robotics article this week, Klaus-Robert Muller tested whether an AI could compete against real human players on the ice. Hi, Klaus.

17:07 Klaus-Robert Muller: Hi.

17:07 SC: What makes curling a good challenge for machine learning?

17:11 KM: Well, curling is in the real world. Usually people that use AI in the context of games, play these games in virtual worlds. So Go is one example for that. So since curling is in the real world, it's actually facing challenges like uncertainty and continuous action spaces and little data.

17:34 SC: Right. The physical world, the real world is constantly changing and full of uncertainty. A robot can't be constantly running something over the ice trying to feel how it's changing.

17:46 KM: So the ice quality is something that is a magic thing. [chuckle] It changes all the time because the big stones go across.

17:54 SC: So this isn't about making a robot that's good at getting stones into the right place, like good at hurling them, physically positioning them. It's a lot more about how to decide what to do, where to place those stones, which shot to take and also taking into consideration what the other team is doing, where their stones are, the quality of the ice, [chuckle] is this some of the questions you have to consider?

18:19 KM: Yes, absolutely. So we have one strategic part in curling which tries to do the best move that is most harmful for the opponent team but in the whole process, there's a large uncertainty. So you cannot reproducibly, because of the ice quality, get the stone always to the same point. It's impossible. Therefore, there's a intrinsic uncertainty in every throw and that makes strategic have to consider uncertainty when planning.

18:49 SC: Let's get back to this idea of the reality gap. So you might have your AI do chess in a virtual gaming room or poker. Even against real players, it's not facing the world. And then I assume that you tried things out in a virtual curling world. How did you approach this gap? How did you try to close the gap between what the AI might experience in a virtual curling competition and in the real world curling competition?

19:16 KM: So if you like to model our stone continues its path, then you use very complicated mathematical equation so called partial differential equations. These make some assumptions and unfortunately they are never true. So there's a huge gap between the reality and what the simulation tells us, so we need to correct for this. This is one aspect. If we throw a stone then in the best possible strategy that we can think of, we also have to be aware that the stone may not come exactly to the point where it's targeted. So there's this uncertainty that makes us consider, "Under the large variants of the stone, is this still the best possible move?" In other words, the best possible move would be a move which works well in one of a thousand times, then it's not a good move to win the game.

20:12 SC: So we're talking about how you use machine learning to attempt to solve this problem. Can you talk about some of the options? What's the framework that you were evaluating that you wanted to test in this situation?

20:26 KM: Let's split up the machine learning in the several parts that it has. The first part is strategy. We can't train deep reinforcement learning like in the Go game because it is a very complex game with the opponent doing unpredictable things. So that's number one. Number two is we need to throw the stone. Number three, we have to estimate the uncertainty. Number four, we have to correct for the simulation that is slightly wrong. And number five, the ice quality changes over time, so we have to practically adapt the robot all the time to the changing conditions. So that's a very hard way, and our innovation is basically taking the uncertainties into account and taking the change into account.

21:15 SC: So that was the focus of this work, was figuring out how to handle the uncertainty in a situation like this?

21:20 KM: Yes. How to handle the uncertainty in the strategy and how to compensate for the non-stationary of the world.

21:29 SC: Let's go to the arena. I've got some audio here from the competition that will help put us in that place. It's a busy environment, there's people and noise. Can you describe the setup? Who is here and what do the players do? What do the robots do?

21:45 KM: The first thing that all players do, humans and robots, is they check out the ice. They can take a very small set of throws, test how fast the ice is in principle. And after having calibrated the human throw or the robot throw, basically everybody's ready to go. People take turns. The one team starts, places the stone, then the curling robot comes into the arena, it looks with his extended neck, where the camera looks where is the stone, where is the house, where's the play field, and locates everything, and then computes a strategy where to place the next stone. And then he does the throw, meaning you get the coordinates and the speed and the curl that the robot must achieve, and then that robot releases the stone at the hog line and then the stone goes and hopefully to the position that it is intended for.

22:44 SC: Are there sweepers? I know I've seen in competitions before a person will travel along with the stone and change its path to some extent with a broom.

22:54 KM: In principle, we have sweeping robots, but we didn't use them because our sweeping robots are not to the level that one would expect. So all the games are played under the wheelchair rules where there's no sweepers, except for one where we had a mixed team. I think one important thing to stress, is usually the press always wants to tell the stories about humans against robots and robots being so much better than humans. The interesting thing is that we played a mixed team, the robot was the thrower and there was a human skip, so the head of the team, and the skip was directing the sweepers, the human sweepers. And so that was a very, very good team. I can tell you that.

23:37 SC: That mixed team, that's very interesting. I would say that the video I've seen, the robot looks a bit like a wheelchair. It holds the stone low down on the ground under a squared-off surface and then, let's go. It's not graceful like a person, but it seems like it's pretty good. What were the results of the competitions that you ran?

24:00 KM: Altogether we played nine games, the first five against a high school team. This was basically to have the robot learn and to calibrate the AI. There were different development steps in the quality of the AI. Then there were four tournament-grade games against national level teams, and we won three out of four.

24:24 SC: Were there different strategies that the robots took on depending on the team or the game that they were playing?

24:31 KM: There's more than a dozen different throws that you can do at any moment. It's a nightmare because you always have to decide which one to choose and you have to predict whether or not given the uncertainty that the stone arrives at the place safely. And since you want to win,

you really want to place that stone where it belongs, and all the throws have a different probability to arrive.

24:54 SC: People that are good at this game, these are not amateurs, these are high-level players, they're not that successful at placing the stone exactly where they want. This is just literally a difficult thing to do.

25:05 KM: Absolutely. There's actually a difference between the X and Y coordinate. The Y coordinate would be along the long ice path; the X coordinate would be towards the shorter side of the field. So you can be very precise in the X coordinate, but the Y coordinate is a challenge. In other words, the stone can travel plus/minus 2 meters to the designated position, which is the reason why sweeping is so useful. Because if the stone is not fast enough, then you can make it faster by sweeping.

25:39 SC: So if you look at those numbers, how good people are at getting the stone to where they want, and the numbers from the robots that you used here, are they comparable?

25:48 KM: They are pretty much comparable. There's about one and a half meters in Y direction. In X direction, it's very precise, something like 30 centimeters.

26:00 SC: In this game where you were facing off humans and robots, were the people playing like they would with other people? Do you think that they were trying very hard basically to beat these robots?

26:11 KM: I think generally in the beginning, the humans did underestimate our robot. And two human teams play against each other, there's a lot of psychology going. Like in a tennis match, one player really has this aura of being able to win. Our robot cannot read auras of people, so just their...

26:33 SC: They don't know what their confidence level is, their demeanor, anything like that.

26:35 KM: Yes, exactly. So the robot will just do what it's learned to do, namely place a stone where it wants, and it will do this very reproducible, whereas humans can get nervous.

26:47 SC: They get the crowd effect.

26:49 KM: Well, it's that, but I think it's like with every game, if you start losing, then also your confidence fades, which accelerates the process. I think that the human teams generally were very open, and sometimes they also cheered for the robot when it could do a really good move. So they had mixed feelings about the robot, I guess.

27:11 SC: When you line up all the problems you have to solve that you're trying to take on with this kind of challenge for an AI, it really seems very applicable across a broad range of topics. What do you think are some parallel situations that might be helped by your investigations?

27:27 KM: I think whenever you go to the real world, basically if you do robotics or something

alike, then you always have a few data points, you always have non-stationarities and you always have uncertainties. That is a standard challenge of everything that is out there, and so in a way, we provide a framework that can be a basis for further progress.

27:51 SC: Alright, thank you so much, Klaus.

27:52 KM: It's my pleasure.

27:53 SC: Klaus-Robert Muller is a professor for machine learning at the Technical University of Berlin. You can find a link to his science robotics article at sciencemag.org/podcasts.

[music]

28:05 SC: And that concludes this edition of the Science Podcast. If you have any comments or suggestions for the show, write to us at sciencepodcast@aaas.org. You can listen to the show on the Science website at sciencemag.org/podcasts. On the site, you can find links to the research and news discussed in the episode. And of course, you can subscribe anywhere you get your podcasts. The show was edited and produced by Sarah Crespi with production help from Podigy, Megan Cantwell and Joel Goldberg. Jeffrey Cook composed the music. On behalf of Science Magazine and its publisher, AAAS, thanks for joining us.