Practical Significance | Episode 63: Real World Data Science—Who's Writing the Story

Donna LaLonde: Well, welcome everyone to *Practical Significance*. We have a real treat this month. We have colleagues who are involved with *Real World Data Science*. And so, we're going to start by asking Monnie, Willis, and Annie to introduce themselves and tell us a little bit about their day jobs. And then we'll dig in so that we learn more about *Real World Data Science*. And Monnie, I'll start with you.

Monnie McGee: Great! I'm Monnie McGee and I'm an associate professor of statistics and data science at Southern Methodist University in Dallas, Texas, home of the Mustangs. My main research areas are compositional data, sports analytics, text analytics, and I'm really interested in the uses of generative AI in education. Another hat I wear is the chair of the ASA Committee on Publications. And I do lots of other things.

Donna LaLonde: And you serve on the editorial board for *Real World Data Science*.

Monnie McGee: Which is why I'm here.

Donna LaLonde: Yes, I should have said that at the beginning. And your colleague representing the ASA on the editorial board is Willis. And so, Willis, tell us about your day job.

Willis Jensen: Hello! I'm grateful to be here and have a chance to talk a little bit about *Real World Data Science*. So, I lead a business intelligence and analytics team at CHG Healthcare. We deliver data analysis, business dashboards, data science work for a variety of business teams and executives that ranges from simple descriptive statistics, data visualization, advanced models, machine learning—all about driving better business decisions.

I've been in the industry for the past 20 years or so after getting a PhD in stats and really love the practical applications of statistical methods. And on the side, I'm an adjunct professor of statistics at Brigham Young University, where I . teach freshmen and sophomores an introductory statistics class and have a lot of fun doing that.

Donna LaLonde: And then, Annie, your day job is what brings you to the podcast, or at least a part of your day job. So tell us a little bit about what you do.

Annie Flynn: So, thanks for having me. I'm the Head of Content at the Royal Statistical Society, which is basically the UK equivalent of ASA. That basically means I'm responsible for ensuring that everything the Society publishes across all our different publications and different digital platforms and channels is aligned with our policy objectives and our strategic goals and kind of tailored to achieve maximum impact across our audiences.

I spend a lot of time working with our editors and editorial boards across the different publications and with our members to produce stories and write articles and all that kind of thing. And a big part of this is serving as managing editor of *Real World Data Science*, which on a day-to-day basis involves a mix of editorial direction, working with authors and contributors, and obviously the editors on the board, commissioning and producing content, writing a little bit of my own, and thinking strategically about how we can make data science knowledge more accessible and useful.

Ron Wasserstein: Thanks, Annie. Let's just jump right in then to *Real World Data Science*. I want to ask you what inspired its creation and then how has it evolved since it launched?

Annie Flynn: So *Real World Data Science* was created by my predecessor, a person called Brian Taran, who's a great writer and editor in this statistical space. And he wanted to fill a gap he felt existed between theory and practice.

There's a huge amount of excellent material out there on statistical methodology, but far less that shows what actually happens when those methods meet real data and real constraints and real decision-making environments.

The RSS wanted to create a space that focused specifically on applied data science—how it gets done in practice, what challenges arise, what lessons people learn along the way, all that sort of thing. And we have a suite of academic journals, including a newly created one called RSS Data Science and AI. And I bring that up because we feel like that journal is vital as a means of leveraging the depth and the permanence of traditional academic publishing to enshrine the foundational theory and validated methodology of data science.

But then we also wanted to provide the opportunity for practitioners in the field to be responding more in real time to new developments and new ideas. Obviously, it's such a fast-moving world. And what we always like to say about *Real World Data Science* is that it bridges the gap between rigorous analysis and real-time relevance. That's kind of our little tagline. And then since the launch, it's evolved in a few important ways. I think first of all, the scope has actually narrowed a little bit. Initially the temptation was to be an inclusive space for sort of all people and all things, and we really wanted to be accessible.

And I think that focus on accessibility led us in the first instance to create a significant amount of content for students and early-career scientists and that kind of world. And we've since realized that that kind of content is actually already really well done elsewhere, and that actually what the world needed more of was high-quality tutorials and case studies of new, cutting-edge methodology that's really relevant to established practicing data scientists.

So obviously we still want our content to be inclusive and accessible, but I think our focus is now a bit more squarely on that established practitioner. And then another recent focus has been creating more content for our audiences in the States. And Monnie and Willis have been a really welcome addition to the board in that perspective. And that's definitely a direction we're thinking about a lot more.

Ron Wasserstein: That's great. Well, then let's just switch over to Monnie and Willis. I'm going to ask both of you—maybe, Monnie, I'll start with you—what makes *Real World Data Science* different from other data science resources that are available?

Monnie McGee: One of the first articles that we published was called "Putting the Science Back in Data Science," so talking about using science to drive the data analysis rather than the other way around. And of course we want things that are statistically sound, but we also want them to be context-driven and making sure that we're answering the question that was posed.

One of the things we kind of want to have there is beyond the clean data set. I guess, for lack of a better term, data don't come textbook-ready. And having people talk about their need to clean—or misadventures in cleaning, perhaps—are important issues that need to get out there. And I think that's something that usually doesn't get put in a journal, but we need to have places for that so people can understand that you can't just riff off an article in an hour or something like that.

There's a lot that goes into it and all these decisions that have to be made, and these decisions sometimes change the focus of the study, but they're all good decisions. So those kinds of things are what we're interested in. And I believe ethics is also an important feature of this—the fact that ethical data science is good data science. And that's something that I would love to see features on as well.

Ron Wasserstein: Well, that's definitely putting real talk into *Real World Data Science* for sure. So Willis, what are your thoughts about what makes *Real World Data Science* different from the other data science resources that are out there in this information era that we're in?

Willis Jensen: There's so much content out there, and generative AI is only making it harder to find good content. And as Annie mentioned earlier, there's a lot of content out there geared toward the beginner. And I think it's really trying to provide higher-quality content that's vetted, curated by experts.

You can trust the content because you know someone else has looked at it, reviewed it by real data scientists—people actually doing this work in industry and practice. So that to me, the quality of the content is going to be so much higher than anything else that you can find out there.

Donna LaLonde: So Willis, I'll stay with you. Following up on that, what types of submissions make you stop and say we need to publish this?

Willis Jensen: I've been on editorial boards for academic journals and it's a very different feeling of what you're looking for because you're not looking for the most novel type of approach or cutting-edge statistical methodology necessarily. I mean, certainly we would welcome that as part of it, but it's really about the impact of the approach. We're thinking of examples and case studies—did it answer a question? Did it impact a decision? What was the impact of that?

And something that has a high impact on an organization or business or solving a problem or answering a question—those are the ones that made me feel, oh, okay, this is something we need in here because it's practical, it's real, and you can really quantify that impact.

Donna LaLonde: And Monnie, you've mentioned some topics that you are particularly interested in, but what is it about the submission that says, "Oh my gosh, we have to publish this!?"

Monnie McGee: I really like the idea of the honest-failure submission—the ones that when things didn't go according to plan—because, you know, a lot of the journals and other websites are full of things that went well, or at least the finished version of things, the data analysis. So I'd like to publish pieces that talk about the road there. Like we tried this state-of-the-art LLM, it failed miserably because of X, and we used some statistical methods to show that this is the reason that it failed and that we had to make this shift to fix it.

And it's very rare to see that; most people don't want to talk about their mistakes, but I think it's really important to have those things out there. One thing I ask when I read an article that's submitted, I say, "So what?" If the answer is, "So I can rank on Google," you know, I don't really want that. But if the answer is, “So a practitioner can avoid a catastrophe,” that's an article that's much more suited for *Real World Data Science*, I believe.

Ron Wasserstein: So Annie, in your role as managing editor, how do you balance this distinction that we're talking about here—between technical depth and balancing it with accessibility for readers at different skill levels? You're going to have all kinds of people at all kinds of places coming into this resource. How do you try to strike that balance?

Annie Flynn: It's a really important question for us and it's something we thought about a lot in the context of what I was saying earlier about whether we wanted our focus to be as inclusive as possible or more squarely on a more experienced practitioner. And I think the starting point of that is that we want all of our content to be as accessible as possible in terms of how it's written—clear language, well-explained context, good structure—and that's all things we can help authors with, but not necessarily introductory in its technical level. So really we're focusing on material that gives practitioners something they're less likely to find in standard learning resources. That's all the things that Monnie and Willis just mentioned in terms of what they look for in an article: the focus on application, the opening of the black box a little bit with the honesty of the path to get to the result, all those kinds of things. And then I think how we would balance depth with accessibility in pieces like that is through explanation rather than simplification.

We always encourage authors to make their reasoning explicit and to define their assumptions and explain their decisions, but without stripping away the technical substance. So a good test for us is: could a practitioner from a neighboring field potentially follow the gist of the content, even if it's not that exact specialism? And if yes, we're usually in the right sort of place.

The other thing I'd say is that we work quite closely with contributors during editing to improve clarity and flow so that technically rich pieces are still really readable and engaging. And I think if an author has a great piece of work or a great method that they want to explicate, we can help with the writing side of things. Always the goal is not to dilute the content; it's to make really strong, practice-focused material that's usable by as broad an audience as possible within that practitioner demographic.

Ron Wasserstein: So thanks for that. Not easy targets to hit, but you are all working hard on that, I know. Willis, do you have anything you'd like to add with regard to this balance of technical versus accessibility?

Willis Jensen: Definitely. Just putting on my data scientist hat is what I do. I've been working in the space for 20 years, and that's true for many of the others on the editorial board, and they're doing data science day to day and have been doing this for a long time.

So it's really just think about my own work and my own job and say, would this be helpful to me in my job? Does it make sense? Can I understand it so I don't have to do a lot to step outside of that data scientist hat or practitioner hat that I'm putting on? Use that as a tester, validate to say, yeah, this has got a good balance because it works for me. And it doesn't read like a journal article or an academic publication.

It reads a little bit more like a blog post or something that's engaging, shorter than a traditional journal article that you can get through in a reasonable amount of time and get the point and use it.

Ron Wasserstein: That makes sense, Willis. Monnie, you have to do this kind of balancing juggling act in all kinds of phases of your career. So how are you bringing that to *Real World Data Science*?

Monnie McGee: What we're looking for is something that is accessible, but that doesn't mean it's dumbed down. It's a clear path to some complex topic. And Willis just mentioned that these articles tend to read more like blog posts, so it's more like a narrative than it is a technical journal article. And I think a strong narrative is the best way to bridge different skill levels and to balance that accessibility with the technical content and the technical details.

When you're telling a narrative or telling a story, they feel like they're necessary steps rather than just dry data, and I think that helps. Questions like Willis was asking about, well, how does this work for me—it makes the why accessible and makes it rigorous for experts too. And one thing I want to point out too is that you notice Willis has a very different job description than I do.

And we have members on the editorial board that are industry members and academic members and I think even some government members. And I think that balance on the editorial board is helpful because we do have different levels of understanding and different backgrounds.

So if somebody can't understand it who’s been assigned to read an article, then that's a flag for us to try to figure out what the issue is and try to make it better and make it more accessible without losing the rigor.

Donna LaLonde: So, Monnie, I'll stay with you and then go to Willis. I'd like you to put your ASA hats on maybe a little bit more firmly now and think about what kind of content you're not seeing enough of. What emerging areas would you like to see addressed by submissions? And I'm going to say in particular by submissions from ASA members?

Monnie McGee: Well, I am an educator and one of the things I would really love to see is a classroom-ready, end-to-end case study. So that we do have narratives on there that tell a great story about a project, but educators kind of need to see more than just about what happened—maybe some sort of teaching kit or something like that.

I know we don't want to really focus too much on telling people what a neural network is, for example—there are a lot of different places to get that done—but to have more of the end-to-end, how you would structure this in a class sort of thing would be helpful. And teaching those communication and collaboration skills I think are really important, and that's been brought up in other areas as well.

And then maybe some things about sports, of course—that's one of my areas. I would love to see that: how that works, how people get the data and analyze the data, move forward with that data, and find real-world applications for that data. And maybe a little more about ethics, or how you balance that with some things that you might be asked to do. And the use of generative AI, of course, is one of those big topics that we talk about a lot. It's like: when is it appropriate, when is it not appropriate, how can you use it in a way that is ethical and smart and helps you with your work productivity, but at the same time doesn't give everything to AI—that there's a human in the loop, so to speak.

Donna LaLonde: Thanks for that. So Willis, keep your industry hat firmly in place and speak to your industry colleagues in the ASA. What should they be thinking about writing about? What areas of submissions would you like?

Willis Jensen: Increased case studies, examples—Monnie mentioned that—but any type of area where there's a case study, whether it's business, whether it's healthcare, whether it's education, wherever data science is being applied, we'd love to see that wide variety of application in there: case studies and examples. And ethics was mentioned already; certainly welcome things like that—socially responsible approaches to data science and whatnot.

I found as an industry member, sometimes it's been hard to get things published in traditional journals if I have a case study or an example, because those journals are looking for something that's novel and new and not the standard run-of-the-mill regression or machine learning method or neural network or whatever it is. And here we would welcome those examples.

It's not necessarily about the novelty of the method as much as it is maybe the novelty in the application or the novelty in impact or whatever it may be. And I guess one other area I would mention that's emerging is that we've been doing a lot of these methods and approaches where the data scientist is throwing every possible feature into a model and aiming for accuracy in that model, which is very valuable for certain situations.

But I'm finding a lot of end users—and this is true for generative AI as well—are getting a little more nervous about the black box nature of these models and we want to understand the science behind it. We want to understand how those models are working and why they work the way they do. So we talk about explainable methods or explainable AI. So examples or case studies or research into how do we make these models more interpretable, how do we communicate that to a decision maker and end user of that model to where they feel more comfortable using it.

It's less of a black box to them without peering into the deep technical details of something that they may not understand. So just approaches or methods to make models easier to interpret and understand, and they become less of a mystery for people that are using them.

Ron Wasserstein: Thank you both. So, Annie, putting on your editor hat and talking to me, say, as a potential contributor—walk me through what the submission and editorial process would look like from my perspective as a contributor.

Annie Flynn: We have tried to design the process to be as supportive and transparent as possible and also familiar for people who are comfortable with technical workflows. And that kind of touches on something which is quite fun and different about the site, which is that we support a range of different formats.

So you can submit to us with a Word document or any kind of normal text file, whatever you like to use. But we also support code-rich and reproducible documents. You could use tools like Quarto, RStudio, VSCode—you can work locally and submit to us HTML drafts, QMD files, all that kind of thing. And I think going back to what I said about my predecessor Brian, who founded the site, that was a really crucial piece of his original vision because I think he wanted people who are working at the coal face, you might say, of data science to be able to directly translate their work and their way of working to submitting something to the site.

So actually I would say we don't get as many submissions in those kinds of formats as we would like. We'd really welcome more of those. People do tend to submit with just a normal Word doc, which is also absolutely fine, as I say. Well, you'd start normally with getting in touch with us with an idea, and then we would work with you to shape it into a content brief and think about the audience and the key messages and the scope, and set you up to succeed before you go ahead and produce your first draft.

And then after that, the editorial review phase is maybe more transparent and collaborative than people might be used to if they're used to submitting to academic journals or other sorts of platforms like that, because you get clear feedback from the board and kind of work with them to refine the article. There's usually at least one revision round—I would say usually more than one for technical pieces. And I always tell authors when they get their first round of feedback: it's absolutely normal to have a lot of comments.

Obviously, we're really intent on being as rigorous as we can be and making the piece the best it can be, but it's all very collaborative. And then you'll always get to preview the final version, of course, before it goes live to make sure you're happy with it. And then we'll always support you with promotion and sharing. So to sum up, it's really about balancing it all—being very rigorous, but also being very author-friendly and well suited to practitioners sharing that applied way in the format I used to.

Ron Wasserstein: That's helpful, especially that you are trying to help people succeed as opposed to weed people out or whatever. So I get my piece published, as it were, on *Real World Data Science*. What can you tell me about how I would go about citing that or what you're doing to make that work discoverable?

Annie Flynn: So every article closes with a linked citation and we promote all the articles across different social channels. So we've got a LinkedIn, a Mastodon, an X, a BlueSky, and we try and get our articles out into the world sort of as loudly as possible through those platforms. And we also encourage authors to share themselves and all their own networks and we'll basically support that in any way we can.

Willis Jensen: If I could just jump in here real quick—so there's no paywalls on the website. This is openly available for anyone that is finding content. So it's there and available and it's quick and easy to get something out there. Maybe in the past, from an academic or traditional ASA perspective, we've looked down on citations to websites and whatnot because maybe it's not the same standard as a journal article.

But I believe because of the curation and review process that we have with *Real World Data Science*, this is a valid, high-quality citation. It's gone through a review process. This is more than just your typical website link that you don't know how high of a quality that is—it's something different because it's *Real World Data Science* and vetted and curated here.

I think for practitioners or those who are looking to submit articles and some of the things in there, I know maybe in the academic environment there may be challenges with getting something that counts as credit toward that. Our target is the practitioner as opposed to the academic.

And so for that practitioner, you can get your word out, you can get your name out, you can get your brand out as a data scientist that's doing great work. And look—you've been published here in this great article online. It's about that publicity and marketing as opposed to that traditional academic citation.

Ron Wasserstein: That all sounds really good. Thank you. If I had anything to contribute, I would absolutely submit. After hearing all that, thanks for providing a forum that gives people the opportunity to make contributions that are not standard for journals but that are important to be made as well.

Donna LaLonde: And we can certainly put this in the show notes, but *Real World Data Science* is also highlighted on the ASA website. So if you are an ASA member and you're looking to find it, it is right there for you. So encourage you to investigate if you have not. I'm going to ask all of you to think about the person—maybe Willis, I'll start with you because you were describing maybe the industry person who hasn't taken the plunge to write yet—what advice would you give for that person who is an experienced data scientist but has never published before?

Willis Jensen: The advice I would say is: put aside everything you know about the traditional publishing process or what you may have done when you were trying to get your dissertation research published in your first journal or whatever it is. Maybe you're not doing that anymore. This is different. This is not an adversarial referee-writer relationship where we're trying to dock you for not getting the grammar perfect and everything in there.

It's very much a partnership, and Annie referred to that partnership of we're in this together—we're trying to get high-quality content that's available. So even if all you have is that idea and you don't have something fleshed out, just share with us the idea and we can work with you and help you to get something that would be really valuable—as opposed to feeling like you have to have it perfect right from the get-go. Just get something out there, and as an editorial board we'll work with you and help you to get there.

And Annie, I will say from experience—and having published a couple of articles in there—Annie is an expert in this and she brings in all of that knowledge of how to make content engaging and relevant and following the styles that we want and wordsmithing and everything there. So we rely on her and she does great with that. So I'm grateful to have Annie as a partner as someone who has actually done a couple of publications for the website.

Donna LaLonde: Well, Annie, I think I have to go to you and ask—what would you add to that?

Annie Flynn: That's so nice of you to say. Thank you so much. Willis hit it on the head, really. I would just reiterate: start from your practice. Keep it simple and start from your practice. You don't need to write like an academic, as we've said—just clearly explain whether it's a problem you're trying to solve.

Explain what that problem was, explain what you did, the decisions you made. And I think exactly as Willis said and as Monnie's touched on at various points, it's the honest reflections that are so much more valuable than perfectly polished theory. And we have lots of resources to help you shape nascent ideas into really valuable pieces. So yeah—just go for it and get in touch.

Donna LaLonde: And Monnie, what would you like to add to what your colleagues have shared?

Monnie McGee: I think one of the things that Annie's mentioned before, and Willis as well, is you don't have to submit your polished article. You can submit an idea or an outline. You can just brainstorm something that you want to do and submit that. And because we have an iterative process and we want to help people get stuff out there, if the idea is good and we believe that the audience is there for the idea that's presented in the article, then that's fine.

You know, obviously we want something that we can understand—it can't be complete stream of consciousness—but we also work with people, and we've said this before, a much more supportive sort of review process. And if you are in for that, willing to be coached—we've had suggestions before, like people have submitted an article and we've said, well, really, this should be two articles.

So, you know, we help people break things up and help people make things more accessible. That's where we are. So if you don't have experience with writing, this is a really good place to start.

Willis Jensen: We're not limited to just writing formats. If you are willing to be interviewed in a video to talk about your career experiences or an example, we've done that on the website as well. So we're not limited to just written. If you're more comfortable just telling a story or sharing a story or experiences about something, we would welcome that as well.

Ron Wasserstein: I love that. I love all of that. The two things that really stand out to me are that it's a collaboration and also that it's not a scarce resource in the journal world. You know, there's only typically so many articles that are going to be put in an issue and so good luck if you don't make the 10% cut or whatever. But in this situation, you are bringing people in, collaborating, and not operating as a scarce resource.

So let's look ahead. You know, that's always great when you're asking people to predict things. Nothing can ever go wrong with that. But it's not so much a prediction as a vision—what you hope for *Real World Data Science* over the next few years. And Annie, since you're the editor, we should start with you on that question.

Annie Flynn: The thing that I would like to see a bit more of, or what I would like to focus on going forward, is the community side of it. I would love for some of the pieces to build connections between practitioners and research and even decision makers in the spaces that the content's referring to—starting ongoing conversations.

So, for example, we've published a few pieces recently. Willis co-authored a really valuable one around the issue of data quality. And after the first one we published, some people wrote in with insights on that. More articles along the same lines of responding to the original argument, building on it, all that sort of thing have been coming in, and now data quality has become a bit of a hot topic for us.

And that's where we can have some real impact—if we're starting conversations and other people are picking them up. Maybe it's a little bit grandiose to say that I hope to start a data quality movement, but even just raising awareness of issues like that by getting more and more people talking about them is, I think, a way we could have some real impact. That would be cool from my perspective.

Ron Wasserstein: Thanks, Annie. Monnie, what's your thinking?

Monnie McGee: We mentioned earlier that often websites or a link to a website is not viewed at the same academic rigor level as a journal link. And there's some truth to that. But one of my visions for *Real World Data Science* is that it would not be seen that way because it is a curated, edited website—that it would be a place where people would go to find content that they can trust because they know that other people have evaluated the content and worked with the authors to try to make it as accessible and technically appropriate as possible.

And other things I mentioned before about this facilitation of end-to-end problem solving and the fact that data work—the messy stuff that we often do to get the data in shape—is really part of the process. As an academic, I always tell my students, well, the data work is 80% or 90% of what you're doing, but we often sweep that under the rug when we publish an article. And I know I'm guilty of it too—I tell my students the data work is 80%, but I did all that and here's your data set. That's not very helpful to them about that really not-fun stuff that has to be done in order to get data in shape.

And those decisions that you make while you're doing that can have an impact on the final interpretation or the final outcome. And to be honest with those decisions. And you know, I also see RWDS as applications for classical statistical concepts to modern things like AI architectures—like a frequentist analysis of LLM hallucination rates or something like this to help us understand these black boxes that we're dealing with, or something similar for neural nets or deep learning, whatever the application du jour is—to help people understand what those are and help people understand, from the policymaker side, that uncertainty is a part of it.

Like you said, it's really hard to make predictions about the future and you're almost always going to be somewhat wrong, but to kind of make people more comfortable with that notion of uncertainty.

Ron Wasserstein: Thanks. And Willis, let's hear from you about your hopes or vision for *Real World Data Science*.

Willis Jensen: Annie and Monnie have really shared a great vision and examples. Whether we see or not, we agree with all of what they've shared in there, I think the only thing I would add is: as statisticians and the ASA, sometimes we see data science as this separate field or maybe we're nervous about this newer field out there, or we've been doing all of this stuff with all the new hotshot data science people doing stuff that we statisticians have been doing for years. And I think over time I would hope to see more merging or synergy of that as a discipline, if you will.

And the ASA being involved with *Real World Data Science* I think is just another step in some of that merging—to say there's this field of data analysis and working with data and the practices and tricks and tips and things that we do to do that successfully to add value or similar. Whether you call yourself a statistician or whether you call yourself a data scientist or whatever it may be.

And *Real World Data Science* is a place that holds a lot of that content and best practices and tips and tricks to be successful no matter what area of that broader data analysis field you're in. So, I hope to see that in the future. We'll see. And then I guess the ultimate standard of success will be is this content the training content for the next-generative AI model because it's recognized as the gold standard of content and high-quality content.

Donna LaLonde: Wow. Well, it's going to be hard to top that, but this has been such a fantastic conversation. I'm sad that we're coming to the end, but I'm excited that I get to ask all of you to share what's on your reading list, your watching list, your listening list. I love to be able to add to all of my lists. So, Annie, I'll start with you.

Annie Flynn: Well, I was really excited about this question, and I wondered if anyone that you'd asked has ever mentioned Foundation, the Apple TV show.

Donna LaLonde: I don't think we have had that recommendation.

Annie Flynn: Well, I've been dying to speak about it to other people in statistics. So basically, it's based on Isaac Asimov's novel, which is called *Foundation*, and it's all about the idea of psychohistory, which is a fictional scientist that uses math and statistical modeling to predict the behavior of large populations over periods of time.

So, it's a sci-fi and it's set in space, but it's all about kind of like the rise and fall of empires on like a galactic scale. But the whole premise is statistical because it's all about how you can't predict individuals, but you can model aggregate behavior, which is of course a statistical idea, and then you can kind of predict different—following a model of the theory of empires—you can sort of predict what different civilizations will do.

And I think what makes it really interesting for people that spend their time thinking about statistics is that it plays with themes that we think about all the time—prediction versus uncertainty and model assumptions and how systems behave at scale and kind of the limits of prediction or how models influence decisions people make or over-influence them. And obviously it's dramatized and sci-fi and silly but also a little bit thoughtful. So, I was excited to recommend that one. It's really good.

Donna LaLonde: That's great. Thank you. And definitely add it to my list. Willis, what about you—reading, watching, listening, all of the above?

Willis Jensen: I've always got multiple books at a time that I'm working through. I love learning. So, I tend to read more nonfiction—history and science and whatnot. So right now, coincidentally, I don't necessarily always read statistics-related stuff, but I'm in the middle of *The Theory That Would Not Die*, which is a book that is about Bayesian methods.

And it's got a historical bent of how Bayesian methods were started by Thomas Bayes and all of the other scientists and researchers who were involved in transforming Bayesian methods to what they are today, and different applications—how it was applied in World War II to crack the Enigma code, and just different things.

So, it's super fascinating to get that stats-history mix in there. And then I would just say, as we're recording right now, the Winter Olympics are starting. So, I'm super excited to watch the Winter Olympics and all of that—especially because the Winter Olympics will be coming to my hometown in 2030. I live in the Salt Lake City area, and we had the Olympics here a little while back. And so, I'm excited to have the Olympics come back. So, I'm looking forward to that.

Donna LaLonde: Well, that is wonderful. And Monnie, I'm going to guess the Olympics are on your list as well. But what else are you watching, reading, listening to?

Monnie McGee: Yes, the Olympics are definitely on my radar. I'm very fond of figure skating in particular. Hopefully we'll be watching that soon. And as far as podcasts go, I listen to Hidden Brain and The Daily a lot. I know those are probably pretty popular. And then I'm going to say something that's kind of on the other end of the spectrum from these intellectual things—*Foundation* and *The Theory That Would Not Die*.

Disney recently came out with a screen adaptation of Percy Jackson and the Olympians, and I have been rereading the books and watching the miniseries because I used to read that book to my children and now that they're out of the house, I'm kind of nostalgic for that. And so that's been kind of my guilty pleasure.

When I go to bed at night, I read a couple chapters of one of those books. And as far as music's concerned, I'm kind of an old-fashioned person. I listen to the radio in the car, so we have a couple of great radio stations that play 80s, 90s, and 2000s music. And that's kind of what I listen to while I'm driving around. And if I'm listening to Spotify, I really like David Gray. That's another one of my favorites.

Donna LaLonde: Well, these have been fantastic recommendations and a fantastic conversation. So, thanks to all of you for joining us. And now we've come to the other tradition of *Practical Significance*, which is Ron's top 10. So I'll turn it over to my colleague Ron for the top ten.

Ron Wasserstein: Thank you, Donna. I haven't developed a taste for true crime podcasts, but I know they are hugely popular. Perhaps it would help if we developed such podcasts for misuse of statistics. Donna and I already have a podcast, but in case any *Practical Significance* podcast listeners are thinking of starting one, here is my "Top 10 List of True-Crime Podcast Ideas."

10 **Serial P-Hacker**—A gripping investigation into a researcher's relentless testing until something, anything, came back significant.

09 **Cherry-Picked: A True Data Crime**—The shocking investigation into a meta-analysis that somehow only found studies supporting the researcher's hypothesis.

08 **Bad Model**—An algorithm to predict patient outcomes became a cautionary tale about training on biased data.

07 **Missing and Imputed**—A deep dive into a research scandal where "imputed" really meant "made up" yet nobody noticed for nearly a decade.

06 **My Favorite Model Fit**—Partners in life and in statistics, Harriet and Sean share wine and stories about gloriously overfitted models they've witnessed in the wild.

05 **Dirty Data**—He said the data was clean, the documentation was perfect, and the measurements were validated—investigators would soon discover that all three were lies.

04 **Cold Case-Control Studies**—Host Marge Inovera reopens decades-old medical studies to uncover the selection bias that was hiding in plain sight all along. (Thank you, Car Talk Guys.)

03 **The Murder of My Assumptions—**Each week, podcast host Bill Melater asks statisticians about that moment when they discovered that all their assumptions were all violated from the start.

02 **The Dropout Rate**—When 60% of participants mysteriously vanished from the clinical trial, one biostatistician refused to let the results be published without answers.

01 **P-Value Junkie—**The shocking true story of scientists so addicted to $p<.05$ that they forgot what their research questions were.

Well, that's it for this episode of *Practical Significance*. We look forward to continuing the conversation next month. Thank you for listening to this edition of *Practical Significance*, the podcast of the American Statistical Association. A new episode will be coming your way next month from Amstat News, the ASA's monthly membership magazine.