Exploring the Impact of Statistics in Automated Driving Systems

Everyday use of automated driving systems (i.e., self-driving cars) seemed far off even a decade ago. However, it is on its way to becoming a reality. New cars are routinely sold with driver assistance features that fall along the continuum of self-driving technology. Further advances that move us closer to the ultimate goal of full automation are being tested on public roadways. There are predictions that self-driving cars will revolutionize, and in some cases eliminate, entire industries and professions.

Along with the excitement comes uncertainty about the scientific aspects and policy implications. How can we know whether these vehicles are safe, and how safe must they be before the public accepts them? How should data be gathered and analyzed to improve the technology? How should driver privacy be factored into the use of such data? An accessible overview of the state of self-driving cars is provided by *Wired* magazine.

This paper grew out of discussions of the 2019 ASA Impact Presidential Initiative Task Force, which was established as a 2019 ASA presidential initiative and aimed to identify important areas in which statistics can have a major impact. Autonomous vehicles were chosen as a topic of focus, given the potentially vast implications of autonomous vehicle technology for science and society, the data challenges, and many uncertainties involved with the development and deployment of autonomous vehicles.

The goal of this paper is to provide an overview of the relevant issues, assess the current state of statistical research and practice in this domain, and identify opportunities for statisticians to have an impact in this area. The interview questions were developed by the task force, and the paper was edited by the working group chaired by Susan Paddock of NORC with contributions from David Banks of Duke University, Fred Hulting of General Mills, Inc., and Jennifer Van Mullekom of Virginia Tech.

We sat down with three experts to ask a series of questions related to automated driving systems: **Maria Terres**, data scientist at Waymo; **Feng Guo**, a professor of statistics at Virginia Tech and the Virginia Tech Transportation Institute; and **Nidhi Kalra**, senior information scientist at the RAND Corporation.

Introduction Questions

Tell us about your work with self-driving cars.
Guo: My main research area is traffic safety modeling. I have a joint appointment with the Statistics Department at Virginia Tech and the Virginia Tech Transportation Institute, a leader in automated driving system (ADS) research. [ADS is the Society of Automobile Engineers' (SAE) preferred term for self-driving vehicles. SAE defines six levels of automation in J3016 standards, ranging from level 0—no automation—to level 5—fully autonomous operation.] I have opportunities to work on dozens of projects related to key issues associated with ADS such as crash risk assessment, driver interactions with ADSs, ADS testable cases, and driving system safety evaluation.

Kalra: I've been working on self-driving car policy for over a decade on topics covering a variety of social, economic, and sociotechnical issues. First and foremost among these issues is safety—how do we think about self-driving safety—but also environmental issues, regulations, accessibility—who will this technology help, who is it less likely to help—and labor issues.

Terres: I am a data scientist at Waymo, the Alphabet company born from the Google self-driving car project. I am part of a cross-functional team of data scientists who collaborate with engineers across the company to help compute metrics related to the safety and quality of our self-driving system. We bring together signals from our real-world driving, simulation, and external data sources to evaluate the self-driving system's performance.

How did you start thinking about self-driving cars?

Guo: It is a natural extension of my research. I have done a great deal of research on driver behavior, which is a major contributing factor in the majority of crashes, and self-driving cars can potentially mitigate or even eliminate human driver errors. I also worked on safety evaluations of some key aspects of the enabling technology for ADSs, for example forward collision warning and lane departure warning systems.

Kalra: I have a background in robotics. When I was at Carnegie Mellon, the faculty there had been working on self-driving for 10–20 years prior to my starting and it was just part of the water in which everybody swam. It was when one of the Defense Advanced Research Projects Agency (DARPA) Grand Challenges were taking place. And so that was my background in it, but I personally began when I arrived at RAND. It was my first day, and I was in line to get my picture taken, and the person standing behind me was James Anderson, a lawyer at RAND. We just struck up a conversation and I said to him, "Self-driving vehicles—they're coming and they're going to need a lot of really sharp legal brains." And we just started talking about how we need a really multidisciplinary approach. From that first day, we began really a very long collaboration between us on policy issues.

Terres: I joined Waymo in the fall of 2018, and at first glance, it probably appears to be an abrupt deviation from my path up to that point. My academic background (undergrad through postdoc) was all focused on environmental applications, from plant ecology to estimating the impacts of global warming and air pollution. After my postdoc, I spent a few years at The Climate Corporation (TCC), where I developed models to recommend fertilizer rates to farmers, aiming to reduce the over-application of nitrogen without reducing end-of-season yield. While data science for self-driving technology has a less direct environmental impact than these previous roles, my work at Waymo is similarly motivated by a desire to help improve the world around me and create a better future for the next generation.

To answer the question more directly, a few years ago, I started to hear more and more about self-driving technology. I became excited about the potential to improve the day-to-day life in our society by improving road safety and the efficiency with which we get around. By chance, I received a recruiting email from a self-driving car company, and I became excited about the possibility of playing a role in the self-driving industry. Less than a year later, I went on the job market and got offers from multiple self-driving companies. I knew I wanted to work for a company where I could have real-world impact, and the self-driving industry offers that opportunity in a way that most other industries cannot.

What has been the most surprising aspect of self-driving cars for you?
Guo: The astonishing development in sensor technology and artificial intelligence (AI) in ADSs. Self-driving car research goes back decades. However, it is only very recently that image processing, LiDAR, and radar—combined with advancements in AI and sensor fusion—make self- driving cars feasible.

Kalra: I have been surprised at how easily people and society have taken what was once science fiction, and now people are like, "Oh, yeah, self-driving cars, that's coming, that's around the corner." The ease with which something goes from imagination to reality, even while not many people currently have these vehicles, is amazing.

What are the greatest unknowns related to self-driving cars?
Guo: Can a self-driving car provide the promised safety, security, and reliability at massive scale?

Kalra: How is this going to shape our future? Let's imagine 1992, when people were first starting to use the internet and we had this idea, this sort of dim idea, that we could send letters faster. And now 30 years later, we can't do anything without the internet. I think, in a similar way, we have this dim idea that we're going to get into a car that can drive us somewhere but it sort of defies the imagination to know how will that capability change how we live, work, engage with each other. We have something that could be transformative in ways that we struggle to imagine.

What keeps you awake at night related to self-driving cars?
Guo: Whether the implementation of novel technologies in driving could lead to catastrophic consequences. For example, a software bug or malicious attack that would affect millions of vehicles overnight through an over-the-air update. An "I, Robot" movie scenario could soon become a reality.

Another one is that level 3 automation, which relies on drivers to resume control of the vehicle when they have not been actively engaged in driving, is a recipe for disaster. A great deal of research has shown that people are not good at doing this.

What are the benefits of self-driving cars?
Guo: Many. Self-driving cars could free billions of drivers from driving tasks and save an enormous amount of human resources; eliminate human error from the driving process, a major contributing factor for safety; and allow optimization of transportation resources at the society level, alleviating traffic congestion and pollution.

Kalra: The benefits could be tremendous. We have a lot of problems in our transportation system that remain unfixed, and driverless cars have the opportunity to transform a tremendous amount of that.

One can imagine a world in which, first of all, traffic transportation becomes much safer because we have really safe self-driving vehicles. We provide mobility to people who don't have it, whether this is older Americans who don't drive or don't want to drive or people who can't afford the cars. We can imagine environmental congestion goes down and environmental impact of transportation goes down, because we have electrified self-driving.

The benefits are limitless. But it's hard for me to answer that question without also thinking about the drawbacks, because the absolute opposite world could occur, as well. We have more road fatalities if the vehicles are not actually as safe as we thought, or congestion increases because the cost of transportation goes down, or the technology exacerbates inequality if it's only available for the wealthy.

Terres: The potential to improve safety on our roads is the primary benefit that I see from self-driving cars. My father has gotten into car accidents for as long as I can remember (thankfully they've generally been minor incidents). Each accident is unique, but they are often due to something small like not noticing the car behind him, and rarely something major like falling asleep at the wheel. I would feel much better about my father's safety if he could simply choose a destination and let the car do the driving; unlike him, the Waymo Driver can see 360° and doesn't get sleepy on long drives.

Of course, self-driving cars have the potential to improve safety for everyone, not just for my own family. According to the World Health Organization, there were 1.35 million deaths world-wide due to vehicle crashes in 2016 and, similarly, 50 million injuries worldwide in 2018. In the US, 94 percent of crashes involve human choice or error. Even a small reduction in that human choice/error percentage would produce huge improvements to Americans' safety on the roads.

There are also a lot of benefits that go beyond road safety, including mobility options for those who cannot drive, safe transportation opportunities for vulnerable groups such as kids and seniors, and a more enjoyable commute for everyone.

Big Picture Issues

What are the most pressing questions in the minds of regulators and policymakers about self-driving cars?

Guo: How to certify and monitor an ADS? Shall original equipment manufacturers (OEMs) self-certify their ADS product? An ADS depends heavily on software, and every new version essentially puts a new system on open road. We need to ensure the safety of an ADS and make recall decisions in a timely manner.

Kalra: Safety is by far and away the most important issue. Are they going to be safe? How safe do they need to be? How will we know? And who's responsible if they turn out not to be safe? The next question I think, especially in some circles, is the labor impact. Whose job is at risk? What jobs will be created, so on and so forth.

What challenges does a mixed fleet scenario hold?

Guo: An ADS needs to be capable of operating in traffic with human drivers. Therefore, the operating algorithm has to consider somewhat unpredictable human driver behaviors. Similarly, human drivers also need to learn how to properly drive near a self-driving car.

Kalra: If by mixed fleet one means some vehicles on the road that are driven by humans and some that are autonomous, this is how it will have to be. It is hard to even imagine a world in which there was anything different. I don't actually see this as a source of that much challenge because, as long as we have different developers developing different autonomous vehicles, it doesn't really matter if we have some of them being human driven and some of them being autonomous.

Are there questions that people should be asking, but are not?

Guo: As a relatively new field, challenges and questions constantly arise. One of the questions I am interested in is how to quantitatively assess the risk associated with an ADS with a limited amount of safety-critical event data.

Kalra: How can we use this technology to achieve transportation goals that have long eluded us? For example, if we are trying to reduce congestion, it makes both theoretical sense—and evidence would suggest—that autonomous vehicles will make it cheaper to drive because the cost of your time is taken away—you can do something besides steering the wheel. That means the latent demand for transportation will induce more travel, which would result in more vehicle miles traveled and more congestion. That's contrary to the policy goal of reducing congestion and reducing gas emissions and other environmental impacts. So, the question is, are autonomous vehicles a good technology to implement congestion control, transportation travel demand policy measures such as mileage-based user fees, coordination—things that might be hard actually in a traditional vehicle but quite easy with the self-driving vehicle?

How should people think about the safety of these cars?

Guo: They should have realistic expectations for ADSs. Although a promising technology, it is unlikely we can achieve the zero-crash goal in the near future. People are not great at understanding risk and have a tendency to overestimate risk when they don't have control over the outcomes. Airplane crashes have a huge impact on public opinion but few notice that more than 100 people die due to traffic crashes in the US *every day.* An important question is how safe is safe enough for driver acceptance of ADS technology.

Kalra: That's a multibillion-dollar question. I wish I had the answer. There are so many questions wrapped up in that. How safe are self-driving vehicles? How safe do we need them to be? How do we know whether safe is safe enough? If we were economic rational beings, we'd say as soon as we can start saving lives on average, we should put these vehicles on the road. So as soon as they're better than the average human driver or as soon as we can start replacing a bad driver with a better autonomous driver, then we're in business. That's not how people work. We have cognitive biases, and also it is about how we value life. I would love for people to appreciate how complicated this issue is. It is so much more than just a reliability rate. There are so many social and psychological components to it that I think we underappreciate.

What are the social barriers or incentives to adoption of self-driving cars?

Guo: Many issues need to be addressed, such as ownership, operational domain, pay for use, liability, and insurance.

Kalra: In the US, we tend to be pretty excited about new technology. There were really no cultural barriers to the adoption of the smartphones, for example. I actually think that's how self-driving cars will be received more so than with suspicion.

But the real test will be when something bad happens, and it will happen—scenarios such as a self-driving vehicle kills child chasing a ball. This scenario happens often enough with human drivers, but it rarely makes the news.

How do those who are not statisticians and are working on self-driving cars think about well-publicized yet rare deaths and serious accidents?
Guo: I can talk about my personal experiences. I have another PhD in transportation engineering. There is a systematic approach for crash investigation and identification of the causes of a severe crash through simulation and accident reconstruction. For pretty much every rare death related to an ADS, detailed crash causation has been identified. The goal is to improve the system so it can properly handle such scenarios in the future. Uncertainty is usually not a top priority in such scenarios. However, the observed crash scenarios are random in nature. Without a proper way to incorporate the associated uncertainty, it will be difficult to prevent future crashes.

Kalra: I think for people who are working on self-driving cars, each tragedy is different, it's very threatening. Every time this rare self-driving death or a crash occurs, everyone holds their breath and says, "Is this the thing that shuts down this industry?" I think it's not being thought of statistically because its impact on the mind of the public is so outsized. I think for those working on self-driving cars who are nonstatisticians, this is not viewed as a statistical issue.

Do you think the public is ready for self-driving cars? If not, when will they be?
Guo: It depends. There are several levels of automation, from level 1 with a simple driving assistance system to fully automated level 5 systems. A car with adaptive cruise control, automatic braking, and lane keeping can be considered a level 2 system, and the public generally embraces these technologies well. I think the public will be ready when ADS technology is ready.

Kalra: There's no monolithic public. There are some people who are all ready, who are gunning for it. There are people who will never be ready. Somewhere in the middle, I think people are cautiously optimistic that this could be a useful technology. And as the pilot studies become more prevalent and prove their worth, I think this will start to become normal. So, some people are at one end of the tail already. Some people at the other end are never going to be ready but, in general, I think we're moving toward more experience and more exposure.

Past/Current Statistical Impact

What are the areas in which statisticians can have the greatest impact on advancing self-driving car safety or technology?
Guo: Statisticians can play a key role in many areas. We can help in various ADS phases, including system development, system testing and certification, and post-market surveillance.

Kalra: Safety, hands down. Understanding how we make use of multivariate data, data from different sources. Data from different modes of testing such as simulating real-world driving tests. How do we put this data together to make inferences about safety? That's a huge and complicated statistical problem. This is where statisticians are so critical.

Terres: At Waymo, statisticians are advancing many aspects of self-driving technology, from computing metrics that help assess safety of the current software, to assessing subtle changes in self-driving motion planning, to optimizing fleet allocations. Our data scientists work in collaboration with engineers across the organization to ensure that changes to the self-driving software are validated and informed by data.

Anyone working very focused on a particular task runs the risk of losing sight of the big picture in which that task is embedded. This can be problematic regardless of the industry one is working in. However, when developing a self-driving platform, the big picture is particularly important due to the inherent interconnectedness of the system; a change in one part of the software could have trickle down effects throughout the software stack. The cross-functional nature of my team gives us a unique bird's eye view of the self-driving performance overall. This enables us to help identify priority areas for future work, separating persistent patterns in the data from the variability around those patterns.

What contributions are statisticians uniquely equipped to make?

Guo: We are trained to deal with uncertainty and make decisions with uncertainty. Statisticians can play a leading role in predicting and assessing crash risk under different scenarios, extracting useful information from massive amounts of driving data, testing whether an ADS meets standards, and detecting an abnormal number of crashes for fleets.

Kalra: Let me add one more, which is helping the public understand statistics. Statistical significance and confidence are very basic concepts to statisticians. Another huge contribution is helping people understand what the data show.

Terres: Driving performance metrics that acknowledge the inferential limitations induced by finite data sets. Particularly while self-driving technologies are being tested on a relatively small scale, it is incredibly important to understand the uncertainty in any extrapolations being made to a broader population. For the non-expert, it is easy to rely on point estimates or make generalizations from anecdotal evidence, and it is just as easy to be led astray by these approaches. A statistician can provide context to these point estimates and anecdotes, ensuring that the resulting inference is grounded and data-driven. The unique influence of a statistician is particularly apparent for situations where a nonstatistician may not even recognize there's a need for statistical expertise; this would include things like multiple testing, Simpson's paradox, and causal inferences drawn from observational data.

Give an example of how you (or statisticians in general) are making an impact on self-driving cars.

Guo: Forward collision mitigation is a critical component of ADS. I have been involved in projects using different data sources (e.g., naturalistic driving study data and fleet data) to evaluate the efficacy of collision mitigation in real-life driving scenarios. The findings of these studies provide information for assessing the safety of self-driving cars.

Kalra: Our report, "Driving to Safety" (Kalra and Paddock, 2016), asked a really important question: How many miles of driving would it take to demonstrate autonomous vehicle reliability? The assumption had been that if we test drive cars enough, we'll eventually get enough data to be able to make inferences about their safety. This paper used STAT101 techniques to answer a question that actually stymied the field. There was no prior analysis on this; nobody knew how to figure this out because nobody was calling a statistician. I finally called a statistician and the answer was, like, "Oh yeah, we can do this." It was amazing to me that this is a question that was being talked about in conferences and the self-driving community. And people were like, "Wow, this is so hard!" And this is where the value of having a statistician was just so apparent, because the community is asking a reasonable question that has a fairly straightforward answer and just was not equipped to know where to look for this answer.

I say this in no way to disparage the importance of what we did—it was a huge contribution—but people don't know statistics. I knew enough to know where to look for the answer, but it is telling that people with PhDs in so many relevant fields had not answered this question, and I was shocked.

Terres: Everyone in this industry wants to ensure that self-driving technology is released responsibly and safely. At Waymo, improving road safety is our mission and our highest priority. Within that safety-focused setting, statisticians are key experts, providing context around the inherent uncertainty associated with our current and future performance. Our expertise in uncertainty and random variation enables data-driven decision-making and rigorous evaluation of current self-driving performance. In this way, we help to continuously improve the safety of the self-driving system.

Methodological Expertise and Needs

What statistical methods or ideas are most important in this domain?
Guo: Just to name a few, statistical learning, rare event modeling, and epidemiological and experimental design are important. This is a multifaceted and fast-evolving field that is in constant demand for novel and advanced statistical methods.

Terres: Very few of our problems can be solved using an out-of-the-box solution, and instead our solutions must be derived from first principles. As such, a strong foundation in statistical thinking is critical to working as a data scientist at Waymo. On an average day, we will encounter problems that require an understanding of how to handle biased sampling algorithms, observational data, experimental design, hypothesis testing, probability theory, causation vs. correlation, and more.

At Waymo, consideration of uncertainty is an integral part of our operations for statisticians and nonstatisticians alike. Statistical fluency is common enough at Waymo that most individuals can draw conclusions from a confidence interval. But uncertainty is also discussed beyond the context of this STAT101-style setting. For example, uncertainty is readily apparent in the way we think about driving scenarios that the self-driving car might encounter. Imagine the self-driving car approaches an intersection. We know where the self-driving car is and where it is planning to go, but there may also be many other moving objects present, including vehicles, pedestrians, cyclists, cardboard boxes, etc. The precise path for the self-driving car will depend on the movements of those other objects (e.g., if we want to guess where all of these objects will be several seconds from now, it is no surprise that there will be some uncertainty in our guess). Some of these objects may be somewhat predictable. For example, a parked car with no driver nearby is unlikely to be in a different location. Other objects may be predictable but with greater uncertainty. For example, a car approaching the intersection with a blinker on is likely to make a turn, but it could also continue straight. This kind of uncertainty is commonly discussed at Waymo across all levels of statistical expertise.

What features of the data are unique to self-driving cars?
Guo: Self-driving cars rely on big data. High-frequency, high-dimensional, and high-resolution sensor and vehicle kinematic data are the driving force for an ADS.

Kalra: With self-driving cars, we actually have continuous data about driving. And so now we get to introduce some really interesting ideas of what makes a statistically meaningful event. Sure, we can take a step back and talk about lane departure or running a red light—these are also events—but what about a swerve or an unnecessary acceleration? We now get into a place where we can make inferences from a continuous data set, which is really an incredible opportunity for thinking about transportation safety. For example, in one of our reports on developing a safety framework for measuring safety, we talk about Roadmanship, which is the idea that you're a good driver above and beyond avoiding crashes. What does that mean from a statistical standpoint? How do you define good Roadmanship? And how do you use continuous driving data to measure it? There's an opportunity here to do some really interesting stuff.

Terres: This is the first time in history that self-driving cars are being developed, which means that many aspects of our data pose unique and unsolved challenges. In particular, I would call out that our day-to-day work often involves (a) sample sizes $\ll \infty$, and this is combined with (b) unbelievable depth of information for each sample.

In any given moment of driving, we have data from our sensor suite, which includes lidar, radar, and cameras covering 360°; data on what other objects were in the scene, where those objects were a moment earlier, and where they go a moment later; characteristics about the roads, traffic lights, and signs; and everything the self-driving car was thinking in that moment. That said, the number of driving moments that we can capture is naturally limited by the number of cars being deployed (or simulation scenarios we can run) at any given time.

At Waymo, we have already driven more than 20 million miles on public roads and more than 15 billion miles in simulation across 25 cities in the United States. While this doesn't compare with the sample sizes seen at big tech companies (e.g., Google, Facebook, Netflix, etc.) where they have nearly infinite data and can rapidly deploy randomized experiments across large swaths of a population within a few days, our depth of data for each data point combined with the relatively small number of data points is one of the key ways in which our self-driving data are unique.

I would encourage anyone interested in getting a better sense of our data to visit *waymo.com/open* and explore the Waymo Open Dataset. It's one of the largest self-driving data sets ever released for public use. This data set contains nearly 2,000 diverse driving segments with associated labels and sensor data, all available for public use and research. This is only one slice of the data we have, but it provides a unique opportunity for students and researchers to look at the data and challenges that come when developing self-driving technology.

What are the biggest methodological challenges to solving the most important problems related to autonomous vehicles?

Guo: How to ensure the system can handle unexpected scenarios. In real life, a crash happens typically because of an unexpected scenario. It is a "big data, rare event" problem. ADS development is usually based on normal driving and a limited number of crash scenarios. The crash scenarios will surely not be an exhaustive list of all possibilities. Limiting the system to a specific operational design domain could reduce the number of unexpected scenarios but will not eliminate the problem. A major challenge is how to generalize an ADS algorithm to scenarios not included in the system development.

Terres: The biggest three statistical methodological challenges I would call out are: (1) quantification of subjective performance; (2) extreme value or rare event statistics; and (3) high-dimensional data with a complex dependence structure. These are all nontrivial problems that require careful statistical consideration.

In many settings, it is straightforward to identify a quantitative metric that aligns with the question we want to answer. For example, a simple metric for ad effectiveness is click-through rate. However, many of the self-driving behaviors we want to measure do not have even simple quantitative metrics associated with them. For example, imagine a self-driving car approaches a stop sign. Did it do a good job stopping? What if I gave you two versions of the self-driving car's behavior in the same setting and it stopped in both versions. How would you decide which one was better (maybe one stopped too harshly, another stopped too long, etc.)? Defining these kinds of metrics requires creativity and a thorough evaluation to confirm that the proposed metrics are consistent with our subjective expectations.

In many settings, an algorithm will be evaluated based on its ability to capture average behavior, with many edge cases effectively ignored as spurious rare events. However, in the context of self-driving cars, it is important to understand the tail performance of our technology. This includes quantification of tail quantiles, as well as estimation of rare event frequency. An example of tail quantiles would be time-related metrics. For example, what's the 99th percentile of our stopping time in a particular setting? It is not enough that our average stopping time is within a certain limit; we also strive to be confident that our longest stopping times meet designated thresholds.

Examples of rare event estimation are perhaps more obvious. These would include estimating the frequency of injury collisions and comparisons to data sets and estimates based on studies of human-driver performance. Understanding the self-driving car's safety performance and our uncertainty about that performance is vital to help ensure we understand what rare rate estimation can and cannot tell us.

In my academic career, I focused on modeling settings where there were inherent interdependencies in the data due to spatial and/or temporal proximity. Accurate inference in these settings requires that the dependencies are accounted for in the statistical procedure. The same is true at Waymo, where our data is high-dimensional with a complex dependence structure. Handling our data sets requires consideration of multivariate spatial data (e.g., location of the self-driving car, as well as other objects in the scene, all embedded in a high-resolution map), multivariate temporal data (e.g., tracking the behavior of any object in the scene across time), and network style interactions between all these features (e.g., how fast a car drives will depend on how fast the car in front of it is driving). Any metric or model we develop will be influenced by these dependencies, and out-of-the-box statistical tools are simply not going to account for all of them.

Growing the Depth and Breadth of Statistical Impact

What additional expertise is valuable to solving statistical problems on autonomous vehicles?

Guo: Domain knowledge is important. Knowing the key research questions and challenges facing ADS development and evaluation will help statisticians to quickly provide solutions. Sufficient computing skills to handle big data are essential to conducting data analysis efficiently. ADS development typically involves an interdisciplinary research team, and being able to communicate with nonstatisticians is also critical.

Kalra: There's a field of research on how people understand science, technology, and engineering. I have to believe this is an important area of study in a way that would allow the hard science of self-driving safety to be converted into an individual's understanding of safety at a more intuitive level. So, this is a combination of social psychology, engineering, public policy, and statistics, and pulling this all together to understand how the technology is understood and used by people is a really important area.

Terres: Data scientists at Waymo work cross-functionally with diverse experts across the organization. As such, skills like effective communication and collaboration are an absolute requirement for the job; a brilliant statistical technique is irrelevant if no one is aware it exists or is willing to use it. Being able to explain subtle statistical details to someone who barely remembers their basic STAT101 course requires some practice, and I would personally argue that teaching experience as a graduate student is a great place to start.

What are barriers for statisticians working in this area, or for those who would like to work in this area?

Guo: Proper domain knowledge, understanding of the key challenges, and access to ADS driving data are a few barriers for statisticians working in this area.

Kalra: It's natural in self-driving to have mechanical engineers, electrical engineers, safety experts, epidemiologists, people with backgrounds in machine learning—but the word "statistician" doesn't often appear in the parlance in this community and that may be a barrier. I think it was really a branding problem more than anything else, because the second anyone says, "I'm a data scientist," the response is, "Come on over—we need you!"

How can a statistician move into this area? How can one get involved in working on self-driving cars?
Guo: Building a successful collaboration with a good team is the quickest path to moving into this area. ADS research involves many fields beyond statistics, such as system design and data collection. To get an appropriate level of domain knowledge is beneficial. Of course, solid knowledge of statistics is a must.

Terres: Data scientists at Waymo have a variety of backgrounds, including statistics and mathematics, as well as bioengineering and other scientific disciplines. When we are considering candidates, we look for a strong statistical foundation, strong communication skills, the ability to handle complex data, and an enthusiasm for our mission to develop self-driving technology and improve road safety.

I would recommend taking advantage of opportunities to gain statistics experience, whether that be through internships, research, or volunteer work. That hands-on experience will sharpen your skills and enable you to really shine. If you're interested in self-driving opportunities specifically, be sure to research the industry before interviewing; you should have an idea of the challenges we encounter and how you might approach them. If you find one of us at a conference, please feel free to reach out and ask us questions. I personally like to attend conferences, and I'm always happy to chat with students interested in hearing about life as a data scientist at Waymo. Finally, as we continue to scale, we're always looking for talented individuals to join our growing team, so check out our open positions at *waymo.com/joinus*.

Reference

Kalra, N., and S. M. Paddock. 2016. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation Research Part A: Policy and Practice* 94, 182–193.