Teaching and research statement

Teaching

I have taught in a wide range of non-traditional formats: short courses at regional, national, and international conferences, and webinars for a geographically diverse audience. Students in these classes have a choice and will not tolerate a poorly taught class. I take it as validation of my teaching quality that I have been re-invited over and over again to give talks by the same organizations (American Society of Andrology, The Analysis Factor, International Research Conference on Complementary Medicine, Medical Library Association, Midwest Society of Pediatric Research).

I don't have (and don't believe in) an overarching philosophy of teaching. Good teaching comes from the little things, and I attribute my teaching success to three little things: finding compelling examples, using humor to make a point, and spending a large portion of my limited teaching time on small group exercises. I want to add some commentary about the differences between the types of classes I have taught and classes taught in a more traditional format. I also want to mention some of the special challenges that I have faced with teaching in an online (webinar) format.

Compelling teaching examples

It takes a lot of time to find compelling teaching examples. They need to be vivid and memorable. They need to reflect the practice of statistics in the real world, but at the same time they need to avoid obscure details that only specialists in an area can follow.

In a book chapter that I wrote for Big Data Analysis for Bioinformatics and Biomedical Discoveries, I was faced with a challenge to find a compelling data set to illustrate the R programming language in action. Most data sets used in Bioinformatics, however, are narrowly focused and require a fairly specialized knowledge to fully comprehend. I chose a data set that had broader appeal and which was easily followed by anyone with a reasonable medical background, Son et al. Database of mRNA gene expression profiles of multiple human organs, Genome Res 2005. 15:443-450. This study took tissue samples from 19 different organs of 30 people (158 samples total) and fed them to a DNA microarray. This provided expression levels of almost 19 thousand genes. This made the data set large enough to qualify for a big data analysis. More importantly, the comparison was concrete and easy to discuss. How does gene expression differ in the heart versus the liver, or in the pancreas versus the spleen? Also important to me was that the data used in Son et al was freely available for download at the journal's website.

Another compelling data set is mortality on the Titanic, a large ocean liner that in 1912 struck an iceberg and sunk. One of the first things you might do with this data set is to calculate a two by two table of gender versus survival. The Titanic sunk during an era where society really did believe in the mantra "women and children first" though perhaps more so among first and second class passengers. The crosstabulation of gender and survival would look something like this.

```
##          No Yes
## Female 154 308
## Male   709 142
```

You can bring these numbers to life by pointing out that Kate Winslet was in the upper right corner of the 308 women who survived and that Leonardo DiCaprio, sadly, was in the lower left corner of the 709 men who died. The data set has an interesting feature: the relative risk of death is 2.5 times higher for males than females, but the odds ratio is much higher, almost 10. It also allows you to study a rather complex interaction of gender and passenger class. I've used the Titanic data set in a tutorial article about measures of risk (Simon SD. Understanding the odds ratio and relative risk. J Androl. 2001. 22(4):533-6), in a short course on logistic regression, and for some of the homework assignments in the Introduction to R, Introduction to

SAS, and Introduction to SPSS classes. I take as validation of this approach the fact that other groups (most notably Kaggle) have also adopted the Titanic as an instructional data set.

Compelling examples are not limited to data sets. In my book, Statistical Evidence in Medical Trials, I talk about blinding with specific attention to blinding in surgical trials. Surgery is easily understood: you cut someone open, take something out, and sew them back up again. But it requires special challenges to ensure blinding, such as the use of extra large bandages to cover up the size of the incision. I make special reference to a study of Parkinson's Disease where fetal cells were injected directly into the brain. The study had a control group, and in order to maintain the blind, the control patients also had their heads shaved, were put under an anesthetic, and the surgeons actually drilled into their skulls. The only difference was that nothing was injected after drilling if you were in the control group. Needless to say, this was a highly controversial study with one ethical expert writing a critical article with the title "I Need a Placebo Like I Need a Hole in the Head." The gruesome thought of skull drilling makes this example memorable.

There is no more compelling example relating to research methodology than the efforts to find a treatment for AIDS. Controversies surrounding AIDS research has changed the way we approach research today. When I lecture about surrogate outcomes, it is easy enough to come up with examples where reliance on surrogate outcomes has led us astray. The classic example is the CAST trial which showed that the use of anti-arrythmic drugs in patients with heart rhythm problems actually led to an increase in mortality versus placebo. It is harder to find examples where reliance on surrogate outcomes has helped us. AIDS trials, however, do clearly illustrate the value of surrogate outcomes. In the first studies of anti-retroviral treatments for AIDS patients, scientists originally advocated the use of "hard" endpoints like mortality and opportunistic infections. But AIDS advocates pushed for surrogate outcomes like changes in CD4 cell counts because hard endpoints took too long and required waiting until you had accumulated a sufficient number of bad events in the untreated group. It is clear that reliance on surrogate outcomes reduced the number of deaths in the clinical trial itself and allowed much earlier FDA approval of these treatments.

Using humor to make a point

My audiences are often apprehensive. Will this guy speak a lot of Greek and formulas to me? A bit of humor very early in the talk will often allay some of those fears. I try hard, however, to use humor as well to make a point. Humor comes from an exaggeration of a basic idea to levels of absurdity, which you can use profitably to emphasize your basic idea later in your talk.

In a lecture on sample size justification for a short course on grant writing, I started with a (fictional) story of a researcher who gets a six year, ten million dollar NIH grant and writes up a report summarizing the research saying "This is a new and innovative surgical approach and we are 95% confident that the cure rate is somewhere between 3% and 98%." This becomes a repeated touchstone for the rest of the talk. The agency that you are seeking to get grant money from wants some assurance that the results will be informative when you finish your work, and not a confidence interval that goes from 3% to 98%.

The previous joke is an original of mine, but I'm not above stealing other people's jokes. There is a classic about two statisticians travelling on an airplane and one engine after another explodes. After each explosion, the pilot announces that everything is okay, but the flight is delayed further and further after each explosion. After the third explosion, one statistician turns to the other and say "Boy I hope that last engine doesn't explode" [dramatic pause] "or we'll be up here forever!". I use this joke when I teach linear regression, because it illustrates a dangerous extrapolation beyond the range of observed data. Later, I introduce the intercept term and offer an interpretation (the estimated average value of Y when X equals zero). Then I point out that such an interpretation is a dangerous extrapolation when you have no data near the value X=0. In fact,the estimated travel time of a jet with X=0 engines is also an intercept term requiring the exact type of dangerous extrapolation.

I also use humor to assess my audience as well. At the very start, I'll tell a corny joke about degrees of freedom (a joke so bad that I won't repeat it here). I look at who laughs and who doesn't. When most of my audience laughs, I compliment them by telling them that if you understood that joke, you won't have any

problems with the rest of my lecture. When most of my audience is quiet, I tell them that they are going to learn a lot today (and I teach things a bit more slowly).

Humor can be overdone, and students are not listening to you for entertainment. I try to keep my jokes short and get them out of the way early. But a little bit of humor does seem to help a lot.

Spending time on small group exercises

I hate to do small group exercises. I really hate them. When you are giving a short course at a research conference, you have a limited time frame, usually three to four hours. Small group exercises can easily eat up a third to a half of that time. I'd much rather be talking because there's so much to teach and so little time. Even so, I include small group exercises in almost all my short courses. The feedback I have gotten from students has been that the small group exercises are the best part of the class.

Constructing a small group exercise is not easy. You have to find a problem that students can tackle, so it can't be too hard, but it has to be challenging enough to force them to use some of the things they have just learned. It has to be something that everyone in the group feels that they can contribute. And it has to be something that each small group can summarize to the entire class in five minutes or less.

The reason that small group exercises work is that students have a variety of different experiences and can learn from each other. The more experienced students benefit because they have to understand the concepts at a higher level in order to articulate these concepts to the rest of the group. The less experienced students benefit because they get a chance to hear about the same concepts a second time and from a different perspective.

In the short course on grant writing, we gave each group a different research paper (actually, just the abstract from the paper) and asked them to suggest a new study with a different research design that builds on the paper's findings and that an agency might want to fund. Showing just the abstract is a big no-no in Evidence Based Medicine, but we apologized and explained that if we handed them a full paper they would run out of time before they even finished the research paper. Each group then appointed a spokesperson to summarize the original paper (just the PICO: Patient group, Intervention, Comparison group, and Outcome measure) and then propose their new research design. Once the spokesperson was done, we asked if anyone else in the small group wanted to add anything and opened it up for comments and questions by the other students. Getting our students to visualize a new research study was one of the big successes of our short course.

In a short course on setting up an independent statistical consulting practice, I was actually able to get two small group exercises squeezed into a four hour format. The first exercise was pairing up students and asking them to share with their partner where they hoped to be five years from now. This served as an ice breaker and got students to think about what a career as a consultant might mean for them. The second small group exercise took some fictional accounts (loosely based on reality) of bad client interactions. The groups had to decide on the best course of action, and this quite honestly might involve just walking away from the project completely, an option that is always available to an independent consultant. The short time frame did not allow each group to summarize their findings to the rest of the class, but the students still found the small group exercises to be valuable.

In a short course, Statistics for Medical Librarians, that I have taught in an online format and at multiple regional and national meetings of the Medical Library Association, I used small group exercises to dissect statistics found in actual peer-reviewed publications. I teach concepts like confidence intervals, provide a non-technical interpretation, and then ask students to do the same in small groups. I show them actual papers (actually abstracts again to save time) that show how a confidence interval might look "in the wild." Then each group has to summarize the abstract using a PICO format and then interpret the highlighted confidence interval in that abstract. All of the abstracts that I choose are from Open Source journals so the librarians can track down the full articles after the class ends, if they are so inclined.

Online teaching

I wanted to mention a bit about the special challenges associated with online teaching, because I see a transition to online teaching almost everywhere I look. I have done webinars, a form of online training, for several several organizations. I've learned a lot from these webinars, though they are not the same as teaching a semester long class in an online format.

One lesson I have learned is that you have to work much much harder at establishing a rapport with your students. In every webinar I have taught, I do not have the option of requiring the students to make themselves visible on a webcam, and only in a few did I have the option of having students ask questions orally rather than through a chat box. This makes the webinars less fun, quite frankly, than a short course taught to a live audience. But I have embraced the webinar format because it allows the participation of a large number of students who are otherwise geographically isolated.

To make up for not seeing and often not hearing the students, I try to make myself seen–using my own webcam when possible. I also take extra time to more directly encourage student participation. It starts with a joke. I introduce the joke with an exhortation. It is difficult, I point out, to say something funny and then hear total silence on the other end, so I after I tell this joke, I want you to type something in the chat box. A "haha" if you liked the joke or a "groan" if you hated it. Then I read the chorus of responses after the joke. It reinforces the connection between me and my audience and it lowers the barrier for students to use the chat box later for more serious comments.

When you ask for questions during a webinar, you need to wait long enough for people to type their questions. And you need to follow up your answer with another exhortation "Did that make sense" because you don't get the feedback from the knowing nods or the glazed eye confusion.

A webinar format demands a good Powerpoint presentation. I do not always use Powerpoint for live talks and short courses and subscribe to the admonition of many that "Powerpoint is Evil." For a webinar, however, if you don't give students a Powerpoint slide to look at, it makes it that much harder to keep a student's attention.

I've taught two traditional courses (Introduction to R and Introduction to SPSS) in an online format. These are in an asynchronous format. Students listen to mp4 files where you lecture and show Powerpoint slides and live screen shots of R and SPSS as those programs grind through their computations. The one issue I've noticed with the asynchronous format is that if one student asks a question, the other students would not normally get to benefit from hearing that question and hearing my response. They also lose out on the opportunity to add something to my answer if they can. So it is very important to share questions and answers that otherwise would only come to you privately via email.

Neither Introduction to R nor Introduction to SPSS lends itself well to small group exercises, but I do recognize by watching others at my current job that small group exercises take a much different form in an online format. In particular, they require much more written communication than oral communication, and the effort, especially for an asynchronous format, is spread out over a longer time frame. This can actually be an advantage because you need to explain yourself more clearly in a written format and the slower pace encourages more thought as well.

No doubt that I will encounter other issues as I teach more classes in an online format. I learned a lot about online teaching at various talks at the Joint Statistical Meetings, and plan to keep my eyes and ears open for other people's experiences.

Research

My best research efforts have been collaborative. I feel like these efforts represent something that neither I nor my co-authors could have developed by ourselves. I've had many wonderful research collaborations over the years. I want to highlight tow of them:patient accrual in clinical trials and with mining information from the electronic health record. I also need to describe my efforts to help others be successful in their research endeavors.

Patient accrual

In 2006, I gave a journal club presentation on how to use control charts to monitor the process of patient accrual in a clinical trial. Too many studies, I hate to say, fail to meet their sample size requirements because researchers grossly underestimated the amount of time it would take to recruit patients. I discussed this control chart in the context of a clinical trial, but it really applies to any prospective research study that collects data from human volunteers.

Byron Gajewski, one of the other faculty members at the journal club, suggested that this problem might be better solved with a Bayesian approach. That turned my research around 180 degrees but it was worth it. His suggestion led to a very profitable avenue of research for the two of us.

The beauty of a Bayesian approach is that it requires the specification of a prior distribution. The prior distribution represents what you know and what you don't know about how quickly volunteers will show up on your doorstep asking to join your study. You can think of the prior distribution as a way to quantify your ignorance. You choose a very broad and variable prior distribution if you know very little about accrual patterns. This might be because you're new to the area, you're using a novel recruiting approach, and/or there's very little experience of others that you can draw upon. You choose a very narrow and precise prior distribution if you've worked on this type of study many times before, your recruiting techniques are largely unchanged, and/or there's lots of experience of others that you can draw upon. What you don't do, even if you are very unsure about accrual, is to use a flat or non-informative prior. A flat prior would be like admitting "I don't know: the study might take ten weeks or it might take ten years and I think both possibilities are equally likely." Someone with that level of ignorance would be unqualified to conduct the research.

The very act of asking someone to produce a prior distribution will force them to think about accrual, and that in and of itself is a good thing. But the advantage of specifying a prior shows during the trial itself. As the trial progresses you get actual data on the accrual rate, and you can combine that with your prior distribution, as any good Bayesian would, to get an updated estimate of how long the trial will take. Here is where the precision of the prior distribution kicks in. If you have a broad prior with high variance, then even a little bit of bad news about accrual during the trial itself will lead to a drastic revision in your estimated time to complete the trial. You will act quickly, either by adding extra centers to a multi-center trial, hiring an extra research co-ordinator to beat the bushes for volunteers, or (if the news is bad enough) cutting your losses by ending the trial early for futility. If you have a narrow prior with low variance, then you've done this trial often enough that you don't panic over a bit of bad news. If the data keeps coming in and it shows a much slower accrual rate than you expected, then you will eventually reach a point where you need to take action. But there's a cost associated with a premature overreaction that a precise prior will protect you against.

Dr. Gajewski put one of his graduate student, Joyce Jiang, on the trail, and she contributed several additional publications after completing a successful dissertation defense of her extensions in this area. I worked very closely with Drs. Gajewski and Jiang, and found an interesting theoretical contribution to Bayesian data analysis that was hidden in their work.

One of the problems with getting researchers to produce a prior distribution is that they sometimes are wrong–spectacularly wrong. If you have a strong prior attached to a prior that is sharply at odds with the actual accrual data, you'd like to find a way to discount that prior distribution, but you'd like to keep that strong prior for the precision it gives you when the prior and the actual accrual data agreed with one another. They came up with a very clever solution. Attach a hyperprior to the precision of the prior distribution. If the accrual data and the prior are in sync, the precision stays high. But if there is a serious discrepancy between the accrual data and the prior, the hyperprior shifts and leads to a much weaker prior distribution.

I dubbed the method they proposed the hedging hyperprior, and suggested that it might work in other Bayesian settings as well. It turns out to be equivalent to the modified Power prior proposed by Yuyan Duan in 2006, but the formulation of the hedging hyperprior is both simpler and more intuitive. I have presented a simple example applying the hedging hyperprior to the beta-binomial model and am preparing a manuscript for publication.

The work that Dr. Jiang did on her dissertation was not just limited to the accrual problem but included an additional Bayesian application to a validation model using expert opinion. The strength of her work in these two areas led to her appointment to a post-doctoral fellowship at Yale University. Dr. Gajewski and I continue to collaborate with her on these models. We have four peer-reviewed publications and an R package so far, and plan to collaborate with other researchers in this area.

Closely related to my research on patient accrual is an effort to audit the records of Institutional Review Boards (IRBs). Too often, researchers fail to obtain the sample size that they promised in the original research protocol, mainly because subject recruitment takes longer than expected. It is very easy to compare the protocol submitted to the IRB to the final report. In the study of 135 submissions to one IRB, more than half of the researchers failed to reach their enrollment targets and the average shortfall was more than 50%. I have approached many other IRBs to ask to replicate this work, but none have shown any interest. But I plan to continue to ask anyone who works on an IRB to help me.

Mining the electronic health record

In January 2016, I was offered the opportunity to work on a research grant funded by the Patient Centered Outcomes Research Institute. The grant supported the Greater Plains Collaborative, a consortium of ten academic health care centers in the midwest. It was run out of Enterprise Analytics, located at Kansas University Medical Center. I jumped at the chance and dropped much of my other work to focus on this grant.

My assignment, derived through discussions with the head of Enterprise Analytics, Russ Waitman, was to develop a develop a phenotype of breast cancer from information in the electronic health record (EHR) and validate it against information in the breast cancer registry.

Such a phenotype would have great value in identifying patients for prospective clinical trials. The advances in high throughput genome sequencing and the linkage of that information with the EHR allows for exploration of novel precision medicine options. Developing a phenotype from the EHR is fraught with peril because information in the EHR on basic issues like diagnoses and treatments is often coded inconsistently. Having a link between the EHR and a tumor registry provides an external validation of the accuracy the EHR phenotype.

The EHR at KUMC (and the other sites in the Greater Health Collaborative) is stored in an Oracle database and is accessible through an i2b2 system that is very easy to use. My work, however, required access to the full database as well as some of the metadata. With the help of Dr. Waitman and the SQL expert sitting across from me, Dan Connolly, I was able to pull information directly from Oracle.

I used a big data model, LASSO regression, to predict whether a patient was in the breast cancer tumor registry and set up sparse matrices as input to better manage the size of the data sets. The breast cancer cases were compared against three separate control groups and in spite of the massive size of the independent variable matrix (more than 45,000 columns), this model ran in under ten minutes. The resulting sensitivity and specificity were very high, putting to rest concerns that the EHR data might be too incomplete or inconsistent to produce an accurate phenotype. The LASSO regression model could easily be run for other tumor types, and just as quickly validated. I have presented these results at a local research conference and plan to submit a peer-reviewed publication soon.

An interesting side effect of this research is also worth mentioning. I had a passing knowledge of SQL prior to my work with Dr. Waitman and Mr. Connally, but I had to quickly learn and apply a broad range of SQL programming to get the data into R and the LASSO regression model. SQL is a fairly easy language to learn, but many of the students in the Bioinformatics program at UMKC do not have even a cursory knowledge of SQL. So I am partnering with a database expert at UMKC to develop a new class, Introduction to SQL, that will cover some of the basic skills that an researcher would need to query data stored in a relational database. This class is not intended to teach someone to become a database administrator, but rather a competent user of other people's databases. It will parallel to a large extent classes that I have already helped develop and teach: Introduction to R, Introduction to SAS, and Introduction to SPSS.

My work on the PCORI grant has been transferred a different grant and I plan to work with partners at Truman Health Center and Children's Mercy Hospitals to develop more research utilization of EHR at their locations. I also have been asked to help develop an analytics platform simplifies data mining of the EHR through a standardized library of functions interfacing SQL databases and R. This library would pull appropriate meta data descriptors as well, expanding the types of analyses available to the end user.

Helping others with research

I am a great believer in the Harry Truman quote "Anything is possible if you don't care who gets the credit." A major portion of my career has been helping others become successful researchers. This work is quiet, behind the scenes, and often leads to very little recognition for me. But I enjoy watching someone developed from a scared and timid person starting out with their very first research study to someone who has learned enough that now he/she is mentoring others.

A large part of my work is helping people who are struggling in their graduate studies. It might be some extra tutoring for that seemingly impossible statistics class. More often, though, it is guiding students through the difficult process of writing and defending their dissertation. I did this for free for doctors, nurses, and other health care professionals that I worked with at Children's Mercy Hospitals and Clinics. After I left that job, I started my own consulting business, and I got lots of referrals to graduate students. They typically are struggling with their disserations and with a dissertation committee that was not giving adequate direction on the data analysis. For a dissertation, you can't do the data analysis for them because they have to be able to explain their work during the dissertation defense. You have to teach them those things that they didn't pick up in their earlier statistics class and teach them so thoroughly that they can offer a clear and convincing presentation of the analysis that they ran. You have to help them anticipate the types of questions that they might get and how best to answer those questions. Perhaps the most important thing is to get them a sense of self-confidence that they are working on an important problem and that they have a solid and defensible plan for solving that problem.

Another big portion of my work is helping people write their first grant. The first grant is almost always for an amount too small to support a statistician as a co-investigator. But these are the grants that need to most help and support. I help with the research design, the sample size justification, and the data analysis plan. But my work is not just limited to that. If the specific aims are vague, if the literature review rambles incoherently, or if the research budget is unrealistic, I offer gentle suggestions to fix these problems. I try very hard not to overstep my bounds; I don't know the science as well as they do. But I can often provide valuable feedback by looking at things from the perspective of an outsider who has seen hundreds of grants in dozens of different scientific fields.

I've taken many classes in grant writing to better understand the whole process and to improve my ability to work with new researchers. But recently I have taught the classes myself. In 2012, I co-developed a class in grant writing for researchers in Complementary and Alternative Medicine with a prominent statistician working for what was then called the National Center for Complementary and Alternative Medicine and with another statistician working at a major college of Chiropractic Medicine. We repeated this class in 2014 for the same conference. I provided lectures on designing a pilot study, justifying your sample size, and what a scientist should look for in a collaborating statisticians. I gained the most benefit, however, by listening to the lectures of the other statisticians. One of the best was a discussion about insuring consistency between the specific aims, the research design, the data analysis plan, and the budget. Another excellent talk was on hiring a data manager and preparing a solid data management plan (including budgets!).

When you include the small internal grants that provide seed money for new researchers, I have helped write hundreds of grants, too many to track. I do list those grants where I am on the budget, but these represent just the tip of the iceberg.

Finally, I help new researchers navigate the daunting process of getting Institutional Review Board (IRB) approval for their studies. At Children's Mercy Hospital and Clinics, I was housed very close to the people who ran the IRB and got to know them very well. It would have been a conflict of interest for me to serve on the IRB, because I would be reviewing protocols that I helped write. But I did work extensively with the

IRB, answering technical questions, accepting referrals from researchers whose protocols had glaring problems with scientific rigor, and providing training courses on ethical issues associated with research. Statisticians, I believe, are key players in assuring the ethical conduct of research. Often researchers are barred from the optimal research design by ethical constraints and our job is to help find an alternative design that meets the needs of the researcher while still protecting the rights of the research volunteers.