Things they did not teach you in graduate school. J Hanley. SSC Student Conference, Montreal, June 2, 2018.

james.hanley@mcgill.ca

Bienvenue a tous, and thanks for attending. A big thank you to the excellent organizers. I picked this topic because much of what statisticians use in practice is learnt 'on the job.' 32 / 32

I will draw on some older material. At the Irish statisticians' meeting in 2005, they limited me to a 4 minute after-dinner speech, as they were in a hurry to get to the bar. When I put the full version on my website in 2009, I dedicated it to 2 of the several great mentors and colleagues I have had. I only observed Fred Mosteller, on your left, for 3 years while I worked in Boston, but his attention to good communication and professionalism left a lasting impression on me. I had met Steve Lagakos 4 years earlier, when I started in Buffalo. Here is a clip of Steve taking us

on his sailboat into the US side of Lake Ontario. The date is interval-censored, somewhere between 1975 when I grew the beard, and 1977 when we moved to Boston. I will have more on Fred and Steve later.

I will also tell a story involving this mentor, who introduced me to ROC curves, and is still going strong.

But I would not be here today if it weren't for this person. In 1969 he offered me a PhD slot in SUNY/Buffalo. I accepted it, but then a few weeks later when Waterloo offered me a slot, I took it instead. But, as it happened, Marvin gave us courses as an adjunct professor at Waterloo, and recognized my name, and asked why I switched. I made up some excuse. During my PhD, when I was having doubts about what I was working on, he gave me very good advice: "it's not what you do for your PhD that matters, but what you do afterwards". Then, in 1973, when

I finished the PhD, he gave me a job in his clinical trials unit. He moved his team of clinical trial biostatisticians to Boston in 1977. So, I was with him for 7 years in all. He taught us all to not be 'tame statisticians' who gave the ECOG and RTOG doctors the p-values they asked for, but to be an equal scientific member of a research team. And it was he who gave me the confidence I have today.

There were several topics in that long written piece I have on my website.

Today, I will cover bits from the underlined ones. They mainly deal with communication, and my earlier years and embarrassments. Then I will try to explain the strategies I have used over the last 52 years, since going into statistics at age 19, to truly learn and understand statistical concepts and principles. I'm not talking about **machine** learning; I am talking about **my** statistical learning. And I will finish with a few remarks on practising as a statistician. 78 / 482

In that 2009 piece I tell how, after 2 weeks coding in FORTRAN, I still had not computed the (exact) power calculations for the "sample size considerations" section for my first RCT with the ECOG. Stuart Pocock, who spent 2 years with Zelen, joining a year ahead of me, taught me my first real-world lesson when he asked me if I had heard of the Normal approximation to the sampling distribution of the difference of two proportions. 77 / 559

My next lesson was about **communication** rather than **calculation**. I sent my interim report on the results of an ongoing trial (a comparison of chemotherapies for advanced colon cancer) to the oncologist who was principal investigator. He phoned Zelen, my boss, to say that my analysis "had to be wrong." I had written that the "estimated median survival" in arms A and B was 7 and 9 weeks respectively. He kept saying "look at the written protocol: it can't be!" Finally, after several futile attempts to explain my analysis, and why it was correct, and how the Kaplan

Meier method deals with censored observations, etc., we realized that he was referring to the *entry* criteria. One of these stated that (in order for there to be an adequate trial of the therapy), only "those patients with an *estimated* survival of more than 12 weeks" would be accepted in the trial. I remember telling him that it wasn't my problem if the oncologists couldn't estimate very well. But there is an important professional lesson, even today, from my communication difficulties back then: those of you who will work in an interdisciplinary setting should be aware that

the same technical term can have different meanings for different people.

Incidentally, the lack of the word "estimated" made trouble for a McGill colleague twenty years ago when a very savvy Québec politician asked her to show him the list of the *4046* people who suffered health problems or even died because they stopped taking their drugs properly after they were required to pay for them. If you are going to report an *estimate* from a statistical model, make sure to say "an *estimated* 4046." Better still, say "approximately 4000" (or "an excess of 4000") so people will understand that you are not able to be precise, or to say which ones are the excess

cases.

In 1978, when I was already 5 years post-PhD, Steve Lagakos asked me to speak about my work on bivariate survival curves at an invited session at the American Joint Statistical meetings (JSM). He had to drag me off the podium after I went way past the allotted time. I was only just over half my way through my overly-busy and overly numerous transparencies, and had not rehearsed. 68 / 938

Steve later told me about going with Fred Mosteller to Washington to a Food and Drug Administration hearing on Red-Dye Number 40, a food colouring, to present their re-analysis of the data on its effect on rats. The afternoon before, in Washington, Fred had rehearsed and rehearsed; then they went to dinner, and Steve thought they were finally done for the day. But on the way back from dinner, Fred said, "let's rehearse just once more time." They rehearsed in Fred's hotel room. Fred had had his trousers ironed/pressed and didn't want to wear them in case they

would get wrinkled for the big hearing the next day. So Steve had to listen to Fred rehearse his talk in his "1940's style drawers that extended down past his knees." After I heard that, I decided that if Fred Mosteller [almost 65 at the time, with 40 years' of experience in academia, and a "television veteran" from NBS's Continental Classroom in the early 50s] needed to rehearse his talks, then so should I!. In 1980, Fred published this piece on giving a lecture or a talk. It, like Student's 1908 paper on the t distribution, should be required reading for everyone under 65 –

and over 65. You much more sophisticated young people of today can still learn something from the writings of this dean of American statisticians.

I learned another important lesson in communication from two other Ivy League professors. I, together with Barbara McNeil, a Harvard Professor of Radiology, sent in a paper to The Journal of Chronic Diseases in 1981 on "Maximum Attainable Discrimination." I was trying to show, from data she and I had collected on every CT scan of the head ordered at the Harvard Medical Area in 1978, that—contrary to those who thought that the results of the scans were predictable from patient signs and symptoms—scans of the head were being used only for the subtle cases. I had used

a saturated model, and still could not show how neurologists could be more efficient and avoid some scans. Barbara had already tried to get me to only use a maximum of 12 (rather than 35!) slides for 10-minute talks, and didn't like my written style either . She knew I didn't always listen to her advice; so she left me submit the paper without any serious editing on her part. I still remember the review, signed by Alvan Feinstein from Yale. "I stopped reading this 28 page paper at page 6, because it was so unclear. I think that what it proposes has already been done [he referenced

a paper of his own], but the paper is so badly written that, **instead of re-inventing the wheel**, it appears that they have **re-invented the ellipse**." After Feinstein's review, I took writing seriously.

But, as much as I like writing, I am still not great at it. A few years ago I sent a first draft of a manuscript to Maurice McGregor, former Dean of Medicine at McGill, and the person who got me involved in evaluating the harms and benefits of cancer screening. His main comment was in the form of a question: "Have you given this anywhere yet?" in other words had you given a talk on it? I said no, and he smiled. "It shows". He learned from experience that the best way to select and organize the material for a manuscript was to first give it as a talk. It's too bad I

didn't know him before giving that disastrous talk at the JSM meeting in 1978. His trick for timing a talk: He wrote out the words by hand. And when he get to 8 pages (about 1200 words) he knew he was at his limit for a ten minute talk. That's why you see the word counts in my version of the 'lyrics' for this presentation. 179 / 1583

The embarrassing story that I describe in most detail in the written piece is the 10-minute telephone consultation with a collaborator from U de M, Just as I was leaving to go on holiday one August in the early 1980s. We didn't have email then. She asked for my help with the sample size determination for a separate project. It concerned the drop-out rate in exercise classes, which was at that time (and probably is even now) close to 50%. She told me the intervention was "a simple psychological intervention, delivered at the well-known times when competing priorities

start to matter, and motivation drops." To elicit a 'delta' I asked what would be a worthwhile effect, and she said, "bringing the dropout rate down to 30%." I had done this calculation so many times in my seven years of clinical trial work that I didn't even have to reach for my calculator. I told her straight off that if she wanted to have about 80% power for that sized delta, she would need about 95 in each arm. She said "perfect, we have about 200 subjects in all." I will let you read the story for yourselves, and I suspect you would would not be as naive as I was.

Just one bit, from the written piece, related to our article on our Titanic study for the Christmas Edition of the BMJ in 2003. Despite its light-hearted nature, and limited usefulness other than for teaching lifetables, this BMJ report received more media attention that all of my other work put together.

However, the media behaviour was very instructive. First, despite the way they fawn over you and butter you up, **the media are not your friends**. Thus, if you wish to have your message reported accurately, take the advice of my (late) media-savvy colleague Abby Lippman: write it out (preferably in suitable sound bites) and keep repeating it as the answer to all the questions. Second, the reaction of one print journalist was particularly interesting. She called me to get the co-ordinates of the three survivors. I told her that we only dealt it aggregates, and directed her to

the Internet site where I had extracted the raw data. But it was her next sentence that was the most telling: "it's a pity you didn't find a difference, so we could have had a story."

Interestingly, the other print journalist turned the "null" result into a story, by saying that "the results turned Darwin's theory of the survival of the fittest on its head." Until then, I had not considered how much extra **filtering** the results scientific studies undergo before they are reported to the general public, and indeed to other scientists. By the time selected instances of our 'over-exact p-values'—often reported to several decimal places—reach the public, they have lost most of their original meaning. 81 / 2059

Communication is important, but before we can communicate, we must truly understand, and so one must continue to deepen one's learning and understanding. Some might call it 'continuing' education, but that I think of as **really** understanding the concept **for the first time,** – even if it is 25 years later. By this heading, I don't mean **machine** learning or AI. I mean **my** learning. 65 / 2124

And here are some concrete examples of what I mean by minimalist. First study and understand the simplest and smallest possible case. On average, a 'sum of squares divided by n' underestimates. Unless you think and dream in matrices, it is hard to really understand the structure of $\sigma^2(X'X)^{-1}$ in p dimensions. Why not start with $p = 1$ dimension or parameter? Why even include an intercept? This example, on my wesbite, has 2 datapoints and one β. Can you think of the smallest regression you can run that still gives a standard error for the $\hat{\beta}$? When I first met

GEEs, I learned what they really did using the smallest possible dataset I could imagine – with $n = 3$.

I am also a big fan of realistic props and visualizations. I list some here, but let you explore them yourselves. My favourites are probably the Elevators, travel examples, and the massive Ocean Depths database (with $n = 933 \times 10^6$) that one can only sample from. 47 / 2301

Voici un autre dans un journal pour les gens plus jeunes , qui illustre le lien entre le discret [Poisson] et le continu [Gamma]). This equivalence was aLways just math to me for 40 years. I include also a link to the English back-translation and some extra historical notes. 50 / 2351

Of the various animations I have made, this one on the Bridge of Life, imagery that goes back 3 centuries, was the most fun. Making it also taught me a lot about stochastic processes. It's getting much easier to make these, so use them when they can be useful. 49 / 2400

I put these under the heading of Unity/Logic but the 1st 4 are again my way of conserving neurons. Why have different approaches for what seem like different contexts but all have the same structure? Textbooks treat these 2 sampling variances under different chapters (means and proportions) when in fact an empirical proportion is just a mean of Bernoulli observations. Likewise, whereas software developers like to sell separate statistical modules for different applications, a lot of sample size, precision, and power considerations can be unified in one framework. They act

as though one of the biggest statistical breakthroughs in my early years (GLMs , 1972) never happened. Try to save neurons whenever you can. I made the same point about 'survival analysis' at the summer school earlier this week. If one uses an incomplete data framework, it does not need to be a specialized topic. The point of the last bullet is that we should also get in the habit, when teaching statistics to newcomers, of referring to a random variable as y, and a mean as ybar, so that when we move on to regression, it will be clear that x is not treated as a random variable, but

rather as selected values at which to measure/record the random variable y.

CURIOSITY Just taking prof's or textbook's word for it, and memorizing it, or being able to produce a mathematical proof, wont help you to truly understand. For example, early on we memorize the rule of thumb that an expected counts of 5 or more are need in chi-square calculations, or Gaussian approximations. I tell biostatistics and statistics students that epidemiologists and other subject matter people use statistics the way they drive cars: when they break down or need fixing, they take them to the mechanic, who understands the wiring, the plumbing etc. You statisticians

need to know the wiring and what is under the hood, and should not rely solely on rules of thumb or electronic packages to help you diagnose problems or to know when it is safe to forego some assumptions.

For years, I could never mathematically understand why measurement error in 'x' flattens the y/x slope, so I made up a simple example, involving errors in measuring Fahrenheit and Celsius, to learn from.

As for curiosity, when I started to teach regression, I had several questions. Why is it called regression? why is the correlation coefficient called r? When were the first regressions fitted, and how?

I found some answers in Stephen Stigler's wonderful book on the history of statistics. It takes a while to get oriented to the language, and to the orientation. But why, in this second regression line ever fit, is the parents' height on the y axis, and the children's height on the x axis? 53 / 2864

It's the same in the data table he used to reduced the observations to 9 pairs of medians, which he fit by straight-edge (a ruler)? And why did he multiply all female heights by 1.08? Couldn't he have included an indicator term for male in a 2-x regression, or just fit parallel lines by eye? 55 / 2919

It's this last question about his multiplicative rather than additive model that set me off on a search for the raw data? With Stephen Stigler's directions, Beverly Shipley, a graduate student in London, found the raw data in a Notebook Galton had put in the archive. Here are 3 of the 7 pages of data, one row per family, sorted first by the father's height, and within it, by the mother's height, and within each row with the children's heights arrayed horizontally within sons and daughters separately.

I'm about five foot six and so my family would also be in the last page. And we would fit Galton's law.

Some people, when they don't like the law, what do they do? [I am not advocating you do this in your professional career!]

But going back to my question about his multiplicative rather than additive model. When I re-analyzed the raw data both ways, using the additive and multiplicative models, the fit was only very slightly better with Galton's way. But it does remind us that our homoskedatic errors model is almost never correct. Most times, heteroskedasticity and non-Gaussian residuals do not matter, but on Tuesday I will show you an applications where they do very much matter.

For this last part, I don't have a lot to say, except to be professional, continue to work on your communication skills and confidence. Thinking for yourself helps with confidence. I will end with a very recent example of putting thinking before techniques. Because this case is still working its way through the judicial process, instead of putting it on paper into the 'lyrics' I will just tell it orally. 70 / 3196

To end up, then, thank you very much again for inviting me, and for indulging me as I looked back on a career I have thoroughly enjoyed. I think you have chosen a good career, and have the potential to not only do a lot of good, but derive a lot of happiness from doing it. 56 / 3252