Transcribed from a lecture delivered as part of a CrossFit Health event at CrossFit Headquarters on June 7, 2019:

Thanks very much. It's a pleasure to be here. Thank you very much for the invitation. Thank you for your hospitality. It's great to be here. I've done lots of talks like this, but I must admit, I've never had dogs running around the room before, so this is a first for me. Thank you very much.

This is the subject of my talk: “Ten percent of the time, it works every time.” The title was given to me by my son who did his MD at the Mayo, and he's also a stand-up comedian. And there is a line in Anchorman — who's seen Anchorman? Very good. So then you know the line. There's a line in Anchorman where he has this fantastic new aftershave that's supposed to attract women, and he says 60% of the time, it works every time. So, that's my title, because when it comes to preclinical research, about 10% of the time, it works every time.

So, my focus is on basic laboratory preclinical research, and I will try and make it relevant to you. Is there anyone in the room that's actually lived that life: basic preclinical research? Okay, that's very helpful. So, here is my agenda: First, I'll present my disclosure information, I'll talk about what the problem is, tell you why I'm still committed to research, tell you what to look out for, and then I'll give you two real-world examples from some of the most famous scientists in our country — I won't identify them — and then I'll talk about the implications. This is not just something that's esoteric. This really has impact in terms of public health and clinical studies. And then finally what we could do about this.

So first, my disclosure information. I'm currently the chief — this is my advertising slide, okay? I'm the chief executive officer of this small company in Australia called BioCurate. I consult for a number of startups and biotechs around the world, and for a number of years I was the chief scientific officer for startups, one in — I told you I've never had dogs in the room before — one in California, and one in Malvern in Pennsylvania. For 10 years, I was the head of the hematology oncology research group at a biotech company on the West Coast called Amgen, the world's largest biotech company, and that was when the issue of scientific reproducibility was highlighted — highlighted to me. I continue to hold stock in these companies, and before that, for 20 years I was a physician scientist in Australia.

My research has been focused on growth factors that regulate blood cells, and one of those is G-CSF, and I'll talk about that. Normally, I would say I'm not going to discuss the off-label use of drugs. Today I will. I won't advocate for it, but I will discuss the off-label use of drugs. This is my last advertising slide. BioCurate was created by the top two biomedical research universities in Australia, and it was created to take their discoveries and turn them into something useful. It would be
nice if we made money, but we don't have to make money. The real focus is to have impact so that that research actually does something useful for humankind.

The fact that these two universities came together was unprecedented. They've been rivals, and they recognized that they could achieve more together by collaborating. So, that's exactly what they've been doing. The two universities together are world class. They ranked in the top 10 in half a dozen different areas here, and we commenced operations in 2016. We've really been running for a year. We've built the staff, we've got the labs and so on in place, and we've got a dozen projects or so that we're currently managing. So, that's the end of the advertising.

So, what I wanted to focus on initially is what the problem is in preclinical, basic research, and the problem is really simple. It's one of perverse incentives, and I'll present that throughout this talk.

So, I told you my job is to take research discoveries and turn them into something useful that will have an impact for humankind. The biggest challenge we face is that most research in the laboratory can't be reproduced. For 10 years, I was the head of the research group at Amgen. I mentioned the hematology oncology research group, and at the end of the 10 years, I decided to look back and see what our success rate had been and to my horror found that 90% of the time we were unable to reproduce the top papers published by the top laboratories from the most famous scientists. 90% of the time we couldn't reproduce the work. It was actually worse than that because — excuse me — because our standard operating procedure was to go back to those laboratories and ask them to do the experiments themselves, and they could not do so. And it wasn't just some small aspect that they were unable to reproduce. It was the fundamental finding, the title, the main point of their piece of work, they were unable to reproduce. Sadly, Amgen's experience wasn't unique. Bayer Health had reported this beforehand, and since then — since we published this work, many groups have come out saying, “This is our experience too,” and that most of the information published in the top scientific journals can't be reproduced.

The problem is there is no metric for quality when it comes to judging research. The metric we use is the journal in which the work is published and everyone in my area wants to get papers published in Nature, Science, and Cell. They're the top journals. And the other metric we use are the number of citations. So, to give you some sense of that, there was a piece of work done by the Medical Research Council in the U.K., the U.K.'s equivalent of NIH, and they reviewed 94,000 publications and then discovered that most of the time, most of them aren't cited at all. So, on average, they were cited two and a half times — 2.8 times. They regarded citations beyond four as being highly cited and very highly cited beyond eight.
So, I'm a product of this system. I've got over 23,000 citations. I understand cognitive dissonance because I still want to get my papers published in these journals. I've got a number of papers that are cited over a thousand times. The papers that I'm going to discuss with you are in the same league. So, when we talk about very highly cited, the ones I'll be discussing with you are very, very, very highly cited. The problem is focusing on the journal and the citations assumes that that is a surrogate for quality; it is not. If I published a paper saying the Earth was flat, it would be very highly cited. It would be cited thousands of times. There is no metric for quality. On the other hand, in industry, you take your discoveries into the clinic — the clinic is unforgiving. Patients either respond or they don't, so there the metric of quality is: Does the drug actually work?

The cost of academic research waste in the U.S. is estimated to be $28 billion a year — not by me but by others that have looked at this more closely. That's an enormous amount of money. The fundamental problem we have is the academic research cycle. So, this is where I've lived most of my life. What we really want is a paper published in the top-tier journals: *Nature*, *Science*, and *Cell*. On the clinical side, it's the *New England Journal of Medicine*. That's where you want to be published. What that does is it secures my next research funding so I can publish my next paper in *Nature*, *Science*, and *Cell*, so I can get my next research grant, so I can publish my next paper. You get the idea.

Earlier on in your career — As you can see that doesn't really apply to me anymore. It did once upon a time — those papers meant that I got promoted. I got a tenured position at a university. It also means that I become famous, so that I then stand up and give lectures like this in front of my colleagues. This is the metric of academic success, it does not include a metric for quality.

Behavior is driven by the perverse incentives with few or no negative consequences. Every scientific paper that I publish, every presentation I give is undeclared self-promotion, self-advertising. That is the truth. So, we would like to think that when I've published my scientific paper, it sits out there in some objective, dispassionate manner. It's not. It's actually self-promotion and self-advertising, and when I stand up and I'm on important committees, that is without doubt self-promotion, self-advertising. It guarantees that I'm more likely to get my next research grant. And we get what we incentivize.

So, I told you I'm a product of this system. Why do I stay in research? Because science delivers. So, when we first published our paper in 2012, we received threats, hate mails, insults. In fact, there was a quote that I really like in a book called *Rigor Mortis*, written by Richard Harris from *NPR*, and there, perhaps the most famous cancer scientist in the world, Richard Weinberg from MIT, said, “To my mind, the Begley paper” — that's me — “was a testimonial to the silliness of the
people in industry, their naivety, and their lack of competence.” Thank you. It's very hard not to take that personally.

So, as a result, I think it's important to present where I'm coming from. Where I'm coming from is, the results that I'll present to you do not challenge the validity or the legitimacy of the scientific method. I'm not talking about fraud. I won't use that word again unless you ask me questions. What I'm talking about is scientific laziness, sloppiness, ignorance, exaggeration, and desperation. That's what I'm talking about. I believe that the vast majority of investigators do want to do the right thing. They just can't help themselves. It's the perverse incentives. The fact that this debate is occurring within the scientific system, to my mind, confirms the strength of the system. Although some people I believe deliberately want to misinterpret what I'm saying, I'm not anti-failure. We learn more from failure than from success. So, I'm not anti-failure, and I'm not anti-academia. I'm not anti-university. What I'm really talking about is fundamental to human behavior, and again, we get what we incentivize.

So, what I'm about to describe to you must be held in tension with the fact that the advances in medical treatment have been truly outstanding. We have every reason to remain optimistic. When I was in medical school, the treatments that are currently available for cancer today were unimaginable. The treatments that we've got for inflammatory bowel disease, rheumatoid arthritis, ankylosing spondylitis were unimaginable. The advances we've made in terms of medical treatment are outstanding. So, you have to hold that intention with what I'm about to tell you. What I'm about to tell you is really about the opportunity cost. So we've made enormous advances, but we could have done so much better. I also believe that this could be immediately impacted if not solved by the funding agency if they had the courage to do so.

So, why am I still committed to research? Because science delivers. If you look at the increase in worldwide life expectancy, it's increasing around the world, including in the countries where we know that there are real problems: Asia, Africa, and so on. There's a worldwide improvement in terms of infant mortality. Remarkable results, especially when you look at countries like Ethiopia and South America, and so where we've seen tremendous improvements in terms of maternal and infant mortality. So, science delivers.

I believe that science is the most effective tool humans have ever developed for driving progress. It's inherently self-correcting, although I know what Max Planck had to say about that: A scientific truth does not triumph by convincing people. It triumphs because people die and the next generation grow up, and they're more familiar with it. And it generally tolerates self-critique. I told you I'm a product of the system, I am within the system, and what I want to do is generate constructive
criticism so that the system changes. I have taken the position that I want to bring about change from inside rather than change from outside. And constructive criticism is about finding something good and positive to soften the blow to the real critique of what really went on. That's what I'm trying to do.

So in the next section, I want to tell you what to look out for if you were to spend your time as I do reading scientific papers. So, when we published our paper in 2012, I had a pleading email from a postdoctoral fellow saying, “You've got to tell me what those papers are, because I might be doing that work. I might be working on that very project.” Well I can't do that, because the only way that we could go into the laboratory of the original investigators was to sign a confidentiality agreement, so they can tell you who they are, but I can't, with one exception. I'll tell you about a specific example at the end, because that's now all been made public, thereby doing away with that confidentiality agreement.

So, I sat down to look at the papers that we could reproduce versus the ones that we couldn't, and these are my six rules. So, the first thing: Were studies performed blinded? They're almost never performed blinded, but they should be. Were all the results shown? I'll give you some examples. That's seldom the case. What happens, in fact, is people cherry-pick the data. They choose the data they like and present the results that they like. Were the experiments repeated? Even in the top-tier journals, it is staggering how many times one experiment is presented. The experiment's only ever performed once. Were there positive and negative controls? Almost never. Were the reagents validated? Almost never. And was the analysis appropriate? Very rarely.

So, my wife is not a scientist, and when I sat down and said, “These are the six rules that I —” she said, “I learned that in grade school. Isn't the way every scientist does every experiment like that?” Well it's not. The reason it isn't is because it's easier to get the results you want if you know what you're looking for. That's not research, but that's what we're incentivizing within our current system.

Now, the answers to all of the next dozen or so slides I'm going to show you are here. If I was king of the world, and in case you noticed I'm not, I would insist that every experiment was performed blinded. All we did when we went to the investigators' laboratories and asked them to reproduce their experiments was took their reagents, put them behind our back, and gave them back to them, and asked them to do the experiment again, and they could not get the same result. As far as I'm aware, none of those papers have been retracted. None of them, have people said, “Our original findings do not stand.”

So, let me give you some examples. Now you're not scientists, so I'll try and walk you through this. What you're looking at on your left is what's called a western blot.
What we do when we do that is we take cells and we scrunch them up, and then every protein in the cell is run out on a gel, something of the order of nine or so inches in size. So, there are 30,000 proteins that are ran out on the gel, and they all then are separated based on size. So, if you imagine taking all of the people in this room and ordering them in terms of height, the very tall people would be at one end of the gel, and the very short people would be at the other end of the gel. So, that doesn't help us a lot, because there's 30,000 people in that lineup. So, we want to know which one we're really interested in looking at. So, the way we do that is we use an antibody, and antibodies can be very very selective. And out of those 30,000 people, or proteins lined up, an antibody can detect the very one that you're looking for. And then you're able to say, “Okay, that's the person, that's the individual, that's the protein in the cell that I'm interested in.”

So, you can see in the panel on the left, there are probably a dozen different bands there of different intensity. So, the antibody that was used there isn't detecting the one band, the one protein in 30,000. It's not detecting the one individual in that lineup of 30,000. It's detecting a dozen. But I can hide that from you by cutting out a sliver of the gel and pretending that actually all of these weren't there, and the focus of your attention should be on just on that one band. So, that's frankly dishonest, but that's what happens all of the time. I can cut out all of the other material and present — pretend that that antibody was able to detect the one protein that we're interested in.

So, investigators do this all of the time. When we generated this data in 2006, we contacted Upstate, one of the antibody manufacturers. They actually make the antibody on the bottom, and they took it off the market. This one here. We contacted Santa Cruz that make this antibody and told them about their data. You can still buy that antibody. It's still available on the market even though it doesn't do what it's supposed to do. In fact, that antibody is so bad, it can even detect the protein that it's supposed to be protecting but it doesn't in mice that lack the gene for that protein. So, that's shown here on the left. These are mice embryos you can see are stained brown. That's because the antibody is detecting the protein. But the protein we're interested in isn't even made in that mouse. The gene for that protein's been totally destroyed. So, it tells you that actually, what you're detecting is one of these extraneous bands here, not even the one that the investigators present as being the one of interest.

So, that's pretty good evidence that this antibody is no good, shouldn't be used, and you would think it should be done away with. But in fact, in 2015, there was a paper published in, again, one of the journals that we all want to publish in. This one's Cancer Cell. That very antibody had been used to stratify women with breast cancer based on expression of the — what's — the protein is called the erythropoietin receptor. So, this is complete nonsense. This antibody doesn't even really detect the
erythropoietin receptor and we said that in 2006, and here we are in 2015, people using that very same antibody to stratify women with breast cancer. This should never find its way into the scientific literature.

The other thing that is very common, and I'll show you some examples, is that as humans, we think in a linear manner. So, the way that we represent that graphically is we have time on the horizontal axis, number of days or weeks, and whatever we're interested in on the vertical axis. So, whether or not we're talking about building houses or making tablets or birthday cakes, we do it one at a time. So, for me it might take weeks to make a birthday cake. For you it might be a matter of hours. We would show that with time on the horizontal axis, and the number of birthday cakes on the vertical axis. But cells don't behave like that. You all know that if you've got one cell it divides to give rise to two, and two gives rise to four, and so on. If cell growth — proliferation — was linear, it would mean that of the four cells that you had lined up here, only one of them could divide, but that's not the case. We know that cells divide, and this is called an exponential function: One gives rise to two, gives rise to four, gives rise to eight, and so on. The appropriate way of representing that is on a scale that's exponential or logarithmic. So, instead of the units being steps of one — from one, to two, to three, to four — the steps should be steps of 10 — from 10 to 100, to 1,000.

Now, the reason for going through this is because this difference looks quite dramatic from 800 to 900, or 8,000 to 9,000, if you use steps of one. But if you actually convert it over to steps of 10, which is what it should be because cell proliferation — cell growth is an exponential, not a linear function, you can see there's almost no difference in those lines and suddenly looks much less impressive. So, I won't present my data that way. I'll present it on the left, because it might fool you into thinking I've actually got something important.

It's also important to even realize that numbers that look quite different are actually the same. So, 512 sounds different to 1,024. The truth is that's one cell division, so wait a few more days and you'll have 1,000 cells. So, there's really no difference even between 500 and 1,000, and if you put that on a logarithmic scale, that would be obvious.

So, what I'm alleging is that the reviewers, the editors of the so-called top-tier journals, grant review committees, promotion committees, and the scientific community repeatedly tolerate poor quality science. If you found a study that fails one of my tests you'll find that it fails multiple tests throughout that study. Famous scientists in top institutions are repeatedly given a free pass, and now what I'd like to do is give you some specific examples.

This is the most important message: Be skeptical. Be skeptical.
So, I'd like to go through two real-world examples, and I'd ask that no one copy these. All of this is in the public domain, but I'm not going to identify the investigators because we were unable to reproduce 47 papers. They came from 46 different laboratories. What we have is a systemic problem. It's not as though there's one or two labs behaving badly. What we have are perverse incentives that are all-pervasive. So you could, if you wanted to work out which labs I'm talking about — I don't want you to do that. It misses the point. — You could easily focus on these labs and say, “They're the problem.” They're not. We have a systemic problem. These are purely examples.

So, the first example I'll give you is from a very famous investigator who's got about 400 publication. That's pretty impressive, frankly. Almost a quarter of them in the top-tier journals: *Nature*, *Science*, and *Cell*. Thirty-six of the papers have been cited over 500 times.

So, as Richard Feinman said, “The first principle of science is that you must not fool yourself,” and you're the easiest person to fool. The second is you shouldn't try and fool someone else.

I wanted to take this from what might be the sort of esoteric and take it into the real world, what it means in terms of public health and clinical trials. So, poor-quality basic research drives poor public policy, and I'll give you one example of that. So, the year before last, I was asked to speak at the annual meeting of the Society for Environmental Toxicology, and they gave me the position statement from the Endocrine Society, this here, and asked me to review it. This is called “Introduction To Endocrine Disrupting Chemicals A Guide For Public Interest Organizations And Policy Makers.” The background to this was papers like this here, that talked about the reduction in penis size and plasma testosterone concentrations in juvenile — juvenile alligators.

Now, I remember this very clearly. I'm not sure anyone else is brave enough to put up their hand and say they remember it too, but this got an enormous amount of press when alligators' penises were shrinking. The thing that intrigued me was how do you go out and measure an alligator's penis? But someone was doing it, and there was a problem. So, again, when I say this to my friends they say, “That's because, Glenn, you've got a problem the rest of us don't have.” Anyway, this is the story: Alligator penises are shrinking in the American alligators, and it was blamed on endocrine disruptors.

Now, you may not remember the specifics of the alligator penises, but you probably do remember that we're supposed to be getting too much estrogen from chicken and all of those other things, and that's responsible for decreased male fertility and
so on. That's the background behind this statement from the Endocrine Society talking about endocrine-disrupting chemicals that are altering male penis length in alligators, male fertility, and increasing the number of females. And in this policy it says, “If estrogen or an estrogenic-endocrine disrupting chemical is added to eggs that are incubated at a temperature that normally produced males, instead you'll get females.” So, I went through this document, and I'll take out two papers just to — as examples of how, frankly, nothing in this public policy statement can be believed.

So, this is the guide for policymakers, recommendations based on poor-quality data. So, this is the first experimental data set that I've selected to show you, and the conclusions are that Propranolol, which is a beta blocker — It's used to lower the blood pressure, and they use trace amounts of beta blocker in this study — that Propranolol, “affected significantly offspring production of exposed females, enhancing reproduction.” So, the data that they showed is in the panels that you can see above, and for those of you that are quick, you'll see that there are problems in how this data is presented. First, it doesn't start at zero, so typically, if you've got a graph, you would begin at zero. In fact, it starts at 60 on the left and 90 on the right. The other thing that's interesting is the size of the error bars, which actually look quite reasonable if it had started at zero, but it started at 60, and it's going to 70, so there's really almost no change in those error bars.

What the investigators want us to believe is that there is a change from the open bars to the shaded bars, that there are more females being generated in this system as a response to Propranolol, the beta blocker. The other thing that I'll point out is that they're using a thousand-fold difference concentration here versus here, but they don't tell us that. If you correct the data so that it actually starts at zero rather than 60 or 90 — I've attempted to do that in the panel on the left, and you can see that really it looks much less impressive. There's probably no difference between any of those if you start at zero rather than 60 or 90.

The other issue is that on the previous slide, in panel A and panel B, these controls are actually the same controls, done in a different experiment on a different day, but they are the same controls. So, therefore, it should be completely legitimate to take the controls from one panel — they are the same controls after all — and swap them over onto the other panel to see whether or not, if the controls had been different on Monday versus Tuesday, you've got the same results. And instead of an increase, if you do that, there's a decrease. So, if you transpose the control from panel A to B, the result is completely the opposite. So, the conclusion that low levels of neuro-active pharmaceuticals alter reproduction is frankly not supported by the data they present. They've interpreted it the way they want to do so.
The second experimental data set is one of my favorites. It focuses on the red-eared slider turtle. Now, until the Society of Toxicology had asked me to review this, I'd never even heard at the red-eared slider turtle. I might have recognized it if it was crawling up my leg because red-eared slider turtles don't usually crawl up my leg, but other than that, I had never even heard of the red-eared slider turtle.

Comment from the audience. Followed by laughter.

For those that didn't hear, there was a comment for me that I have a very small penis, which is probably true, but everything is relative. How can one know?

Now, where were we? We were talking about the red-eared slider turtle. So, the authors claim that more female red-eared slider turtles were treated with estrogen treatment, and if you look at the data that I've circled here, which is from their Table 2, it's true. There are more females: 18 is different to what they expected, 11.5. Not a lot in it, but there, it might be true, and 24 is probably different to 17.

So, here there are two experiments where observed was actually greater than expected. So, this was the basis for their claim that estrogens can turn male turtles into female turtles. That's Table 2. If, however, you look at Table 1, there are more examples of the converse. There are eight examples where it went in the wrong direction and six examples where it went the way they wanted it to. Worse than that, there's no evidence of a dose response. So, here are two different estrogens that have been combined. This concentration of Estriol stays the same. Estradiol increases, so as the concentration increases, if this was genuine, you'd think there'd be more and more and more females born. In fact, the opposite happens: As you increase the concentration, there are more males born, so that, too, weakens their argument.

They claim in their paper on 43 occasions that this effect of one estrogen and another estrogen shows synergy — that is, one plus one equals three. So there's cooperativity between the two. So, you get more than you would expect. Here is the data that they show, but the critical controls are not shown. So, what they've shown here is the green increasing to the blue, and it probably does. But, what they haven't shown is the effect of this agent alone. It is shown in that Table 1 that I mentioned earlier, so if I take the data from Table 1 and put it on this graph, the alleged increase is even less than what you would get simply adding the two together. So, not only is there no synergy; it's not even additive. So one plus one doesn't even equal two; it equals one point five. So, their claim is not supported by the data.

However, they use that data in the report to say, “The first and most important experiment proving that there's no threshold for these endocrine-disrupting
chemicals took place in the 1990s — “and I won't tell you this, but it was actually my experiment. I won't tell you this, which I misinterpreted and didn't support the claim I wanted to make, but I won't tell you that — “In the red-eared slider turtle, it is temperature that determines whether the individual will develop as a male or female.” This is where it gets amazing: “Similar to the X and Y chromosomes — how they determine sex in humans. With that exception, sex chromosomes versus temperature, the remaining biological processes are remarkably similar between turtles and humans.” Forgetting that humans don't have a shell, turtles live for 200 years, they lay eggs, they don't even have a diaphragm, they breathe by flapping their arms, the only turtle that I can think of that's related to humans is this one.

The other direct consequence of poor laboratory research is the impact it has on clinical studies. So, this was a study that was published in Nature and cited 8,000 times. That's remarkable. That is a very, very highly cited paper, but again, citation is no surrogate for quality. Most of the time I told you I don't identify the investigators. These investigators have identified themselves, so I'm happy to tell you about this story.

It was published in Nature. The senior author is Piero Anversa, and it came from Valhalla, New York. What they did here was generated heart attacks in mice, and then they took bone marrow cells and injected them into the region of the heart attacks — the arrows here and here — and then claimed that those bone marrow cells helped the hearts repair. And here, the pictures to try and show that the heart is repaired as a consequence of injecting bone marrow cells into the heart. This is a remarkable claim, a remarkable claim, which is why this paper has been cited 8,000 times. This was mind blowing at the time.

Now, my computer has just frozen, so I assume we've lost power. No? Okay, so then on the next slide, which I now can't show you, but someone might be able to help me. Help is on its way? Okay, thank you.

So, this was a remarkable finding. It was something that we really wanted to try and understand, and then the next paper that they published was showing that you could achieve the same effect — Thank you. It's on its way. — You can achieve the same effect by not taking the bone marrow cells and injecting them into the heart but simply getting the bone marrow cells to move out of the bone marrow into the blood, and the bone marrow cells would find their own way to the heart attack and repair the heart attack. Now, the molecule that they used to do that is called G-CSF. That was something that we discovered during my Ph.D. in the 1990s, and I was responsible for the G-CSF franchise at Amgen. At the time of these studies, the G-CSF franchise was worth $4 billion a year in the U.S. And G-CSF is actually approved for patients having chemotherapy, because it helps them recover from the chemotherapy, and it's approved for moving bone marrow cells into the blood.
for bone marrow transplantation. Because it's easier to collect the bone marrow — Thank you. — the bone marrow cells from the blood than from the bone marrow.

So, thank you. We — thank you. That's great. Thank you very much.

So, this is the study I was just telling you about, where these same investigators took G-CSF, injected it into the mice. The G-CSF is well demonstrated to move the bone marrow stem cells into the blood, but then in this model, they are able to find their way to the heart and repair the heart attack in these mice. The data looked very compelling. Here are the mice treated with G-CSF, and they're surviving. Here are the mice that were untreated, and they're dying. The graph shows the number of surviving mice after their heart attack. And then they showed some lovely pictures showing that the G-CSF-treated mouse shows healing of the heart attack. The non-treated mouse shows ongoing damage and results in death.

As I said, these were stunning claims and of great interest, because the G-CSF franchise was part of my responsibility at Amgen. G-CSF is not approved for heart disease. Amgen would have been extremely interested in proving that it was useful in heart disease, because it would have increased the franchise from $4 billion a year to $40 billion a year. It would have been enormous. So, we went to Anversa's lab on several occasions, and it was obvious that he could not reproduce his own data in blinded experiments. After that, about a thousand papers have been published supporting Anversa's work. Last night I checked, and I found six papers that were unable to reproduce it, but the vast majority said Anversa's study could be confirmed.

These findings were so dramatic, they triggered clinical studies around the world trying to repair heart attacks using G-CSF or bone marrow. The bone marrow was harvested and injected into hearts, or it was injected down stents for people that had had heart attacks. People were given G-CSF to move their bone marrow cells into the blood to find their way to the heart and repair the heart attacks. So, I've got several slides that look like this. These are the first dozen or so studies that were using bone marrow or blood stem cells for heart disease. And what I want to illustrate is that almost none of these had actually got a final report listed, even though that's a requirement for the studies. But very few of them have actually got a final report listed. Some of them are ongoing. In none of these studies was the company that stands to make the most from this being true supporting the study. So Amgen stood to gain an enormous amount by this being true, so you would think Amgen would have a vested interest, and it would in terms of proving this to be the case. But none of these studies were supported by G-CSF.

Here's another list. Again, you can see that three of these studies are still listed on clinicaltrials.gov as enrolling. And two on this list: ongoing. Again, none of these
were supported by Amgen. Although, I can tell you that there was enormous pressure from the marketing group for us to support these studies, but the data wasn't there, so fortunately, we were able to make the case that we shouldn't do so, and so on. And then there are a number of studies that have gone beyond heart disease to look at the effect of G-CSF in Alzheimer's, brain injury, and so on.

There are several meta-analyses that have been performed. A meta-analysis, we'll hear about later, is to take the data that's been presented from clinical studies, put it all together, amalgamate it, and see if we can make sense of it from that. Now, most of the studies have still not been reported. The reason I suspect they haven't been reported is because they were negative, but there are a dozen or so that have been reported. And several meta-analyses have looked at these from every different direction and have concluded that there is no benefit.

The meta-analyses did not contemplate that the underlying hypothesis might have been wrong. Unfortunately, G-CSF treatment alone is associated with increased blocking of stents after someone's had a heart attack. They get a stent put in to reperfuse the heart, but G-CSF, probably, although it's not statistically significant, increases the restenosis rate. There was an increased rate here of 50% with cell infusion, G-CSF alone, but in the controls, 30%. In this study, these two are actually both stopped, because there was concern about increased restenosis within the stents. Ten percent of patients in this study had second heart attacks and death after G-CSF.

In this study, the restenosis rates were similar. So, there's no statistical evidence that G-CSF was actually doing harm. There's clear evidence that it wasn't having benefit, and it was probably doing harm. This is why preclinical research that's happening in the laboratory is actually important. It has real-world implications.

So how did we get here? Piero Anversa created this whole field. He promised to restore damaged hearts. Harvard said his lab fabricated results. They fired him. They recommended that the last 31 papers be retracted. The original papers that I told you about still stand, because they weren't performed at Harvard. They were used by Piero to get his position at Harvard, so they haven't been retracted, but the last 30 papers have. And Harvard paid back at least $10 million to NIH for the funds that NIH had provided Harvard over the years.

Interestingly, Piero Anversa said, “I've done nothing to deserve this,” and I think honestly he's telling the truth. He's just doing what the system encourages. We have a problem of perverse incentives. If you get your paper published, you can get your next grant to get your paper published to get your next grant. He was just doing, I think, what everyone else was doing. I visited his lab, I told you, on several
occasions. I don't think he was doing — not that I'm proud of it — but I don't think he was doing anything different to what many people do.

There are some important lessons here in terms of investigators that want to do their own studies or drug repurposing studies. I think the most important thing is that investigators must independently reproduce the key preclinical data. Do not trust preclinical data, even if it's from top-tier laboratories, first-class institutions, published in top-tier journals. It must be reproduced. Independent clinical studies, even from other first-class institutions, don't provide validation or confirmation. But where if the company that stands to most benefit from this isn't involved? There's a reason why they're not involved. And as an aside, there's little opportunity for any commercial return when you're attempting to repurpose drugs.

So, what should we do? We have a systemic problem. We have a system that encourages, certainly tolerates these sorts of poor-quality sciences. The principal responsibility rests with the investigator in the institution. When we first wrote up our paper in 2012, I sent it to Fran Visco. Some of you may know her. She's the president of the National Breast Cancer Coalition. And I sent the paper to Fran for her comment, and I really laid the blame at the feet of the journals, and Fran, who's a really smart lawyer, called me up and said, "Glenn, you've got it wrong. It's the responsibility of the investigators and the institutions, not the journals." And she was right. It's the responsibility of the investigators and the institutions.

To solve this will require a multi-pronged approach. Patients expect, certainly deserve more. People have referred to this as a reproducibility crisis. I don't think it is. I think it's an innovation opportunity. Simply take the money from the bad actors and give it to the good actors, because we have got good actors, and we would be much better off. I typically present this to scientists and the first thing I say is, "We should read papers before we cite them." The scientists laugh.

We should read papers before we cite them. Typically what we do is read the title and the abstract and then decide whether or not we're going to cite them. And if they're from famous investigators in top-tier journals, they get cited. It's not good enough. We as scientists need to do better. We should refuse to cite papers of poor quality. We should refuse to accept the journal as any surrogate for quality. We should focus on methods rather than the results. If we were making motorcars, we'd be put out of business. Ninety percent of the cars are crashing. It's not acceptable. It's time we started focusing on how we make motorcars.

We should do things properly ourselves. The journals could play a role. They could blind the reviews of the editors and the scientists, so if I'm a famous scientist, and I submit my paper, the editor knows I'm famous. They want me to keep submitting my papers to their journal so they can sell their advertising pages. So, they're not
going to reject my paper. They should pay reviewers. Mostly people do their reviewing with a glass of red wine, sitting in front of the television late at night. There's no incentive for quality of reviews. I think we should limit the number of publications that a scientist can publish in any one year. The journals don't like that. It would put them out of business. The people that are publishing three papers a month wouldn't like it, but if I could only publish two papers a year, I'd want to be sure that they were right.

The institutions have got a big role to play. We have good laboratory practice. We have good clinical practice. We have good manufacturing practice. It's time we had good institutional practice. The institutions should hold their investigators responsible for the papers they publish. There should be compulsory annual training for principal investigators and trainees, random review of laboratory notebooks. The reason most of us don't cheat on our tax returns is because we're scared the tax office, the IRS, might come and check us out. The same thing should happen in the research laboratory. There should be random reviews of laboratory notebooks. So, if I say an experiment was done blinded, you can check and see that it really was done blinded.

The institutions should be much more honest and accurate in their publicity statements. I used to hate Monday morning oncology clinic because the patients would come in with Saturday's newspaper, and the front page almost always, of Saturday, was, “We've cured cancer, again, in mice,” and the patients would come in wanting to know when they could have this new treatment. Institutions should do more to tell the truth, and if any of you are on the boards of institutions, this is something that you should hold them to.

Governments and funding agencies should fund quality rather than quantity, should demand good institutional practice, and require licensing of biomedical scientists. Can you believe that tax advisers, plumbers, lawyers, builders, electricians, pathologists, make-up artists, pharmacists, professional lobbyists, strict policy all require licensing, just not scientists. It's time that was changed.

In conclusion, we don't have a reproducibility crisis, we have an innovation opportunity. Consumers, patient advocates, the press should demand quality. I shouldn't be able to get my paper on the front television story or on the front page, of the — Do people still have newspapers? — on the front page of the newspaper. I shouldn't be able to do that. The journal should know that it wasn't performed blinded, it wasn't repeated, there weren't controls, the reagents — And it should never get taught. We all are participants in the system.

Thank you very much for your attention, I'm happy to try and answer your questions.
I'll take one from the dog.

**Question:** Thank you. I'm curious with all of the challenges that you've outlined, and your solutions, and the willingness of review boards who approve slides that we may categorize as sloppy, have you thought about creating some sort of independent rating system, a “Begley barometer” of sorts of good science, or a reference that people could use as a counter to that to begin addressing the issue?

**Begley:** So, I have, but I won't do it. Someone else will. So, when we started talking about this in 2012, I told you, I got threats and hate — hate mail. Things have changed enormously over the last seven years, and there is a genuine desire — sadly not so much in the U.S. I don't think, but elsewhere — to try and address this question. And so, for example, I think probably the U.K. is leading the way. So, I've presented several times to U.K. funding agencies, and they are certainly committed to trying to find a metric for quality. The challenge is that this is so deeply rooted, and we don't have a metric for quality. There are people out there trying to do things — I received — I told you last night I checked that paper in terms of how many confirmatory and how many studies refuted it. That was possible because a colleague sent me an algorithm that they've just created and is publicly available that allows you to search on that and find out how papers are being judged by the community. Because a citation might be because I reject this paper, but I still have to cite it. So, there are moves afoot, I think. I'm optimistic. I think it'll happen, but personally, not fast enough. I want to see this happen in my lifetime. I don't know how much longer I've got left, but I want to see it happen in my lifetime.

**Question:** Yeah, um, a couple questions. Do you think from your perspective, it's clearly a cultural and systemic problem.

**Begley:** Yes.

**Question (continued):** Are these people fooling themselves, or are they trying to fool others? Are they aware of what they're doing, or are they unaware? And then the other suggestion I was going to make — back in my deep history I wrote a book on cold fusion, which was called Bad Science, and one of the ideas that we had that circulated in the physics community after that — every few years you create a completely bogus paper with a bogus result like the bone marrow paper, and then you put it out there, and then everyone who confirms a result can get thrown out of science, like a way to clear the bottom of the barrel, and I'm just thinking it may not — listening to your talk — it may not be as unwarranted as we thought back then.

**Begley:** So, the reason that I want scientists to be licensed and accredited is so they can be unlicensed. I don't see any value in the licensing itself, except it's in a threat.
Dr. Glenn Begley: Perverse Incentives Promote Scientific Laziness, Exaggeration, and Desperation

So, that if I do those sorts of things, then I could lose my licensing. I'd no longer be able to have graduate students, I'd no longer be able to write grants, and I wouldn't be able to stand up and give talks. So, I agree with you. I think we need — at the moment there are no consequences for these sorts of behaviors. In terms of your first comment, are they genuinely self-deluded or trying to delude someone else? If they're trying to delude someone else, I call that fraud. And I don't talk about fraud. I think mostly people are convinced that their data is real and that they're telling the truth. And they do that because they refuse to do the experiments blinded. They're convinced that this is right and that they're seeing what is real truth. That's true, I believe, for most scientists. They're self-deluded. They're not trying to dilute the rest of the community. The few that are are caught and recognized, but that's only the tip of the iceberg. What I'm talking about applies to 90% of scientists. The deliberate fraud, I think, is probably only two to five percent of scientists.

Question: Oh, but you can do fraud without manipulating the data.

Begley: You can.

Question (continued): You can do malpractice instead of misconduct using the necessary controls — you don't have the blind —

Begley: Yes. Yes. So, all of that's true, and so what I'm saying — so, I'm not calling this fraud, and there are several reasons why I'm not calling it fraud. A big one is because this is very common. This is what — in our experience, 90% of papers look like this. So, if I say to you, “You're committing fraud,” there's an emotional response to that. I'm not committing fraud. I'm doing, like Piero, I'm doing the same as everyone else. And I think to use the term “fraud” is so emotionally laden, that it's not helpful. So, I'd rather call it sloppy or sick science, because everyone's doing it, and as soon as I use the word fraud, the barriers go up, and they're not listening anymore. But when I talk about this, you're not the problem. It's always someone else in the room that's the problem. If we talk about fraud, it's a different conversation, but you could argue that many of the things that I'm talking about might be best counted as fraud. I've deliberately stayed away from that, because I think that that's such an emotionally laden term, but you know, I don't know that I'm right. I'm just telling you the way I've decided to play it.

Question: Fascinating — fascinating discussion here. You go through and mention a number of different fundamental errors that they make using — not using log scale, using arbitrary measurements, small sample size, poor-quality data, not understanding the limitations of the data. And when I kind of run down that list, there's there's one overriding feature, which is, it seems to me that the people that you're talking about are fundamentally quantitatively illiterate. And so, the question, and — and I say this as a quant, and I look at all of the errors that are being made, and saying, “If the people that
are undertaking the experiments in fact had the same quantitative literature — quantitative understanding that you have, if you give them the benefit of the doubt that they're acting honorably, then they take very different steps.” So, the fundamental question is: Is the problem more in line with not having a solid quantitative background as part of the general training for this type of scientist?

Begley: I definitely think that is part of the problem, but it's only part of the problem. So, when I give talks like this, typically I get to speak to all the students and postdocs as well, in a closed room and just talk to them. They have a much better understanding about statistics, you know, randomization, blinding, pre-specified power calculation. They have a pretty good understanding. The real problem is with the principal investigators, the people of my generation, who are running the show. And it's not fair to expect the students and postdocs to hold the principal investigator to account. The principal investigator may never have learned it, may well know it, but at the same time know that there is pressure to continue to publish in the top-tier journals, and it's that imperative that institutions should help them with. Because the investigators often know what's the right thing to do. They just can't help themselves. But the institutions are complicit in this too. Because if I'm at an institution publishing in these journals, then the institution gets more grants, more NIH funding, and so on, and gets higher rankings in the international scale. So, the institutions don't want me to stop behaving like that, because then they would fall in the rankings. And the rankings are critically important in attracting students from China and India who bring with them the money that pays for the research that pays for the papers that pays for the research that pays for the papers to attract the students. So, the institutions are complicit in this, and I don't know if that answers your question.

Question (continued): So, so basically you're saying it's a — it's a confluence of the economic incentives together with the — together with the — the lack of quantitative literacy on the part of an older generation that gets transferred into current research still?

Begley: Yes. The youngsters, the 30- and 40-year-olds, really want to do things differently. But they're constrained by a system that my generation has helped build, sadly.

Question: You talked mostly about tier-one institutions and top-tier journals.

Begley: Yes.

Question (continued): How pervasive do you think this problem that you've identified quite nicely here goes down into the majority of scientists worldwide that would publish in B-level, C-level, D-level and can't be in that 12% funding to NIH and the big granters?
Dr. Glenn Begley: Perverse Incentives Promote Scientific Laziness, Exaggeration, and Desperation

Begley: So, I don't have hard data to answer the question. My impression is that the tradesmen's journals are better quality, because the top-tier journals want a flashy sexy story, and they want everything towed up — tied up with a ribbon and the bow on top. The tradesmen's journals, the society journals, are much more willing to accept a story that's not quite as complete. That's my impression. Having said that, the last two examples I gave you with respect to endocrine-disrupting chemicals were in journals that you've probably never heard of, and each of those papers has only been cited 20 or 30 times. So they're very low-tier journals compared with the others, and yet, the problems that I presented to you are the same as that you'd see in the top-tier journal.

Question (continued): It extends all the way down all the way to the bottom?

Begley: Yeah, I think so, but I think it's — that I think the middle-ranking journals, the society journals — you know, the Society of Microbiology — and I think they're probably better quality. My impression.

Question: I'm actually a Ph.D. student working on a dissertation in gerontology, and I was wondering — I have P.I.s who kind of are in the old realm, and I was wondering if you have any recommendations for people like me trying to kind of butt up against that system?

Begley: So, it's really tough, right? So, I've — It's really tough. I think my recommendations are: Remember why you started this in the first place. Everyone started this because Auntie Mary had cancer, and Uncle Harold had a heart attack. That's why you went into it in the first place. Remember that. Don't forget that. The second thing is: Be true to yourself. You don't have to play by this system. There have been enormous changes over the last half a dozen years as a result of the sort of paper that we published. The system is changing. It hasn't changed, but it is changing. And people of your generation are much more committed to bringing about change than the people that currently control a system, who have a vested interest in preserving the system that allowed them to get to the top. Don't get conned. Don't get caught up in it. That's all I could say. And good luck. I hope your work goes very well.

Question: Yeah, hi. I have three points, so please bear with me. First licensure: There's almost no evidence, in fact I don't think there is any at all, that licensure in the other fields you mentioned has actually improved quality, whereas we do know there are many examples of the threat of taking someone's license away being used to silence people who opposed dominant failing paradigms, such as Tim Noakes in South Africa, Gary Fettke in Australia, so why would you believe that licensure would have this positive impact when it has obviously failed to do so elsewhere? That's point number one. Point
number two: Fraud. If these were truly honest mistakes then when the investigators found out that their results were not duplicated, or when the companies found out that their antibodies were not effective, then they would change their behavior. They would retract their papers. They would stop selling the ineffective product. But in almost none of the cases that you identified did anyone change their behavior after finding out that what they were doing was wrong. So, does that not suggest a very deliberate intention and therefore fraud behind the behavior? Point number three: Life expectancy —

Begley: You're expecting me to remember all this?

Question (continued): Licensure doesn't work. They're probably committing fraud. And life expectancy's actually going down. Life expectancies decreased in the United States for the past three years, so you were suggesting that the rise in life expectancy was justifying the investment and science that we've been making, but doesn't it perhaps appear that, especially in health, the impact of science is starting to have more of a negative effect, especially in the margins, the decrease in life expectancy being due largely to nutrition-related lifestyle diseases — strokes, Alzheimer's, increases in deaths from those causes, as well as from the opioid epidemic, which itself is partially result to — the result of scientific fraud and also the health-care system? So, might not the negative effects be starting to outweigh the positive effects?

Begley: Okay, so dealing with the last one first: So, the time scale on that was 200 years, and there is no doubt that over the last 200 years, life expectancy has increased. Is it starting to go down in Western societies? All the things you nominated are correct. So, certainly smoking doesn't help, diet doesn't help — adverse diet doesn't help and certainly — and you're right about the opioid crisis. That is certainly starting to turn things around, but still the overall trend in the last 200 years, not just in first-world, but in third-world countries has been an improvement in lifespan and maternal and infant mortality. So, it's not to say we're there, and there is still room to be done, and in wealthy societies we have additional opportunities to spend our money. Many of them are not good for our health. In terms of the retractions or the non-retractions, and is this deliberate or not? That's a fair comment. I told you that none of the papers have been retracted, that we know the investigators were unable to reproduce, and your comment is fair. That probably does speak to motivation. The first comment in terms of licensure: I don't think there's a single solution to fix this problem, and I see licensing just as one approach to trying to address this. The comments you make about how it can be used to silence people is true, but if what I'm saying is true and that 90% of the papers that are published are unable to be confirmed or unable to be substantiated, there is a real problem. The cost in the U.S.: $28 billion a year. These are real problems, and I would offer licensing as just one part of a solution. To give you an example of how absurd it is: I'm a physician, I'm licensed. In the morning I can go and see patients and maybe see half a dozen patients in the
morning. In the afternoon I can work on a drug that might affect 20 million patients, 30 million patients — a vaccine that might affect hundreds of millions of patients. In the morning I'm licensed, treated as a professional. In the afternoon I'm not. There is really no accountability. If I misbehave in the afternoon, I can continue to do that tomorrow and the next day. If I misbehave in the morning, I can lose my license and no longer continue to see patients. It seems that the balance is wrong. I don't think that licensing is the solution, not at all. But at the moment, I can do the sorts of things I'm telling you about. I can get elected to the National Academy, I can continue to get NIH funding, I can continue to train students and postdocs. There is no consequence. So, I don't think it's the solution by any means. I do recognize the potential concerns, but I think there is a real problem that needs to be addressed, and I see it as part of the solution. If you've got something better, I'm happy to hear it. I'm not married to this, but I do think we need to do something that puts negative consequences into the system. We don't have — we have almost none at the moment. I'm open to — I'm happy to listen — I hope you know —

Moderator: Do you need the mic back, Russ? Sorry, we've got a question over here.

Question: So, you brought up meta-analysis. And I think that's — the technology that we have now is kind of exposing a lot of these pre— prior studies that were not fraudulent, but they were, they were erroneously misconstrued. So, do you see multi— meta-analysis being kind of the way we can kind of expose these and kind of bring back real, real science and help?

Begley: I'm reluctant to speak on that when there are people that are expert in this field that are going to speak later today. I think — I think it will help. Again, I don't see that there's a single solution. I think that certainly it's helpful in terms of trying to get to what the real truth is.

Question: This is a little bit of a follow-up to Russ’ question. It seems a big problem with a lot of this is a lack of accountability, and you know, the question then becomes — you sort of talked about the individuals and the institutions being responsible, but they sort of have a, you know, disincentive to want to play well, because they lose their funding. So, do you think some of the accountability should go to an even higher level like the NIH, which is spending the public dollars —

Begley: Absolutely —

Question (continued): — and, I mean, should there be some systems built into the NIH to require evaluation of the quality of papers and a requirement that all these parts of the studies, the negative controls, etc., are an absolute requirement for any paper that is funded by the NIH, and, you know, that they also have a review board to, you know,
evaluate both the individuals, the institutions, and the journalists, and the work that's being put out?

Begley: Yes. Absolutely. I had on my slide — so I said, “I don't think there is a single solution.” I think there is one single solution. That's the funding agencies. If they took this seriously, I think the funding agencies could — I think the funding agencies could deal with his overnight. So, a few years ago, everyone became an AIDS researcher. Why? Because that's where the money was. A few years later, everyone became a breast cancer researcher. Why? Because that's where the money was. If NIH or the other funding agencies around the world said, “This will be a requirement for ongoing funding,” it would change behavior. There is no doubt, because researchers follow the money. It would absolutely change it. I'm disappointed that to my mind NIH hasn't taken the steps that it could to address this. There are people within NIH who are certainly committed to this, but at more senior levels, there's more of an ambivalence, and the changes that I see in my view were primarily a result of Congress demanding that something actually happens. But I think NIH could change this overnight if they wanted to and lead the world.

Question: (continued): And just a brief follow-up to that: Do you think there is a role for criminal and financial —

Begley: I think that, yes —

Question (continued): — implications in the people that do this? Because I think that would be a big disincentive for a lot of these professors who are driven and motivated by the system. If they know they're going to go to jail instead of just sitting in their nice New York apartment, and say, “I feel like I'm being, you know, wrongly accused,” while they're sitting in a beautiful flat in New York City, it's a slightly different game.

Begley: Yeah actually, he wasn't in his apartment. Sadly, he was in his son's apartment. He didn't end up covered in glory as a result of this. But I agree, there should be consequences. This is public money. It's not my money. It's the community's money that I'm privileged to spend. And if I squander it, then yes, I agree people should be held — and the institution should be held accountable and responsible. At the moment, we have really no consequences for bad behavior. That's why the licensing, I think, would — it's not the sole solution. We've got to have some consequences in the system for bad behavior. There are almost none.