Informal Discussion Transcript
Concurrent Session 6A: Mortality Improvement Approaches

Presented at the Living to 100 Symposium
Orlando, Fla.
January 4–6, 2017

JEAN-MARC FIX: One of the great things of being at this symposium, and I hope you took advantage of it, because it's almost the end, is that you get to interact with people from all different areas. If it wasn't for this symposium and a couple of their volunteer work, I would never have talked to a pension actuary, let alone interact with one.

Larry Pinzur here worked as a consulting pension actuary over 30 years before focusing on actuarial RMD [required minimum distributions]. He is an active volunteer on a number of Society of Actuaries committees, as I mentioned, including the Retirement Research Committee, the Retirement Plans Expense Committee (RPEC) and the Longevity Advisory Group, on which I had the pleasure to make his acquaintance. He has a PhD in math from the University of Illinois at Urbana-Champaign and is a Fellow of the Society of Actuaries. Let's welcome Larry Pinzur.

LAURENCE PINZUR: Thank you. Good morning. We are coming around the home stretch here. I do have a special announcement, though. There was a—you might have missed it last night—a minor miracle on ice. The World Junior Hockey Championship in Montreal, Team USA versus Team Canada, and Team USA tied it up 4–4 in the third period and won 1–0 in the shootout, so good for Team USA. Sorry, Jean-Claude.

I also want to mention that all of the opinions I am going to be presenting today are just mine. I am not representing the SOA, RPEC, Longevity Advisory Group and certainly not Aon Hewitt. But hopefully, you find this stuff interesting, and if you have questions, you may not want to wait until the end.

So here's the agenda today. You will notice throughout the material is that coming up with more effective mortality improvement models or approaches is hard, for a number of reasons. We are going through some of those challenges right up front. I also want to cover some basics. Most of you in this room are going to be familiar with the basics, and it's one of the things, actually, if you weren't familiar with the basics, I probably should have been giving this talk at the beginning of the symposium, rather than at the end, because maybe it would have helped you with some of the others. I will focus almost entirely on deterministic models in the U.S., U.K. and Canada, and I'll explain a little bit why I am focusing on deterministic models when we get to that section. The last thing, before the questions, I want to just point out, as you dig into mortality improvement models and as you look around the world, who's doing what, and you're trying to develop them for your own country and for your own specialty, it turns out that there really are some difficulties in assessing the effectiveness. At least there are currently, and I'd love it if someone would be able to come up with better ways to assess the effectiveness of some of these models.

So here's the challenge. This is probably my favorite quote of all time: "Prediction is very difficult, especially about the future," by Niels Bohr. By the way, some people think it's Yogi Berra. It's not. It is Niels Bohr, a famous Danish physicist and older brother of Harold Bohr, for those of you who are familiar with the mathematician Harold Bohr. This is the challenge, or the first part of the challenge. What we are looking at here—these are age-adjusted mortality improvement rates for ages 50 through 95 and for males. I'm going to be focusing a lot on male charts. The female charts look very similar, and I would rather not spend time going through both.

Being a pension actuary, I am most focused on ages 50 through 95, and this chart was actually taken from the SOA's Scale MP-2016 report. The mean is right around 1 percent. The standard deviation is 1.6 percent. Actually, the standard deviation for females, it is actually higher, around 1.8 percent. And you just look at this, and it's looking almost like white noise, there's so much bouncing around here. And this is part of the problem. How do you, with this type of volatile series, try to figure out what's going to happen in the future with any amount of accuracy? So that really is the challenge. Another—a couple of other things to notice, with a mean of 1, a standard deviation of 1.6, we'd expect—just think about it—about a third of the time that the change would be higher than 2.6 or lower than –0.6. That's a third of the time, and in fact, I just quickly counted up the numbers above 2.6 and the peaks and the troughs lower than –0.6, and it turns out there were 11 above and 11 below. That's 22 out of 64 years, which is pretty darn close to one-third. So it's a challenge, and what can we expect to do—maybe that's the better question: What's the best we can hope to do?

Then the answer may be to try to get the long term right and get something close in the near term. But no way do we think we can really start nailing this on a year-by-year basis. Here's another picture of the challenge: This is the actual heat maps. By now, I'm hoping you are familiar with the heat maps, where this is showing mortality improvement where we have time going along the bottom—this is from 1951 to a little after 2010—and ages going from 35 up to 95 here. This is actual Social Security data, and Cindy MacDonald in the prior session here mentioned that Social Security does smoothing internally by calendar year. And that's why you get this corrugated type of look, because within calendar years, which are going up and down on vertical here, there is that smoothing going on. If you looked at the raw Social Security rates and you did the picture, it would be much more pixelated. It would be very hard to see any patterns whatsoever. Now, if you sort of squint your eyes, you might be able to see some patterns growing here. But don't squint your eyes for too long, I will just show you the type of graduation, or at least RPEC's—that's the Retirement Plans Experience Committee. This is the type of graduation that RPEC has performed, and we're going to get into more details about some of the issues around graduation.

But what you see here, there's a whole bunch of things: once again, males, for the U.S. population and heat maps showing rates of mortality *improvement*—remember, not base mortality rates. This is the year-over-year improvement in mortality. So think of it as the first derivative of base mortality rates. These base mortality rates are usually these tiny little things close to 0, and that is one of the reasons we get as much volatility, with this first derivative and these tiny little things.

Just a couple of things I want to point out. I could probably spend the rest of the session talking about this chart alone. We've got these vertical things going on where there's periods of very low levels of mortality. Think of this as topographical map—that's the best way to think about it if you haven't already—where you're looking from 30,000 feet up. And the blues and the purples are going underwater, and the purple is getting really deep, so that's bad levels of mortality improvement. The greens, that's sort of like the grasslands and you're in the 0 to 1 percent range of mortality improvement, the average long-term level. The yellows are the foothills, and the reds and the oranges are the mountains. So you'll notice we went from a period in the 1960s of really low levels of mortality improvements to what I call Medicare Ridge, over here. Medicare came in around 1965, and it's pretty amazing to me how uniformly by 1975, mortality improvements had gotten so much better in the United States. This was for males and for females. When you look at it, you see the same thing on the female display.

A couple of other things, going horizontally: Those are what we would call age effects. But you'll notice, if you look across, there is nothing that is sort of consistent across ages. The only thing you'll notice is that up here, once you get into the 80s and 90s, mortality improvement is getting lower, which is the age gradient effect that Steve Goss was talking about yesterday.

Finally, the other features you should be looking at are along the 45-degree lines, and these are the cohort effects. There is a clear cohort effect in the United States population for men and for women. There is this golden generation with relatively high levels of mortality improvement born around 1935 in the U.S. The corresponding golden generation in the U.K. actually turns out to be born around 1930. There's also what I call Death Valley here, and that's the baby boomers. So, many of you in this room are with me here in Death Valley. Hopefully, we can break out of Death Valley at some point.

There are other features that I'll just quickly point out. You'll notice that there's a lot of variation, a lot of ups and downs, undulations here under age 50. I think to a large extent, that's just because the underlying mortality rates are so small there that any sort of changes that you get are going to be exaggerated. But the other thing to notice is that with so many of the other conversations that we've had over the past few days, age 50 seems to be the time when things start going wrong with the human body. And you almost can put a line here at age 50, and that's where a lot of the bad stuff starts happening. So, there's lots of interesting things here.

Stochastic versus deterministic—I think everyone in this room understands that they both start with historical experience. With stochastic mortality models, you build in a certain distribution or probably distribution, so that when you're all done, not only do you have a model and you can figure out what mortality improvement rates might be, but you can also quantify the future volatility. I have listed here some of the most widely known examples. The ones I am most familiar with are Lee-Carter and CBD [Cairns-Blake-Dowd] family of models. MI stands for mortality improvement throughout the entire presentation. We've got Lee-Carter, CBD; there's a whole family of CBD models which you might be familiar with.

Deterministic models—when I say they produce a single answer, the single answer may be a two-dimensional table of mortality improvement rates. It's not very easy. And this is one of the big drawbacks, probably the biggest drawback, of deterministic models, is that there's really no good way, no easy way to get a good distribution of how bad things could get, using the deterministic model, other than just doing shocks—saying we're going to push one parameter up a lot and push it down a lot and see how big the range is over time. Some of the more familiar deterministic models are the CMI models, CMI standing for Continuous Mortality Investigation in the U.K. If you haven't gone to the CMI website and you're in this symposium, you really need to log on and take a look at what CMI is doing. It's pretty amazing. And then RPEC, the Retirement Plans Experience Committee. We are going to go through all of those models in some detail.

Why am I focusing on deterministic models? The answer is just that it's not that they're better; the vast majority of pension actuaries in North America live in a completely deterministic world, especially when it comes to mortality assumptions. They may do a little stochastic stuff on interest rates, but with mortality assumptions, they still live in a completely deterministic world. For better or for worse, that's the way it is. And that's what practicing pension actuaries are using right now. U.S. government minimum-funding requirements are deterministic; accounting requirements are deterministic … for pensions. I know that other disciplines within the SOA, especially with PBR, are getting into more stochastic stuff, especially on the interest rate side, not so much on the mortality yet.

A little bit on the age-only versus two-dimensional: Age-only says that we're going to have a projection scale, and the only thing that's going vary is by age. One example is scale AA, and it says when you've got scale AA, you have one factor here. If you want to project a q_x value in 2016 for some time into the future, you take that q_x at 2016, and you multiply times this AA factor. So if AA_x is 2 percent, you multiply times 0.98. However, many times you need to get out to the future year. That's one-dimensional.

Two-dimensional means there is an age variable and a time variable. One example is the MP-2016 scale. MP-2014, MP-2015 and MP-2016 are all two-dimensional, and we'll see examples of that. Here the difference is that if you want to go from a point in time—let's say 2016 forward to the future, like 2018—you still have the same number of factors as you had above, but each of these factors is slightly different, because there's a time dependency as well.

So in either case, this is static versus generational, which I'm going to talk about for a second. In either case, you end up with a two-dimensional table of mortality rates, whether it's age only or a two-dimensional projection. Once you've got a starting point of mortality rates—let's say, in 2016—you can produce a two-dimensional table here. And that's what I've done just for example here. Once you've got your two-dimensional table of mortality rates, you can do two things. You can go just [go] down a column, and that's called a static calculation, static annuity. I picked 2023. If I put this in 2016, that's how you would calculate, for example, a *period* life expectancy, using a static table like that.

Contrast that to generational, where each year you move down diagonally. You move over one year in time and one year in age, and so this has been the generational, and this corresponds to *cohort* life expectancy. I'll just mention, as a pension actuary, why I put the static annuity shifted over a little bit. That's a typical trick for pension actuaries. If you want to approximate a generational annuity using a static annuity, you've got to push the static calculation out by a number of years—typically, something around the duration of the underlying annuity.

So now I'm going to get a little bit more into the deterministic models. Thank goodness, I have Jean-Marc here to pronounce this wonderful saying by Paul Valery, are you going to do it for me?

JEAN-MARC FIX: [*in French*] Tout ce qui est simple est faux, mais tout ce qui ne l'est pas est inutilisable.

LAURENCE PINZUR: I think that sounds so much better in French than in English. "Everything simple is false, but everything complex is unusable." And that's the balance that I think most of us are trying to come up with [for] mortality improvement models; it's what we're trying to do. You know that if you make it too simple, you're probably missing something potentially big. If you make it too complicated, nobody's going to end up using it. It really is a balancing act here that we need to accomplish. So very quickly, I'm just going to go through—we'll be going into each one of these in a little bit more detail.

I am calling these "SOA mortality improvement scales." They are published by the SOA but are sometimes in conjunction with the academy and very often need to be blessed by the regulatory authorities and the NAIC [National Association of Insurance Commissioners]. So I just put them under the heading of the SOA mortality improvement scales. One of the things I wanted to point out in this discussion is that there is no uniformity yet. One of the things I'm hoping for is that we can try to, at least, get some better framework eventually where it makes sense to have mortality improvement models that can be applied in different situations for different applications and there's a rationale for why one might be a little different than the others. Right now, there is no or very little rationale why these are all different.

I'll just quickly name them right now. One of them, the first one, is used for life insurance products, is a scale used in conjunction with AG [Actuarial Guideline] 38. It does not have a name, as far as I know. I am just going to call it AG38 MI [mortality improvement] in the rest of this paper. For life insurance, we'll look at what scale AG38 MI actually looks like. Each year it gets updated, so we'll look at AG38 for 2016.

Another scale is used for individual annuity products, scale G2; look at that. Statutory reserving for group annuities uses a Scale AA which is quite a bit different than what is used for pension plan obligations. I'll just mention that the pension plan obligations Scale AA was used up until about 2011—from 1994, actually, until 2011. Scale BB was sort of a transitional model that was used from 2012 to 2013. Finally, starting at 2014, we started looking at using two-dimensional mortality improvement scales for measuring pension obligations, and those were MP-2014, MP-2015 and MP-2016.

So here's the AG38 MI once again used for life insurance products. I just want to point out that this turns out to be the same for 2015 to 2016, because there wasn't enough difference to warrant a change. And it's rather straightforward, I would say on the less complicated side. The rationale behind it is for life insurance purposes, so you're not projecting mortality improvement way into the future. Generally speaking, for life insurance purposes, you're taking some base mortality rate from some time in the past, and you're trying to just bring mortality to the current date or maybe slightly beyond. Specifically in the AG38 paper, they say you shouldn't be using scale AG38 MI for anything other than short- or near-term projections, because it was never intended to do that. But you'll see it's got a fairly high age gradient going back to age 0, where the mortality improvement rates are approaching 2 percent for males and females. It's a long period here where it's 1 percent for females, exactly 1.2 percent for males, and it drops eventually to 0 percent at 95. It's based, by the way, on a 10-year retrospective look at Social Security, actual Social Security history, and then a 20-year prospective look at what Social Security is projecting under their Alternative 2 (their middle-cost alternative), and sort of an average of those two things. And that's how you come up with a one-dimensional scale; that's how AG38 MI was developed.

Scale G2 is for individual annuities, so this one you need to be projecting into the future, potentially very far into the future. The rest of this table was, once again, based primarily on what the future Social Security assumed Alternative 2 rates are projected to be. We're going to talk about Social Security rates, and it's going to be interesting to see that potentially gets you into trouble if you believe in cohort effects. I'll show you why that is when we get to the Social Security slide.

Continuing on, this one looks a little different now. This is Scale AA. Once again, this is the current statutory basis for group annuity reserves. It looks way different than all the stuff, and this one has looked way different right from the start, when it was published, in 1994. Let me just also say it's used for statutory reserves right now for group annuities, but it is also the basis for many government calculations, required calculations for pension plans. Scale AA is still used by the IRS for minimum funding standards. Now they just proposed something new starting in 2018, but at least for 2017 Scale AA, it's still used for minimum funding standards and for minimum lump sum provisions from a pension plan. So it is still being used. The background for this? This is based on Social Security and Civil Service Retirement System historical mortality improvement over the period from 1977 to 1993. Nineteen seventy-seven to 1993—just keep in your mind [that] this average central year is 1985. So for whoever is still using this scale, it is based on mortality improvement from about 30 years ago. I think it needs to be updated.

I always wondered about what was going on with Scale AA ever since it first came out in 1994. I knew how they did it, but I just didn't quite believe it, especially for a one-dimensional scale. Let me just go back a minute: For a one-dimensional scale, it's saying if you're up here at 2 percent at age 50-ish, 55, for every single year in the future, we think you should get 2 percent reduction ad infinitum at age 50. And similarly, if you're down here, you should get only 0.5 percent ad infinitum. And it just didn't make sense to me. And by the way, the actual numbers dropped considerably below 0.5 percent. The developers of Scale AA just decided that they didn't want the scale to go less than 0.5 percent at any given point in time. It turns out if you go negative, obviously, and you start projecting, that you eventually get some weird crossing over, so—.

So this is the slide that eventually made sense to me about Scale AA. I'm bringing back that historical heat map again. I put a white box around 1977 to 1993, where they picked up this information. Their Scale AA for males—this is for males at this point, and just know, look what's happening. If you just sort of mentally in your mind's eye average what is going out up here at 95, you see it's a lot of purples and blues, and they just push that. That's why that low number is down here.

You do the same thing at this high level around 55, that's what's generating the 2 percent. And finally, if you just mentally average those purples here over that period of time, that's the 0.5 percent. So what finally occurred to me is what Scale AA did and is still doing [is] it shows cohort effects that were in effect around 1985 and just assumed they were going to be in effect perpetually. To the extent that Scale AA is working now for any calculation is just pure dumb luck, in my mind.

Scale BB was a complete transformation. This is a transformational scale, and it may not look that way, but BB was released in 2012. RPEC [Retirement Plans Experience Committee] had realized we needed to get away from Scale AA. We wanted to go to a two-dimensional scale, but we also realized that not only were the pension actuaries not ready for a two-dimensional scale without any warning, but we also understood that the valuation software really couldn't handle two-dimensional easily yet. So what we did with Scale BB is we said, "Okay, we want to move away from Scale AA and head toward a two-dimensional scale, but we wanted to then translate this two-dimensional scale back to something that works in current valuation systems." And that's how we came up with Scale BB.

I'm going to jump to how we did this and maybe come back to this slide in a second. What we did to develop Scale BB [was] we created first this two-dimensional heat map and a two-dimensional scale. We then created, using the two-dimensional scale, on a generational basis, deferred-to-age-62 annuity values that were based on RP-2000 mortality rates, which at the time were the most current base mortality rates, and at 6 percent interest.

When you think about it for a second, when you think back to your Jordan's *Life Contingencies*—I know I'm really dating myself—think back to *Life Contingencies*. If you have a table of deferred-to-62 annuities as of 2012 and you know the interest rate, you can back into the underlying mortality rates that reproduce those annuity values. So that's what we did with Scale BB. The way you do that, by the way, [is] you start at the old ages, the oldest ages, you figure out what mortality improvement rate you need at age 104, because at 105 it's 0, but at 104 you roll back into the BB at age 104. Once you have the BB at age 104, you compare an annuity at 103 to an annuity at 104, and you can back into BB_{103}. And you just sort of bootstrap your way back, and you come up with a whole sequence of age-only rates.

I'm going to go back for a second. I've called them mortality improvement rates, but they're really not mortality improvement rates, and RPEC actually stated that in the Scale BB report. Nobody remembers that. Whenever they think of that Scale BB, they just say, "Oh, these are the mortality improvement rates." That's not the case. This is a mathematical artifact that, if you plug this in with RP-2000 base mortality, produces annuity values that are darn close to those calculated with a full two-dimensional projection. So at the time we went out with BB, we said, "We're going to do that transition. We're going to give you a two-year warning that we're going to go to BB. But in the interim, you can use this one-dimensional scale without changing your software."

So what I've done here is just to aggregate all these various scales—at least all the age-only scales—on one graph, and you see at these key ages, the stuff. Just consider right now, Scale AA is the outlier. So it's in the long green dashes—clearly an outlier. I'm not going to spend any more time on this one. But you'll notice for the others in this key period, there's a lot of consistency basically from about 55 to 60 on. Also, I forgot to mention that the G2 scale is higher than the AG38. G2 specifically included a little bit of an extra margin there, between 0.4 percent and 0.2 percent, so there's an extra little bump there, and you actually can see that right here. Lo and behold, even the Scale BB—the one that is really not mortality improvement rates, that we sort of backed into—was pretty darn close, especially at these older ages, because that bootstrapping process produced similar rates. Then BB eventually drops back to a much lower number to make everything sort of work out. So this is for males; same sort of picture for females. Once again, Scale AA is definitely an outlier, and I think you now understand why.

So two-dimensional scales, RPEC: I am going to talk about the RPEC 2014 model, because this is really the first two-dimensional scale that the Society of Actuaries has published. This is the basis for all of MP-2014, MP-2015 and MP-2016 now. The conceptual framework is based on CMI's current model. And that's what I am going to be talking about today. I want to talk about the CMI model. I want to talk about the one that's used today, not the one they have recently proposed in terms of some of the changes. Even though there are a number of proposed changes, the fundamental concepts are much the same.

There really are three key concepts: That near-term mortality improvement ranges should look something like what the recent mortality improvement rates were. They don't have to be exact but it should look something like the recent rates. Also, that long-term rates that are based on—and I'll put these in quotes—"expert opinion." That's what they say: "expert opinion," which we know will be all over the lot, even among this group of experts. Finally, there is a transition from the near term to the long term that's a nice, smooth blending with cubic splines.

There is one big difference between what we've done in RPEC 2014 and what CMI has done. CMI has a methodology that includes a very explicit split of age, period and cohort effects that was (and is) still quite complex—and a bit of a "black box," even though they've done a very good job of explaining a lot of the guts of that process. The RPEC 2014 model attempted to simplify that portion of the CMI model. We're just going to claim anything that's happening on a 45-degree angle; we're just going to call that a cohort effect, basically. And maybe we simplified it too much. But that's what we did, and in fact, one of the projects that the Longevity Advisory Group is working on right now is looking at potentially more efficient ways to do that APC [age, period and cohort] split. Then RPEC will be able to see whether it makes more sense to get a little bit more involved with explicit APC splits.

I mentioned this already—the smoothing, the last part here. The smoothing from near term to long term involves cubic splines, both in the RPEC model and in the CMI model. This is what Scale MP-2016 looks like. Notice that [the] white dashed line, for example, is where the actual data ends and where the projection begins; the light-gray line is 2016. You'll see a continuation of many of those cohort effects. The question is, Has that cohort effect ended at this point? There's some indication that it's certainly slowing down. It looks like this period effect is slowing down in general. And so the question is, longer term, are we going to have sort of a resurgence of the cohort effect, a positive cohort effect, for those born in 1935? Unfortunately, the baby boomers are still in trouble here, stuck in Death Valley.

Next one I want to talk about is the Social Security model, and this is important, because I don't believe it's very well known. I don't think it's ever, or only rarely, shown in public what the Social Security model looks like, because Social Security is a two-dimensional model. This is what the Social Security looks like when you put it in a two-dimensional heat map format. I'll just mention that mortality is just one assumption of many that the Social Security Administration uses to assess the ongoing viability of the system. I think it was Mark or Steve yesterday who mentioned the importance of the fertility rate assumption. So it's just one of many, but it's clearly an important assumption and one that gets studied quite a bit, and in fact that [gets] used in a lot of other areas.

There are three alternative SSA assumption sets: low, intermediate and high. Everything I've talked about today has been intermediate. Long-term rates of mortality improvement are based primarily—and Steve Goss went through this in some detail—on in-depth cause-of-death analysis. They have a much stronger emphasis on this age gradient over time than at least RPEC does. And there are good reasons for that, because they are looking much more important across an entire population of projections, whereas many of us are focused on ages 50 to 95.

One big difference is their interpolation methodology. For every single age, their starting point is just an average over the prior 10 years. And once they've got their starting point, there's a very steep, very quick convergence to the assumed long-term rate. And it's purely horizontal, so let's just go and look at it.

This is the picture, once again, so the history is exactly the same with cohorts and everything. And so here are the issues. If you look at the starting points here—first off, when you see that yellow bit over here, that's—and once again, in your mind's eye, just try to put a little box around this last 10 years. All of the starting points here are based on what's happened over the past 10 years. In addition, there is a convergence. Almost half of the convergence, I believe, happens in the first five years to the long-term rate, so it's not a gradual convergence. If you're higher than the long-term rate, you come down very quickly toward the long-term rate. If you're below the long-term rate, you're swinging up very quickly to it.

The other thing here that is most obvious, I think, is the fact that the projections are purely horizontal. There is no recognition of cohort effect. And I'm not saying that's wrong; it's just somewhat surprising to me. At a minimum, it would be interesting to see especially, for example, what is going on with the baby boomers. Right now, the baby boomers are expected to hit a big, relatively high level of mortality improvement, using the Social Security assumptions over the next few years.

And that's the problem that I have with overreliance on SSA with Scale G2 or maybe the AG38 MI scale. On the Social Security projections, Social Security is just doing horizontal stuff, when, if you believe that some of these cohort effects are real, you're not going to be reflecting that, if you are just looking at what is going on with Social Security.

So here's a quick summary. I'm not going to go through all of this. But the highlight here is there's just a whole bunch of stuff out there just with Social Security and within the SOA. There's not a lot of rhyme or reason why we use one for one and then something else for another.

The biggest discrepancy, in my mind, is the fact that today we could be using this basis to value a pension plan, using Scale MP-2016 and a two-dimensional table of long-term rate of 1 percent. Then if we terminate the plan, we need to purchase group annuities. The next day, for statutory purposes, the purchased annuities must be valued under Scale AA. Now, I understand for this, there's probably lots of things going on for pricing, but for statutory purposes, there's a disconnect there, in my mind. So one of the things that Longevity Advisory Group has on its to-do list is potentially—it doesn't have to be a precise framework, there can be flexibility in it—to have some more consistency around mortality projection scales. I'm hoping that we can come up with something that's flexible enough that it can apply to lots of different applications.

I want to quickly make sure I have time to go through models in the U.K. and Canada. Let me go quickly, as many as you are familiar with CMI. By the way, I have used the new CMI colors on this chart. So for those of you who are color-blind, please let me know if this is better. We've already switched, so if it's not, I'm not going to change back at this point.

By the way, I forgot to add the time scale; this is 1976 to 2016 on the bottom here. I forgot to put them on, and age is 20 through 100. And this little thing, I think my cursor was on right before the 4 here, and I took a snip of this picture. That's not a 14, it's just a +4 percent, but you'll see a very strong cohort effect in the U.K. What they do—and this is the very explicit APC splits that they do in each period cohort, plus they have a residual; I'm not going to go through this—what they do is combine the A [age] and P [period] components, and once again, this is the current CMI methodology and not the newly proposed. They do an interpolation to 100 percent of the assumed long-term rate. And by the way, if you're using the CMI method, you must come up with a long-term rate assumption; they don't make any recommendation. You have to come up with at least one assumption, so whatever it is, your age and period pieces are combined and converged to the long-term rate. Your cohort piece then converges to 0. Then they put the history together with the future projections. Then they add them together, plus the residuals. And that's how you come up with their projection.

This, for example, is an example of the CMI 2015 model with an assumed long-term rate of 1.5 percent. Much stronger cohort effects, by the way—considerably stronger, at least this is under their "core" layer; there's also an "advanced" layer.

This is the Canadian Pension mortality improvement scale; the CPM just came out a few years ago. The assumed long-term rate is equal to what CPP and QPP assumption is for 2030 and beyond: a flat 0.8 percent through age 82 that then grades down to 0 by 115. The gender, age and specific rate decreased linearly between 2012 and 2030, and you'll see there is no recognition of cohort effect, so this is a picture of what CPM-B looks like for males. You'll see it is sort of an averaging going on up through age 2010, based on the average mortality improvement rates, and then it's purely horizontal, that graduation going from 2010 all the way to 2030. So this is what the heat map looks like for the CPM-B mortality improvement scale.

It’s interesting, and we saw this yesterday with Jean-Claude and Annie, that when you look at Canadian heat maps, there seems to be a fairly strong component of cohort effect from the males but hardly any at all for females, which was interesting. But as acknowledged in the CPM paper, the research team did some analysis and they said there was not a materially big impact of using just this simplified approach, rather than building in an additional cohort effect.

I want to spend a little bit of time assessing models. This is another one of my favorite quotes, and in fact, I can read this one. “Essentially all models are wrong, but are some are useful.” George Box was a very famous English statistician who did most of his best work in the mid-20th century at Princeton and the University of Wisconsin. These are some desirable features, I think, and I think most people would agree. Desirable features from a mortality improvement model would be stability, and by stability, I mean when you add another year of data, you don’t want a disproportionate impact on whatever you are using it for. If you’re doing pension valuations, you don’t want one more year of data to increase your liabilities by 1.5 percent. We don’t want to create a drastic change.

Historical fit is important. Forecast accuracy is how well the model predicts the future, and there are various ways to look at that. I think it’s almost impossible to predict the very near term accurately, but the midterm and the long term: How well is it doing?

You also want some general smoothness. You don’t want extra parameters that you really don’t need or [that are] not doing anything. Plus just a general simplicity. What I’ve found is that the relative importance of these features may very well depend on your application. It may be different for long-term pension valuations versus some other application where you really need to get a better handle on what is happening short to midterm. So that is one of the things I’ve learned, is that the balancing of these considerations and how you want to weight one versus the other, I think it will depend on the application you are going to use it for.

And in fact, the last bullet here, I think, may be obvious. What almost always happens in these types of situations is that when you try to improve one factor, one feature, you typically decrease the effectiveness of another. And once again, it becomes a balancing game.

I want to spend a little bit of time on stability and fit. Those are two big ones that are almost always at odds. When you increase stability, it means that you are going to lose something on historical fit. We'll see some pictures of that in a second. It is almost always inversely proportional.

This is especially true in something called the direction of travel. That is, if you read the paper, the most recent CMI model being proposed, that basically the direction of travel—or even to try to pick up the current slope of the mortality improvement when you are jumping off— or if you are just going to assume that it's a flat, zero-slope assumption—that's one way that MP-2016 model has changed from MP-2014 and MP-2015. We were picking up a direction of travel in our models. You'll notice that in MP-2016, we said that was adding too much instability. If you flatten that out and assume zero slope in the starting-off future MI values, that adds considerably to the stability of the model. Interestingly, in the newest proposed CMI model, they included an option to reflect some direction of travel.

Then there's the whole smoothness-versus-fit controversy. It's probably best just through this picture here. I just took a snip of what the old Social Security looks like, and I'm just going to show various slides here of how much smoothing does. I think this is the best example up here, that on this one, where you have relatively granular smoothing; if you need to know this stuff that's going on here and you don't think that it's noise, it's important to see that. And yet for long-term pension valuations and stability is critical, this may be just fine. So I'm not sure; I think I'm convincing myself there's not one graduation that works best for every actuarial application. I think that there may be a range.

Long term rate—and I'm running out of time, want to make sure we leave room for questions—the long term rate was another hugely important assumption here. The first question is, How long-term is long-term? Because if you want sort of a target out there in terms of where things are going to, not necessarily flatten out, it's really just saying, "I can't predict much further out, so I'm just going to make a certain assumption as to what an average rate is going to be—1 percent or something." There's quite a bit of consistency. For Social Security, it's about 25 years. For MP-2016, it's 20 years for cohorts and now 10 years for the horizontal component. But 20 years is pretty consistent, and there's a whole bunch of ways of looking at that. How much reliance should you place on the history? I'm thinking some, but not a heck of a lot. Cause-of-death analysis, like Steve Goss is talking about for Social Security, is important but maybe not the only way you need to be looking at that. This whole thing about expert opinion: What is expert opinion? Who is the expert? And what is really known about mortality improvement?

The big question in my mind is the unknown unknowns. In that category, for example, I might put "Can we slow down the aging process?" I think a known unknown is the obesity problem. But as an unknown unknown, I would probably put "Can we slow the aging process, and if we can, how much can we slow it down, and what would the impact be?" So much of this stuff is basically a shot in the dark. I am hoping we are getting closer and doing something in the interim that makes sense.

You know, let me stop here. We only have forecast accuracy, which turns out to be—you can read in CMI Working Paper 91 why forecast accuracy, it sounds like such a logical thing, but we don't seem to have good metrics to measure forecast accuracy, and I tend to agree with them. And it's turning out to be a problem.

And finally, there is something on inherent volatility, which just means that when you take first derivatives of something that's volatile, you get a lot of instability, you get more instability, and then when you take the second derivative, which is the direction of travel, you get even way more instability. So this whole thing about direction of travel, this whole page is taken from Paper 91 again. It's basically saying, "If you really want to do direction of travel, just understand how much volatility goes into your model."

So, questions? [*Applause*] Thank you.

JORDI POSTHUMUS: Jordi Posthumus from Hannover RE. You said there was the one heat map you could spend the whole talk on. To me, it would be very interesting to look at that heat map side by side with a heat map showing the amount of medical expenditure and research expenditure targeted at procedures and drugs that are typically used at the various age groups. I would imagine one of the primary drivers of mortality improvement, we'd like to hope, is this medical expenditure and medical research. The extent to which you could actually disentangle that progress from what we view as a cohort effect would be illuminating. Because I think it's easy for us to say this is a cohort and it has its own attributes. But I wonder if a lot of what we think of as a cohort effect isn't really just because that's where the money and research is. And if money is there, that's where they are driving mortality down or extending and delaying death as much as possible. So just a thought.

LAURENCE PINZUR: I've never seen anything like that, but just to support your point, this thing over here—and this is my opinion again—that's mostly statins driven—well, a combination of statins and the impact of finally people starting to smoke less. You know, smoking—over here is where the Surgeon General made his statement. It takes a while for that to kick in. But I think, for sure, much of this yellow, orange, and red is statins related.

JEAN-MARC FIX: I was going to make the comment that it's not necessarily medical research that makes big changes in mortality but also public health programs, like we see with Medicare or the campaign against smoking, that have much more of an impact than the amount of money that is spent on direct medical spending. But I think it's a good exercise anyway to try to explain those peaks and valleys and try to see if we can find a logical explanation for why that happened. And that should color your thinking about, you know, is that going to happen again or is something else.

JORDI POSTHUMUS: Yeah, exactly. I totally agree with you. It is the amount of money you actually put down on the table for dealing with these things that would drive this as well. And the other part is to the extent that you think about projecting a cohort forward. I think it is also fair to recognize the biggest markets or opportunities for making medical money will change over time as the population distribution changes. To accurately predict the future hot spots for mortality improvement rates, I think you need to consider for which future age groups there will be the most money to be made, as that changes over time, and it's where medicine will concentrate its efforts.

LAURENCE PINZUR: Yeah, I'd love to see more spending to help the Death Valley group here.

JEAN-MARC FIX: And I think, as part of the cohorts, keep in mind that there is significant research to show that not all birth cohorts are equal, because of childhood exposure. And that is a permanent effect or long-term advantage.

SAM GUTTERMAN: An additional wrinkle, as a follow-up to the previous comment, is the effect that cycles in mortality have. An obvious example in the heat chart is what happened in the

United States at attained ages 30s and 40s because of the dramatic effect that the HIV/AIDS epidemic had during the last few decades. It is fairly obvious, at least in retrospect, that after the intense period of mortality deterioration, there would be corresponding improvement when the effect of the epidemic waned. An analogy occurs in the area of interest rates. Fixed investments will tend to revert to the previous mean rates or trend, although with uncertain lag. But of course, it is easier to conclude that there is this reversal of sign in mortality trend looking backward, as in the middle of that epidemic, it was quite difficult to project the course of that epidemic. Nevertheless, once advancements in the field were public, one could have expected that there would be some sort of "correction."

There's a related current issue in the United States at attained ages in the 40s and 50s, particularly involving the effects of opioid overdoses and suicides: Are these temporary blips or permanent changes in mortality levels? If the former, when will the reversion to the previous levels occur? And while it does occur, will it be a cohort or a period effect? It does seem logical that some recovery or offset to the deterioration will occur. Thus, blind extrapolation or continuation of the recent trend may adversely affect the accuracy of mortality projection at those ages.

JEAN-MARC FIX: I think it is important to consider the regression to the mean on the peaks and valleys you're talking about. As for valleys overall, people's motives and action will change to address those valleys, and naturally that will tend to get the people back to normal to the extent that we can do it. And so far, we've always been successful in doing it. The question is whether or not that is going to happen or not in the future. There are mitigating factors that society itself will take to counter those valleys and bring the people back to the normal trend. So in the years following the valley, you have this—in a way, artificial—peak because mortality improvements were extremely depressed and societal efforts were made to correct that.

LAURENCE PINZUR: I was just going to mention, you know, this is HIV/AIDS in this really dark purple here. The question is the continuing pattern here. At one point, I was thinking it was sort of the residual effects of HIV/AIDS on that community. There were some treatments that were successful. On the other hand, what could be going on here is that, regardless of HIV/AIDS, there would have been something bluish here, and then the HIV/AIDS crisis made it purple. The other thing, just along with the opioid issue, I think what's going on here may be an evolving or starting of another Death Valley. At least, I think some of this is the opioid issue. These are the folks who are around 35 and 40 right now, and we know there is a serious opioid issue. That could be driving some of that.

ROLAND RAU: My name is Roland Rau from the Max Planck Institute, University of Rostock. Part of my question was already addressed here; the HIV/AIDS epidemic might be part of the Death Valley. The other thing, which I doubt a little bit, I think you call it the "Medicare ridge." I think for several reasons it's probably not the case. Medicare arrived at age 65 and not at age 35. Wouldn't the driver start a few years earlier? And the main point is that this kind of pattern—I looked at various countries from the Human Mortality Database—this kind of strong period effect followed up by a cohort effect can be observed in many, many countries even without the introduction of Medicare, so I'd rather think that is due to treating and preventing circulatory diseases, where people then benefit for the next decades for those birth cohorts.

LAURENCE PINZUR: Thank you. Yeah. "Medicare Ridge" just sounded so good that I probably keep using it, but I agree with your point.

JEAN-MARC FIX: Maybe can see Larry's bias for pension, and only older people count.

LAURENCE PINZUR: And I think I mentioned, if you draw this imaginary line at age 50 here, there is still a lot of pluses and minuses going on under age 50, and that's as I said, there's good reason for that. Some of that is sort of the overcorrection/undercorrection type of stuff that Sam was mentioning. But there's still a fairly strong and continuing level of yellows and oranges here over age 50 post-Medicare, but your point is well taken.

JEAN-MARC FIX: Some of the things that I wanted to point out—some work that the SOA is doing. There's a project trying to decompose the age period cohort effect in the U.S. What we're noticing is that there is "below age 50" and "above age 50," more or less a kind of an opposite pendulum effect here. You see that the purple valley is next to a mountain, so that as soon as one goes up, the other goes down. Overall it might look neutral, but if you look at age by age, it's not so obvious.

So those are the issues that are complicating the modeling that we are facing. We have not seen that being discussed in other venues. The more we look at things, the more we realize there's quite a few factors at play in those projections.

LAURENCE PINZUR: Well, thank you very much. [*Applause*]