# RECORD, Volume 28, No. 1*

Colorado Springs Spring Meeting
May 30-31, 2002

## Session 16L
## Medical Science and Life Insurance

**Track:** Product Development

**Moderator:** MR. BRADLEY T. ROUDEBUSH
**Panelists:** MS. ANNA HART†
DR. ANTHONY MILANO‡
MR. JEREMY STANLEY§

*Summary: Advances in medical science, particularly therapies developed as a result of stem cell research, are expected to have a significant effect on mortality rates. Attendees learn the most important advances in medical science and current research that may lead to additional breakthroughs. Potential effects of those advances on life insurance products and the life insurance industry are explored.*

**MR. BRADLEY T. ROUDEBUSH:** Today's panel from the medical underwriting and actuarial disciplines has a little bit of everything for you. At the end of the session, I hope you at least feel a bit more educated on the current direction of medical research and how we think it applies to the life insurance business. I hope you gain some practical tools to translate some of these findings to your day-to-day work.

Dr. Anthony Milano is currently vice president of medical research at what is now known as Generali Re. Jeremy Stanley is a software research developer at BMA. Anna Hart is an independent underwriting consultant from Texas.
I'm Brad Roudebush. I've held a number of positions at Northwestern Mutual for 18 years, having spent the last five years assisting our medical directors in their research for setting underwriting status.

---

*Copyright © 2002, Society of Actuaries

†Ms. Anna Hart, not a member of the sponsoring organizations, is a risk consultant at A. R. Hart Consulting in Eastland, Texas.

‡Dr. Anthony Milano, not a member of the sponsoring organizations, is medical research vice president at Business Men's Assurance in Kansas City, MO.

§Mr. Jeremy Maninger Stanley, not a member of the sponsoring organizations, is a software research developer at BMA Insurance in Kansas City, MO.

**Note:** The chart(s) referred to in the text can be found at the end of the manuscript.

In addition to having what we hope to be highly educational and entertaining presentations, this panel has one other thing in common. We each, in some capacity, serve on the Mortality and Morbidity Liaison Committee (MMLC).

I mentioned liaison committee, which is the liaison among the Society of Actuaries, Association of Home Office Underwriters and American Academy of Insurance Medicine. This includes actuaries, underwriters, MDs, and specifically those who have some connection to the mortality studies committees in each of those bodies. And there is a committee, a mortality-related committee, in each of those studies.

From the actuarial side, this committee has had a representative from the Individual Life Experience Study Committee for a number of years. Currently that person is Tom Rhodes.

The MMLC's purpose is to produce experience studies that support the risk management information needs of those involved in pricing and underwriting individual life and disability products in the United States and Canada.

Also, it wouldn't hurt to have those studies in case of some lawsuit against a company over some contentious rating or decline. You'd certainly like to have the actual experience part rather than the actuarial judgment part to back up a lot of our decisions. If you're into analogies, I think that the MMLC is to impaired lives as the individual life experience committee is to the standard ordinary studies—what we traditionally think of as the Society studies. If you think those studies are behind, at least they have something to study. The bane of this committee's existence is lack of data.

Now to get away from this soliciting once every ten years, a small group of companies is coming out with impairment studies. In 1996, this committee began what we hope to be a more dynamic system to deliver more timely results from a wider base of companies. It uses the computing facilities of the Medical Information Bureau (MIB), just like is done for the regular SOA studies. But the impairment study capture system (ISCS) just uses the regular contributions that companies make to MIB part of your ordinary daily business.

MIB maintains all the appropriate security, turns around and gives those contributions back to the companies, and basically asks them did this person die? Did this person lapse? The company gives it back to MIB and, voila, impairment study.

The first tangible result of this new system was realized last year with the highly regarded study on the codes relating to alcohol abuse and elevated liver enzymes, what we call the AALE study. The study was really intended to prime the pump for the ISCS system. It eventually involved, I think, about 35 companies that contributed to the study.

Here's a high-level excerpt on it; the numbers aren't that terribly important. I'll point out a couple of them later here. It shows the experience of those codes compared to the recently developed 1990-95 Basic Table, which was a great table to get eventually because that's almost exactly the exposure period of the impairments for this study.

This is one of the studies published last fall in the *Journal of Insurance Medicine*. A good result was that people for whom companies didn't rate as standard experience, turned out to be almost 100 percent of expected. This is a great pattern also of what you might expect. Young people, alcohol abuse: There was actually a driving code also included in there and elevated liver enzymes. You could probably expect that experience to be worse than, say, a 40-year-old or 50-year-old with those same codes. Overall, lives that were rated with these impairments experience about 150 percent of the 1990-95 table. It's a good study.

Now, a similar study is planned for elevated blood pressure later this year. That fills about a 30-year gap of intercompany meaningful experience on that impairment. Just to mention, this committee in prior incarnations was also responsible for the 1979 Build and Blood Pressure Study and the lesser-known Single and Multiple Impairment Study of the 1980s. So this committee goes back a long way.

In fact, the committee would like to crank out at least one study like this a year. Like I said, I hoped that the regular workings of contributing companies would do just that. But since its inception, very few additional companies have signed on, and contributions from existing companies have lagged a bit.

So, based on a small number of companies, i.e., few exposures, few deaths, there really are very few impairments right now on which we could do a credible study. There are a couple of them that we're certainly going to try to throw together in the near future.

The classic argument that a committee like this would try and use to think about contributing to it is that no one company's experience is going to be credible. None of your companies probably does enough impaired business for its experience to stand on its own. So, we have to get together.

A scarier tactic might be that we feel the future ability of the whole industry to classify risk or defend ourselves against lawsuits could rely on a system just like this. Either way, if you have any connections back at home, we urge you to talk to your studies areas, see if you are contributing to this. If not, ask why, and try to generate support. We have a Web site, www.mmlc.org. It has any contact information you'd ever need, and of course you can come up and talk to us after the session.

That takes care of my plug for the MMLC and why we don't have large studies of

impaired insured lives. Did I mention that our other presenters will not be basing their results on large studies of insured impaired lives? Well, they won't. Does that mean that what they say should be totally ignored? Of course not.

For the rest of my time, I'd like to present an abstract of a paper that was submitted to the *North American Actuarial Journal* and has received at least favorable initial reviews. It deals with how to take the results seen in the kind of settings you're going to hear about, the clinical and public health domains, and how you might translate them to the insurance risk appraisal world.

Let's talk about why that is not as straightforward a job as you might think. First of all, a clinical study may have a totally different mix of things not being studied in an insured study, things like other impairments going on. That's selection bias. On a related note, even if we did know that information, just from a clinical article, we might not have enough information to mix those other things the way we want it. In traditional actuarial terms, we couldn't come up with the right expected basis. There just isn't enough information from, say, a clinical write up.

In terms of protocol, I mentioned earlier an elevated liver enzyme study. Well, words like elevated, abnormal and significant can mean different things, certainly, in the clinical world versus an insured world. And we have to make sure that the results are comparable.

Finally, and mixed in with all of this, there's no guarantee that the level and pattern of excess mortality in the outside world, due to some impairment, is going to look the same for insurance applicants who have that same impairment.

Now, the accepted methodology for translating that excess today is commonly referred to as the excess death rate (EDR) model or method, first set forth by the American Academy of Insurance Medicine and specifically Dr. Singer. It's pretty straightforward. Dr. Singer said that when you look at a clinical article, convert the extra mortality that's being described into an absolute difference; don't use the mortality ratios that they use. Mortality ratios in the outside world won't be applicable when put on top of insurance experience.

Then simply add that extra to some unimpaired insured basis, and that's a bit ambiguous, to get what you think the deaths from this impairment would be in your insurance world.

Again, put simply, the extra you'll see in a clinical setting on an absolute basis will be the same as you experience in an insurance cohort for that same impairment. It's pretty simple.

No real empirical evidence today supports that model, although it makes sense intuitively. There are plenty of other models that could work, and for those of you familiar with clinical literature, you might be thinking: I wonder if the Cox model

would do a better job of explaining excess mortality when you take clinical and apply it to the insured world.

In a paper authored by myself and Dr. John Klein of the Medical College of Wisconsin, this EDR model is compared to others. By the way, for any of you still taking exams, Dr. Klein is coauthor of the big yellow survival analysis textbook used in Course 5.

In that paper we tested these other models versus the EDR model. How did we do that? We started with NHANES I data. Raise your hand if you know what NHANES is. That's one of the big public health follow-up surveys done by the government. In NHANES I, the baseline survey was done around 1972, so it's a great study in that we have a long time to follow up.

Basically, we treated this like a clinical study. There was the survey itself, and there was a medical examination done as a part of that survey. It looked a lot like this insurance underwriting process. And from the examination that was done by doctors, as part of NHANES, they identified three commonly seen impairments in underwriting. They were hypertension, diseases of the endocardium, which is a heart valve problem, and chronic obstructive pulmonary disease, or lung stuff.

For each of those impairments, we came up with the excess mortality estimates for this entire cohort. NHANES started with about 14,000 adults who went through this medical exam, just like you'd like to do when you look at a clinical article. Then we took those estimates and saw how they held up in an insurable subset of this data. As it turns out, about 75 percent of this cohort we deemed to be insurable based on some build and blood pressure cutoffs and some other impairments that were coded.

Technically, we did this by doing repeated Monte Carlo simulations, where we split the data in half to do the estimates, and then ran an actual to expected on the other half.

Of course, for the different estimates we used in each model that would be model specific, it's not always the "q". It could be the hazard rate; it's not always something that we add on. It could be something that we multiply by, in simple terms. But in the end, the EDR model did perform the best, having the smallest cumulative difference of actual to predicted over all durations, so that made us feel good.

Let me make a very explicit conclusion here, tying the first part of my talk to what I just talked about. If there is little data on actual insured lives, we believe the use of publicly available data, such as that in NHANES or Framingham or whatever you can get your hands on, could help bridge the gap between clinical literature and its application in insurance underwriting. By identifying, if possible, typically insurable lives from what are generally healthy populations, we can continue to explore and

validate other models, until some day we hope to get rid of the actuarial judgment and use actual experience.

Anna Hart is here to discuss the current direction of medical research and how that applies to the insurance world. Anna is the independent underwriting consultant in a small town in west Texas. She has worked mostly in reinsurance for 11 years. Her areas of expertise include structured settlements, critical ailments, emerging medical technology and expert testimony. Anna is here to discuss current trends in medical research, and I think that you'll learn a lot.

**MS. ANNA HART:** I'm going to take a little bit of a different turn in my talk. I do want to make a comment on what Brad said in his paper, that he used the NHANES data. Since that time, they have done two additional studies in NHANES 2 and in NHANES 3, which are all using a similar population over time and asking similar questions in, I believe it was NHANES 3 that actually went back and got additional medical information and lab data, which they're going to use to do some comparisons from the original study information.

I'm going to talk a little bit about medical science and life insurance. My objectives are basically to provide you with an awareness of what is happening in the medical science research field, the impact of those advances in medical science on mortality and the potential impact on life insurance products and within the life insurance industry. Some of this is very forward thinking and probably will not impact us in the very near future, although I'll give some specific examples where they will.

To start with the purpose and goals of medical research, the primary goal is to understand the disease process, to find more effective ways to treat these diseases, and then ultimately to find a cure. In the disease process, it was originally felt that a lot of times, there was just an A to B to C and the process was very specific. As time has progressed medical science has gotten so much more technological, and you'll see this later in what Jeremy and Dr. Milano have to say. There are many more intricacies, and these have become much more apparent with the human genome project and finding out where each disease process falls on one person's genome. In terms of disease, a definition just for everybody is an impairment of health or conditioning of abnormal functioning. It encompasses both mental and /or physical states. And it could be influenced by multiple factors, not necessarily just what you think of in terms of hypertension and diabetes going together. But you have the impact of the lifestyle factors: things like physical activity, whether you smoke or not (which, of course, is the primary one), and interaction with other people. We have a long list of what becomes important in just basic disease process.

Stress has become identified as having one of the major impacts on disease in general. I read an article this morning in the paper that talked about people who are working in companies and how so much downsizing has happened that vacations aren't taken by these people very much anymore.

The impact that is being thought of as a result is that so many more people will become sick because they will burn out and stress out if their companies are asking their employees to do so much more. I'm sure all of you are being asked to do more and more as so many people around you are leaving. I've been there and experienced that.

In the current direction of medical research, there are various areas, some of which are more today thinking and some of them are a little forward thinking. Genetics has become a very big area of research that is very political. There is also stem cell research. I know you all have heard that before, and I'm going to give you some specifics on what it implies and what it includes.

Some of the drug therapies that are used in some of the advanced cancers (chemotherapies) are not very traditional. Some of the things they give you in chemotherapy are what you kill rats with, basically. So they're modifying some of these drugs to use on humans. And then there are vaccines that are being looked at, for example, in Alzheimer's and AIDS.

In genetics, you have the human genome project, with which they have actually defined the whole human genome. And recently in one of the science magazines, the actual person whose genetic DNA was used admitted that it was him. I'm not sure that I'd want my DNA on a map for someone to see, but he admitted that it was sort of a secret as to who it was originally supposed to be.

There are also biomarkers and genetic mutations. We use some of the biomarkers in some of our screening for prostate and some of the cancer tests. And then I go to screening tests and what their value is in the industry. You talk about prostate-specific-antigen testing, and most companies start at age 50. I saw something that came down from the medical field, that the screening had been so effective and that the therapies that were used were actually encouraging patients not to be tested every year, but to come in every so many years, because of the impact of that additional information.

Both examples show the impact of genomics. That's a new word. I actually did a lot of research on this. One of the areas in which I've spent a lot of time is research and development. It is defining the molecular mechanisms of complex biological processes. That's like going down on the nitty gritty level and watching what happens inside your body move around. It looks at the stages of the disease process from stage one, stage two, stage three, or of course stage four, which tends to be the most deadly.

It identifies the biomarkers that may cause a specific disease. There are certain mutations in the genes; there are certain things that happen to genes as a result of things that happen in your life. It could be stress. Some things are just genetically oriented, and some things may be a familiar part of your family history as well. And then there are the effects specific genomes can have on the diseases that we

looked at.

But the goal of genomics is to develop individualized treatment for specific diseases. Some of those being looked at are diabetes and heart disease, that is new treatments for some of the diseases that we as the insurance people have deemed unacceptable in the past, as not insurable. The next goal is to look at what the survival strategies of cells are. And you think about why some people die before other people and what makes that longevity in one person versus another, and that's a survival gene, basically. It's not a will to live exactly; it's just a strength of your basic DNA.

Basically, genomics looks at DNA and then move down through the messenger RNA. This is all a little process that you take on a pictorial step down. Then, from the messenger RNA, it looks at the individual protein that is in the cell, then the cell physiology of what exactly makes up that cell, and then at the phenotype of specific identity of each cell.

It's very complex; they make some really pretty pictures that make it look simple, but the runs that the scientists use make it very, very difficult. That's why they do what they do.

Stem-cell research is a little different area, and it goes a little beyond genomics in general. So I want to talk about what a stem cell is, because you've heard about stem-cell research, and there are a lot of companies that are going into competition about actually obtaining big globs of stem cells. What are they talking about when they say that? Then I'm going to look at how they're used, what they're used for, some of the issues, which are highly emotional and ethical for some people, the implications, and some of the expectations that we can see from them.

A stem cell is a generic cell that makes exact copies of itself forever. It just keeps going and going and going. The value in that is, you're not looking at one cell and then it's gone; you're looking at a never-ending supply for use in research. In stem-cell types, we get into some of the ethical issues, which I will touch on very lightly. Embryonic stem cells are obtained from aborted fetuses, which is a very political issue, and also from in vitro fertilization. And then there are those that can produce cells for almost all the tissues in the human body. That is where most of the exciting part of research is. It's not limited to one disease process; it's limited to only as far as your imagination or your ability to research is.

Adult stem cells are less versatile. They're more specific to cell types, individual cell types, but they are found in adults and in children, not in embryos. Stem cells can produce specialized cells for various tissues throughout the body, for example, in the heart, brain and liver. A lot of this is being used to look at the possibility of regenerating liver tissue in hepatitis C or severe alcoholic cirrhosis, even in some of the liver cancers.

These cells can be maintained forever, which is again the added advantage. Again, they're used in Alzheimer's, diabetes, heart disease and molecular oncology, which is the area that makes up cancers, for some of the drug therapies. I think on the Food and Drug Administration (FDA) site, they give a list of all the approved drugs every year. So all the new drugs and what they're for are out there for everybody to see. The FDA controls this application and improvement process and then lists it available year by year.

There are a lot of foreign drug company trials and availability of drugs. The FDA is a little bit tighter in its process of approval than some of the foreign companies. I remember seeing an application one time from a man for a melanoma vaccine. It was very experimental. I had never seen that before, but he didn't get the drug in the United States. It was from a German company, and he had a lot of money and a lot of influence in order to get that or be part of that very specialized clinical trial.

This leads me to clinical trials. I'm sure some underwriters see a lot of people who get into some of the disease states where they talk about there being no real cure for what you have, and they will try you on a clinical trial. I want to give you an idea of exactly how they work, which is kind of interesting and leads to the approval process, as some of the drugs that are out there being looked at now will eventually get to your drugstore, and, hopefully, to some family member who will need it.

How do clinical trials work in theory? What is the timeline and the number of people involved? In clinical trials, there's the basic research. They may do animal studies, and then these become a formal proposal, patient recruitment, and then patient consent.

Now, you've got to be very careful when you get into a clinical trial, if you're actually in one. There are some actual research facilities that are not as thorough in following you as an individual and really maintaining the ethical guidelines. Because they tend to be funded by the drug companies, they have a vested interest in that drug performing well. So there, you really have to be careful when you're involved in that kind of thing.

In the timeline there's phase one, which is basically a safety stage. They use very small numbers of healthy volunteers, and that can last up to one year. In phase two, there's an increased participation, maybe several hundred volunteers, and they have a research group and a control group, so they do a comparison between the two. That may last several years. I'm sure everybody's seen "the phase 2 trials of this drug" and then "the phase 3 trials."

A lot of times the government—or if it's something that would really impact the industry or impact a lot of lives—a lot of the drug companies will try to push the FDA process to get past some of those stages quicker. Phase three is basically effectiveness testing by comparing the existing treatments and using a placebo. An example of that, which is interesting, was the research done recently on

depression. They were comparing one of the depression drugs and placebo, and it turns out that the placebo performed better than the actual drugs, and that just freaked the drug companies out.

So they actually said that this is too small a comparison, and maybe it wasn't the right length of time for that person, but it shows the power of the human mind, to a certain extent. When that person found out they were on a placebo and went off it, their depression immediately came back and they went running to the doctor for the drug, because they really felt that that was what had made the difference or would make the difference in their lives. I thought that was interesting.

And then comes FDA approval in phase four. Interestingly, I don't think a lot of people hear about phase four testing. That's looking at continued testing up to ten years, and if you look at some of the adverse reactions on some of the drugs that come out, that's part of the ten-year process. It could be one person that has an upset stomach, or one person whose hair falls out, but always, during that ten-year period, they're accountable for all the adverse effects that can result. Some of the liver damage and some of the things that really come across during that follow-up testing have actually led to removal of the drug from the market in some areas.

The numbers involved are interesting. There are a lot of categories of clinical trials out there, which means that the drug companies are paying a lot of people. There are 23 categories and 18,661 trials, which I find amazing. And of the 80,000 clinical trials in 2001, worldwide, they involved 20 million people. You think: Well, clinical trials are not going to impact the insurance industry; 20 million people worldwide is a lot of people.

So, you may have actually written on someone who, in an application or in some kind of area that you're involved in, is involved in some kind of trial. These include some of the new therapies or drug development, which is always ongoing, new blood tests that will identify disease states earlier, chemotherapies that are directed to specific cell treatment, and then vaccines.

The implication of all this is that there will be improved health, patient health, which will ultimately go down the line. This means new applications and more insurance and more money and more profit, which are all of our goals. We classify disease from fatal to chronic. I think that's the direction that a lot of people thought AIDS would go in, and has, in a sense, early on. I guess when the viatical companies used to make all their money off the AIDS patients, they had a one- to two-year life expectancy and survival. And then all the drug cartels and therapies and regimens came out, and now you're looking at a ten- to 15-year disease in some people. And there's new pricing involved. While the insurance industry has not embraced it as a new opportunity necessarily, it is evident that some of the things that are being done out there will offer us new opportunities in the future.

We can also transfer clinical and evidence-based research, which is what I've been

talking about, and then apply that to our current underwriting and see where we can find a place for that. An example is in Alzheimer's. There were some vaccine clinical trials, although you don't hear a lot about this, in the United States and Europe in late 2001. But they halted the trials in February 2002 and the reasons were varied; some of the people were having very good results and some were not. There were a lot of political issues and too forward thinking too quickly, I think sometimes.

One of the issues that arises in all these various areas, stem-cell research and oncology issues, is the financial part of this, i.e., who's going to pay. It becomes an issue of fairness versus profit. There are a lot of people out there who would love to have a cure for their disease, and be in some really high-level clinical trial that gives them access to a drug that no one else had, but it tends to be determined by some other factors, and that is not fair.

The drug companies usually benefit from the financial side of this because they're going to make money because they're going to have a new drug on the market. So they'll have exclusive availability for a certain number of years. The ethical issues are another thing. There's the morality of what is being done in medical science and medical research. How far do we go to find cures for everything? Do we put people through treatments that are worse than the diseases themselves? In some cases, they are.

And then there is the availability of it, which is so limited. And that's another ethical issue. Who do you make available? Who has access to things and who actually knows? I mean, if a person in a small town were to look for clinical trials, that person would die before he or she gets enrolled, because that person is not in one of the major research centers. There are very few; they're very selective throughout the nation.

And then there's the political issue, which you know if you ever sit in the Senate and the House and listen to some of the debates. Like with Michael J. Fox and Parkinson's disease. He made an impassioned plea for more money for medical research.

Well, who gets access to do those kinds of things? They do impact a lot of people. And hopefully, if someone at that level can actually make that kind of plea, then the government may take that kind of plea. Then the government may actually put more money into research, and eventually the rest of the world will benefit from that.

There are the expectations. There's the medical community and what their expectations are, and that's tied politically to a lot of things. There's a general population: What's out there for me? When is it going to be here? When can I see a cure for some of these things that I see my family dying of and my friends dying of? When are you going to find a cure?

There's the insurance population that's going to say, "Hey, this is a new drug I'm on and it's got proven 'viability' and 'efficacy', and when are you going to give me a better policy?" Because I have gone out and gone beyond what is available to most people, and I have some proof it's going to work; I'm in remission.

And most often we do not offer coverage on those people, because we're looking at long-term studies and how large numbers of people can translate into something that you as actuaries can actually put on the line. Because you have to sign the papers at the very end of the day.

Concerning the insurance industry as a whole, there are companies out there that are actually looking forward, thinking, "And how can I expand my base?" Because it's a very small market now, in terms of big companies that are surviving. And some of the small ones have to find a niche where they can actually make a place for themselves.

And how do these advances translate? Well, they translate into prevention and treatment, and then there is an impact on life expectancies. We are beginning to see that, and I think as time passes we'll see more of that.

Looking ahead, how will the life insurance industry be affected? Eventually, there will be new products, and there will definitely be new markets. I think that we will start offering insurance to more people at advanced stages of disease because there has been more therapy and more progress made in the medical field.

And then there will be, as a result, improved and more accurate underwriting, because we'll actually have critical information on which to base our decisions, and that will make everybody's products better.

There's great potential available as a result of medical science research and advances. But unfortunately, the impact will probably not be immediate for our industry. But for those proactive and very research-oriented companies, there may be an advantage for you if you start now. I'll give everybody one statement: Let's maintain a healthy balance between skepticism and gullibility and find what's really appropriate and do what we can really do to go that direction.

**MR. ROUDEBUSH:** Well, the way that I look at it, by what you've just heard, you've made it through your undergraduate career. We're now sending you on to grad school with our last two presenters. It would be impossible for me to chronicle all the achievements of Dr. Anthony Milano. His experience seems to have touched every phase of medicine: clinical, military, health care management, the dollars of medicine and insurance medicine.

His M.D. is from New York Medical College, and he also has a master's in Public Health from Boston University. He was a colonel in the United States Air Force and a flight surgeon. He's familiar with the area; he's actually taught leadership courses

at the Air Force Academy in Colorado Springs, nearby. His experience is heavy on the education; he's been director of medical education at a number of hospitals. He joined insurance medicine in 1984, having worked now for the Old State Mutual Companies, New England Financial and most recently BMA, now known as Generali Re. He's currently vice president of medical research for Generali Re. He was the keynote speaker at the American Medical Association national convention in 1995. And from the things I've seen from him, clearly his strength is evidence-based insurance and finding ways to extract data and present it in ways that users like us can use. That introduction does not do him justice, but we've got to move.

Capably assisting him today is Jeremy Stanley, software research developer, also at Generali Re. Jeremy comes to us from the academic world. He is an undergraduate from Wichita State, and he also did some work at Cornell as a national science foundation undergrad. During that time, he wrote a couple of papers, "Boundary Value Problems for Partial Differential Equations, and "The Analysis of Differential Equations on Fractal Domains." He made a couple of mistakes on that one, but I'm not going to call him on that. He continues to work on his master's at Wichita State while being employed at Generali Re where he's come up with all kinds of neat things, one of which we're going to see today. The title of their presentation is "Enhancing Life-Table Methodologies With Neural Net Applications." First up is Tony.

**DR. ANTHONY MILANO:** Today we're going to talk about, as you can see, enhancing life table methodologies in a way that makes sense for all of us, with neural network applications.

The mission of our project has been to coherently integrate, unify and leverage population medicine science with traditional actuarial life table methodology, with nonlinear computational intelligence. I admit that this has had great cache for me in its ability to compellingly link and to integrate and to unify these various disciplines.

We have a series of goals and objectives. We will very briefly review our traditional 100-year role life-table methodology. We will then apply neural network and learning. We'll look at neural network applications for large database analyses and the potential for future research in the area. Finally, we'll try to come to an integrated and cohesive, coherent conclusion.

Our first goal was to create expert systems. We planned to do this from large statistically analyzed databases, and we hope to optimize the predicted value of mortality information. We obviously hope to improve risk appraisal, in terms of its accuracy and consistency and timeliness.

Just as a brief caveat, we're going to look first of all at a traditional life-table methodology. We selected colon cancer because I happen to be working on revising the whole database with Dr. Singer of the National Cancer Institute, as we speak. I

was working on colon cancer at the time, so this lent itself very well to our analysis of the SEER database for this particular project.

The mortality domains that we selected were site, histologic type, sex, ages, stage or extent of disease, the grade, which is the severity of disease, and treatments, which vary in duration since diagnosis. As you will note, we selected only eight of 96 variables: Basically, male 65- to 74-year-old cohort, regional spread, moderately differentiated, site-specific surgery of one sort or another, and at least 15 years of duration of follow-up. That tends to answer the questions, what happens to our applicants in the years 12, 13 and 14, and what is the projected mortality?

So we selected a cohort very briefly, of 1,075 patients, and Table 1 is all survival data from the Seer program, the number that died, each annual interval of the numbers left to follow-up, the observed, the expected, and the relative interval survivals with standards of error for each of the intervals.

Table 1

Surveillance, Epidemiology, and End Results — National Cancer Institute

**S6WM-5, Sigmoid Colon Adenocarcinoma (8140), Ages 65-74, Regional, Grade 2, Rx Site Spec Surg-30**

**SURVIVAL DATA**

| | | Alive at Start | Died | Lost to Follow-up | Observed Interval | Observed Cum | Expected Interval | Expected Cum | Relative Interval | Relative Cum | 2 S.E. Ob: Interval | 2 S.E. Ob Cum |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | < 1 yr | 1,075 | 65 | 1 | 0.9395 | 0.9395 | 0.9644 | 0.9644 | 0.9741 | 0.9741 | 0.0145 | 0.0145 |
| 2 | 1-<2 yr | 1,009 | 99 | 56 | 0.8991 | 0.8447 | 0.9619 | 0.9276 | 0.9347 | 0.9106 | 0.0192 | 0.0223 |
| 3 | 2-<3 yr | 854 | 95 | 61 | 0.8846 | 0.7472 | 0.9589 | 0.8895 | 0.9226 | 0.84 | 0.0223 | 0.0273 |
| 4 | 3-<4 yr | 698 | 92 | 44 | 0.8639 | 0.6456 | 0.9559 | 0.8502 | 0.9037 | 0.7593 | 0.0264 | 0.0307 |
| 5 | 4-<5 yr | 562 | 53 | 46 | 0.9017 | 0.5821 | 0.9521 | 0.8094 | 0.947 | 0.7192 | 0.0257 | 0.0323 |
| 6 | 5-<6 yr | 463 | 31 | 40 | 0.93 | 0.5413 | 0.9477 | 0.7672 | 0.9813 | 0.7056 | 0.0242 | 0.0332 |
| 7 | 6-<7 yr | 392 | 37 | 38 | 0.9008 | 0.4876 | 0.943 | 0.7237 | 0.9553 | 0.6738 | 0.031 | 0.0342 |
| 8 | 7-<8 yr | 317 | 19 | 41 | 0.9359 | 0.4564 | 0.9381 | 0.6793 | 0.9977 | 0.6719 | 0.0284 | 0.0349 |
| 9 | 8-<9 yr | 257 | 24 | 35 | 0.8998 | 0.4107 | 0.9323 | 0.634 | 0.9652 | 0.6478 | 0.0388 | 0.0361 |
| 10 | 9-<10 yr | 198 | 22 | 26 | 0.8811 | 0.3618 | 0.9258 | 0.5881 | 0.9517 | 0.6152 | 0.0476 | 0.0373 |
| 11 | 10-<11 yr | 150 | 13 | 25 | 0.9055 | 0.3276 | 0.921 | 0.5421 | 0.9831 | 0.6043 | 0.0499 | 0.0383 |
| 12 | 11-<12 yr | 112 | 14 | 22 | 0.8614 | 0.2822 | 0.9151 | 0.4963 | 0.9413 | 0.5687 | 0.0688 | 0.04 |
| 13 | 12-<13 yr | 76 | 6 | 23 | 0.907 | 0.256 | 0.9085 | 0.4508 | 0.9984 | 0.5677 | 0.0723 | 0.0416 |
| 14 | 13-<14 yr | 47 | 5 | 14 | 0.875 | 0.224 | 0.8997 | 0.4062 | 0.9725 | 0.5513 | 0.1046 | 0.0452 |
| 15 | 14-<15 yr | 28 | 2 | 14 | 0.9048 | 0.2026 | 0.8906 | 0.3628 | 1.016 | 0.5585 | 0.1281 | 0.0499 |
| 16 | 15-<16 yr | 12 | 0 | 12 | + | + | + | + | + | + | + | + |
| | The median survival time = 6.769905. | | | | | | | | | | | |

Looking at that over 1,000-patient database, we're able to construct very quickly for ourselves, using just an Excel spreadsheet, cumulative survival out to 15 years (Chart 1). The blue line is the observed cumulative survival. The pink line is standard population survival, and the yellow line is the relative survival, which we'll talk about in a moment.

Now, that data had to be converted in a standard actuarial life-table sense to mortality. This is what a mortality table, very briefly, looks like (Table 2). Basically,

again, it's our annual intervals after 15 years, the observed durations. The raw data is the number alive at the start of each interval, the numbers lost to follow-up, and the exposures in person-years, which is part of the standard actuarial approach to determining mortality (Chart 2).

Table 2

Surveillance, Epidemiology, and End Results

NATIONAL CANCER INSTITUTE

**ANALYSIS OF CANCER MORTALITY & SURVIVAL BY AGE, STAGE, GRADE, HISTOLOGIC TYPE, Rx & INTERVAL FROM DIAGNOSIS**

(Patients diagnosed from 1973 to 1998) **ANTHONY F. MILANO, MD, MPH** **Date:**

**TITLE: Table No.** **S6WM-5, Sigmoid Colon Adenocarcinoma (8140), Ages 65-74, Regional, Grade 2, Rx Site Spec Surg-30**

**SOURCE:** Observed & derived data; SEER Survival matrix source table: SEER*Stat 4.0

| Interval | Duration | Alive at | Lost to | Exposure | Number of Deaths | | Mortality | Annual Mortality Rate | | | Observed | Expected |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Start-end | Start | Follow-up | Person-years | Observed | Expected | Ratio | Observed | Expected | Excess | Interval | Interval |
| | (Years) | | | | | | | | | | Survival | Survival |
| | | | w | $E = l - w/2$ | | d' = q'E | 100d/d' | qi = di/Ei | q'=1-p' | EDR = q-q' | | |
| No. | t to t+ch t | l | | E | d | d' | | q | q' | (q-q') | p | p' |
| 1 | 0-1 | 1075 | 1 | 1074.5 | 65 | 38.25 | 170 | 0.0605 | 0.0356 | 0.0249 | 0.9395 | 0.9644 |
| 2 | 1-2 | 1009 | 56 | 981.0 | 99 | 37.38 | 265 | 0.1009 | 0.0381 | 0.0628 | 0.8991 | 0.9619 |
| 3 | 2-3 | 854 | 61 | 823.5 | 95 | 33.85 | 281 | 0.1154 | 0.0411 | 0.0743 | 0.8846 | 0.9589 |
| 4 | 3-4 | 698 | 44 | 676.0 | 92 | 29.81 | 309 | 0.1361 | 0.0441 | 0.0920 | 0.8639 | 0.9559 |
| 5 | 4-5 | 562 | 46 | 539.0 | 53 | 25.82 | 205 | 0.0983 | 0.0479 | 0.0504 | 0.9017 | 0.9521 |
| 6 | 5-6 | 463 | 40 | 443.0 | 31 | 23.17 | 134 | 0.0700 | 0.0523 | 0.0177 | 0.93 | 0.9477 |
| 7 | 6-7 | 392 | 38 | 373.0 | 37 | 21.26 | 174 | 0.0992 | 0.057 | 0.0422 | 0.9008 | 0.943 |
| 8 | 7-8 | 317 | 41 | 296.5 | 19 | 18.35 | 104 | 0.0641 | 0.0619 | 0.0022 | 0.9359 | 0.9381 |
| 9 | 8-9 | 257 | 35 | 239.5 | 24 | 16.21 | 148 | 0.1002 | 0.0677 | 0.0325 | 0.8998 | 0.9323 |
| 10 | 9-10 | 198 | 26 | 185.0 | 22 | 13.73 | 160 | 0.1189 | 0.0742 | 0.0447 | 0.8811 | 0.9258 |
| 11 | 10-11 | 150 | 25 | 137.5 | 13 | 10.86 | 120 | 0.0945 | 0.079 | 0.0155 | 0.9055 | 0.921 |
| 12 | 11-12 | 112 | 22 | 101.0 | 14 | 8.57 | 163 | 0.1386 | 0.0849 | 0.0537 | 0.8614 | 0.9151 |
| 13 | 12-13 | 76 | 23 | 64.5 | 6 | 5.90 | 102 | 0.0930 | 0.0915 | 0.0015 | 0.907 | 0.9085 |
| 14 | 13-14 | 47 | 14 | 40.0 | 5 | 4.01 | 125 | 0.1250 | 0.1003 | 0.0247 | 0.875 | 0.8997 |
| 15 | 14-15 | 28 | 14 | 21.0 | 2 | 2.30 | 87 | 0.0952 | 0.1094 | -0.0142 | 0.9048 | 0.8906 |

It also then develops the numbers observed and expected deaths, mortality ratio and the excess death rate that you heard Brad Roudebush talk about.

As you can see, from that last large table, we constructed quickly interval mortality rates of both observed and expected. And you see a great deal of variability in the observed mortality, and that is something that we hope to fix with neural network applications (Table 3)

Table 3

Surveillance, Epidemiology, and End Results — National Cancer Institute

S6WM-5, Sigmoid Colon Adenocarcinoma (8140), Ages 65-74, Regional, Grade 2, Rx Site Spec Surg-30

| Duration | No. Alive | Exposure | Number of Deaths | | Mortality | Mean Ann. Mortality Rate/1,00( | | Excess Death Rate | Cumul. Surv. Rate | |
|---|---|---|---|---|---|---|---|---|---|---|
| Start-End | at Start | Pt.-Yrs | Observed | Expected | Ratio (%) | Observed | Expected | /1000 | Observed | Expected |
| t to t+ch t | l | E | d | d' | 100d/d' | q | q' | (q-q') | P | P' |
| 0-1 | 1,075 | 1,074.5 | 65 | 38.25 | **170** | 60.5 | 35.6 | **24.9** | 0.9395 | 0.9644 |
| 1-2 | 1,009 | 981.0 | 99 | 37.38 | **265** | 100.9 | 38.1 | **62.8** | 0.8447 | 0.9277 |
| 2-5 | 854 | 2,038.5 | 240 | 89.48 | **268** | 117.7 | 43.9 | **73.8** | 0.5821 | 0.8096 |
| 5-10 | 463 | 1,537.0 | 133 | 92.72 | **143** | 86.5 | 60.3 | **26.2** | 0.3618 | 0.5858 |
| 10-15 | 150 | 364.0 | 40 | 31.65 | **126** | 109.9 | 86.9 | **22.9** | 0.2027 | 0.3594 |

Finally, using automated methodologies, we are able to summarize our data into five quinquennial groups of ages—actually three in this particular representation. These are age group zero to five, five to ten, and ten to 15, in which we are looking at the exposures in patient-years, which give us some idea of the validity of the database that we're analyzing.

There is also the number of deaths observed and expected, the mean annual morality ratios, and the excess death rates for looking at that population—from which, by the way, it is very easy to calculate what the select mortality ratios are, using the 1990-95 select and ultimate tables.

We'll describe it very briefly. Basically we have our observed data in the number of intervals by duration; the number of lives at the start of each of those durational intervals, using the SEER database again, from the National Cancer Institute; the exposures in patient years which are from the SEER database, which are equal to the number of lives minus the number of withdrawals divided by two, as you see; and the number of the deaths in the SEER database directly calculated. From that, we then calculate the mean death rate, which is basically a proportion of the deaths to the number of survivals, or to the exposures in patient-years.

And we calculate the expected death rates, which are a product of the exposures in patient-years, by the expected mortality rate for that interval. And so we'll have then a mean death rate, and also, using contemporaneous population tables, what the expecteds are.

The mortality ratios are easily derived. You don't use this very much for underwriting, but the excess death rates are also very quickly and easily derived, and these basically are what we use to move forward in terms of assessing the risk, because those are absolute numbers.

With that, I'd like to leave this to my compatriot, Jeremy, and he's going to show you something that we're very excited about.

**MR. JEREMY MANINGER STANLEY:** Now that we've seen the traditional life-table approach, I'm going to throw up a demo—basically, a prototype that we've created. This is based on SEER data as well and the similar problem that was addressed right before with the life table. Here we have a set of inputs: age, diagnosis, site, grade, stage, all the way down to the diagnostic confirmation. And we've also got an output here that I'll illustrate in a second.

To basically put this in a frame of mind for you, this is an interface for a neural network to produce predicted annual mortality, in a way similar to what life tables do, but with more predicted values. So if I come in and click on a 70-year-old, now let's look first at "localized." Localized in stage obviously means the cancer hasn't spread to a regional state, and the predicted annual mortalities are all around 4 percent, relatively low. But if you watch, we go from localized to regional, and we jump from four percent still in the first year, to 12 and 10 percent in the years 2 through 5.

Here I can compare against the descending population mortality to get my excess death rates per 1,000 individuals. So, we have basically taken neural networks in the background of this and gone from the input to these, with a higher predictive value than you would get by doing this with a traditional life table. But then we still compare this to a population mortality in the same way that you would with a life table.

So now I'll talk about how this is being done. We'll start with neural network computation and learning, and basically our goal here is to provide tools, so here's a caliper. We want to provide tools to assist in risk appraisal.

So what do I mean? Well, if you start with a group of people who are all seeking insurance, then the traditional way of analyzing the risk of these individuals would be to group them into two groups, for example. And say that one of these groups has 75 percent of population mortality and the other group has 125 percent of population mortality. So, maybe these are your preferred standard classes or something like that.

What we would like to do instead is break these groups down and identify a more precise mortality assessment or risk assessment for each individual. For half of them, we can offer improved pricing—that is, some of these people actually will experience mortality less than the group that they were lumped into. For the other

half, we can offer improved risk selection, so we can identify those that were actually at a higher risk than the group that they were taken into. So in both of these cases, it's good for the insurance industry.

So, how do we foresee this actually occurring in the underwriting process? If you take one of these individuals, that person will have certain characteristics, like this one might be 48, male, with cancer of the colon that's been surgically treated. And an underwriter comes into the picture and analyzes these inputs for this impaired individual, and probably consults some sort of a mortality guide put together by the insurance industry and develops a rating.

What we'd like to do is not take the underwriters out of this picture, but instead provide them with a tool. So we'll take all of these inputs and feed them directly into the computer, via the underwriter, and then the underwriter and the computer will interact to come up with a more precise rating. Obviously this is very vague, and so I should clarify, what I mean by the individual interacting with the computer.

Well, in order to do that, let's zoom in here on the computer and see what's going on. So if I zoom in and crack it open, then you can see that the inputs come into the computer, feed into a black box, and the black box produces some sort of risk assessment. Here I'm talking about a black box; it's just a general system, so it could be anything.

So we're done; this is how it's done. That's obviously not acceptable. We want to look at what's going on inside of the black box. So I'll zoom in again, and open up the black box. And for the purposes of this presentation, what's inside of the black box is a neural network; so this is a nice segue to talk about.

What are neural networks and what do they do? Basically, neural networks start with some sort of an input, and it's fed into an input neuron, which serves the purpose of branching this piece of information off to the rest of the neural networks. So, here is a set of hidden neurons. Each of these hidden neurons actually does some sort of a computation that I'll talk a little bit about. And in some sense, these are what solve the problem.

In Chart 3 I've listed two hidden neurons, but there can be anywhere from one to millions of hidden neurons, although it becomes computationally very difficult as you increase the number of hidden neurons.

Then the output of these is collected into an output neuron, and it produces some sort of an output (Chart 4). So you can think of this as similar to what happens in your brain. Input maybe some sort of a sound; the input neuron could be your ear sending it off into the brain. The hidden neurons would be the brain's processing ability to make some sort of a decision, and the output might be something coming

out of your mouth in response. So, it propagates from left to right through this neural network as a signal.

So how do these things calculate? Basically, some of the neurons actually do computations, and those are the hidden and output neurons. They take multiple signals in, they add them together, they send it in a nonlinear way and then send another signal out. And this is modeled after the way neurons actually work in the brain, to some small degree. We don't want to make airplanes with wings that flap with feathers, right? So we don't want to copy biology entirely, but we can draw from biology to make good models.

So, for the overall structure we take some sort of an input, where here it comes from the real line from negative one to one, then that's sent into the input neuron, which is connected to the hidden neurons. But there's also a bias connected to the hidden neuron to give us one more degree of flexibility in the model. Each one of these connections is not just a connection; they're weighted connections, and these rates correspond to how your synapses grow—connections between each other based upon stimulus.

And so here, some of the weights are thicker than others, which means they would be larger in magnitude than others. And some are black and some are red. Black would be positive and red would be negative. Then once it passes through that layer of hidden neurons, there's also another bias there. We've got another set of weights connecting the hidden neurons to the output, which then produces some sort of an output signal.

So that's more or less how neural networks function in real time. I mean that's exactly the process that was happening in that prototype that I showed you. So now we've got these weights, and those weights are really the free parameters in the neural network. And those are what make the neural network able to fit some sort of a model. So we need to look at how the neural networks learn, how they find out what weights they need to fit the model. That's called back propagation.

So the general idea of the model is, maybe we've got some sort of a database here, a training set taken from something like the SEER database. We've got agents on the horizontal access and percent of mortality in the vertical. So you see that there's a lot of variation, and this is similar to the slide that Dr. Milano showed, actually drawn from the database with all that variation.

We'd like to be able to interpolate or smooth that out, so that for an individual at 72, we can go directly up to what the actual morality would be like. A more precise estimate of that, in SEER, would be 8.2.

Okay, so how do we train these rates? Well, we start off with an individual age, and that age has an associated mortality that we could take from the database. And we take the age and feed it directly into the neural network, and we get that forward

process of how it calculates. Now, this produces a mortality output figure, and it's wrong at this point in time. There's some sort of an error there; it has missed.

So why is it wrong? It's wrong because when we start with a neural network, we randomly initialize all of the weights, so it just produces things somewhat at random. We'd like to go from that random initialization to a model that actually fits the data. How do we do that? We start by identifying each weight's contribution to the error. That can be done in a mathematical way called back propagation, and not only can we identify its contribution, but we can actually decrease it. It's got a grading descent algorithm, for those of you that remember advanced calculus.

But this is still a problem, because we don't have just one individual; we've got a whole set of ages here, each one of which has an associated mortality. The neural network will make predictions for every single one of these ages. Just because we reduce the error on one of the ages, it may change the rest and probably will, so that the error might be magnified someplace else. So this is a problem. How do we get our global models here?

Basically, the solution is to take each age, present it to the neural network, and train it to fit each age. So it is like one of those games where you try to use the hammer at the arcade, to knock the little goblins or that pop up their heads. You do this long enough, and eventually they all will stay down. Every time that you do this randomly, we call this an epoch, and it may take tens of thousands of rounds of these epochs, which makes your arm very tired to accurately train the model.

So, here's an example of something taken from an actual training process showing you how neural networks fit data.

You've got a population in reality, which is the black curve (Chart 5). And sampled from that reality are population statistics. Notice that there's no error here, just for simplification. And we've got a neural network too, that's going to come down and fit these, and these are gradually taken from the actual training process.

And you can see that the neural network slowly starts to figure out the overall pattern and fit it (Chart 6). Notice here at the bottom that I said each animation frame represents 100 epochs. We've got 36 data points here, so 36 data points times 100—every single time the frame switches, which is faster than your eye can perceive it. It's an immense amount of computational effort to train one of these neural networks. To put it in perspective with today's computational speeds, I did this one on my laptop in about four hours.

Here's another example in three dimensions, so it illustrates that neural networks aren't just a simple algorithm and can approximate any nonlinear function in any number of dimensions (Chart 7). And you can see that it's an odd process, but it starts to fit part of the data and misses some other part of the data. So you can see it fits the right side very well, but now it's stuck in what's called a local minimum.

So, right here it really hasn't fit the data very well. What's nice is that when I said to randomly present the training samples, that causes sort of a stochastic search of the weight space that will allow it to slowly work its way out of these local minimums where it thinks it has found a solution. So, now you can see that what I said is actually happening, and this red is starting to inch its way up, and as it approaches the top, it gets faster and faster. Suddenly there's so much difference here that it can make a big change rapidly and improve the model.

So I stopped the training. I didn't stop the training there, but I stopped taking the images because it's a pretty time-consuming process. So I'll go ahead and show you how close it goes after all this training.

So, here the red is the neural network and the blue is the model, and you can see that the fit is really quite close. And the reason that it has missed in some of these places is because we have a finite number of hidden neurons. As you increase the number of hidden neurons, you increase your model's ability to be accurate.

So now let's look at how we take these neural networks and apply them to a large database analysis with the goal of doing risk appraisal. Here is the methodology that we've developed. At first your goal is to select mortality-correlated variables, so you don't want to include into your model something like whether or not they're left-handed or right-handed. That has nothing to do with mortality.

Then we went to rank order by mortality contributions, so we know which variables contribute the most mortality. Then there is a need to classify variables for neural network applications. I'll speak more about all of these steps later. And then finally we actually create multiple neural networks for this appraisal.

First, we selected the mortality-coordinated variable. The SEER database has data from 1973 to 1991, collected by the National Cancer Institute as a part of our National Institute of Health. There are over three million cases of cancer in the database, 96 data fields per case, and that's everything from place of birth to cause of death, for almost 300 million individual pieces of data that could potentially be analyzed to predict mortality.

So, the quick solution to selecting mortality correlated variables was that we took the SEER database and created an actual Access database, just because it was a quick thing to do, and looked at all 250,000 colon cancer cases. And we linked Access to the Excel pivot tables and visually inspected the change in mortality across all 96 variables. Then we correlated with what most people in the clinical world perceived as being the mortality contributors to colon cancer.

Let me first talk about the drawbacks of this method. Obviously, it is subjective and unranked; that is, any time you're doing something visually or using the existing expert knowledge, it's a bit subjective.

The variable interdependencies are unaccounted for, so we can't tell whether or not grade and stage are really predicting the same mortality—that is, maybe we could use just one of them.

And the solution is to utilize statistical and computational intelligence techniques that will be more rigorous, that will allow us to rank order all 96 variables by predicted value, and that will account for interdependencies like grade and stage. There is a set of methodologies that has been developed for application to other problems that can easily be converted to do this. I really don't have time today to talk about them, but if anyone has questions after this, I'd be happy to talk about it.

So, our initial guess is this: that obviously age must play a significant role in contributing to mortality. Stage and grade are measures of the cancer severity and should be very high, but once you get beyond that, it's a little bit tough to tell which ones should be where. And obviously, we will rigorously accomplish this before we create functional models.

So, for classifying variables for neural network application, basically, there are two types. Type one variables have well-ordered domains. And what I mean by that is something like age; age is well ordered. Either age one is less than age two, or age one is larger than two. So basically, you can take all ages and set them out in a linear line and it makes sense, the position for each individual age.
So, that allows for is interpretation across age. To give you a graphical example of what I mean, here I've got three ages, 63, 64 and 65, and I've used survival time in years in the vertical axis. So for a 63-year-old from the database, you would expect to get a distribution of survival times going up in the vertical axis. You would expect the same thing for 65-year-olds.

But, suppose for some reason, your 64-year-olds weren't very well represented within your database. It would be hard to identify where that mean value is to get a predictive result. So, interpolation allows you to intelligently lay a curve across the domain, to get a more precise answer for where the 64-year-old ought to lie by utilizing the data from the 63- and the 65-year-old.

So, other variables aren't well ordered, and those are the type two variables whose domains are not well ordered. For example, therapy is not well ordered; it's very difficult to see whether surgery is less than radiation or surgery is larger than radiation, particularly when you have multiple types of surgeries and multiple types of radiation. There's no good way of ordering them together.

So, we have to create multiple neural networks. For all of the surgery cases, we'll create one neural network. And for all of the radiation cases, we'll create another neural network. Also, there may be some ways of getting beyond this column, but I really don't have the time to talk about those today either.

So, then the final step is to create multiple neural networks for risk appraisal. We will train these by the type two subdivided data. We'll take all of the surgery data and train one neural network, and all the radiation data and train another network. We'll calculate the mortality with the type one inputs like age, stage and grade, all of which can be interpreted across.

Then we will incorporate the top mortality contributors. So, you always want to include age into your model, but you may not necessarily want to include diagnostic confirmation. Now, why is that? The reason is that we want to vary the depth of the variables we incorporate so that the individual data points have some statistical validity. And there's a balance there that will have to be fine-tuned to maximize the predictive value of the model.

I'll give you an example. Here's the actual neural network that we used in the prototype, feeding in age, stage and grade; 12 hidden neurons predicting mortality from year one to year five.

The reason we have to vary the depth is because if you start with a Seer database, there are almost three million cancer cases. You think, "Well, I can include all 96 variables and each data point will have at least 50 deaths or something like that, and we should be fine."

But the problem is that once you get to colon cancer, you get 270,000 cases. And as we go down from sex to diagnostic confirmation, the number of cases represented in the database drops, usually by a factor of two, sometimes minisculely, but sometimes even by a factor of three. And at the end you still have to feed age, stage and grade into the neural network.

Okay, so what are our preliminary results? Here is just a very quick example of what the neural network can accomplish. At the top, I've got training data, and then this is the neural network and its data. Basically, you've got age, grade and stage, so the ages are grouped into two-year increments to consolidate them a bit.

I've selected just grade two regional cases. The mortality goes across for each one of these. So, for a 65-year-old, the third year mortality would be about 0.09. And then I compare this with the neural network. Going across the bottom, you can see that each one is within about one percent of each other. So 0.9 to 0.8, 0.09 to 0.08, 0.1 to 0.1, 0.19 to 0.2, 0.10 to 0.10. So, this is really quite close, but in some cases there is a more substantial variation. So here we've got 0.12 and then 0.10, so a 0.02 variation.

So you're asking yourself why has this happened. Why is the neural network giving me something different than what I got from the database? The reason is the interpolation. If we look, going left to right in the data, we go from 0.05 to 0.12 to 0.10, which is quite a change; you know it's not that smooth. Similarly, going from top to bottom doesn't make much sense. Why should 67- to 68-year-olds have

higher mortality than 69- to 70-year-olds? So it's not very smooth. That's because the data points have randomness to them, so there's going to be some variance there.

Whereas, if we look at the neural network going from left to right, it fits far more closely. Similarly, going from top to bottom, it fits more closely. Keep in mind here that we're also interpolating not just across age, but across grade and stage as well. And those would be multiple dimensions here.

So, in some sense, it's difficult to get a graphical feeling for why the neural network is producing better results. But what you can fundamentally do is look at the predictive value of your model, and I'll talk a little bit about that toward the end of the presentation.

This is a stacked graph comparison of the statistics versus the neural network (Chart 8). Basically, each graph is laid on top of the one before it. And the reason that I put this here is to show that visually the shape and height of the two curves from the statistics of the neural network are quite similar. So, it can split the model overall very well.

Okay, so what is that actual predictive value? We wanted to present a third-party application of neural networks to convince you that we're not making all of this up. So we chose the PROCAM Study, which is a comparative neural network application and analysis in the clinical world. Basically, they took a Munster Heart Study in Europe, looking for a prediction of coronary heart disease. And they have neural network algorithms and their standard PROCAM algorithm, which is our linear algorithm, probably analogous to life tables or a linear regression, and they compared the two. And these are the inputs that they used, which aren't really critical.

So, I'd like to show you the actual results. And here is a graph of the sensitivity and specificity of different models (Chart 9). So, you can see, here is the low-density lipoprotein cholesterol, which corresponds to this blue. Then just the low- and high-density lipoprotein would be the green; the standard algorithm is the red; and the PROCAM neural network is the yellow. For every single one of these specificities, you can see that there is an order increasing in predictive value and sensitivity for each one of the specificities across the board. So, like in all cases, neural networks were substantially better in terms of their predictive value.

So, what is our work at this point in time? Our goal here is to compare, in some sense, the predictive values of a set of different methods for predicting mortality. Starting at the most historical and, in some sense, the simplest base, would be the life tables. What are the predictive values of the life table? Then we'll move up to linear regression, and logistic regression, which is quite similar to linear regression but using a logistic function instead. Then neural networks are the point we're at

right now, and we haven't really compared all of these in a publishable way, but we intend to do that.

Then there are other avenues that are developing in very recent years. One is training neural networks with evolutionary algorithms, which is kind of a hybridization of two computational intelligence techniques. Evolutionary algorithms allow you to find global minimums faster and with more accuracy than back propagation does. Then the support vector machines are an entirely new concept for our model creation that tend to have a better generalization ability, so they tend to have better predictive value. So most definitely compare all of these.

I should quickly talk about what I mean by comparing the predictive values. We'll compare the predictive values as means. We'll start with all of our data, say the 250,000 colon cancer cases. And we will randomly select 80 percent of that data. Then we will train the method using the 80 percent, so we'll create some sort of a model from that 80 percent and we'll compute the method predictive value on the remaining 20 percent. That will give us a feeling for how well the model performs on data that it's never seen before. Then we will repeat these three steps to obtain 30 predictive values for each model and each parameter set in that model, because there may be some ways of substantiating different models. So, for example, there is the number of hidden neurons you use.

So, why do I need to do this 30 times? Why isn't one sufficient? The reason is that there is randomness in the way a neural network learns; that is, it's initialized randomly and then it approaches the data in a sort of random sense. So every time you run a neural network, you may get a slight variation in the predictive value. Also, I am randomly selecting 80 percent of the data each time, so that may affect how the neural network performs or how the method performs. In this distribution, I will actually be able to then compare the predictive value distribution by looking at their mean, by looking at their standard deviation, and by doing something like a Student's *t* test. So, this will be able to tell us exactly what we expect to get when we apply the method to all 100 percent of the data.

So, in conclusion, we have to provide enhanced sensitivity, specificity and predictive value, which are your measures for accuracy in both survival and mortality; and now, test efficiency, which will provide more correct mortality answers; consistency, so the users of these systems will have identical mortality information, which is a nice advantage; then protective value, we'll be able to look at the cost-benefit analysis of the cost of getting the input, and what are the benefits and predictive value of using those in the most advanced methods we can find.

So, these are some cutting-edge technologies. There have been a lot of scientific advances, as recently as two years ago, in how to train neural networks and how to do support vector machines. Many of these technologies were initially developed a

long time ago, but they haven't been practical. Part of that is because of technological advances. Obviously, you wouldn't be able to train a neural network ten years ago with any sort of speed that you can now, even on a laptop like this.

And to our knowledge, these technologies have never been applied to the insurance industry with respect to mortality prediction. And at Generali Re, we are actively researching and documenting exactly what we find with each method and prototyping systems that will have immense value, we hope.

**FROM THE FLOOR:** This question is for Jeremy. This is kind of interesting. The evolutionary algorithm and the global minimum, I didn't quite understand that. What did you mean like adjustments for trend?

**MR. STANLEY:** What I mean is that there may be different models that will vary drastically among the parameters that will have high predictive values. And if you were to look at the parameter space, which would be all of the weights, then there might be ten different parameter sets that would give you a high predictive value. But out of those ten, there will be one, and that's what I call the global minimum; that will have the highest predictive value.

Evolutionary algorithms allow you, with infinite computational resources and time, to always find the global minimum. So it doesn't necessarily mean that we will always find the global minimum, because we don't have infinite time.

**FROM THE FLOOR:** Is there a way or how would you adjust for trends over time? If you had a medical advance, how does that get in there?

**MR. STANLEY:** I think that's a two-part question. We can predict trends over time if it is a standard-of-living change, something that is gradual and expected to be continuous. Medical advances will always require underwriters' knowledge and medical directors' knowledge; you can't predict when cancer is going to be cured. So if that ever happens, obviously the models we built today won't be useful.

**MR. DAVID WYLDE:** You've used these mostly on impaired risk studies like colon cancers. Has there been any thought to using them for preferred risks?

**MR. STANLEY:** That is definitely a possibility, and these techniques can be used on any prediction where you have a large amount of data. In fact, there are a variety of techniques that can be used and have been used in the past to do just what you're saying. I don't think anyone has ever looked at it in detail with the models and some of the advanced models that we're using. But it certainly is possible. I think we started with the impaired risks because that was the data that we had to start with.

**FROM THE FLOOR:** I think everyone will agree with me, most insurance products are smoker-nonsmoker-based. You did not incorporate smoking into your model. Why not?

**MR. STANLEY:** That's an excellent question, and it's a similar question to why we didn't do this with an insured population. The answer is that the data is not there. As Brad Roudebush mentioned, when you have something and you're missing data, there are standard linear ways like what Dr. Singer did to correlate your general population to a subpopulation. So we expect to utilize those and to look at their predictive value. I think there are also perhaps some more advanced nonlinear methods of doing that sort of analysis, and we'll definitely look at those in the future.

**DR. MILANO:** The other aspect is, since I started analyzing the SEER database a half a dozen years ago in its yearly update iterations, the number of variables has increased from something over 50 to, in just five or six years, to now 96 variables. And so the Seer database, which is a wonderful way to look at and to employ computational intelligence, is on a yearly basis expanding the number of variables that are input. Furthermore, the trends go back a minimum of 20 years so that we can find and determine trend data, which is exactly what we want to do for our older applicants.

**MR. RICHARD BERGSTROM:** I don't really have a question. I just wanted to first of all commend the panel for what I thought was an exceptional presentation. It was enlightening, intelligent, very forward thinking, and we need more panelists like this. And just from my position as chairman of the MMLC, one of the long-term goals of the committee is to be able to somehow provide a basis of data to offer any applicant some kind of coverage. I think we all know that about 90 percent of life insurance applicants are offered some type of standard type of coverage; six percent are offered some kind of substandard or rated coverage; and four percent are declined. It's the four percent that we're working on right now.

The reason these applicants are declined is because we don't have enough data upon which to make an intelligent decision. So, what you've seen up here are ways that we are working toward, long term, being able at least to come up with some kind of a definitive way to be able to offer these four percent coverage, such that anybody who applies at some point in time could be offered some kind of a coverage.

Chart 1

Surveillance, Epidemiology, and End Results

NATIONAL CANCER INSTITUTE

**SIGMOID COLON ADENOCARCINOMA (8140), M, Ages 65-74, Regional, Grade 2, Rx SSS-30**

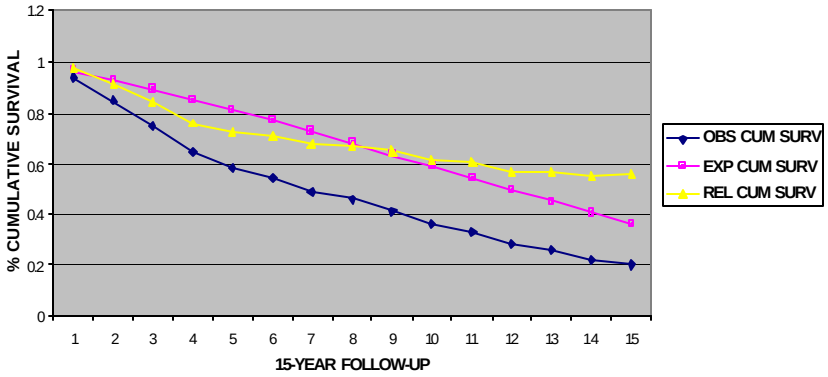



Chart 2

Surveillance, Epidemiology, and End Results

NATIONAL CANCER INSTITUTE

**SIGMOID COLON ADENOCARCINOMA (8140), M, Ages 65-74, Regional, Grade 2, Rx SSS-30**

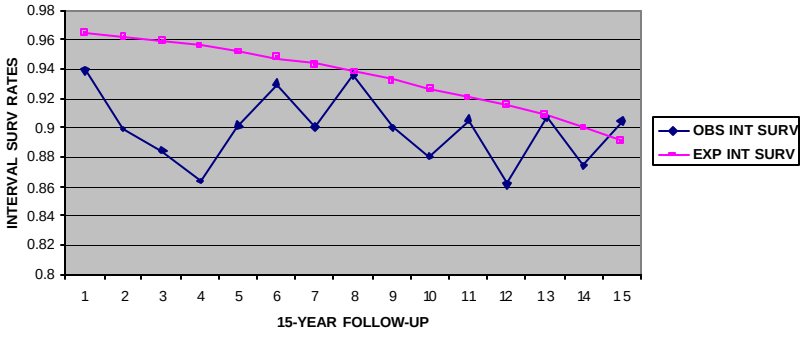


Chart 3

## Neural Networks

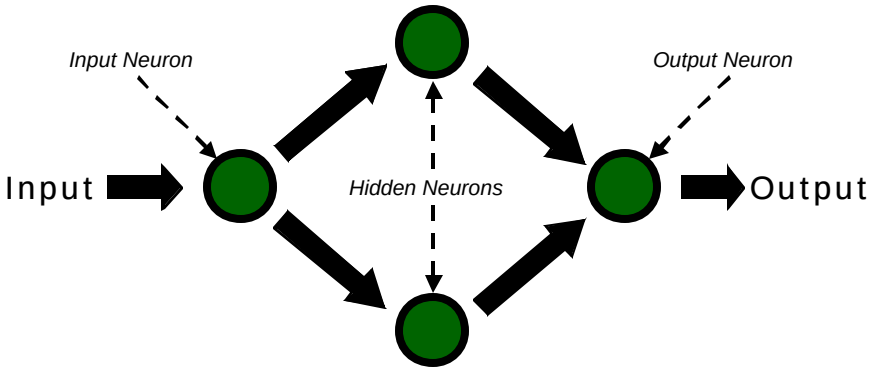



Chart 4

## Neural Networks

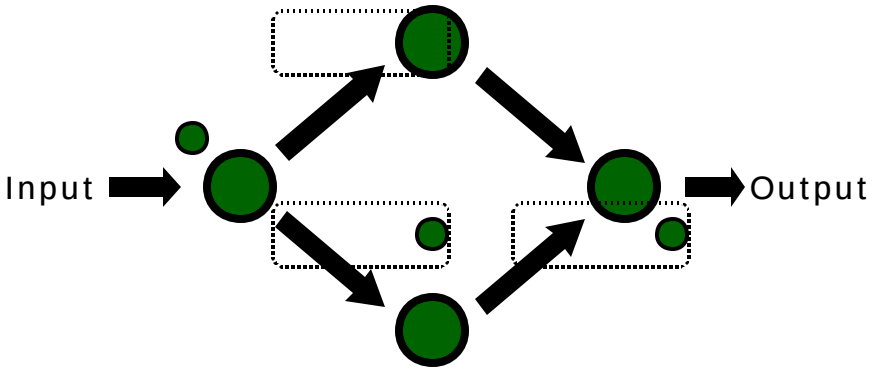
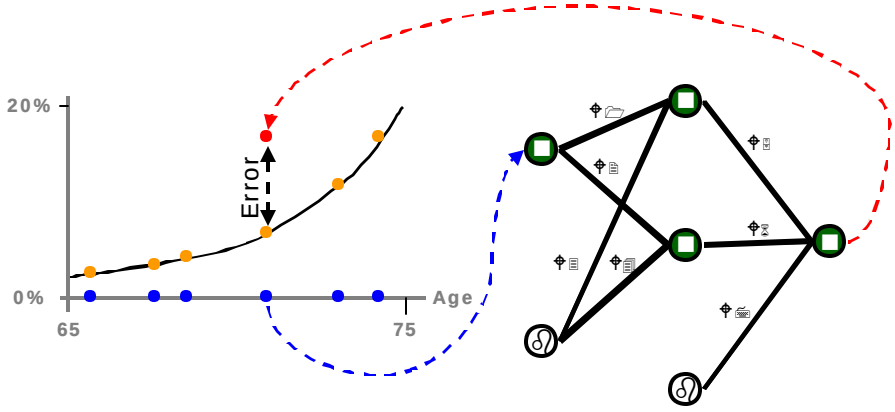


Chart 5

# Back-Propagation

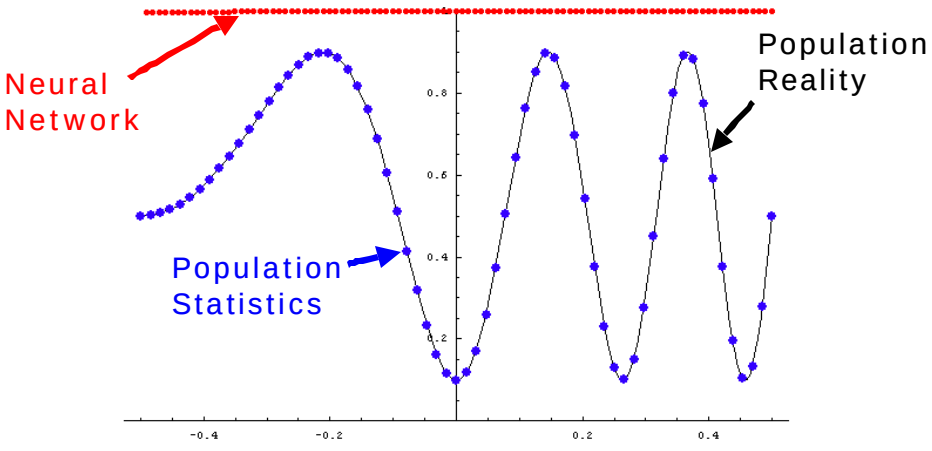



**Error** = ☜☎☤🗁🗃 ☤🗎🗃 ☤🗏🗃 ☤🗐 🗃 ☤🗄🗃 ☤⌛🗃 ☤📻🕑

☤♓'s contribution = $\frac{?☜}{?☤♓}$

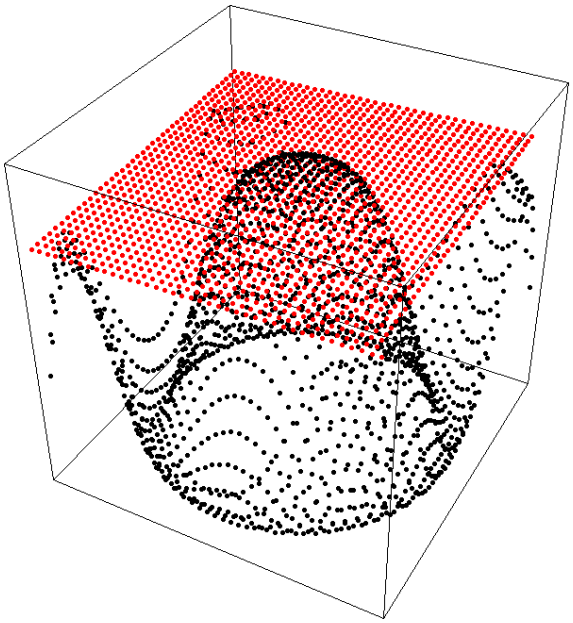
Chart 6

# Back-Propogation




(Each Animation Frame represents 100 Epochs)

Chart 7



(Each Animation Frame represents 10 Epochs)

Chart 8

# Results

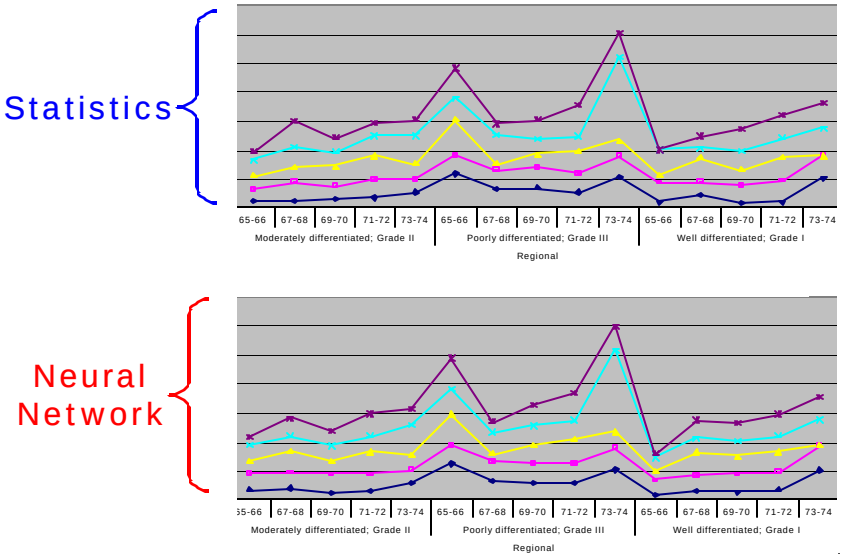
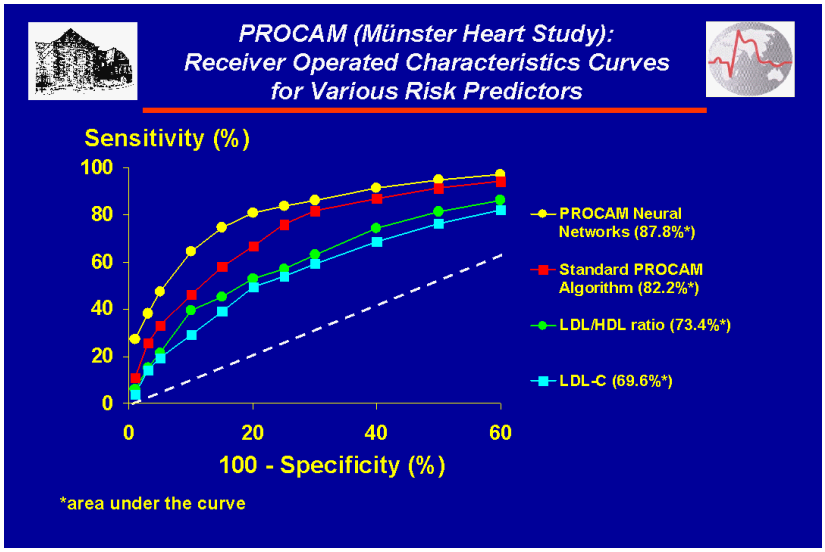


Chart 9


PROCAM (Münster Heart Study):
Receiver Operated Characteristics Curves
for Various Risk Predictors
Sensitivity (%)
100
80
60
40
20
0
0
20
40
60
100 - Specificity (%)
*area under the curve
PROCAM Neural Networks (87.8%*)
Standard PROCAM Algorithm (82.2%*)
LDL/HDL ratio (73.4%*)
LDL-C (69.6%*)