Discussant Comments
Concurrent Session 1A: Data Sources and Analysis

Thomas Edwalds

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

THOMAS EDWALDS: Thank you, Kai. One of the main roles of the discussant at these sessions is to stimulate discussion and questions from the audience. Clearly, my role here is completely superfluous. I'll do my best to try and contribute something to the discussion today.

The first thing that occurred to me when I was reading these papers is the motto of the Society of Actuaries, which is that the work of science is to "substitute facts for appearances and demonstrations for impressions," a quote from John Ruskin from the 19th century. When we're thinking about advanced-age mortality, which is the topic of this whole conference, one of the first questions is, What are the facts? We can't substitute facts for appearances until we know what the facts are, and in order to determine the facts, we need accurate data. Some of the basic questions include what are the current mortality rates at these advanced ages; if they're changing, how are they changing; and what are the factors that might affect mortality at really advanced ages. All these are important questions that we'd like to know the answers to, and we need data to address them.

We've got our standard actuarial formula for central death rates; it's just the deaths divided by exposures. But the question is, How do you count the deaths, how do you count the exposures? And these are not easy questions to answer. All three of these papers show how much difficulty there is in actually getting accurate data so that we can answer these questions.

The Human Mortality Database is a gold standard for mortality analysis. It's taking the data from vital statistics and trying to get the best of that data, the most you can get out of it. They take exposures from census data and deaths from vital records. I thought it was significant that they used country-specific experts in assessing the quality of the data and understanding what's going on with it.

At the high ages, they do use extinct generation and survivor ratio methods to get those mortality rates at the high ages, and as Ms. Barbieri pointed out, they may have good rates at 80s and 90s, but above 100, it gets really questionable. I thought it was significant that the Human Mortality Database was launched online in 2002, which was the exact same year that we had our very first Living to 100 Symposium. That's a complete coincidence, but it's based on the fact that at that time, we had growing awareness of how important it was to try to actually understand what's

going on with mortality at these high ages.

When I started my career as an actuary, the tables ended at age 100. Nobody cared about what happened after age 100, because nobody sold insurance after age 70, and the effect of mortality after age 100 would just be rounding error in calculating prices or reserves for life insurance. But by the time we got to the turn of this century, we started to realize that there was a growing cohort of centenarians, and this has some important implications for public policy and for our insurance products, so we actually need to know something about it.

The Human Mortality Database, as Ms. Barbieri pointed out, is a widely used reference for mortality, because they put great effort into creating high-quality data, and it is thoroughly documented. So they make it very clear how it's put together, and many papers presented at this conference have used Human Mortality Database as a data source, including one at this session here.

One of the drawbacks of the data set, as Ms. Barbieri pointed out, is that it's got very limited data elements. The data elements are gender and age and then country, and in some cases you've got subpopulations within country, but that is pretty much the extent. They are extending it to include cause of death, so if that can be done successfully, it would be a big addition, although any of us who have worked with cause of death realize that it's very dicey as to whether you can actually get accurate information on cause of death. Even from the original death certificates, that's not necessarily correct information.

One of the other papers here, the last one by Mr. Andreev, uses the Human Mortality Database, adding yet another citation for that work. I have to admit that when I looked at the title of this paper, I was very skeptical about measuring regional mortality and breaking it down by state from the death registrations and census data. In constructing the mortality rates, they're using the almost extinct cohort method, as he explained pretty thoroughly. One of the assumptions with that is that there is negligible migration. At the country level, I'm willing to believe that. It's somewhat less believable at the state level. Just anecdotally, I had an aunt who lived to be 91. At 87 she moved from Denver to Milwaukee, because her son wanted her to be closer. The mortality estimates are somewhat sensitive to the estimates of the numbers surviving for those non-extinct

cohorts, but as Mr. Andreev pointed out, a lot of the cohorts are fully extinct, which helps stabilize that somewhat.

As has been demonstrated numerous times, age misreporting is a huge problem for advanced-age mortality. The example that Mr. Gu and Mr. Andreev ran through in their paper is a hypothetical simulation looking at the effect on mortality rates from an assumed pattern of misreporting. If you think about the *dx* curve and the modal age of death, their assumption of symmetrical age misreporting will have a tendency to push the deaths away from the mode. So above the mode, the reported age at death will tend to be higher, and below the mode, the reported age at death will tend to be lower. That's driving the mortality pattern they found of the age misreporting causing depressed mortality rates at the high ages and the elevated mortality rates at the younger ages. Other papers presented at previous Living to 100 Symposia have shown similar results concerning the effects of age misreporting on measured mortality rates.

The title of the paper promised to look at level and trend of mortality by region, and as I said, I was very skeptical about this. They measured mortality at exactly two points in time, the 1960s and current, and the mortality trend was actually not measurable from this data, which is what I would have expected. That's because the data from the 1960s really weren't usable for determining the level of advanced-age mortality in the 1960s.

Now, the current data do seem quite plausible, although I am not persuaded that mortality in Hawaii is really all that good. I can think of several reasons why. First of all, it's a small sample. Out of all of our states, it has one of the smallest populations. I think that migration could well be an issue there, and considering the fact that it wasn't even a state until 1959, I question the accuracy of a lot of those records, although I'd be willing to experiment by moving to Hawaii to see if I live longer.

I was very skeptical about measuring level and trend of mortality, but it turns out that the paper is really a data-quality investigation. It's using the Human Mortality Database as a benchmark for good-quality data, since it seems to be the best possible, so if we observe a similar mortality pattern in our state-level data, then we can have confidence that this is good data. If we observe a different pattern, we do not have confidence that the data is any good. So although the

working title of the paper said they were going to measure a trend in the level of mortality by state, what they really measured was a trend in data quality. There's a pretty clear demonstration that the data in the 1960s were of poor quality, and those mortality estimates really can't be relied on as a good indication of what mortality was at that time. However, because of the trend in quality, we now have better-quality data, and current data are actually a much better indication of what the current level is.

One of the things that this study was doing is taking a Human Mortality Database type of approach, because they are the same data sources fundamentally, and trying to add a data element to it by breaking it down by state. The title of the paper talks about regions. There wasn't really an aggregation by region. I might have had a little more confidence in regional groupings of the data, with actual aggregations of data across regions as a way of demonstrating that the observed mortality differences really make sense. It's still good work, and the graph that was shown in the slides with the states colored in by life expectancy range gives a sense of where these regional clusterings might be.

The other paper—the one in the middle that I skipped over on centenarian pedigrees—I thought was an exciting approach, because this is now a completely different data source. It is not using the vital statistics data directly, but it's building off of the centenarian studies. It starts with people that we know live to be 100 and then tries to build out information about these individuals. They used FamilySearch.org as the primary source for building out this information, so they are using online resources to help get a better understanding of these individuals who have lived to be 100, including a lot of data elements. They showed both a manual and an automated way of constructing these pedigrees. I found that to be very interesting and very exciting.

The major idea here was to be able to analyze the heritability of longevity, but the approach would give you the opportunity to look at other variables as well, because you're gaining all this information about the people who are in the study. The ideal situation in this data set, as well as in any of the data sets, is that if you get a birth certificate and a death certificate on the same person, and you're certain of it, you know how old that person was when she or he died. This is the ideal situation, but we don't actually find it happening all that often. All of the effort that went into

constructing these pedigrees, as described in the paper, illustrates how hard it is to expand the data elements, to get more information about the individuals, so that we can understand the factors that might affect mortality at these high ages.

The Human Mortality Database has 80 billion life years exposed. In the centenarian pedigrees data set, there are about 65,000 life years exposed. It's a lot of effort for a relatively small number of cases, but the good news is that when you get those cases, you have robust data on those individuals. The automated methods for creating the pedigrees can also be useful in accelerating the process of expanding this data set. I know that the Gavrilovs have done a lot of work with matching of data. I think that anything that helps that process move a little faster is very useful for giving us more information.

I do want to congratulate all of the authors. I think these are excellent papers, and in particular, they're helping us develop accurate data, so that we can do advanced-age mortality analysis.